Tesis Doctoral

Modelado predictivo en flujo de datos de procesos con deriva de concepto y su aplicación al turismo en Canarias

Programa de Doctorado en Empresa, Internet y Tecnologías de las Comunicaciones

Juan Antonio Guerra Montenegro

Las Palmas de Gran Canaria

Mayo de 2022

D. JOSÉ ALBERTO RABADÁN BORGES, COORDINADOR DEL PROGRAMA DE DOCTORADO EMPRESA, INTERNET Y TECNOLOGÍAS DE LAS COMUNICACIONES DE LA UNIVERSIDAD DE LAS PALMAS DE GRAN CANARIA, INFORMA,

Que la Comisión Académica del Programa de Doctorado en su sesión de fecha __/__/____ tomó el acuerdo de dar el consentimiento para su tramitación, a la tesis doctoral titulada "Modelado predictivo en flujo de datos de procesos con deriva de concepto y su aplicación al turismo en Canarias" presentada por el doctorando D. Juan Antonio Guerra Montenegro y dirigida por el Doctor D. Javier Jesús Sánchez Medina y el Doctor D. David de la Cruz Sánchez Rodríguez.

Y para que así conste, y a efectos de lo previsto en el Art º 11 del Reglamento de Estudios de Doctorado (BOULPGC 7/10/2016) de la Universidad de Las Palmas de Gran Canaria, firmo la presente en Las Palmas de Gran Canaria, a __ de _____ de dos mil _____.

"La cosa más misericordiosa del mundo, creo, es la incapacidad de la mente humana para correlacionar todos sus contenidos ... algún día, la unión del conocimiento disociado abrirá vistas tan terribles de la realidad y de nuestra temible posición en el mundo que nos volveremos locos por la revelación o huiremos de la luz hacia la paz y la seguridad de una nueva Era Oscura."
Howard Phillips Lovecraft.

A mis padres, por inculcarme conocimientos y la sabiduría para saber cuándo aplicarlos.
A mi pareja, por aportar la luz más cálida y brillante en las horas más frías y oscuras.
Y a mi abuela, porque cada día que pasa me demuestra que ser joven está en la mente.

AGRADECIMIENTOS

A mis directores, Javier Sánchez Medina y David Sánchez Rodríguez, y a mi tutor de tesis, Agustín Sánchez Medina, por darme la posibilidad de probarme a mí mismo y de demostrar mi valía ante un camino tan largo y tortuoso, pero a la vez tan edificante e iluminador, y por depositar su confianza en mí.

Al *Centro de Innovación y Sociedad de la Información* y a todo su personal, por ser mi hogar durante toda mi trayectoria predoctoral y por facilitarme los medios y el apoyo necesarios para hacer frente a los retos que se me han presentado.

A la *Escuela de Ingeniería Informática*, por recibirme siempre con los brazos abiertos y por los buenos recuerdos que me evocan no solo sus pasillos, sino también su alumnado y docentes.

A la *Universidad de Las Palmas de Gran Canaria* y a la *Agencia Canaria de Investigación, Innovación y Sociedad de la Información*, por la inversión realizada en mi formación académica.

A la *Universidad del País Vasco / Euskal Erriko Unibersitatea*, en especial a Javier del Ser Lorente y a los compañeros de *Tecnalia*, por acogerme durante mi estancia y tratarme como a uno de los suyos.

Mención especial reciben ocho personas sin las cuales este periplo habría sido mucho más complejo. A Eduardo y a Fabio, por contribuir a que no perdiera la cordura gracias a nuestras salidas, nuestras charlas y nuestras risas. A Jessica y a *Puli*, por sacarme de la monotonía y enviarme a mundos de ensueño con sus bellos y fantásticos juegos de mesa junto a todas esas charlas, piques y risas que han conllevado. A Atteneri y a Alba, por permanecer tan cerca a pesar de estar tan lejos. A Echedey Martín, el pipiolo con el que he cazado estrellas y perseguido sueños más allá de un mar de luces de neón y de lluviosas ciudades en la noche. Y a Enrique Santiago Souto, por su apoyo incondicional y por su amistad más allá de todo deber y de toda distancia.

Y, por último, a mi familia. A mi abuela *Juanela*, que cada día que pasa me enseña que ser joven es algo que está en la mente y que nunca debes dejar atrás, y a mi abuela Manuela, que me mostró que la rectitud y la perseverancia no van reñidas con el cariño y la felicidad. Y en especial a mis padres, por apoyarme en mis proyectos, por creer en mí y en mis decisiones y por soñar junto a mí mis sueños. Gracias a toda mi familia por mostrarme un apoyo y un amor más allá de lo imaginable.

Y a Lorena, por ser mi compañera, mi confidente y mi aliada. Por mostrarme siempre la dirección correcta y por iluminar mi camino a través de la noche de la incertidumbre.

Mi más profundo agradecimiento a todas las personas implicadas.

ULPGC
Universidad de
Las Palmas de
Gran Canaria

Tesis Doctoral

Modelado predictivo en flujo de datos de procesos con deriva de concepto y su aplicación al turismo en Canarias

Programa de Doctorado en Empresa, Internet y Tecnologías de las Comunicaciones

Directores:
Dr. D. Javier Jesús Sánchez Medina
Dr. D. David de la Cruz Sánchez Rodríguez

Las Palmas de Gran Canaria
Mayo de 2022

Tabla de contenido

1. Resumen. 19

2. Introducción. 21

2.1. Motivación. 22

2.2. Hipótesis de partida. 23

2.3. Objetivos. 24

3. Estado del Arte. 27

3.1. Inteligencia Computacional. 28

3.1.1. *Machine learning*. 28

3.1.2. Toma de decisiones. 32

3.1.3. Computación natural. 33

3.1.4. *Data Stream Mining*. 34

3.2. El sector turístico y la Inteligencia Computacional. 35

3.2.1. Gestión y estimación de ingresos. 45

3.2.2. Sistemas de perfilado y recomendación. 52

3.2.3. Previsión de demanda turística. 57

3.2.4. Previsión meteorológica y evaluación de riesgos medioambientales. 61

3.3. Las condiciones meteorológicas y su influencia en el sector turístico. 62

3.4. Tráfico y turismo: relaciones y nuevas tecnologías. 65

4. Metodología. 67

4.1. Desarrollo de un modelo de regresión lineal incremental y adaptativo. 67

4.2. Desarrollo de un modelo de predicción basado en la Teoría de la Resonancia Adaptativa. 72

5. Aplicaciones. 76

5.1. *Data Stream Mining* aplicado a la predicción de la velocidad máxima del viento en las Islas Canarias. 76

5.1.1. Conjunto de datos y análisis exploratorio. 78

5.1.2. Experimentación y resultados. 80

5.1.3. Discusión de resultados. 87

5.2. Una aproximación basada en la resonancia neural para tratar flujos de datos no estacionarios de tráfico. 88

5.2.1. Descripción del conjunto de datos: la autopista M-30 de Madrid. 89

5.2.2. Experimentos y resultados. 91

5.2.3. Discusión de resultados. 97
6. Conclusiones. 98
6.1. Contribuciones teóricas. 98
6.2. Contribuciones prácticas. 100
6.3. Retos y desafíos: Inteligencia Computacional y Sector Turístico. 101
6.3.1. Inteligencia práctica. 102
6.3.2. Fusión de datos de múltiples fuentes. 102
6.3.3. Aprendizaje dinámico y *online*. 103
6.3.4. Modelos de datos encriptados. 104
6.3.5. Anticipación a sesgos en los datos. 105
6.4. Limitaciones y futuras líneas de investigación. 106
6.5. Conclusiones y recomendaciones para el sector del turismo. 109
6.5.1. Inversión en nuevas investigaciones basadas en *Data Stream Mining*. 110
6.5.2. Convenios con organismos públicos para la utilización de datos. 110
6.5.3. Inversión en infraestructura vial. 110
7. Referencias. 112

1. Resumen.

Esta tesis tiene como objetivo analizar los procesos del modelado predictivo en flujos de datos sujetos a deriva de concepto o deriva conceptual (del inglés *concept drift*), utilizando para ello diferentes metodologías englobadas dentro de la Inteligencia Computacional (del inglés *Computational Intelligence*), así como su aplicación al tejido turístico en las Islas Canarias.

En primer lugar, se ha realizado un estudio exhaustivo del estado del arte respecto a la aplicación de metodologías de *Computational Intelligence* dentro de toda la industria turística y hotelera, dando a conocer cómo se están integrando estas metodologías en este sector y qué se está haciendo para solventar los problemas del mismo, además de proponer una nueva clasificación de dichos esfuerzos en base a su utilización y sus casos de uso. Dicho estudio puso de manifiesto la ausencia de investigaciones dentro del área del turismo relacionadas con la predicción de fenómenos meteorológicos adversos o con la predicción de condiciones de tráfico no deseadas en zonas turísticas. Este hecho, unido a una carencia de investigaciones de este tipo dentro de las áreas turísticas del archipiélago canario, propició el desarrollo de diferentes metodologías basadas en *Computational Intelligence*, y más concretamente mediante minería en flujos de datos y aprendizaje *online*, con el objetivo de generar modelos predictivos capaces de adaptarse al *concept drift* y de proporcionar nuevas predicciones basadas en datos recibidos y analizados en tiempo real.

Teniendo esto en cuenta, se propone la aplicación de dos metodologías diferentes de *Computational Intelligence* basadas en el aprendizaje *online* con el objetivo de crear modelos predictivos resistentes al *concept drift* presente en datos no estacionarios. Como primera aproximación, se han analizado datos recogidos por la AEMET en todas sus estaciones climatológicas del archipiélago canario, con el objetivo de realizar predicciones con una hora de antelación respecto a fenómenos climatológicos adversos. Esto se ha podido lograr mediante el uso de un modelo de regresión lineal adaptado al aprendizaje en tiempo real. Como segunda aproximación, se ha utilizado un modelo predictivo basado en redes neuronales adaptativas resonantes (del inglés *Adaptive Resonance Theory*, o *ART*) para predecir, con una hora de margen, condiciones de alta densidad de tráfico en el caso particular de la autopista M-30 de Madrid, la cual es conocida por sus condiciones de circulación caóticas y aleatorias.

El objetivo de esta tesis persigue, mediante la proposición de las metodologías previamente descritas, dotar tanto a instituciones públicas como a organismos privados de nuevas herramientas o aproximaciones para resolver problemas que puedan degradar la calidad de los servicios prestados en el sector turístico y hotelero, a través de la aplicación de metodologías de *Computational Intelligence* sobre datos recogidos y analizados en tiempo real, y revelando los beneficios que presentan estas metodologías al ser adaptadas para hacer frente al *concept drift*.

2. Introducción.

En los últimos años, el interés de la industria turística por la utilización de metodologías orientadas al análisis de datos mediante aprendizaje automático se ha visto notablemente incrementado debido principalmente a la gran cantidad de datos que esta industria es capaz de generar. Estos datos, aunque poseedores de un gran potencial, resultan difíciles de tratar debido a su gran tamaño y a las relaciones ocultas inherentes a su naturaleza, por lo que para su análisis es necesario utilizar la potencia de las metodologías actuales de *Computational Intelligence*. Según (Siddique, 2013), estas metodologías pueden ser definidas como el conjunto de metodologías y métodos computacionales inspirados por la naturaleza con el objetivo de resolver problemas complejos para los cuales los modelos matemáticos o tradicionales no pueden hallar solución por algún motivo. Dicho conjunto de metodologías, según el organismo internacional *Institute of Electrical and Electronics Engineers* (IEEE), se compone de redes neuronales artificiales, sistemas difusos y algoritmos evolutivos, y están basadas en el raciocinio de los seres humanos teniendo en cuenta información inexacta o incompleta, generándose acciones adaptadas a dicho razonamiento.

Dentro de la *Computational Intelligence* se encuentra el *machine learning*, el cual fue definido por (Mitchell, 1997) como la utilización de algoritmos computacionales que son mejorados automáticamente a través de conductas evolutivas basadas en los datos. En los recientes años, la utilización del *machine learning* ha aumentado exponencialmente debido a su extrema utilidad a la hora de realizar actividades complejas de forma automática sobre grandes cantidades de datos, las cuales han comenzado a estar disponibles de manera relativamente reciente (Kamel et al., 2018). Desde el análisis exploratorio de los datos hasta el análisis predictivo o prescriptivo, esta metodología resulta notablemente útil para comprender las relaciones internas y la estructura de los datos analizados, además de generar predicciones precisas en base a los patrones encontrados en dichos datos, las cuales resultan de gran utilidad para los responsables de la toma de decisiones dentro de las organizaciones públicas y privadas. Dentro de esta rama ha surgido una nueva metodología, basada en la recogida, el preprocesamiento y el análisis de datos en tiempo real para generar modelos predictivos adaptables a cambios dentro de los propios datos. Esta nueva metodología, conocida como *online learning* o *Data Stream Mining* (minería en flujo de datos) (Bifet y Kirkby, 2009; Gama y Gaber, 2007) se presenta como la evolución natural del *machine learning*, y cambia de manera radical la forma de entrenamiento y aprendizaje de los modelos dentro de esta área debido a que sigue un modelo de entrenamiento *test-then-train*, o entrenamiento precuencial (Dawid, 1984). Este modelo está basado en probar primero el rendimiento del modelo generado para luego entrenarlo con nuevos datos recibidos en tiempo real. Esto se hace en lugar de utilizar un gran conjunto de datos o *Big Data* para realizar el entrenamiento del modelo y luego probarlo con nuevos datos o con datos previamente reservados del bloque de *Big Data* anteriormente mencionado, como se venía haciendo habitualmente (Krawczyk et al., 2017). Este nuevo modelo de entrenamiento *test-then-train* resulta más eficiente, ya que es más realista que el paradigma clásico en el que se asumía tanto una capacidad de computación “infinita”

en términos de CPU y memoria como una estabilidad estadística del fenómeno que se pretende modelar (Krawczyk et al., 2017).

2.1. Motivación.

La motivación principal de esta tesis es abordar el estado de adopción de metodologías de *Computational Intelligence* dentro del sector turístico canario y de cómo pueden influir positivamente dentro del mismo, mejorando la calidad de la estancia turística e incrementando el valor turístico del archipiélago.

En los últimos años, la aplicación de metodologías de *Computational Intelligence* dentro del campo de la industria turística se ha visto incrementada de manera notable, desvelando resultados positivos dentro de áreas como la predicción de la demanda turística (Kamel et al., 2018) o el pronóstico del consumo energético (Chen, Tan, y Berardi, 2017). Esto ha aumentado los márgenes de beneficio e incrementado la calidad de dichos servicios a la hora de ser ofertados a visitantes o turistas (Guerra-Montenegro et al., 2021). No obstante, la utilización de este tipo de análisis predictivo mediante *Big Data* en tiempo real en lugar de utilizar históricos de datos continúa siendo algo novedoso dentro de esta industria (Bernard, 2016), generando así la necesidad de desarrollar metodologías basadas en este paradigma, siendo este el objetivo principal de la presente tesis doctoral.

En términos mundiales, el sector turístico en las islas es un factor de dependencia para las mismas, algo que se intensifica en las islas de menor tamaño debido a la alta frecuencia con la que el turismo es el motor económico de estas (Briguglio y Briguglio, 2005). Respecto al turismo dentro del archipiélago canario, es de gran importancia destacar el posicionamiento de esta localización como destino turístico de fama mundial (Ispas y Saragea, 2011) debido a diversos factores, tales como su fácil accesibilidad desde Europa, su oferta de servicios al nivel europeo y la hospitalidad de sus habitantes frente a los turistas (Sánchez-Medina et al., 2019). Dicha hospitalidad es fruto del posicionamiento del turismo como industria principal del archipiélago dado que, acorde a los datos del Instituto Canario de Estadística (ISTAC) en 2018, el sector turístico ha llegado a generar un 40% de empleo y un 35% del PIB de las Islas Canarias, aunque con la pandemia estos niveles han cambiado. No obstante, existe un vacío en la literatura científica donde las técnicas de *Computational Intelligence* sean aplicadas a este sector del archipiélago canario, siendo este el principal objetivo de la presente tesis doctoral a pesar de la ausencia de datos de acceso abierto necesarios para crear modelos predictivos mediante estas metodologías.

Adicionalmente, es importante recalcar el papel del transporte público dentro de la economía turística, ya que el turismo como tal no existiría sin el apoyo de este sector (Le-Klähn y Hall, 2015). Los turistas esperan que los servicios provistos en esta área posean una serie de características, de entre las cuales destacan la rapidez, la confiabilidad y la seguridad (Budiono, 2009; Fellesson y Friman, 2012). Es conocido que tanto la fluidez de movimiento como la oportunidad de acceso entre diferentes zonas

son características intrínsecas al crecimiento sostenible de las mismas, originando que la movilidad tenga un impacto notable en la competitividad y prosperidad de dichas zonas (Duval, 2007; Page, 2007). Esto proporciona una oportunidad lucrativa que depende de los productos y servicios de transporte público ofertados a los turistas, y cuya capacidad de explotación se ve directamente influenciada por el atractivo y el correcto diseño de estos (Albalate y Bel, 2010).

Habida cuenta de la influencia de los factores climatológicos y de las condiciones del transporte público sobre la calidad de las estancias turísticas, unido esto a la ausencia de nuevas investigaciones sobre dichos factores dentro del archipiélago canario, surgió la motivación para desarrollar la presente tesis doctoral con el objetivo de dotar de nuevos mecanismos de predicción y planificación a organizaciones públicas y privadas, realizando también una categorización del estado del arte referente a las metodologías de *Computational Intelligence* respecto a su aplicación en el sector turístico y hotelero para comprobar qué metodologías dentro de esta disciplina podrían ser las adecuadas para resolver posibles problemas o retos dentro de este sector.

2.2. Hipótesis de partida.

De manera sintética, las hipótesis planteadas en esta tesis son:

- Teniendo en cuenta la importancia del clima en un destino turístico, sería posible mejorar las predicciones meteorológicas mediante la aplicación de metodologías de *Data Stream Mining*.
- Dado que el tráfico juega un papel clave en la calidad de la estancia turística, es posible mejorar la predicción de fenómenos de circulación adversos mediante la utilización de algoritmos adaptativos.

Respecto a la primera hipótesis, se pensó en cómo la predicción de sucesos climatológicos adversos podría ayudar de manera notable al sector turístico a planificar diferentes tipos de eventos teniendo en cuenta diversos factores meteorológicos. Fue por ello por lo que, para verificar la veracidad de dicha hipótesis, se escogió un fenómeno particularmente difícil de predecir: la velocidad máxima del viento. Para realizar dicho modelo predictivo, se utilizaría una aproximación basada en *Data Stream Mining* con objetivo de generar modelos robustos y resistentes a los cambios de tendencia dentro de la naturaleza del viento.

Para la segunda hipótesis se pensó en una aproximación diferente. Como se ha mencionado anteriormente, el transporte público juega un gran papel dentro del sector turístico, ya que los turistas esperan que sea rápido, fiable y seguro (Le-Klähn y Hall, 2015; Budiono, 2009; Fellesson y Friman, 2012). Esto, unido al hecho de que la movilidad es determinante tanto para el crecimiento sostenible de unas zonas como para su prosperidad (Duval, 2007; Page, 2007), fue la base de la segunda hipótesis. Para demostrar su veracidad, se decidió escoger la red de carreteras de una zona de gran atractivo turístico con intención de predecir fenómenos de tráfico no deseados, utilizando para ello una novedosa aproximación basada en el funcionamiento del

cerebro humano y de sus redes neuronales, capaz de adaptarse al caótico flujo del tráfico de manera rápida y precisa.

2.3. Objetivos.

De las hipótesis anteriormente formuladas surge el objetivo general de esta tesis doctoral, que es analizar posibles aplicaciones del *Data Stream Mining* sobre diferentes infraestructuras intrínsecamente ligadas a la industria turística, desarrollando para ello una metodología basada en creación de modelos predictivos en *streaming* y en su aplicación a retos propios de los destinos turísticos y contribuyendo así a tres áreas de conocimiento.

En primer lugar, al estado del arte de metodologías de *Computational Intelligence* aplicadas a dicha industria con el objetivo de generar predicciones de gran valor logístico y/o económico. Por otra parte, se contribuye también al análisis en tiempo real de datos climatológicos del archipiélago mediante la aplicación de algoritmos de *machine learning*, con el objetivo de predecir la velocidad máxima del viento para así prever fenómenos meteorológicos adversos con suficiente antelación como para preparar respuestas adecuadas a los mismos. Por último, se contribuye a la predicción en tiempo real de las condiciones de circulación de tráfico propias de un destino turístico de gran relevancia, utilizando como caso de estudio la ciudad de Madrid debido a la inexistencia de datos públicos de tráfico en el archipiélago canario a la hora de realizar esta tesis.

Dentro del objetivo general se enmarcan tres objetivos específicos, los cuales han ido cumpliéndose incrementalmente durante el desarrollo de la presente tesis:

- Realización de un estudio exhaustivo del estado del arte respecto a las metodologías de *Computational Intelligence* en el sector turístico.
- Desarrollo de un modelo computacional predictivo sobre condiciones meteorológicas adversas, aplicado a una zona turística concreta.
- Aplicación de un modelo basado en *Adaptive Resonance Theory* para la predicción de condiciones de tráfico adversas en zonas turísticas.

En primer lugar y para dar respuesta al primero de los objetivos específicos, se realizó un estudio exhaustivo del estado del arte respecto al *data stream mining* aplicado a desarrollos *data-driven* en el campo turístico mediante el uso de metodologías de *Computational Intelligence* con el objetivo de identificar la distribución de las investigaciones y sus posibles vacíos dentro de este sector, culminando en la publicación del artículo científico "*Computational Intelligence in the hospitality industry: A systematic literature review and a prospect of challenges*", realizado por el doctorando junto a Javier Sánchez-Medina, Ibai Laña, David Sánchez-Rodríguez, Itziar Alonso-González y Javier Del Ser, y publicado en la revista *Applied Soft Computing*.

Dado que el clima es un factor determinante dentro de la economía turística (Scott y Lemieux, 2010), como segundo objetivo específico se marcó el desarrollo de

modelos predictivos de las condiciones meteorológicas locales en zonas turísticas de Canarias en base a las señales recibidas por diferentes estaciones meteorológicas, creando para ello un modelo de regresión lineal basado en *Data Stream Mining* capaz de realizar predicciones sobre la velocidad máxima del viento con un margen de 60 minutos y resultando en la publicación de un segundo artículo científico titulado "*Data Stream Mining Applied to Maximum Wind Forecasting in the Canary Islands*", realizado conjuntamente con Javier Sánchez-Medina, David Sánchez-Rodríguez, Itziar Alonso-González y Juan Navarro-Mesa, y publicado en la revista *Sensors.*

Como tercer objetivo específico, se planteó el desarrollo de un modelo predictivo para condiciones de tráfico en una ciudad de gran valor turístico, siendo escogida la ciudad de Madrid por su gran relevancia turística a nivel europeo y por su autopista de circunvalación M-30, conocida por su elevada afluencia de tráfico. Este objetivo se estableció debido a la influencia que ejercen los atascos de tráfico en la calidad de la estancia turística y, por ende, en la satisfacción de los turistas (Jo et al., 2016). Para ello, se utilizó un modelo de redes neuronales basado en la Teoría de Resonancia Adaptativa, o *Adaptive Resonance Theory* (ART), la cual presenta capacidades nativas de *online learning* que la permiten adaptarse rápidamente a cambios en la tendencia interna de los datos. Esta última investigación completa el tercer objetivo de esta tesis.

Adicionalmente, de las conclusiones de estos tres objetivos específicos se extraen diversas recomendaciones que podrían servir como guía al sector turístico a la hora de realizar futuras acciones relacionadas con el *machine learning* aplicado al turismo en Canarias, respondiendo así al cuarto y último objetivo específico de la presente tesis.

A continuación, en el apartado 3 se expondrá la revisión del estado del arte realizada para identificar posibles aplicaciones y *gaps* en la aplicación de *Computational Intelligence* en el sector turístico. En el apartado 4 propone y se describe de forma exhaustiva la metodología aplicada para alcanzar cada uno de los dos objetivos específicos restantes. En el apartado 5 se exploran las aplicaciones de dicha metodología y los resultados obtenidos para cada uno de los casos junto a conclusiones parciales extraídas de los mismos. Finalmente, el último epígrafe se dedica a las conclusiones de esta tesis doctoral.

3. Estado del Arte.

El interés por el *machine learning* ha crecido exponencialmente en las últimas décadas (Kamel et al., 2018), especialmente tras el reciente aumento de la disponibilidad de datos en todos los campos de la ingeniería. Los beneficios del uso de estas metodologías van desde el análisis exploratorio (la comprensión de la estructura y las relaciones internas de los datos), hasta el análisis predictivo (el modelado de los procesos observados con el fin de predecir su evolución futura) y el análisis prescriptivo (la generación de recomendaciones de alto nivel para los tomadores de decisiones o los gerentes).

A continuación, se dividirá el estado del arte en tres subsecciones:

- Inteligencia Computacional.
- El sector turístico y la Inteligencia Computacional.
- Las condiciones meteorológicas y su influencia en el sector turístico.
- Tráfico y turismo: relaciones y nuevas tecnologías.

En la primera subsección, además de proporcionar una definición teórica precisa de la *Computational Intelligence*, también se describen las diferentes familias de *machine learning* y las metodologías contenidas en las mismas. Esta subsección sirve como una breve introducción para facilitar al lector la comprensión de los diferentes términos y metodologías que serán expuestos y mencionados en las subsecciones 3.2, 3.3 y 3.4, donde se expondrá el grueso del estado del arte estudiado en la realización de la presente tesis doctoral.

Tras esto, se muestra de forma detallada el estudio del estado del arte respecto a la aplicación de *Computational Intelligence* y, más específicamente, *machine learning*, en las diferentes áreas dentro del sector turístico. Este apartado también expone una categorización realizada a partir de todos los datos analizados, clasificando todas las metodologías halladas en base a su aplicación dentro de este sector, así como varias tablas donde se clasifican todos los trabajos de investigación analizados para ofrecer al lector una visión global de las investigaciones más recientes, así como de los posibles *gaps* y de las potenciales oportunidades para realizar nuevas investigaciones en el área.

En tercer lugar, se describen los avances más recientes en materia de predicción de fenómenos meteorológicos, aportando así una visión fiel del panorama actual a este respecto. También se incluyen trabajos de investigación realizados en el propio archipiélago canario, los cuales sirven de marco teórico para la aplicación de una metodología basada en *Data Stream Mining* para predecir la velocidad máxima del viento, siendo dicha metodología y aplicación descritas en los apartados 4.2 y 5.1 de esta tesis doctoral.

Por último, se exponen avances recientes dentro de la materia de predicción del tráfico en zonas de interés turístico. También se describen las metodologías más comúnmente utilizadas para resolver determinados problemas, todo ello utilizado como marco teórico para la novedosa aplicación de una metodología basada en redes

neuronales capaz de predecir con precisión eventos de tráfico adversos, tales como ralentizaciones o atascos. Esta metodología y aplicación son descritas en los apartados 4.3 y 5.2 de esta tesis, respectivamente.

3.1. Inteligencia Computacional.

Antes de seguir adelante, y con el fin de proporcionar una definición comúnmente aceptada de la *Computational Intelligence*, se han examinado en profundidad las definiciones previamente establecidas en la literatura. Según (Siddique, 2013) se puede afirmar que la *Computational Intelligence* es un conjunto de metodologías y enfoques computacionales inspirados en la naturaleza con el objetivo de abordar problemas complejos del mundo real para los que los modelos tradicionales o matemáticos no son aplicables por algunas razones. La Sociedad de Inteligencia Computacional del IEEE considera que estas metodologías comprenden las redes neuronales artificiales (ANN), los sistemas difusos y los algoritmos evolutivos. Estos métodos se aproximan a la forma en que los humanos razonan utilizando conocimientos incompletos, produciendo acciones de control adaptables, lo que hace que los sistemas de *Computational Intelligence* sean capaces de aprender de los datos de la experiencia.

3.1.1. *Machine learning.*

Dentro de la *Computational Intelligence* es necesario definir una de sus mayores subáreas, que incluye la mayoría de las aplicaciones conocidas de este tipo de técnicas, como la clasificación, la previsión, la agrupación, la regresión o el descubrimiento de patrones. Esta subárea se denomina *machine learning,* y puede definirse como el uso de algoritmos informáticos que mejoran automáticamente a través de la experiencia mediante la evolución de comportamientos basados en datos (Mitchell, 1997), y a menudo se considera como parte del campo de la Inteligencia Artificial. El *machine learning* se utiliza en un amplio conjunto de campos de investigación, como el reconocimiento del habla, el control de robots o la visión por ordenador (Mitchell, 1999).

Este conjunto de algoritmos informáticos puede dividirse en cuatro categorías principales (Jordan y Mitchell, 2015):

- **Supervisado:** Los datos de entrenamiento se estructuran en forma de pares *(x, y)*, donde el objetivo es producir una predicción *y* en base a un *input x*. Los *inputs* pueden ser tanto simples vectores como elementos de mayor complejidad, como grafos, imágenes o documentos.
- **No supervisado:** Utiliza datos sin clasificar o etiquetar, suponiendo ciertas propiedades estructurales dentro de los datos (probabilísticas, algebraicas o combinatorias), para crear un modelo capaz de generar una salida determinada en base al *input* o a una transformación de este.
- **Semi-supervisado:** Utiliza datos sin clasificar para aumentar la cantidad de datos clasificados en un contexto de aprendizaje supervisado, utilizando

entrenamiento discriminativo para combinar las arquitecturas desarrolladas para el aprendizaje no supervisado con formulaciones de optimización que hacen uso de las clasificaciones.

- **Por refuerzo:** En lugar de una salida correcta *y* para una entrada determinada *x*, los datos de entrenamiento solo proporcionan una indicación de si un resultado es correcto o no. De una manera más genérica puede asumirse que, dentro del contexto de una serie de *inputs x*, una señal de "éxito" engloba a la totalidad de dicha serie de *inputs*, y la asignación de éxito o fracaso no puede ser ligada a elementos individuales de esta serie.

Los métodos de *machine learning* se alinean así con la categoría de modelización (Eiben y Smith, 2015), ya que están orientados a encontrar el modelo. En los últimos años, el *machine learning* se ha aplicado en diferentes áreas, demostrando que puede ser una herramienta útil para resolver diferentes tipos de problemas clasificando datos o prediciendo resultados de situaciones. En el sector turístico y hotelero este enfoque computacional arroja resultados útiles que permiten a los establecimientos hoteleros obtener ventajas competitivas frente a sus competidores (Guerra-Montenegro et al., 2021).

A continuación, se explicarán las metodologías clasificadas dentro de cada una de las cuatro categorías anteriormente definidas para, de esta forma, facilitar la comprensión de los subsiguientes apartados.

3.1.1.1. Redes Neuronales Artificiales.

Las Redes Neuronales Artificiales, o *Artificial Neural Networks* (ANN) *en inglés*, se inspiran en la funcionalidad del cerebro humano y se componen de varios tipos de capas que contienen neuronas. La primera es la capa de entrada, luego puede haber una, dos o más capas ocultas y, por último, una capa final de salida (Wang, 2003). Se inventaron en los años 40, y desde entonces han tenido altibajos en su popularidad. Las ANN se definen imitando la estructura del cerebro humano, con neuronas, axones y dendritas que las conectan. La operación matemática consiste en un conjunto de neuronas dispuestas en capas e interconectadas de una manera determinada. Cada neurona realiza una operación aritmética (normalmente una suma) entre todas las conexiones entrantes, cada conexión ponderada por algún valor. Cada neurona tiene también una función de activación y su resultado se transmite a la siguiente capa o a la salida de la ANN.

Las redes neuronales convencionales, como modelos no supervisados, han demostrado ser una tecnología eficaz para el reconocimiento de patrones estructurales (Wang, 2003). Recientemente, la abundancia de potencia de cálculo disponible combinada con el uso de las unidades de procesamiento gráfico (*Graphical Processing Units, GPUs*) en lugar de las unidades centrales de procesamiento (*Central Processing Units*, *CPUs)*, por sus instrucciones *Very Long Instruction Word* y la consecuente idoneidad para procesar grandes matrices con operaciones algebraicas, está renovando

el interés de la comunidad científica por las ANN, en particular por las incluidas en el área del llamado *deep learning*: redes neuronales convolucionales (*Convolutional Neural Networks,* CNN), redes neuronales recurrentes, memorias de largo plazo y *autoencoders*, entre otras), consistente en general en un incremento del número de capas ocultas dentro de la red neuronal.

Entre las limitaciones de las ANN está su necesidad de un tamaño elevado de datos para el entrenamiento, su tiempo de entrenamiento generalmente largo y una tendencia a sobreajustar sus modelos. También existe otra preocupación sobre la interpretabilidad de los modelos aprendidos de *deep learning*, que en algunos contextos puede pesar mucho en contra de su uso.

3.1.1.2. Árboles de Decisión.

Los árboles de decisión son un caso de aprendizaje supervisado, y son una técnica muy común en la toma de decisiones secuenciales (Ishwaran y Rao, 2009). Son rápidos, producen modelos inteligibles y pueden ajustarse fácilmente para operar en una configuración de aprendizaje adaptativo en línea. Básicamente, un árbol de decisión es una estructura de decisión similar a un diagrama de flujo, compuesta por nodos y hojas. Cada nodo evalúa una condición de una característica particular del conjunto de datos, y dependiendo de la evaluación de dicha condición para cada observación del conjunto de datos, esta evaluación progresa hacia otro nodo inferior conectado para otra evaluación de característica posterior, o a una hoja donde se encuentra la categoría particular que predice el modelo. El aprendizaje de un modelo de árbol de decisión consiste en el entrenamiento de la estructura del árbol de nodos, las características que se evaluarán en cada nodo y los umbrales utilizados para decidir en cada nodo el siguiente. La característica más interesante de este tipo de modelos es su interpretabilidad y que su estructura contiene, como valor añadido, información interesante sobre la importancia relativa de las características. Los árboles de decisión se aplican con mayor frecuencia a los problemas de clasificación.

3.1.1.3. Métodos probabilísticos y bayesianos.

El paradigma bayesiano expuesto por (van de Schoot et al., 2014) interpreta la probabilidad como la experiencia subjetiva de la incertidumbre. En este paradigma, el ejemplo clásico de la experiencia subjetiva de la incertidumbre es la noción de hacer una apuesta". Además, también se afirma que hay tres componentes principales de la estadística bayesiana: el conocimiento de fondo, la información contenida en los datos, y la inferencia posterior que se obtiene combinando los dos primeros componentes (van de Schoot et al., 2014).

La inferencia bayesiana y la probabilística son métodos de aprendizaje que, a partir de la observación de instancias de datos, actualizan la probabilidad de una hipótesis de forma incremental. En este sentido, pueden considerarse dentro del subconjunto de aprendizaje supervisado del *machine learning*, que se aplica

normalmente a la clasificación. Dos de los principales puntos fuertes de las redes bayesianas (el tipo más conocido en este campo), son su capacidad para incorporar el conocimiento previo de los expertos (acelerando el proceso de aprendizaje), y que las salidas no son sólo categorías, sino también niveles de confianza. Además, uno de los métodos bayesianos más utilizados es el *Naive Bayes*, que consiste en un modelo de probabilidad condicional que asume que el valor de una característica concreta es independiente del valor de cualquier otra característica, dada la variable de clase.

3.1.1.4. Aprendizaje basado en instancias.

Los algoritmos de aprendizaje basados en instancias, pertenecientes a la categoría de *machine learning* no supervisado, son muy similares a los algoritmos *nearest neighbor* editados, y también son una derivación del clasificador de patrones *nearest neighbor* (Aha et al., 1991). Estos algoritmos son métodos que se basan en el aprendizaje de los casos pasados para crear modelos sin tratar de generalizarlos. En cambio, la idea es recordar todos los casos anteriores y asimilar cada nueva observación, agrupándola junto a las ya aprendidas.

Buenos ejemplos de esta familia de métodos son los *k-Nearest Neighbors* (kNN) (Silverman y Jones, 1989), un método no paramétrico útil en la regresión y la clasificación que consiste en una entrada de las 'k' muestras de entrenamiento más cercanas en el espacio de características, y la salida es una pertenencia a una clase (clasificación) o el valor de la propiedad para el objeto (regresión); y las redes de función de base radial (*radial basis function,* RBF) (Orr, 1996), que es un caso especial de una ANN simple que utiliza RBF como funciones de activación.

3.1.1.5. Ensembles.

Las técnicas de *ensemble* también están encontrando su lugar dentro del panorama del *machine learning*. Un *ensemble* puede definirse como un conjunto de clasificadores entrenados individualmente (como redes neuronales o árboles de decisión) cuyas predicciones se combinan al clasificar nuevas instancias (Opitz y Maclin, 1999). Mediante la combinación de modelos que pueden funcionar bien en algunos casos, (o complementarse con otros si tienen un peor rendimiento), se construye un modelo mejor que las partes. La forma de combinarlos puede ser tan sencilla como votar o seleccionar la categoría más votada en el caso de un problema de clasificación.

3.1.1.6. Clustering.

El *clustering* de datos puede definirse como el proceso de identificar agrupaciones naturales o *clusters* dentro de datos multidimensionales basados en alguna medida de similitud (por ejemplo, la distancia euclidiana) (Omran et al., 2007), y es un proceso importante en el *machine learning* y el reconocimiento de patrones. El *clustering* utiliza un amplio conjunto de técnicas de *machine learning* no supervisadas,

en las que cada observación no está asociada a una variable dependiente concreta. En su lugar, las observaciones se agrupan por medidas de similitud. Por lo tanto, el conocimiento se extrae en términos de cómo se agrupan las muestras en torno a una o varias características, identificando los posibles modos en el espacio de observación. La principal ventaja de las técnicas de *clustering* es que no es necesario etiquetar las observaciones (no hay que asociar ninguna categoría a cada observación), lo que supone un gran ahorro en términos de preprocesamiento de datos.

3.1.1.7. Minería por reglas de asociación.

El objetivo de la minería de reglas de asociación es extraer correlaciones significativas, asociaciones, patrones frecuentes o incluso estructuras casuales entre diferentes conjuntos de artículos en repositorios de datos (Kotsiantis y Kanellopoulos, 2005). El conocimiento que se persigue en este tipo de técnica supervisada es la extracción de conjuntos de artículos frecuentemente asociados, por ejemplo, cuando se trata de artículos que se compran juntos con mucha frecuencia en el supermercado, lo que da a los responsables de marketing oportunidades en el sentido de avanzar lo que más probablemente necesita un cliente o perfil de cliente concreto. La extracción de estas reglas de asociación puede resultar costosa desde el punto de vista computacional, pero es muy frecuente que sirva para maximizar el margen de beneficios de los minoristas o de otros proveedores de servicios. Una posible aplicación dentro del sector turístico puede consistir en la extracción de patrones entre los diferentes productos o servicios turísticos y las demandas de los distintos tipos de turistas, por ejemplo realizando una estrategia de marketing directo (e-mailing) tras la visita para aumentar la fidelización de estos turistas con el destino.

3.1.2. Toma de decisiones.

Los sistemas de toma de decisiones basados en datos son aquellos orientados a ayudar a sus usuarios a tomar mejores elecciones. Entre ellos se encuentran los sistemas de lógica difusa, los sistemas de recomendación y otros que se encuadran en diferentes categorías, por lo que este tipo de sistemas se sitúan a medio camino entre las áreas de modelización y optimización. Aunque la salida de casi cualquier sistema de *Computational Intelligence* puede ser considerada como una ayuda a la toma de decisiones, aquí nos hemos centrado en los sistemas basados en reglas, que no pueden ser catalogados en ninguna de las otras categorías propuestas. Adicionalmente, el sector hotelero es cada vez más consciente de la ventaja competitiva que proporciona el uso de datos en la toma de decisiones, aunque las enormes cantidades de datos (*Big Data*) dificultan este tipo de toma de decisiones (Phillips-Wren y Hoskisson, 2014). Además, en dicho estudio también se afirma que el *Big Data* es cada vez más importante para los líderes de las empresas porque puede estar directamente vinculado a la generación de valor.

3.1.2.1. Sistemas difusos.

En 1965, se definió un conjunto difuso como una clase de objetos con un continuo de grados de pertenencia (Zadeh, 1965). Estos conjuntos se inspiraron en el modo de razonamiento no discreto que tienen los humanos, donde una variable de decisión puede contener valores vagos (difusos) como "bajo", "medio" y "alto". Gracias a este cambio conceptual fundamental, en los últimos 50 años se ha desarrollado un enorme corpus de investigación en muchos campos, especialmente en la ingeniería de control y la robótica. La flexibilidad y la capacidad de incorporar el conocimiento humano a esta forma de modelar dotan a la lógica difusa de una gran importancia dentro del *machine learning*.

3.1.3. Computación natural.

La computación basada en procesos u organismos naturales está muy extendida y se aplica a tareas de optimización, así como al afinado y ajuste de modelos (Eiben y Smith, 2015). Estos métodos se utilizan a menudo para explorar un amplio espacio de soluciones, que se codifica en forma de miembros de una población que evolucionan (computación evolutiva), o interactúan entre ellos (*swarm intelligence*), para resolver el problema de forma eficiente. Esta disciplina informática tiene interesantes aplicaciones dentro del sector hotelero, donde se ha utilizado para optimizar diferentes áreas como la asignación de recursos (consumo de agua y energía) o la gestión de reservas.

3.1.3.1. Computación evolutiva.

En 1992, se sentaron las bases de los algoritmos genéticos (Holland, 1992), la semilla inicial del vasto campo que es hoy la computación evolutiva. Inspirado en la teoría de la evolución de las especies de Charles Darwin, (Holland, 1992) propuso una abstracción en la que las especies eran posibles soluciones combinatorias a un problema muy complejo y, mediante selección, recombinación (*crossover*) y mutación, dichas soluciones evolucionan para adaptarse al entorno, maximizando una función de aptitud previamente definida.

3.1.3.2. Swarm Intelligence.

(Beni y Wang, 1993) definieron la “inteligencia de enjambre”, o *swarm inteligence*, como el comportamiento colectivo de sistemas descentralizados y autoorganizados, naturales o artificiales. Estos sistemas suelen estar formados por una población de agentes simples que interactúan con el entorno y entre sí. Aunque estos agentes siguen reglas simples sin esquemas de control centralizados, tienden a formar un comportamiento global inteligente. Ejemplos de este tipo de sistemas son las bandadas de aves, los rebaños, las colonias de hormigas y abejas o incluso el crecimiento bacteriano.

3.1.4. *Data Stream Mining*.

Es de gran relevancia destacar un ámbito emergente del *machine learning* conocido como *online learning*, y puede definirse como una metodología en la que los datos llegan de manera continua y cambiante en tiempo real, donde los modelos predictivos deben actuar de manera rápida, adaptándose y haciendo uso de una memoria limitada para evitar que su precisión se degrade con el tiempo (Žliobaitė et al., 2014). Es una rama relativamente nueva dentro del *machine learning* en la que los datos se obtienen a partir de un flujo continuo de datos (o *data stream*) mediante minería de flujo de datos, o *Data Stream Mining*, en lugar de analizar grandes bloques de datos. Esto otorga a los modelos la capacidad de adaptarse a los cambios dentro de los propios datos, haciendo a dichos modelos resistentes al paso del tiempo.

Estamos viviendo un cambio social y económico impulsado por los datos. La disponibilidad de datos en todos los aspectos de nuestra vida se ha convertido en una gigantesca fuente de posibilidades, y se hacen grandes esfuerzos para extraer automáticamente el conocimiento de estos. La producción de datos está viviendo una aceleración exponencial, impulsada principalmente por la omnipresencia de la informática (es decir, los dispositivos informáticos personales), el despliegue de redes de sensores y la hiperconectividad global. Es la llamada Internet de Todo, o *Internet of Everything*, (Bradley et al., 2013), y produce *Big Data* acorde a la regla de las 5V: volumen, velocidad, variedad, veracidad y valor, como enuncia Anuradha (2015) y Demchenko et al. (2013).

Cuando se trata de volumen y velocidad en particular, está claro que hay un gran cuello de botella por delante. Cada vez es más impracticable almacenar todos los datos producidos para extraer de ellos el conocimiento en un momento posterior debido al gran tamaño que pueden llegar a alcanzar. Es necesario desarrollar nuevas metodologías *online,* desarrollar todo un nuevo ámbito de literatura de *machine learning* para construir modelos en tiempo real capaces de extraer conocimiento a medida que llegan nuevas observaciones de datos de forma incremental y adaptativa.

El *Data Stream Mining* (Bifet y Kirkby, 2009; Gama y Gaber, 2007), un nuevo enfoque dinámico de la minería de datos, es la evolución natural del *machine learning* y la minería de datos bajo la presión de la escala del *Big Data* y, lo que es más importante, de la obsolescencia de los conceptos modelados durante el proceso de adquisición de datos. Este nuevo enfoque tiene muchas ventajas, como la reducción en los recursos computacionales necesarios o la reducción del retraso en el aprendizaje y la ejecución de los modelos predictivos.

En el *Data Stream Mining*, un concepto central que explica su propia necesidad es el *concept drift* (Schlimmer y Granger, 1986; Widmer y Kubat, 1996), que se define como un cambio en la distribución de probabilidad de la variable modelada en el flujo de datos recibido. Más formalmente, la deriva de conceptos entre el punto de tiempo *t* y el punto de tiempo *t+1* ocurre cuando la desigualdad de la ecuación (1) es verdadera.

$$\exists t: p_t(X, y) \neq p_{t+1}(X, y) \quad (1)$$

En la ecuación (1), p_0 y p_1 denotan la distribución conjunta en los tiempos *t* y *t+1*, respectivamente, entre el conjunto de variables de entrada X y la variable objetivo *y*.

En otras palabras, el *concept drift* aparece cuando hay no-estacionariedad estocástica en el fenómeno que se está modelando. Por tanto, puede detectarse mediante métodos estadísticos directos (media, varianza y autocovarianza) o mediante métodos prácticos indirectos, como la observación de un empeoramiento estadísticamente significativo del rendimiento de un modelo previamente entrenado.

3.2. El sector turístico y la Inteligencia Computacional.

A la hora de realizar una investigación exhaustiva sobre el estado del arte referente a la aplicación de *Computational Intelligence* en el sector turístico, se analizaron 180 artículos de diversos editores y revistas dentro de la *Core Collection* de la *Web of Science*, utilizando las siguientes *keywords:* "*Hospitality Industry*", "*Tourism*", "*Data Stream Mining*", "*Machine Learning*", "*Computational Intelligence*" y/o "*Deep Learning*", además de filtrar los resultados mediante una heurística de "Búsqueda por tema", la cual busca las *keywords* previamente mencionadas dentro de los *abstracts*, títulos y/o *keywords* de cada artículo. Adicionalmente, se acotó el periodo de búsqueda entre los años 1998 a 2020 para obtener el mayor número de referencias posibles y de esta forma realizar una buena base de datos para el estudio en cuestión.

Tras obtener los resultados de la búsqueda anteriormente definida, se utilizó una herramienta de gestión de bibliografías, concretamente Zotero, con el objetivo de reunir todas las referencias y de proporcionar datos acerca de las mismas (por ejemplo, el número de artículos publicados en un mismo año). Tras esta organización inicial se utilizó una metodología de observación directa y lectura para así analizar, de manera profunda y detallada, todos y cada uno de los artículos con el objetivo de encontrar y clasificar las metodologías de *Computational Intelligence* aplicadas en cada uno de ellos, además de su aplicación sobre las determinadas subáreas dentro del sector turístico y hotelero, así como los resultados obtenidos.

En los últimos años, el uso de métodos basados en la *Computational Intelligence* aplicados a la industria hotelera también ha experimentado un aumento considerable, dando resultados positivos en diversas áreas como la previsión de la demanda turística o el consumo de energía (Chen, Tan, y Berardi, 2017; Kamel et al., 2018). El uso de estas técnicas no sólo genera mayores márgenes de beneficio para la industria, sino que también aumenta la calidad de los servicios ofrecidos a los turistas y visitantes. La elaboración de perfiles y la categorización de los usuarios, la personalización de los servicios para adaptarse a cada perfil de cliente diferente son buenos ejemplos de ello. Sin embargo, el uso de la analítica de datos predictiva con *Big Data* para obtener conocimientos a partir de datos en tiempo real en lugar de datos históricos es todavía un nuevo paradigma dentro de la industria hotelera y turística (Bernard, 2016).

La *Computational Intelligence* es un campo muy amplio que ha alcanzado a casi todas las áreas de conocimiento e investigación. No existe un consenso claro sobre cómo realizar una taxonomía precisa de las técnicas y métodos dentro de este campo, con una gran variedad de enfoques (Pedrycz y Peters, 1998; Wang, 2012), y el foco puesto en diferentes características, lo que motivó el desarrollo de una nueva taxonomía para tratar de clasificar las familias de técnicas de *machine learning* dentro de la *Computational Intelligence* más representativas desde el punto de vista de la industria hotelera. Se puede hacer una división importante entre los métodos de la *Computational Intelligence* en tres categorías principales: modelización, optimización y simulación (Eiben y Smith, 2015). En esta división, cada categoría trata de despejar la incógnita en una ecuación en la que un elemento del esquema datos de entrada-modelo-salida es la incógnita. Sin embargo, el corpus relevante de la literatura relacionada con la *Computational Intelligence* en el sector hotelero gira en torno a la modelización y la optimización. Por lo tanto, resultó ideal el destacar las técnicas que se han encontrado en el corpus bibliográfico estudiado, y cómo, en varios casos, se utilizan metodologías híbridas entre temas. Así, la cobertura de esta clasificación se reduce a los tipos de técnicas y métodos que se han encontrado relevantes para este sector partiendo de una ontología basada en dicha división. La Figura 1 presenta la taxonomía de las diferentes técnicas de la *Computational Intelligence*.

En esta taxonomía, las metodologías de *Computational Intelligence* se han clasificado atendiendo a sus principales objetivos, que son el modelado y la optimización. El modelado de *machine learning* se refiere al descubrimiento de las relaciones entre las entradas y salidas de datos. Por otro lado, la optimización del *machine learning* se centra en encontrar qué entradas maximizan o minimizan la salida de un modelo de *machine learning*. Por último, esta taxonomía da lugar a una clasificación más detallada de los trabajos de investigación estudiados en una de las categorías o subcategorías de la *Computational Intelligence*.

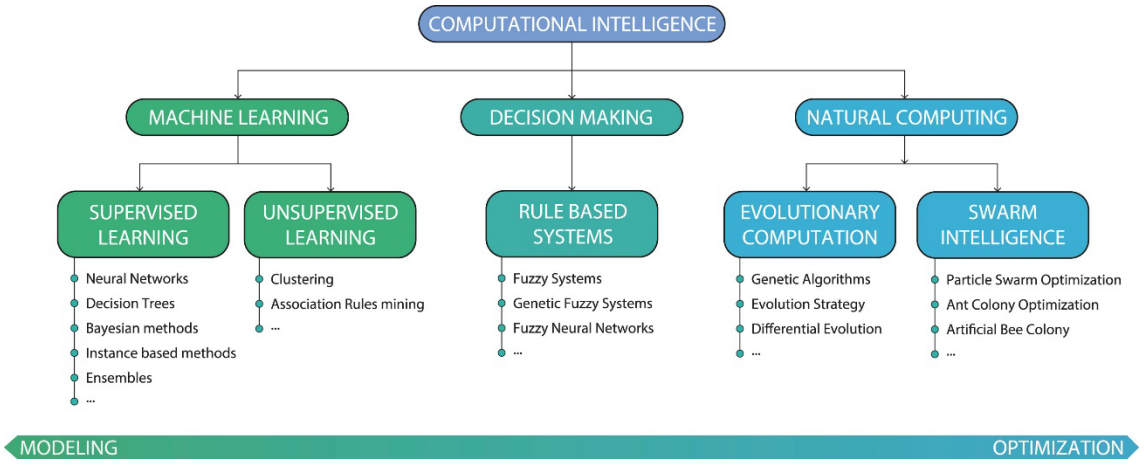


Figura 1 – Taxonomía del machine learning, toma de decisiones y métodos de computación naturales (Guerra-Montenegro et al., 2021).

La cuantificación y distribución porcentual de las técnicas utilizadas en el sector hotelero realizada por Guerra-Montenegro et al. (2021) se representa visualmente en la Figura 2, donde se muestra que, dentro de la *Computational Intelligence*, el *machine*

learning es el conjunto de metodologías más utilizado en dicho sector. Dentro de este ámbito, las metodologías más utilizadas parecen ser los métodos probabilísticos/bayesianos y los métodos de instancia/lineales porque son metodologías clásicas que llevan más tiempo presentes en este ámbito. Sin embargo, las ANN también están empezando a surgir como una metodología de uso frecuente en este campo de investigación, especialmente el *deep learning*, que está jugando un papel notable en las soluciones de *Big Data* por su capacidad de cosechar conocimiento de sistemas complejos (Zhang et al., 2018). La computación evolutiva también se está utilizando en el sector (aunque en menor grado) para resolver diferentes problemas de optimización como la asignación de recursos y la previsión, que son solo algunas de las principales dificultades dentro del panorama hotelero y turístico. Un buen ejemplo de ello es el consumo de energía, un área en la que los establecimientos hoteleros luchan frecuentemente por adaptarse debido a su necesidad de métodos de gestión energética eficientes, necesarios para garantizar su rendimiento y sostenibilidad (Casteleiro-Roca et al., 2019).

Por último, los sistemas difusos y basados en reglas se están utilizando en menor medida en el sector hotelero para predecir diversos tipos de problemas, aunque no suelen ser tan versátiles como los algoritmos de *machine learning*.

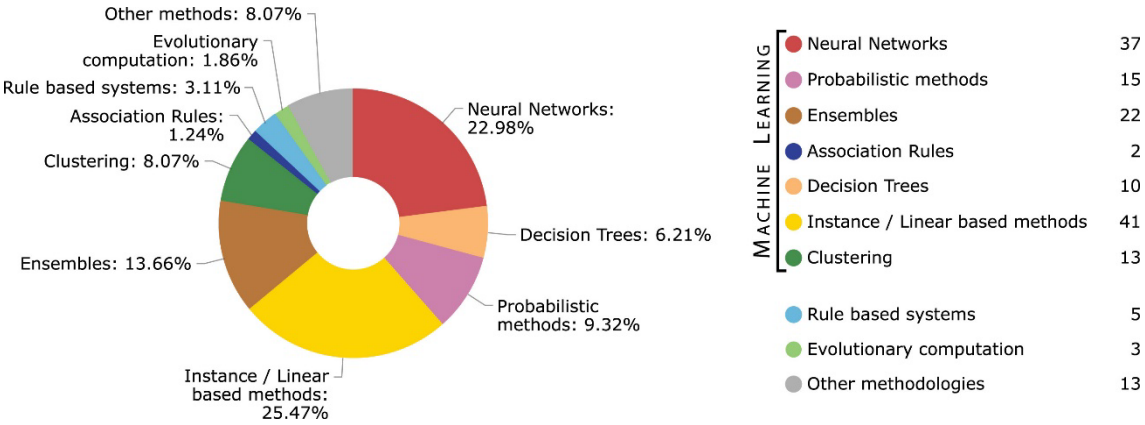


Figura 2 - Distribución de las metodologías revisadas (Guerra-Montenegro et al., 2021).

En la revisión bibliográfica llevada a cabo a la hora de realizar la presente tesis doctoral, se ha creado una categorización de la literatura de la industria turística y hotelera (Guerra-Montenegro et al., 2021). Esta taxonomía es el resultado del análisis basado en el uso de las aplicaciones de *Computational* Intelligence, que puso de manifiesto la necesidad de una categorización en este sector. La Figura 3 muestra dicha categorización del estado del arte, que se divide en cuatro grandes bloques, con 10 subcategorías en total. Los cuatro grandes bloques de la nueva categorización son:

- **Gestión y estimación de ingresos**, que comprende todo lo referente a la gestión de los hoteles, la asignación de recursos, los asuntos relacionados con el mercado y la gestión de los ingresos.

- **Sistemas de perfilado y recomendación**, que agrupa todo lo relacionado con el perfilado y análisis de los clientes/turistas y con los sistemas de recomendación.
- **Previsión de la demanda turística**, que es un área enorme en sí misma, y se centra en la previsión de la demanda, desde la asignación de habitaciones hasta los patrones de ocupación estacional.
- **Previsión meteorológica y evaluación de riesgos medioambientales**, un área que presenta trabajos de investigación sobre predicciones meteorológicas y climáticas y posibles riesgos medioambientales, todo ello aplicado a la industria turística y hotelera.

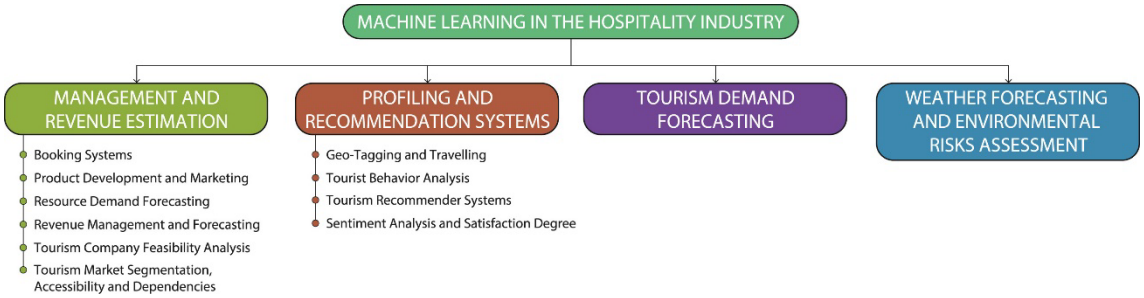


Figura 3 - Categorización propuesta para el estado del arte (Guerra-Montenegro et al., 2021).

Además, la Figura 4 muestra la relación entre esta nueva categorización y el estado de las técnicas, lo que ofrece una visión útil sobre las metodologías y su grado de aplicación en una determinada área de investigación.

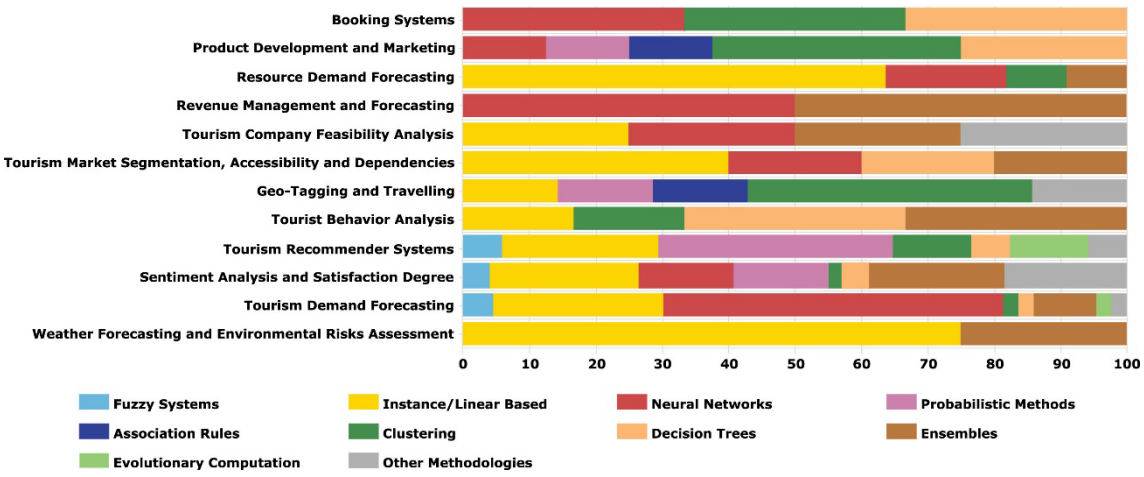


Figura 4 - Distribución de la literatura revisada por área y por técnica de Inteligencia Computacional aplicada (Guerra-Montenegro et al., 2021).

Para mostrar de manera correcta cómo los métodos de *Computational Intelligence* están mejorando el sector turístico y hotelero es de obligada importancia el cruzarlos con el área donde se aplican.

Aunque podría haber metodologías igualmente válidas para realizar un trabajo de estas características, se ha utilizado esta porque un barrido exhaustivo del área de

investigación es posiblemente la estrategia más adecuada para recopilar y no dejar de lado ninguna aportación valiosa de la investigación. Con una cantidad tan grande de publicaciones en el área analizada, podría ser bastante fácil pasar por alto trabajos de investigación que ofrezcan perspectivas de gran valor, creando así la necesidad de un *review* que analice en profundidad la gran mayoría de los trabajos presentes en el campo.

Para crear dicho *review*, se ha realizado una búsqueda profunda con el objetivo de encontrar todos los trabajos de investigación relacionados con la industria de la hotelera y turística y la *Computational Intelligence*. Combinando palabras clave como *"Hospitality Industry", "Data Stream Mining", "Tourism", "Online Learning", "Computational Intelligence"* o *"Deep Learning",* junto con una heurística de búsqueda por temas (título, resumen y/o palabras clave) y un rango de fechas entre 1998 y 2020, se obtuvo una gran cantidad de resultados de la *Core Collection* de la *Web of Science*. Tras estos resultados, un examen exhaustivo de todos esos trabajos de investigación permitió descubrir qué contribuciones entraban en el ámbito de este *review*.

Este proceso dio como resultado más de ciento sesenta trabajos que se relacionan directamente con ciertos métodos de *Computational Intelligence*, con una aplicación en una o más de las cuatro áreas principales identificadas. En la Tabla 1 se muestran todas las referencias analizadas con el objetivo de que dicha tabla sirva de guía y orientación tanto a investigadores experimentados como a nuevos investigadores que deseen aportar a la *Computational Intelligence*, al sector turístico y hotelero, y a la combinación de ambos. Por un lado, permite a la comunidad de la *Computational Intelligence* conocer qué métodos se han utilizado mayoritariamente en cada área del sector, detectando nichos de oportunidad en torno a un determinado método de *Computational Intelligence* que permanece inexplorado en cada área. Para la comunidad investigadora de este sector, esta tabla ayudará a discriminar qué metodologías de *Computational Intelligence* son las más utilizadas para cada subtema, o incluso a medir cualitativamente hasta qué punto se ha aplicado la *Computational Intelligence* en su área específica de interés.

Adicionalmente, en la Tabla 2 se presenta una lista exhaustiva de las áreas y subáreas del sector identificadas, los casos típicos de uso, las metodologías de *Computational Intelligence* comúnmente utilizadas, los problemas no resueltos y las futuras líneas de investigación. Esta tabla pretende ofrecer un análisis comparativo tanto de los métodos de *Computational Intelligence* como de los asuntos relacionados con esta industria, mostrando cómo se están aplicando estas metodologías en las áreas de investigación de la industria hotelera y turística identificadas en este *review*.

Esta tabla se ha realizado mediante un análisis a fondo de cada trabajo de investigación, identificando la información clave que contiene. En primer lugar, se categorizó el tema de la industria turística/hotelera en el que se encontraba y el problema que se abordaba con una determinada solución basada en metodologías de *Computational Intelligence*. En segundo lugar, se realizó un barrido general para identificar correctamente los métodos de *Computational Intelligence* más comunes

utilizados en cada área, junto con los problemas o dificultades no resueltos que dicha área presenta en términos de investigación o ejecución. Por último, se mencionan posibles vías de investigación para ayudar tanto a los expertos en *Computational Intelligence* como a los interesados en el sector a orientar futuros estudios o aplicaciones de metodologías basadas en *Computational Intelligence* o casos de uso dentro de este sector. Por todo ello, el objetivo principal de esta tabla es mostrar la unión entre ambas disciplinas, así como poder descubrir las posibles lagunas y oportunidades futuras de investigación.

Tabla 1 - Revisión bibliográfica de las categorías de Machine Learning, cruzadas con las principales áreas de la industria hotelera y turística.

Paradigma	Familia	Técnica	Áreas			
			Gestión y estimación de ingresos	**Sistemas de perfilado y recomendación**	**Previsión de demanda turística**	**Previsión meteorológica y evaluación de riesgos medioambientales**
Modelado	Aprendizaje Supervisado	Redes Neuronales Artificiales	(Cho y Leung, 2002), (Chen et al., 2010), (Bettin et al., 2011), (Gayar et al., 2008), (Kim, 2011), (Chou Jui-Sheng y Lin Chieh, 2013), (Huang et al., 2013), (Kampouropoulos et al., 2014), (Xue-Bo y Shi-Ting, 2014), (Claveria, Monte, et al., 2016b), (Kofinas et al., 2016), (Lu, 2017), (Porto y Irigoyen, 2017), (Atsalakis et al., 2018), (Al Shehhi y Karathanasopoulos, 2020)	(Claster et al., 2010), (Sharma y Dey, 2012), (Phillips et al., 2015), (Yang et al., 2015), (Bugarski et al., 2017), (Nilashi et al., 2017), (Nilashi, Ahani, et al., 2019), (Cheng et al., 2019), (Chang et al., 2020), (Luo et al., 2020), (Ren et al., 2020), (Shoukry y Aldeek, 2020)	(Kamel et al., 2018), (Xue-Bo y Shi-Ting, 2014), (Pai et al., 2005), (Chen et al., 2012), (Berenguer et al., 2014), (Cankurt y Subasi, 2015), (Claveria et al., 2015), (Wang et al., 2015), (Claveria, Monte, et al., 2016a), (Claveria, Torra, et al., 2016), (Noersasongko et al., 2016), (Sun y Chang, 2016), (Antonio et al., 2017), (Chang y Tsai, 2017), (Claveria et al., 2017), (Folgieri et al., 2017), (Han et al., 2017), (Oger Vihikan et al., 2017), (Rafidah y Ani, 2017), (S. Sun et al., 2017)	(King et al., 2014)
		Árboles de decisión	(Cho y Leung, 2002), (Chou Jui-Sheng y Lin Chieh, 2013), (Lu, 2017), (Emel y Taşkın, 2005), (Ha y Park, 1998)	(Min et al., 2002), (Guoxia y Jianqing, 2009), (Zhang et al., 2011), (Luberg et al., 2012), (Zhang y Zhang, 2014), (Banerjee et al., 2015), (Nakamura et al., 2015), (Kbaier et al., 2017), (Yordanova y Kabakchieva, 2017), (Nilashi, Yadegaridehkordi, et al., 2019)	(Kamel et al., 2018), (Antonio et al., 2017), (Cankurt, 2016), (Akın, 2015)	(King et al., 2014)
		Modelos probabilísticos / estadísticos / bayesianos	(Deng y Li, 2018)	(Kamel et al., 2018), (Claster et al., 2010), (Nilashi, Ahani, et al., 2019), (Zhang et al., 2011), (Nakamura et al., 2015), (Ye et al., 2009), (Weichselbraun et al., 2010), (Shimada et al., 2011), (Hsu et al., 2012), (Wang et al., 2012), (Xu y Zhao, 2012), (Chang y Ma, 2013), (Duan et al., 2013), (Sixto et al., 2013), (Namahoot et al., 2015), (Zhao et al., 2015), (Arruza et al., 2016), (Ebadi, 2016), (Li et al., 2016), (Saputro et al., 2016), (Yuan et al., 2016), (Ren y Hong, 2017), (Belmonte-Fernández et al., 2018), (Hsu et al., 2009), (Guo et al., 2017), (Yi-Chung, 2017), (Haruechaiyasak et al., 2010), (García-Barriocanal et al., 2010), (Bermingham y Lee, 2014), (Li et al., 2015), (Ma et al., 2018), (Sanchez-Franco, Cepeda-Carrion, et al., 2019), (Sanchez-Franco, Navarro-Garcia, et al., 2019)	(Claveria, Monte, et al., 2016b), (Claveria, Torra, et al., 2016), (Claveria et al., 2017), (Wu et al., 2012), (Lei y Lam, 2015)	
		Ensembles	(Lu, 2017), (Belmonte-Fernández et al., 2018)	(Banerjee et al., 2015), (Brida et al., 2018), (Athanasiou y Maragoudakis, 2016)	(Antonio et al., 2017), (Cankurt, 2016), (Pai et al., 2014)	(King et al., 2014)

Paradigma	Familia	Técnica	Áreas			
			Gestión y estimación de ingresos	**Sistemas de perfilado y recomendación**	**Previsión de demanda turística**	**Previsión meteorológica y evaluación de riesgos medioambientales**
Modelado		Modelos basados en instancias	(Chen, Tan, y Berardi, 2017), (Cho y Leung, 2002), (Gayar et al., 2008), (Kim, 2011), (Chou Jui-Sheng y Lin Chieh, 2013), (Lu, 2017), (Yang et al., 2015), (Belmonte-Fernández et al., 2018), (Shi-Ting y Bo, 2014), (Yang et al., 2016), (Zhang et al., 2018), (Hsu et al., 2006), (Romero Morales y Wang, 2010), (Hong et al., 2013), (Lin et al., 2013), (Wang et al., 2014), (Cao y Wu, 2016), (Wang et al., 2016), (Chen, Tan, y Song, 2017), (Li y Sun, 2012)	(Nilashi et al., 2017), (Zhang et al., 2011), (Luberg et al., 2012), (Banerjee et al., 2015), (Ye et al., 2009), (Wang et al., 2012), (Xu y Zhao, 2012), (Zhao et al., 2015), (Arruza et al., 2016), (Li et al., 2016), (Li et al., 2009), (Xia y Peng, 2009), (Zheng y Ye, 2009), (Lin y Chao, 2010), (Shi y Li, 2011), (Yao et al., 2011), (Kasper y Vela, 2012), (Lu et al., 2012), (Tokuhisa et al., 2012), (Jiang et al., 2013), (Hsieh et al., 2014), (Chiu et al., 2015), (Dickinger y Mazanec, 2015), (Sun et al., 2015), (Xiang et al., 2017), (Martin-Fuentes et al., 2018), (Gawlik et al., 2011), (Kumar Duvvur y Romanowski, 2016)	(Kamel et al., 2018), (Claveria, Monte, et al., 2016b), (Pai et al., 2005), (Claveria, Monte, et al., 2016a), (Antonio et al., 2017), (Han et al., 2017), (Rafidah y Ani, 2017), (Cankurt, 2016), (Akın, 2015), (Pai et al., 2014), (Xu et al., 2009), (Cai et al., 2009), (Hong et al., 2011), (Cankurt y Subaşi, 2016), (Chen et al., 2015), (Antonio et al., 2016), (Lijuan y Guohua, 2016), (Liang y Bi, 2017), (Zhang et al., 2017), (Chen y Wang, 2007), (Ali et al., 2017), (Liu y Yao, 2017)	(King et al., 2014), (Gokaraju et al., 2011), (Xu et al., 2016), (Xiao, 2012)
	Aprendizaje no supervisado	*Clustering*	(Chen, Tan, y Berardi, 2017), (Ha y Park, 1998), (Lin y Huang, 2009), (Liao et al., 2010), (Pitman et al., 2010)	(Nilashi et al., 2017), (Ebadi, 2016), (Vu et al., 2015), (Miah et al., 2017), (Gavalas y Kenteris, 2011), (Jiang et al., 2011), (Peng y Huang, 2017)	(Pai et al., 2014), (Shahrabi et al., 2013)	
		Minería por reglas de asociación	(Liao et al., 2010)	(Lin y Chao, 2010), (Junping et al., 2008), (Liu et al., 2013), (Lucas et al., 2013), (Versichele et al., 2014)		
Optimización	Toma de decisiones	Sistemas difusos	(Atsalakis et al., 2018)	(Biuk-Aghai et al., 2008), (Afzaal et al., 2016)	(Pai et al., 2014), (Sakhuja et al., 2016)	
	Computación Evolutiva	Algoritmos Evolutivos	(Kampouropoulos et al., 2014), (Hsu et al., 2006), (Hong et al., 2013), (Wang et al., 2014), (Cao y Wu, 2016)	(Lin et al., 2013), (Biuk-Aghai et al., 2008), (Chen et al., 2013)	(Noersasongko et al., 2016), (Pai et al., 2014), (Cai et al., 2009), (Hong et al., 2011), (Chen et al., 2015), (Lijuan y Guohua, 2016), (Chen y Wang, 2007), (Shahrabi et al., 2013), (Sakhuja et al., 2016), (Hadavandi et al., 2011)	(Xiao, 2012)

Tabla 2 - Áreas del sector turístico y hotelero identificadas, cruzadas con los casos de uso más comunes, las metodologías Computational Intelligence más frecuentes, las limitaciones y o problemas conocidos y posibles líneas de investigación.

Área	Subárea	Caso de uso habitual	Soluciones CI más comunes	Problemas no resueltos	Posibles líneas de investigación
GESTIÓN Y ESTIMACIÓN DE INGRESOS	Sistemas de reserva	Gestionar reservas en un establecimiento. Mejorar sistemas de reserva. Aumentar la velocidad del sistema de reserva.	*Clustering.* Árboles de decisión. ANN	Tiempos de entrenamiento de modelo altos. Lento si se compara con un agente humano.	Metodologías CI más rápidas y reactivas para generar modelos más rápidos. Utilización de o*nline learning* para mejorar la adaptabilidad y la velocidad del sistema.
	Desarrollo y marketing del producto	Desarrollo de nuevos productos. Mejorar decisiones de marketing. Obtención de *customer knowledge.*	*Clustering.* Árboles de decisión.	Tamaño de los datos almacenados. Baja disponibilidad de los datos. Anonimidad de los datos.	Recolección de datos anónima e inocua. Métodos de seguimiento para recoger datos de turistas de manera anónima. El *online Learning* reduciría notablemente el tamaño de los datos almacenados.
	Previsión de la demanda de recursos	Previsión de carga eléctrica. Previsión de consumo de agua. Previsión de consumo de gas.	ANN Métodos lineales/basados en instancias.	Adaptabilidad de los sistemas.	Utilización de *online Learning* para crear sistemas adaptables a largo plazo.
	Gestión y previsión de ingresos	Gasto de clientes. Predicción de ingresos. Previsión de precios.	ANN. *Ensembles.*	Varianza de valores. Dificultad para predecir los ingresos en base a la demanda.	Sistemas híbridos de predicción (Ingresos + demanda turística). Sistemas estacionales de predicción.
	Análisis de viabilidad de empresas turísticas	Predicción de bancarrotas. Gestión de reclamaciones y disputas de litigios. Previsión de servicios recientemente lanzados.	ANN. Métodos lineales/basados en instancias.	Nicho de investigación por explotar.	Experimentar con diferentes metodologías basadas en CI para aumentar el conocimiento disponible en esta área.
	Segmentación del mercado turístico, accesibilidad y dependencias	Identificar segmentos de mercado. Evaluación de la ubicación de nuevos hoteles. Identificar las relaciones entre el precio del hotel y la accesibilidad del mercado.	Métodos lineales/basados en instancias.	Nicho de investigación por explotar.	Experimentar con diferentes metodologías basadas en CI para aumentar el conocimiento disponible en esta área.

Área	Subárea	Caso de uso habitual	Soluciones CI más comunes	Problemas no resueltos	Posibles líneas de investigación
SISTEMAS DE PERFILADO Y RECOMENDACIÓN	Geoetiquetado y viajes	Descubrir establecimientos o atracciones turísticas populares. Ranking de establecimientos. Buscar patrones en la conducta de los viajeros.	*Clustering.*	Modelos de predicción estáticos. No se han aplicado suficientes metodologías basadas en CI como para explorar esta área.	Explorar el uso de algoritmos de *clustering* más eficientes. Aplicar diferentes metodologías de CI para comprobar su efectividad. Usar *online Learning* para crear modelos más estables.
	Análisis de la conducta turística	Predicción de lealtad por parte del cliente. Buscar factores que influyan al cliente. Analizar la distribución de los gastos realizados por el cliente.	Árboles de decisión. *Ensembles*.	Complejidad de los datos. Los datos suelen contener "ruido".	Buscar métodos de preprocesado que "limpien" el ruido de los datos. Aplicar *deep learning* para sobreponerse a la alta complejidad de los datos. Aplicar *online Learning* para crear modelos adaptables al paso del tiempo.
	Sistemas de recomendación turística	Recomendar diferentes atracciones o establecimientos. Crear recomendaciones personalizadas. Crear rutas turísticas basadas en las preferencias del turista.	Métodos probabilísticos.	Gran tamaño de los datos. Velocidad de ejecución de algoritmos lenta.	Optimización de parámetros. Metodologías basadas en *online Learning* podrían reducir el tamaño de los datos y crear modelos adaptivos.
	Análisis de sentimiento y grados de satisfacción	Extraer datos de sentimiento a partir de *reviews*, *blogs,* redes sociales etc. Análisis de imágenes publicadas por clientes en diferentes plataformas. Detección de *reviews* y comentarios falsos.	Métodos lineales/basados en instancias. *Ensembles.*	Tamaño de datos gigantesco. Tiempo de entrenamiento de modelos alto.	Optimización de métodos conocidos. Aumentar la generación de modelos mediante *deep learning*. Metodologías basadas en *online Learning* podrían reducir el tamaño de los datos.
PREVISIÓN DE LA DEMANDA TURÍSTICA	Previsión de demanda turística	Preparar servicios para una carga determinada. Prevenir la sobrecarga de servicios. Predecir las llegadas de turistas. Prever cancelaciones de reservas.	ANN	Área intensamente investigada. Las tendencias de los datos suelen cambiar con frecuencia.	Exploración de metodologías híbridas y *ensembles*. Utilizar *online Learning* para crear algoritmos más versátiles.

Área	Subárea	Caso de uso habitual	Soluciones CI más comunes	Problemas no resueltos	Posibles líneas de investigación
PREVISIÓN METEOROLÓGICA Y EVALUACIÓN DE RIESGOS MEDIOAMBIENTALES	Previsión meteorológica y evaluación de riesgos medioambientales	Predecir ciertas condiciones climáticas o riesgos medioambientales extremos. Planificación de eventos.	Métodos lineales/basados en instancias	Los modelos predictivos quedan obsoletos rápidamente. No se han aplicado suficientes metodologías basadas en CI como para explorar esta área.	Experimentar con diferentes metodologías basadas en CI para aumentar el conocimiento disponible en esta área.

A continuación, se realiza un resumen de cada uno de los bloques identificados (véase Tabla 2) junto con las referencias de los trabajos de investigación revisados con el fin de proporcionar una descripción detallada del estado de la técnica en la aplicación de la *Computational Intelligence* en la industria turística y hotelera.

3.2.1. Gestión y estimación de ingresos.

La industria hotelera y turística está comenzando a explotar gradualmente los beneficios del uso de metodologías de *machine learning* y minería de datos para mejorar sus servicios, o incluso para crear otros nuevos basados en ellos. En esta sección expondremos varios servicios e infraestructuras en los que se están aplicando, a saber: los sistemas de reserva, que facilitan el procedimiento de reserva en sus hoteles; el desarrollo de productos y el marketing, en los que se aplican métodos de *machine learning* para mejorar las decisiones empresariales; la previsión de la demanda de recursos, en la que recursos como la energía deben asignarse de forma inteligente; la gestión y previsión de los ingresos, donde los métodos de *machine learning* se utilizan para predecir y gestionar diferentes aspectos de los ingresos económicos, lo cual es esencial en un mundo en el que cualquier imprevisto puede dar lugar a una cancelación; el análisis de viabilidad de las empresas turísticas, que es de suma importancia para empezar a competir en un mercado en disputa y, por último, la segmentación, la accesibilidad y las dependencias del mercado turístico, donde la *Computational Intelligence* puede utilizarse para identificar diversos parámetros en la distribución del mercado y las oportunidades de acceso.

En esta categoría se explicará con mayor detalle cómo la *Computational Intelligence* es aplicada con el objetivo de predecir ingresos económicos o de gestionarlos.

3.2.1.1. Sistemas de reserva.

En la actualidad, con un mercado global tan grande, es muy importante para las empresas hoteleras proporcionar a los clientes potenciales servicios de reserva en línea, como indican Bettin et al. (2011), donde también se expone que el software utilizado para apoyar la reserva de los clientes debe ser también una "guía" para encauzar las preferencias de los clientes desde la fase inicial en la que dicho cliente establece sus propios requisitos preliminares o elige los servicios, haciendo que la reserva en línea se ocupe de los instrumentos estratégicos para perseguir dos aspectos relevantes del mercado: la fidelidad y captación de los clientes.

Aunque no está centrado en los sistemas de reserva, se describió un enfoque de *clustering* mediante el uso de *k-means* para extraer la inteligencia empresarial oculta en los datos de registro de la web, almacenándolos en un formato de base de datos para procesarlos mediante el uso de *x-means* (una variación de *k-means clustering*), revelando una diferencia significativa entre las necesidades de los clientes y demostrando así que esta técnica es útil para una descripción precisa de las necesidades de información de los clientes que utilizan el sitio web de una empresa turística (Pitman et al., 2010). Un año después, se estudió el uso de ANN aplicadas a los sistemas de reserva online, utilizando un perceptrón multicapa como herramienta de inferencia para aproximar una solución de reserva adecuada y rápida para los clientes que quieren reservar una habitación dentro de un hotel (Bettin et al., 2011). El modelo que utilizaron fue entrenado para resolver automáticamente la asignación de habitaciones en función del número de habitaciones desocupadas y del número de clientes que reservan una habitación. En este trabajo de investigación también se menciona que una RBF podría haber representado una buena alternativa a su metodología. Por último, se aplicaron diferentes metodologías de *machine learning* para pronosticar los precios de los billetes de avión (Lu, 2017), descubriendo que los árboles de decisión *AdaBoost* obtienen un rendimiento satisfactorio frente a la regresión por mínimos cuadrados, la regresión logística, las ANN, los árboles de decisión, los bosques aleatorios y *nearest neighbors*.

En este ámbito, la aplicación de metodologías de *machine learning* está encontrando una limitación en términos de temporalización comercial, ya que el tiempo necesario para entrenar ANN de perceptrón multicapa es notablemente elevado. Para estos casos, una metodología *machine learning* debe ser lo suficientemente rápida como para proponer una solución de reserva adecuada para el cliente en un corto periodo de tiempo con el objetivo de ser más eficaz en comparación a un operador humano.

No hay muchos trabajos relacionados con el uso de metodologías de *machine learning* en este ámbito, pero ha resultado ser un nicho de investigación muy interesante en el que se podría aplicar la previsión basada en *machine learning* para descubrir nuevos servicios que se puedan ofrecer al cliente en tiempo real, mientras se realiza el proceso de reserva.

3.2.1.2. Desarrollo y marketing del producto.

En el sector hotelero no basta con tener un buen producto, sino que hay que mejorarlo a lo largo del tiempo para mantener la ventaja competitiva y fidelización de los clientes. En un entorno tan hostil, hay que aprovechar las oportunidades de alto rendimiento para obtener una ventaja determinante. Se enunció que los directivos de las empresas tienen que explotar los datos generados y recogidos diariamente por sus organizaciones y transformarlos en información y conocimientos útiles de forma automática e inteligente (Ha y Park, 1998). Con el rápido desarrollo de la industria de Internet móvil, los terminales inteligentes personales se han utilizado ampliamente, lo que hace que todo tipo de información en la red crezca exponencialmente. También fue descubierto que la tecnología de minería de datos, que puede utilizarse para obtener información valiosa, está en constante desarrollo (Yuan et al., 2016). Esto permite obtener información de los clientes para mejorar los servicios que ofrece un establecimiento hotelero, lo que se traduce en mejores críticas de los turistas y en servicios renovados.

Uno de los primeros ejemplos en los que se aplicó la *Computational Intelligence* fue en el desarrollo de un algoritmo de *clustering* utilizando un mapa autoorganizado con el fin de descubrir el conocimiento oculto dentro del *Big Data* generado por una tienda *duty free* de un hotel, aplicando este conocimiento a las decisiones de marketing (Ha y Park, 1998). Esto también fue realizado mediante el uso de diferentes técnicas como ANN, árboles de decisión y *nearest neighbor*, junto con otros métodos de aprendizaje estadístico (Cho y Leung, 2002).

Asimismo, la minería y el análisis del conocimiento de los clientes puede conducir al desarrollo de nuevos productos turísticos (Liao et al., 2010), aplicándose el análisis de *clustering* para generar reglas de asociación junto con el algoritmo apriori para la minería de datos de información con el fin de extraer conocimientos útiles sobre el conocimiento de los clientes. También se ha utilizado un interesante enfoque de doble aprendizaje para minar las preferencias de los turistas utilizando C4.5 y árboles de decisión (Zhang y Zhang, 2011). Adicionalmente, se ha aplicado *Computational Intelligence* a la minería de datos de seguimiento de *bluetooth*, permitiendo obtener patrones de atracción turística mediante reglas de asociación (Versichele et al., 2014). Por último, se ha utilizado una versión mejorada de una técnica de *machine learning* conocida como *Latent Dirichlet Allocation* (LDA) para perfilar la actividad turística junto con el descubrimiento de nuevas tendencias (Yuan et al., 2016).

Uno de los problemas habituales en este ámbito es el tamaño de los datos almacenados y el rendimiento de los algoritmos de minería de datos. También es habitual tener datos sin clasificar, conjuntos de datos pequeños o, debido a las leyes de protección de datos, ni siquiera tener datos de entrenamiento. Además, tal como se vio previamente, el uso de metodologías de rastreo para obtener datos tiene ciertos límites en cuanto a la colaboración de los usuarios, lo que significa que la distribución de los dispositivos de registro necesita de la colaboración de los individuos rastreados (Versichele et al., 2014). Las posibles nuevas líneas de investigación podrían incluir

formas de recopilar datos de los usuarios de forma anónima e inocua para crear conjuntos de datos útiles que podrían utilizarse para mejorar esta área de investigación. Esto podría dar lugar a modelos de previsión más precisos que llevarían a un mejor desarrollo de productos y a estrategias de marketing más precisas, aumentando la satisfacción de los clientes y los ingresos.

3.2.1.3. Previsión de la demanda de recursos.

Recursos como la energía y el agua son de suma importancia en el sector hotelero y turístico debido a su alto valor estratégico. Ser capaz de predecir su demanda puede suponer no sólo la prevención de la escasez de agua o electricidad, sino también un aumento de los recursos disponibles que no se van a utilizar, reservándolos para una ocasión más adecuada. En la actualidad, el *machine learning* se utiliza para prever varios tipos de demanda de recursos tales como la electricidad, el agua, el gas e incluso el tráfico.

Existen diversas publicaciones científicas que avalan este hecho. Por ejemplo, las SVM han resultado ser idóneas para la previsión de la carga eléctrica cuando se utiliza la regresión por vectores de soporte (*Support Vector Regression,* SVR) combinada con algoritmos genéticos con el fin de ajustar los parámetros de la SVM para aumentar la precisión de la previsión (Hsu et al., 2006). También se ha aplicado SVR a la previsión de la carga eléctrica optimizando el método con algoritmos genéticos caóticos y añadiendo un componente estacional, creando un modelo de previsión de la carga eléctrica cíclica que arrojó mejores resultados de previsión que el modelo ARIMA (*AutoRegressive Integrated Moving Average*) y otros modelos SVR (Hong et al., 2013). Shi-Ting y Bo (2014) también mencionan que las SVM están siendo ampliamente utilizadas en estos temas debido a la idea de minimización del riesgo estructural. En este mismo trabajo de investigación, se utiliza SVR para analizar varios tipos de datos dentro de la economía del turismo, como el consumo eléctrico y de agua, modelando la demanda de tráfico y, además, los datos mensuales de cantidad de turistas. Adicionalmente, ha sido demostrado que se puede proporcionar una mejor planificación y administración de la energía mediante una previsión precisa del consumo mensual de electricidad, utilizando un algoritmo híbrido de optimización de la mosca de la fruta junto a SVR para prever el consumo de electricidad en una base estacional (Cao y Wu, 2016). La evolución de la red eléctrica en un sistema inteligente, o *Smart Grid*, también ha sido expuesta por Jiang et al. (2013), donde enuncian que la importancia de una alta precisión es un factor clave en un programa de inteligencia energética. En este trabajo de investigación, se presenta otro modelo de previsión híbrido combinando el SVR con el algoritmo de búsqueda de cuco (*cuckoo search*) y el análisis de espectro singular, combinando además estos dos métodos con un modelo ARIMA estacional (SARIMA). Kampouropoulos et al. (2014) expusieron una metodología de sistemas de inferencia neuro-difusos adaptativos (*Adaptive Neuro-Fuzzy Inference Systems*, o ANFIS) combinada con algoritmos genéticos que demostró predecir con éxito las necesidades energéticas a corto plazo. Wang et al. (2014) aplicaron un enfoque de *swarm inteligence* para la previsión de la carga de

energía aplicando esta técnica sobre modelos híbridos SARIMA y SVR optimizados mediante *cuckoo search* y el análisis de espectro singular, que produjo resultados de previsión muy satisfactorios.

Como trabajos de investigación más recientes, encontramos que Wang et al. (2016) utilizaron un modelo de regresión lineal funcional parcial para la predicción de la potencia como metodología principal en lugar de SVR, principalmente debido a la adecuación de la producción de potencia diaria del sistema energético del día anterior como predictor funcional, combinado con las variables climáticas utilizadas como covariables. Adicionalmente, Kofinas et al. (2016) exponen que uno de los componentes más importantes de las tecnologías de la información y la comunicación aplicadas a la gestión del agua es la reciente incorporación de métodos de previsión de la demanda de agua. En su trabajo de investigación, demuestran la idoneidad de las ANN y ANFIS para predecir la demanda de agua. Chen, Tan, y Berardi (2017) propusieron un enfoque híbrido basado en *clustering* para la previsión horaria de la demanda de electricidad utilizando *fuzzy C-means* junto con un modelo híbrido basado en sensores. Además, Chen, Tan, y Song (2017) utilizaron un método híbrido SVR para prever la demanda eléctrica no estacionaria. Por último, Porto e Irigoyen (2017) expusieron un método de previsión del consumo de gas basado en ANN para sectores residenciales capaz de prever la demanda de este recurso con una ventana temporal de 7 días.

Como conclusión, el uso generalizado de las SVM para pronosticar varios tipos de demandas de recursos resalta su viabilidad como método estándar, pero podría haber modelos más adecuados que aún no se han descubierto. Un enfoque interesante para la previsión de recursos podría ser el uso de la minería de flujos de datos con el fin de generar modelos adaptativos que cambien con el tiempo, aprendiendo de los nuevos datos que reciben y creando modelos que puedan ser válidos para periodos de tiempo mayores.

3.2.1.4. Gestión y previsión de Ingresos.

En un sector turístico y hotelero en constante cambio, no sólo es importante prever los recursos, sino también los posibles ingresos que pueden ser generados. Se ha descubierto que el *machine learning* es un mecanismo de previsión de ingresos factible, el cual permite a esta industria predecir sus ingresos monetarios mediante el uso de diversas metodologías de *machine learning* (Gayar et al., 2008; Lin et al., 2013; Al Shehhi y Karathanasopoulos, 2020). Una de las prácticas más comunes en el sector hotelero es la gestión de los ingresos, que se utiliza para ayudar a los establecimientos a decidir la asignación y la clasificación de las habitaciones, siendo esta práctica difícil, pero esencial, para crear presupuestos de ingresos con un alto grado de calidad (Lin et al., 2013).

Un método combinado de previsión de la gestión de ingresos fue desarrollado por Gayar et al. (2008), basado en un módulo de previsión de la demanda que consiste en varios métodos *machine learning*, a saber, suavizado exponencial, métodos de recogida, media móvil, métodos de Holt, regresión lineal y ANN; combinado con un

módulo de optimización y un módulo de decisión humana/conocimiento experto. Lin et al. (2013) también expusieron un método híbrido de previsión de ingresos, basado en una combinación de varios métodos *machine learning*: mínimos cuadrados difusos (*Fuzzy Least* Squares, o FLS), SVR y algoritmos genéticos (GA), que crea su eficaz método de previsión de ingresos estacionales FLSSVRGA. Bugarski et al. (2017) utilizaron ANN combinadas con RBFs y gradiente conjugado escalado para crear un sistema de apoyo a la decisión para la clasificación de los huéspedes del hotel por su gasto adicional en diferentes servicios del hotel. Por último, Al Shehhi y Karathanasopoulos (2020) compararon varios modelos, a saber, SARIMA univariante, ANFIS, *deep learning* basado en máquinas de Boltzmann restringidas y SVM polinómicas suavizadas para pronosticar los precios de las habitaciones en los establecimientos hoteleros de los países del Consejo de Cooperación del Golfo, mostrando que los modelos basados en ANFIS dentro de este ámbito obtienen un rendimiento superior, seguidos de cerca por los métodos de d*eep learning*.

La previsión de los ingresos puede resultar complicada debido a la variabilidad de sus valores, lo que dificulta la predicción de los ingresos generados en función de la demanda. Una línea de investigación prometedora podría ser el uso de más metodologías de *machine learning* combinadas con métodos de previsión de la demanda turística para generar sistemas de previsión híbridos, capaces no sólo de predecir los ingresos en función de la ocupación de las habitaciones, sino también de un enfoque más preciso basado en la temporada definida por la previsión de la demanda turística.

3.2.1.5. Análisis de viabilidad de empresas turísticas.

El *machine learning* y la minería de datos han mejorado la previsión de muchas maneras, pero una de las más interesantes es la previsión del posible éxito o quiebra de una empresa. Esto puede hacerse estudiando casos de fracaso o éxito ya conocidos y comparándolos con una empresa ya en funcionamiento. Li y Sun (2012) expusieron que el desarrollo de modelos de predicción de fracasos de empresas para la industria turística beneficia a los gestores, clientes, inversores y funcionarios públicos al reducir las pérdidas entre las empresas relacionadas con el sector hotelero, y Kim (2011) también afirma que el uso de estos mecanismos como sistemas de alerta temprana o de ayuda a los responsables de la toma de decisiones es útil para predecir la quiebra.

En este campo se han utilizado diferentes metodologías de *Computational Intelligence*. Kim (2011) comparó varias metodologías para la predicción de quiebras de hoteles en términos de clasificación global, precisión de la predicción y ratios de coste de error relativo, comparando las características funcionales de los modelos ANN, logístico, de análisis discriminante multivariante y SVM. Asimismo, Li y Sun (2012) adoptaron un enfoque SVM para corregir las muestras no equilibradas en un conjunto de datos compuesto por empresas hoteleras y turísticas chinas utilizando un enfoque de generación de muestras minoritarias basado en un porcentaje aleatorio de distancia *nearest neighbor*, junto con una SVM *nearest neighbor*. Chou Jui-Sheng y Lin Chieh

(2013) afirman que el esfuerzo, el tiempo y el coste de la gestión de posibles reclamaciones puede reducirse considerablemente mediante la previsión proactiva de las disputas en la fase inicial de la colaboración público-privada, exponiendo así un *ensemble* de varias técnicas de *machine learning* (SVM, ANN y C5.0) utilizadas para clasificar la propensión a las disputas en términos de medidas de rendimiento global, exponiendo también que la SVM es la mejor técnica de modelo único para esta tarea. Por último, Atsalakis et al. (2018) combinaron ANN con ANFIS para predecir el éxito de los servicios recién lanzados en el sector turístico.

Existe poca literatura sobre la previsión del lanzamiento de nuevos servicios y productos, a pesar de que el interés por este tema es cada vez mayor, especialmente cuando se aplica al turismo (Atsalakis et al., 2018). Esto hace que esta área sea un buen nicho de investigación en el que se podrían hacer grandes descubrimientos aplicando metodologías de *Computational Intelligence* no sólo para evitar la quiebra de una empresa turística, sino también para mejorar sus servicios y sus relaciones con otras empresas y partes interesadas.

3.2.1.1. Segmentación del mercado turístico, accesibilidad y dependencias.

Invertir en cualquier servicio del sector hotelero puede ser arriesgado si no se identifican correctamente las necesidades y los segmentos del mercado. Se ha mencionado que, en un mercado en el que las necesidades de los consumidores son heterogéneas, es necesario segmentarlos en grupos de clientes homogéneos en términos de actitudes, comportamientos y demandas para obtener una ventaja competitiva (Emel y Taşkın, 2005). La accesibilidad al mercado también es un factor importante, ya que de nada sirve invertir en un establecimiento hotelero si su mercado potencial ya está lleno de competidores, es decir, no existe nicho de mercado. También existen dependencias cruzadas entre los distintos mercados, y el turismo no es una excepción.

En los últimos años, el uso de técnicas de *machine learning* en este ámbito ha crecido exponencialmente. Por ejemplo, el algoritmo de clasificación de árboles de decisión C5.0 para extraer conocimiento útil para identificar diferentes segmentos de mercado para una gestión óptima de los clientes. Otra aplicación bastante interesante se encuentra en la evaluación y valoración del emplazamiento de un hotel al ser de suma importancia para su prosperidad empresarial (especialmente a largo plazo), debido a la imposibilidad de reubicar el establecimiento y a la elevada inversión realizada, por lo que se desarrolló un enfoque de *machine learning* para la evaluación de la ubicación del hotel mediante el uso de diferentes métodos, a saber, regresión de búsqueda de proyección (*projection pursuit regression*), ANN, SVR y regresión potenciada en combinación con una aplicación GIS basada en la web (Yang et al., 2015). Otro método de clasificación para la planificación de la localización también puede observarse en la investigación publicada por Zhang et al. (2018), donde se utilizan SVM junto con herramientas GIS para controlar el cambio de la cubierta del suelo y el uso de este, realizando la evaluación ecológica para una zona turística determinada. Las

dependencias cruzadas entre mercados son también un factor importante para el refinamiento de las previsiones, y Claveria, Torra, et al. (2016) utilizaron una ANN de perceptrón multicapa junto con un modelo de regresión de proceso gaussiano para mejorar significativamente la precisión de las previsiones. Por último, también se ha investigado la relación entre el precio del hotel y la accesibilidad del mercado mediante el uso de un marco de precios hedónicos, basado en un conjunto de datos multinivel que contiene varios factores de señalización de calidad, utilizando un modelo de regresión lineal de efectos mixtos de tres niveles (Yang et al., 2016).

Uno de los mayores escollos en esta materia podría ser que no hay suficiente investigación al respecto (Claveria, Monte, et al., 2016b), siendo un área de investigación bastante inexplotada donde se pueden lograr grandes descubrimientos. La aplicación de otras metodologías de *machine learning* que no sean SVM y ANN podría tener un efecto sorprendente en la previsión de la segmentación del mercado, donde la regresión de proceso gaussiano se utiliza para aumentar la precisión de la previsión de un perceptrón multicapa. Las investigaciones futuras en esta línea podrían resultar fructíferas en materia de previsión del mercado turístico.

3.2.2. Sistemas de perfilado y recomendación.

En el sector hotelero, al igual que en otros campos, es de suma importancia ofrecer buenos servicios, alternativas e instalaciones a los clientes para no solo mejorar como establecimiento hotelero, sino también para mantener esa posición frente a los competidores dentro de esta industria en constante cambio. Las técnicas de *machine learning* también han encontrado un buen nicho dentro de este campo debido a su variedad de aplicaciones y su potencial en la obtención de conocimientos de *Big Data*. En esta categoría vamos a exponer diferentes aplicaciones de estas técnicas en varios campos, las cuales permiten a la industria extraer conocimiento de grandes cantidades de datos como las reseñas de los clientes, el comportamiento o incluso los medios geoetiquetados.

3.2.2.1. Geoetiquetado y viajes.

El geoetiquetado permite añadir datos de localización a diferentes tipos de archivos, como fotos o publicaciones en redes sociales. Esta información ofrece un gran potencial en el descubrimiento de conocimiento, que puede obtenerse utilizando técnicas de minería de datos y enfoques de *machine learning* (Emel y Taşkın, 2005). Una de las áreas más exploradas en este tema es el uso de estas técnicas para descubrir atracciones turísticas populares mediante la minería de conocimiento de todo tipo de medios geoetiquetados. También se ha combinado la agrupación de datos con algoritmos de búsqueda rápida y picos de densidad para descubrir atracciones turísticas populares utilizando *Big Data* de medios sociales geoetiquetados (Peng y Huang, 2017). Además, Jiang et al. (2013) utilizaron un modelo SVM en combinación con un método

de clasificación para categorizar las atracciones turísticas aplicando *machine learning* sobre un conjunto de datos fotográficos.

También podemos encontrar una aplicación bastante útil de la *Computational Intelligence* al turismo en la investigación de Deng y Li (2018), donde se aplica un modelo *Naive Bayes* sobre un conjunto de datos fotográficos con el fin de crear un modelo que recomiende fotos publicitarias adecuadas para la promoción de destinos. La minería mediante *k-Means* también fue utilizado por Lin y Huang (2009) para obtener datos valiosos de los destinos, que tienen un gran potencial en el área de marketing.

La búsqueda de patrones en el comportamiento de los viajeros también puede lograrse mediante *machine learning*. En Vu et al. (2015) se expone un método de *clustering* basado en cadenas de Markov aplicadas a un conjunto de datos fotográficos, lo que permite extraer los comportamientos de viaje de los turistas mediante el procesamiento de fotografías geoetiquetadas. Además, Bermingham y Lee (2014) aplicaron la minería de patrones secuenciales para obtener información sobre los destinos pasados de los turistas y los posibles destinos futuros. Por último, también se ha utilizado un método de minería de reglas de asociación de muestreo distribuido, el cual se centra en el análisis del comportamiento de viaje del turista en su destino (Junping et al., 2008).

La clasificación de este tipo de datos para extraer conocimiento de ellos suele requerir técnicas de clasificación por *clustering* para procesar los datos de forma más efectiva. Los gestores del turismo llevan mucho tiempo buscando información sobre el comportamiento de los viajeros, especialmente para el desarrollo de productos, la gestión de destinos y el marketing de atracciones (Vu et al., 2015). Esto es importante porque la agrupación espacial es la clave para encontrar atracciones atractivas a partir de datos geoetiquetados (Peng y Huang, 2017).

Se podrían llevar a cabo nuevas líneas de investigación en cuanto a la creación de algoritmos de *clustering* más eficientes, nuevas técnicas de categorización o incluso la aplicación de algoritmos de *machine learning* ya conocidos. Por último, ha quedado demostrado que las SVM y las técnicas de *Naive Bayes* se han utilizado con éxito para obtener conocimiento de los datos geoetiquetados. Además, el uso de técnicas de minería de flujos de datos podría aumentar el tiempo de vida útil de los modelos de previsión obtenidos debido a su inherente adaptabilidad (Deng y Li, 2018; Jiang et al., 2013).

3.2.2.2. Análisis de sentimiento y grados de satisfacción.

Una de las mayores áreas en las que se aplica el *machine learning* al turismo es el análisis de sentimientos. Liu (2012) mencionó que el análisis de sentimientos, también llamado minería de opinión, es el campo de estudio que analiza las opiniones, los sentimientos, las evaluaciones, las valoraciones, las actitudes y las emociones de las personas hacia entidades como productos, servicios, organizaciones, individuos, asuntos, eventos, temas y sus atributos. Con estas técnicas es posible evaluar el grado

de satisfacción de los clientes mediante la extracción de opiniones y comentarios de las redes sociales y los sitios de reseñas. Esto permite a las empresas hoteleras mejorar sus servicios en función de estas reseñas.

Una gran cantidad de técnicas de *machine learning* han sido aplicadas en esta área, pero el 34% de los trabajos de investigación analizados sobre este tema exponen el uso de algoritmos SVM. Por ejemplo, se han utilizado SVM para extraer datos de sentimiento de reseñas, blogs y diversas plataformas (Chiu et al., 2015; Dickinger y Mazanec, 2015; Kumar Duvvur y Romanowski, 2016; Li et al., 2009; Shi y Li, 2011; Tokuhisa et al., 2012; Xia y Peng, 2009; Yao et al., 2011; Zheng y Ye, 2009), o también mediante combinaciones con otras metodologías tales como *kernels* N-Gram, diversas técnicas de Naive Bayes, ANN, reglas de asociación, clasificadores de máxima entropía, C4.5 y bosques aleatorios, JRIP y SVR (Banerjee et al., 2015; Hsieh et al., 2014; Li et al., 2016; Lin y Chao, 2010; Lu et al., 2012; Xu y Zhao, 2012; Ye et al., 2009; Zhang et al., 2011; Zhao et al., 2015).

Las metodologías de aprendizaje profundo también se han utilizado ampliamente en esta área. Ma et al. (2018) aplicaron una CNN preentrenada combinada con el procesamiento del lenguaje natural y una red neuronal recurrente para descubrir la utilidad de las imágenes proporcionadas por los usuarios en las reseñas de hoteles. Shoukry y Aldeek (2020) analizaron el papel del Internet de las Cosas (*Internet of Things*, o IoT) en el aumento de la satisfacción de los clientes dentro de la industria hotelera comparando también una CNN con un modelo de aprendizaje profundo basado en una red SVM y una red neuronal artificial, con la CNN demostrando un mejor rendimiento que las otras dos contrapartes de modelado. Ren et al. (2020) también aplicaron metodologías de *deep learning* para captar una comprensión integral de las preconcepciones reflejadas en las reseñas de los hoteles mediante el análisis de las imágenes publicadas por los clientes. Cheng et al. (2019) entrenaron un modelo CNN profundo con las reseñas de AirBnB para predecir la percepción de confianza de los huéspedes potenciales sobre un establecimiento hotelero. Chang et al. (2020) analizaron las reseñas y las respuestas de los hoteles mediante el uso de la analítica visual, la lingüística computacional y el *deep learning* para detectar las respuestas proactivas de los hoteles, utilizando un sistema de fusión de múltiples características basado en CNN. Por último, Luo et al. (2020) aplicaron *deep learning* para modelar las experiencias de los clientes de hoteles económicos chinos, utilizando un modelo de memoria a corto plazo bidireccional combinado con un modelo de campo aleatorio condicional.

Otras metodologías utilizadas en esta área también arrojan resultados positivos, como se ha visto con *Naive Bayes* y con las redes Bayesianas (Claster et al., 2010; Duan et al., 2013; Hsu et al., 2009; Nakamura et al., 2015; Sanchez-Franco, Navarro-Garcia, et al., 2019; Shimada et al., 2011; Weichselbraun et al., 2010); el procesamiento del lenguaje natural también se ha utilizado para extraer conocimientos en la clasificación de sentimientos (García-Barriocanal et al., 2010; Li et al., 2015; Sixto et al., 2013). También se han utilizado ANN para extraer y clasificar opiniones de clientes en

diferentes entornos (Phillips et al., 2015; Sharma y Dey, 2012). Otras metodologías de *machine learning* que también se han aplicado al análisis de sentimientos son LDA (Guo et al., 2017; Ren y Hong, 2017; Sanchez-Franco, Cepeda-Carrion, et al., 2019; Taecharungroj y Mathayomchan, 2019), C4.5 y árboles de clasificación y regresión (Yordanova y Kabakchieva, 2017; Zhang y Zhang, 2014), regresión lineal localmente ponderada (Gawlik et al., 2011), clasificación N-Gram (Kasper y Vela, 2012), minería de reglas de asociación orientada al contraste (Liu et al., 2013), algoritmos de lógica difusa (Afzaal et al., 2016; Nilashi, Yadegaridehkordi, et al., 2019), *gradient boosting* (Athanasiou y Maragoudakis, 2016), regresión multivariante (Xiang et al., 2017) y procesamiento del lenguaje natural (Haruechaiyasak et al., 2010). Un trabajo de investigación realizado por Nilashi, Ahani, et al. (2019) también muestra cómo el *clustering* puede aplicarse al análisis de sentimientos utilizando ANFIS, combinado con un enfoque de reducción de la dimensionalidad para reducir el tiempo de entrenamiento *offline* del algoritmo. Por último, un interesante trabajo de investigación expone que es posible detectar reseñas falsas aplicando un *ensemble* de varios métodos, como kNN, regresión logística, SVM, bosques aleatorios, *gradient boosting* y perceptrones multicapa (Martinez-Torres y Toral, 2019).

Aunque hay una gran cantidad de métodos entre los que elegir, las SVM suelen tener un mejor rendimiento que cualquiera de las otras metodologías a la hora de extraer el sentimiento de las reseñas y las opiniones. Sin embargo, las metodologías de *deep learning* están empezando a aplicarse lentamente en este campo debido a su velocidad, en comparación con otros métodos. Dado el enorme tamaño de los conjuntos de datos que se utilizan en estas áreas, todavía hay algún nicho para la optimización en este sentido, junto con la aplicación de nuevos métodos y técnicas de *ensemble*.

3.2.2.3. Análisis de la conducta turística.

La predicción del comportamiento de los clientes es un tema bastante interesante en términos de negocios. Aunque parece que en el sector hotelero no es una tendencia, este enfoque predictivo está empezando a crecer como área de investigación. Aplicando técnicas de previsión eficaces, es posible averiguar dónde y qué va a hacer un cliente, lo que puede aumentar el ahorro en muchas de las áreas que abarca un establecimiento de este tipo.

En 2002 se afirmó que, con la creciente competencia en el sector hotelero, es de suma importancia para la supervivencia de un hotel procurar servicios para los cambiantes estilos de vida y preferencias de los clientes, por lo que se propuso el uso de árboles de decisión C5.0 para predecir la fidelidad de los clientes y los servicios más valiosos de estos, junto con la segmentación de la población de clientes y la definición de qué segmento es el más adecuado para los servicios del hotel (Min et al., 2002). Asimismo, Guoxia y Jianqing (2009) utilizaron árboles de decisión C4.5 junto con el método estadístico χ^2 para encontrar los factores que influyen en el consumo de los turistas y la evaluación integral. Las SVM también se han utilizado en este campo de investigación, como se observa en el trabajo de Xu et al. (2009), donde se afirma que la

complejidad, la no linealidad y el ruido de los datos turísticos en bruto pueden crear desafíos para las técnicas de *Computational Intelligence* existentes, por lo que se propone una clasificación basada en SVM con dos técnicas de proyección de características no lineales (ISOMAP y técnica de mapeo probabilístico) para el análisis de datos turísticos. En 2017, las técnicas de *clustering* son utilizadas por Miah et al. (2017) en fotos de Flickr geoetiquetadas subidas por los clientes con el fin de analizar y predecir los patrones de comportamiento de los turistas en destinos específicos. Adicionalmente, en una investigación se ha utilizado un método de regresión no lineal que comprende el operador de selección y contracción mínima absoluta y *random forests* para determinar la distribución del gasto total de los turistas de cruceros en Uruguay (Brida et al., 2018). Por último, Belmonte-Fernández et al. (2018) crearon un sistema de mapeo de interiores mediante el desarrollo de un modelo basado en radiosidad *machine learning* que utiliza un *ensemble* de redes bayesianas, kNN, perceptrón multicapa, bosques aleatorios, SVM y optimización mínima secuencial, que se aplicó a la huella digital *WiFi* para reducir la cantidad de tiempo necesario para crear un mapa de radio en comparación con el modelo tradicional y manual anterior.

El análisis del comportamiento de los turistas está empezando a crecer como un área de investigación importante, pero la complejidad de los datos utilizados para entrenar los modelos, junto con su posible ruido en términos de información, genera una posible subárea de investigación en la que es obligatorio el pretratamiento de los datos para obtener conjuntos de datos limpios que puedan utilizarse para entrenar con éxito los modelos de previsión. También es necesario experimentar con otras técnicas de *machine learning* diferentes a los árboles de decisión o las SVM, como el *deep learning*.

3.2.2.4. Sistemas de recomendación turística.

Biuk-Aghai et al. (2008) mencionan en su investigación que las aplicaciones que suministran contenidos multimedia y los muestran en dispositivos móviles se han vuelto cada vez más comunes en los últimos años. Mientras se planifica un viaje de turismo, es bastante común aceptar sugerencias del planificador de viajes para mejorar la visita a lugares turísticos ya conocidos, o incluso para descubrir otros nuevos que pueden añadirse al itinerario de viaje. Estos factores hacen que los sistemas de recomendación turística sean una creciente área de investigación, en la que se están aplicando técnicas de *Computational Intelligence* para encontrar temas que puedan ser de interés para el turista planificador del viaje.

En este ámbito, las metodologías de *machine learning* más utilizadas son los métodos bayesianos y las SVM. Hsu et al. (2012) utilizaron redes bayesianas junto con un modelo Engel-Blackwell-Miniard y *Google Maps* para un sistema de recomendación inteligente de atracciones turísticas. También se afirma que un sistema de recomendación que recoge información de la web puede acabar teniendo datos duplicados en su base de datos, por lo que diseñaron un enfoque SVM combinado con árboles de decisión para resolver este problema (Luberg et al., 2012). También se ha

expuesto que la mayoría de los sistemas de recomendación turística existentes utilizan enfoques basados en el contenido y en el conocimiento, los cuales sufren el problema del "arranque en frío" y necesitan suficientes datos históricos de valoración y de conocimiento adicional, por lo que se expone un sistema de recomendación que categoriza a los turistas utilizando su información demográfica y luego hace recomendaciones basadas en las clases demográficas utilizando Bayes ingenuo, redes bayesianas y SVM (Wang et al., 2012). Además, se ha demostrado que los servicios turísticos son altamente sensibles al contexto, siendo esta una de las razones para desarrollar un sistema de recomendación de atracciones turísticas basado en el contexto (Chang y Ma, 2013). Namahoot et al. (2015) también desarrollaron un sistema de recomendación sensible al contexto basado en un algoritmo *Naive Bayes* mejorado. También se ha confirmado que las fotos geoetiquetadas en las redes sociales revelan las trayectorias de los turistas y sus preferencias sobre los puntos de interés y las rutas, lo que permite la creación de un sistema de recomendación de viajes basado en el contexto que utiliza SVM y regresión logística binaria como clasificadores, junto con otros métodos de filtrado (Sun et al., 2015). Arruza et al. (2016) utilizaron *Naive Bayes*, SVM y *gradient boosting* como metodologías para desarrollar un clasificador factible para un sistema de recomendación de hoteles. Saputro et al. (2016) desarrollaron un sistema de apoyo para reconocer lugares turísticos en las páginas web basado en *Naive Bayes*. Nilashi et al. (2017) también utilizaron SVR, combinado con ANFIS, para mejorar la precisión predictiva de un sistema de recomendación junto con varias técnicas de *clustering*. Por último, Martin-Fuentes et al. (2018) aplicaron una técnica de clasificación SVM para modelar un esquema de clasificación de alojamientos *peer-to-peer* con el fin de evitar problemas como la asimetría de información y la sobrecarga.

Otras metodologías de *machine learning* aplicadas a los sistemas de recomendación turística son los algoritmos genéticos (Biuk-Aghai et al., 2008; Chen et al., 2013), la lógica difusa y la minería de reglas de asociación (Lucas et al., 2013), la agrupación de datos (Gavalas y Kenteris, 2011; Jiang et al., 2011), LDA + procesamiento del lenguaje natural (Ebadi, 2016) y árboles de decisión + kNN (Kbaier et al., 2017).

Con el creciente uso de sistemas de recomendación para diferentes servicios dentro del sector de los hoteles, una de las subáreas más atrayentes en las que la investigación podría ser útil es en la optimización de diferentes parámetros, como el tamaño de los conjuntos de datos y la velocidad de ejecución de los algoritmos, dado que los sistemas de recomendación suelen ejecutarse en tiempo real. Además, la creación de algoritmos adaptativos que puedan recomendar cosas diferentes en función de las tendencias actuales también podría ser un tema de investigación relevante.

3.2.3. Previsión de demanda turística.

Junto con el análisis del sentimiento de los clientes y el grado de satisfacción, la previsión de la demanda es una de las mayores áreas de investigación en la previsión de turismo y los hoteles. Poder predecir cuántos clientes va a tener un establecimiento o servicio en un momento determinado es notablemente útil si se pretende preparar los

servicios para una determinada carga o para coordinarse con el objetivo de evitar el *overbooking* o la sobrecarga del servicio. Con el auge de las metodologías de *machine learning*, la previsión de estos parámetros se ha vuelto, en cierta manera, más sencilla que hace años, aunque esto depende de las metodologías utilizadas para ello.

El uso de las ANN es algo habitual, ya que supone casi un 54% de las técnicas utilizadas, seguidas de las SVM, que suelen combinarse con las ANN para mejorar los modelos de previsión. En 2005, Pai et al. (2005) investigaron la viabilidad de las SVM combinadas con redes neuronales de retropropagación (BPNN) para predecir con éxito la demanda turística. Unos años más tarde, Chen y Wang (2007) aplicaron un enfoque novedoso en este campo combinando SVR con algoritmos genéticos para encontrar y aplicar los parámetros óptimos para construir el modelo SVR. Además, esta última técnica también ha sido aplicada por Cai et al. (2009), demostrando que una combinación de algoritmos genéticos y SVR, a saber, GA-SVR, es una técnica factible para la previsión de la demanda turística. También se ha aplicado ANFIS en la previsión de las llegadas de turistas con el objetivo de demostrar su viabilidad sobre otros tres modelos, a saber, las series temporales difusas, el modelo de previsión de Gray y el modelo residual modificado de Markov (Chen et al., 2010). También se ha utilizado un modelo de descomposición de modo empírico combinado con BPNN para pronosticar la demanda turística de forma más precisa mediante la descomposición de los datos brutos y la suma de las predicciones de ambos modelos (Chen et al., 2012). En 2013, se propuso un modelo de previsión híbrido para las cancelaciones de clientes, basado en la combinación de BPNN y redes neuronales de regresión generalizada (Huang et al., 2013). En 2014, Berenguer et al. (2014) propusieron un modelo de previsión de ANN combinado con series temporales, pudiendo así predecir la demanda turística de forma estacional. También se ha utilizado un *ensemble* de BPNN combinado con *Bagging* para superar algunos inconvenientes presentados por las BPNN (Xue-Bo y Shi-Ting, 2014). En 2015, Akın (2015) comparó varias técnicas de *machine learning*, a saber, SARIMA, v-SVR y un perceptrón multicapa, para encontrar la más adecuada para la previsión de la demanda turística, siendo la segunda la mejor para este tipo de predicciones. Además, Cankurt y Subasi (2015) utilizaron modelos de perceptrón multicapa y SVR para generar de forma determinista variables auxiliares que perfilasen diferentes componentes de las series temporales, mejorando el rendimiento de la previsión. Siguiendo esta tendencia de las ANN en la investigación de la previsión de la demanda, Claveria et al. (2015) compararon tres técnicas diferentes basadas en ANN: perceptrón multicapa, RBF y redes de Elman, con el fin de comparar su rendimiento, encontrando que las dos primeras técnicas superan a la última, y demostrando también que la dimensionalidad es muy importante para las predicciones a largo plazo. En otro trabajo de investigación se utilizó una máquina de aprendizaje extremo para calcular diferentes variables que mejoran el modelo de predicción final, superando incluso a las SVR (Wang et al., 2015). En 2016, se expuso la importancia del horizonte de previsión en la selección del modelo comparando los rendimientos de SVR con un *kernel* RBF con ANN utilizando un modelo lineal como referencia (Claveria, Monte, et al., 2016b). Noersasongko et al. (2016) también compararon el rendimiento de las BPNN frente a otras técnicas de ANN, como

la kNN y la regresión lineal múltiple, utilizando algoritmos genéticos para optimizar los parámetros de las BPNN, demostrando así que estas últimas tienen menores errores de predicción en términos de error cuadrático medio. La viabilidad de las BPNN para predecir la demanda turística utilizando MATLAB también fue demostrada por Sun y Chang (2016). Las predicciones regionales mediante SVR, regresión del proceso gaussiano y ANN combinadas para generar modelos de previsión más precisos también han sido estudiadas, demostrando su idoneidad cuando los horizontes de previsión aumentan (Claveria et al., 2016a). En 2017, una investigación examinó varias metodologías, a saber, ANN, SVM localmente profundas, junglas de decisión, árboles de decisión y árboles de decisión potenciados, para predecir con precisión las cancelaciones de reservas de hoteles (Antonio et al., 2017). Chang y Tsai (2017) compararon el rendimiento de *deep learning*, SVM y ANN para la previsión del número de turistas, encontrando que esta metodología supera a los otros dos métodos en precisión. También ha sido comparada la viabilidad de la regresión de procesos gaussianos frente a las ANN en un entorno de múltiples entradas y múltiples salidas, encontrando que a medida que la memoria de los modelos aumenta el rendimiento de previsión de la regresión de procesos gaussianos, aunque las ANN que utilizan RBF superan a la regresión de procesos gaussianos para la previsión a largo plazo (Claveria et al., 2017). Folgieri et al. (2017) volvieron a comparar las BPNN con métodos de regresión lineal, demostrando la viabilidad de las primeras como método de previsión de la demanda turística y como herramienta de toma de decisiones. Además, Han et al. (2017) propusieron un posible método de previsión de la demanda turística mediante un modelo combinado de visión cruzada basado en algoritmos BPNN y SVR. Rafidah y Ani (2017) también utilizaron ANN para pronosticar las llegadas de turistas, comparándolas con SVM, siendo este último superado por el primer método en términos de error medio absoluto. También se ha utilizado un modelo de Gray-Markov para predecir los ingresos de turistas extranjeros incorporando redes neuronales en su modelo de previsión (Yi-Chung, 2017). Sun et al. (2017) propusieron un *kernel* basado en una máquina de aprendizaje extremo para predecir la llegada de turistas, mostrando su modelo una mayor precisión en comparación con otros métodos. Oger Vihikan et al. (2017) utilizaron BPNN para predecir la llegada de turistas a Bali. Por último, en 2018 se investigó el rendimiento de siete métodos diferentes de *machine learning* (perceptrón multicapa, RBF, redes neuronales de regresión generalizada, kNN, árboles de clasificación y regresión, SVR y regresión de proceso gaussiano), mostrando que hay diferencias entre estos métodos, pero también que no hay un mejor método en los resultados obtenidos, que fueron analizados por el error porcentual medio absoluto (Kamel et al., 2018).

Las SVM y las SVR también se han utilizado en la previsión de la demanda turística, aunque parece que la precisión de las previsiones de estos métodos suele ser superada por las ANN. Sin embargo, ha habido muchos estudios sobre este tema. Romero Morales y Wang (2010) utilizaron una SVM con una regresión logística basada en *kernel* para predecir con éxito las tasas de cancelación, lo que permite gestionar los ingresos de forma más precisa. Pai et al. (2014) presentaron un útil modelo híbrido de previsión de la demanda turística combinando *fuzzy C-Means* con SVR de mínimos

cuadrados logarítmicos (denominado LLS-SVR), utilizando además algoritmos genéticos para seleccionar de forma óptima los parámetros de su modelo híbrido. Además, también se ha demostrado que los algoritmos genéticos adaptativos combinados con SVR estacional, superan al SVR normal y a las BPNN en la previsión del flujo turístico (Chen et al., 2015). En 2016, Antonio et al. (2016) utilizaron SVM junto con un *kernel* de regresión logística para demostrar que es posible generar modelos para predecir las cancelaciones de reservas con alta precisión. Varios métodos de regresión fueron analizados, incluido SVR, descubriendo que este último supera a los modelos de regresión lineal múltiple y de perceptrón multicapa en la previsión de la demanda turística (Cankurt, 2016). Lijuan y Guohua (2016) también crearon un sistema híbrido SVR con un componente estacional y se optimiza mediante el algoritmo de optimización de la mosca de la fruta, arrojando resultados positivos que posicionan a este modelo como una solución factible de previsión turística. Xu et al. (2016) extrajeron reglas difusas de Takagi-Sugeno a partir de SVM entrenadas para aumentar la precisión de la previsión de la demanda turística, proporcionando además información comprensible para los responsables de la toma de decisiones. Ali et al. (2017) combinaron el análisis wavelet con las SVM, creando un modelo "WSVM" que supera a un modelo normal basado en SVM. Liu y Yao (2017) propusieron una SVM de mínimos cuadrados modificada para pronosticar el flujo de pasajeros en días festivos para el sistema de metro, utilizando un algoritmo de *swarm optimization* de partículas mejorado para optimizar los parámetros. Liang y Bi (2017) expusieron un método SVR de variación estacional para predecir los flujos turísticos, descubriendo que el SVR es más preciso que los métodos de regresión lineal multivariante. Por último, Zhang et al. (2017) desarrollaron un algoritmo SVR híbrido, combinado con el algoritmo del murciélago, para pronosticar el volumen de turistas optimizando los parámetros del SVR con él, creando un método BA-SVR que puede superar al método SVR normal.

También se han aplicado otros métodos de *machine learning* en este ámbito, pero en menor medida. Como se ha visto anteriormente, los algoritmos genéticos también se han utilizado para optimizar diferentes parámetros dentro de otros métodos de *machine learning*, dando resultados más precisos que sus homólogos no optimizados. También se han combinado algoritmos genéticos con técnicas de lógica difusa (Hadavandi et al., 2011; Sakhuja et al., 2016). Hong et al. (2011) también aplicaron algoritmos genéticos caóticos para predecir la demanda turística y superar los problemas clásicos que provoca este método. En el trabajo de investigación publicado por Shahrabi et al. (2013) se combinó el *clustering* con técnicas de lógica difusa junto con algoritmos genéticos utilizando la evolución simbiótica para la función de *fitness*. También se ha comparado la regresión del proceso gaussiano aplicada a la demanda turística con la media móvil autorregresiva y los modelos SVM, añadiendo un componente de dispersión que reduce la complejidad computacional y aumenta su capacidad de generalización (Wu et al., 2012). También se ha aplicado ANOVA y regresión por pasos en destinos turísticos chinos relacionados con el juego, generando un modelo que resultó útil para encontrar los determinantes de la tasa de ocupación hotelera (Lei y Lam, 2015). Por último, como demostraron Cankurt y Subaşi (2016), los

árboles de decisión y sus *ensembles* también se han probado en este campo de la previsión, superando estos últimos a sus homólogos individuales en términos de precisión de la previsión.

No hay gran margen en esta área para realizar investigaciones, pero es posible que tanto *ensembles* de modelos como nuevos modelos híbridos puedan estar esperando a ser descubiertos. Estas técnicas híbridas también han demostrado su utilidad en la previsión de la demanda turística, superando en ocasiones a las ANN y las SVM. Además, la demanda turística es un campo en constante cambio en el que la afluencia de turistas suele cambiar con el tiempo, por lo que el uso de técnicas de minería de flujos de datos podría ser útil no sólo en términos de almacenamiento de datos y optimización del procesamiento, sino también en modelos que puedan adaptarse con el tiempo.

3.2.4. Previsión meteorológica y evaluación de riesgos medioambientales.

La predicción meteorológica, aunque no está directamente relacionada con el turismo, es uno de los temas más recurrentes en *machine learning*. De hecho, posiblemente por su relación indirecta con el turismo, no hay muchos estudios sobre sus posibles aplicaciones turísticas. Ser capaz de predecir el clima y ciertos riesgos ambientales es de suma importancia para los clientes, y podría ser un potencial cambio de paradigma frente a los competidores. Gokaraju et al. (2011) exponen un método que puede detectar con éxito las floraciones de algas nocivas en la costa mediante el uso de un SVM basado en *kernel*, lo cual es importante porque, como se indica en ese trabajo de investigación, estos sistemas proporcionan advertencias para industrias como la de los alimentos de origen marino, las actividades turísticas y los gestores locales de recursos y del medio ambiente. Otro enfoque útil fue el adoptado por Xiao (2012), en el que se pronostican las zonas siniestradas por los incendios forestales utilizando una SVM con parámetros optimizados mediante algoritmos genéticos. Como se ha visto en artículos anteriores, el enfoque "algoritmo genético-SVM" resulta ser más eficaz que el algoritmo SVM clásico.

King et al. (2014) expusieron un *ensemble* de técnicas de *machine learning*, a saber, regresión lineal múltiple, árboles de clasificación y regresión y ANN para predecir los destinos más escogidos por los turistas de esquí a lo largo de seis temporadas, utilizando datos locales, regionales y nacionales para crear el modelo.

Además, también se ha aplicado *online learning* para predecir las condiciones de viento máximo en la región de las Islas Canarias mediante el uso de datos en tiempo real proporcionados por diferentes estaciones meteorológicas que están distribuidas en estas islas (Sánchez-Medina et al., 2019). El clima en las islas Canarias tiende a ser clemente y estable, pero algunas veces se han producido condiciones meteorológicas extremas que influyen en el turismo y la industria hotelera, por lo que este trabajo de investigación es útil para estas industrias en el archipiélago canario.

Estas son sólo unas pocas aplicaciones de *machine learning* sobre la previsión meteorológica y ambiental, lo que significa que todavía hay mucho margen para la investigación en esta área mediante el uso de nuevas técnicas de *machine learning* o la aplicación de modelos ya conocidos sobre las áreas del sector turístico y hotelero.

3.3. Las condiciones meteorológicas y su influencia en el sector turístico.

Una de las potenciales aplicaciones del *Data Stream Mining* dentro del sector turístico es la de realizar predicciones sobre fenómenos meteorológicos adversos o eventos medioambientales potencialmente peligrosos en zonas turísticas, un área muy poco explorada por el ámbito de la investigación, tal como se ha mencionado anteriormente. Una de las metodologías utilizadas para ello son las SVM, usadas para predecir de manera exitosa brotes de algas dañinas a lo largo de la costa del Golfo de México (Gokaraju et al., 2011), enviando así avisos a industrias como la pesquera o la turística para planificar o modificar eventos en función de dichas predicciones. Las SVM también han sido utilizadas para predecir posibles incendios en zonas sensibles a los mismos (Xiao, 2012), algo que podría ser fácilmente aplicable a establecimientos turísticos ubicados en zonas boscosas de alta densidad donde existan temporadas sujetas a altas temperaturas. Otra aplicación mucho más obvia dentro del ámbito turístico es la de realizar modelos predictivos en zonas de esquí para predecir qué días presentarán una mayor o menor cantidad de nieve, algo realizado mediante un conjunto de modelos o *ensemble* formado por regresión lineal múltiple, redes neuronales, y árboles de clasificación y regresión (*Classification and Regression Trees*, CART) (King et al., 2014). No obstante, en el momento de realizar la presente tesis no se halló ningún artículo reciente enfocado en la aplicación de metodologías basadas en *Computational Intelligence* aplicadas al turismo en Canarias o que utilizase *Data Stream Mining* como paradigma de entrenamiento para crear su modelo predictivo.

Las Islas Canarias son un destino turístico de primer orden por muchas razones: su accesibilidad desde Europa, sus servicios de nivel europeo y la hospitalidad de los canarios tras décadas de ser el turismo su principal industria. Cabe destacar como dato de gran importancia que según el Instituto Canario de Estadística (ISTAC), el turismo representa casi el 35% del PIB de Canarias. Otro elemento crucial para el atractivo de este destino es el clima templado durante todo el año, aunque existen excepciones a este clima benévolo. No obstante, es de gran relevancia mencionar que ha habido episodios de clima extremo en los que, en determinada medida, se ha visto comprometida la seguridad de la población y la economía local canaria. La tormenta tropical Delta (Beven, 2006; Seco et al., 2009) de noviembre de 2005, por ejemplo, supuso unas pérdidas de hasta 364 millones de dólares y al menos 7 víctimas mortales directas (León et al., 2006), con más de 225.000 residentes afectados por cortes de electricidad y 12.000 por cortes en el servicio telefónico. La ráfaga máxima registrada en la isla de La Palma fue de 95 mph (152 km/h), y en Tenerife la ráfaga máxima fue de 90 mph (147 km/h).

Este tipo de eventos adversos ha hecho que la predicción de la velocidad del viento a corto plazo sea un campo de investigación muy activo y de especial interés hoy en día, cuando el cambio climático y la necesidad de reducir las emisiones de gases de efecto invernadero son de suma importancia para muchos gobiernos. Por ejemplo, en la Unión Europea, los objetivos de la estrategia Europa 2020 para un crecimiento inteligente, sostenible e integrador incluyen una reducción del 20% de las emisiones de gases de efecto invernadero con respecto a los niveles de 1990, que el 20% de toda la producción de energía de la UE proceda de fuentes renovables y una mejora del 20% de la eficiencia energética (Commission (EC), 2010). Una parte esencial de esto es la energía eólica, por lo que la predicción de la velocidad del viento a corto plazo es un reto importante, en particular en lo que respecta a la optimización del control automático de las turbinas eólicas. Por ejemplo, y en una escala temporal más larga, las predicciones de la velocidad del viento con un horizonte de horas pueden ayudar en el lento procedimiento de encendido y apagado de las turbinas (Bossanyi, 1985).

Para esta tarea, existe un amplio corpus de literatura centrado en el análisis de series temporales: se monitoriza una única variable a lo largo del tiempo y se aprende su estacionalidad, tendencia y autocorrelación para construir el mejor modelo matemático (analítico) posible, que suele intentar capturar directa o indirectamente la distribución de probabilidad de la velocidad del viento. Se han realizado muchas investigaciones en este ámbito. Por nombrar algunos, se han desarrollado modelos autorregresivos de media móvil (ARMA) y de persistencia para la previsión de la velocidad del viento en un horizonte de 10 horas en cinco localizaciones diferentes mediante la transformación y estandarización de las series temporales de entrenamiento (Torres et al., 2005), o también se han desarrollado modelos de Markov discretos de primer y segundo orden para la previsión de la energía eólica y se han aplicado a datos reales (Carpinone et al., 2015).

También se han realizado investigaciones basadas en SVM como la desarrollada por Mohandes et al. (2004), donde comparan esta técnica con los resultados obtenidos con un perceptrón multicapa para la predicción de la velocidad media diaria del viento en Madina, Arabia Saudí. Los experimentos parecen apoyar la superioridad de las SVM en este caso.

El filtrado de Kalman también se ha utilizado con frecuencia en varias investigaciones a este respecto. Bossanyi (1985) utilizó esta técnica para predecir la velocidad media del viento con margen de un minuto en Stornoway con el fin de reducir el error de la previsión de "persistencia". Esta técnica también fue utilizada por Cassola y Burlando (2012), siendo aplicada a la salida directa de otros modelos numéricos para corregir errores sistemáticos en los mismos.

Otras técnicas probabilísticas menos comunes pueden ser las redes bayesianas como las utilizadas por Wang et al. (2018), donde se introdujo una tecnología avanzada de modelado de estructuras de dependencia basada en la cópula regular de vid en el campo de la previsión probabilística de la energía eólica. El modelo obtuvo buenos resultados tanto en los casos de datos completos como en los de datos ausentes, con el

valor añadido de describir las condiciones de previsión; o por Jiang et al. (2013), donde se propuso un modelo bayesiano de previsión de series temporales. Un elemento muy interesante del análisis bayesiano del que se beneficia este trabajo es la posibilidad de incorporar a los modelos el conocimiento de los expertos del dominio. La metodología de los autores incorpora velocidades de viento de alta frecuencia recogidas de los aerogeneradores y aprovecha el concepto de rupturas estructurales. Las máquinas de aprendizaje extremo (del inglés *Extreme Learning Machines, ELM)* también han sido utilizadas por Salcedo-Sanz et al. (2015), donde los investigadores combinan un algoritmo de Optimización de Arrecifes de Coral (CRO) con operadores de la búsqueda de armonía con el fin de seleccionar el mejor atributo meteorológico posible para entrenar una red ELM. Adicionalmente, las limitaciones de los modelos de predicción meteorológica fueron estudiadas por Marrero et al. (2009), donde se centraron en las "condiciones laterales de contorno" de las Islas Canarias a la hora de dar cuenta de tormentas extratropicales poco frecuentes como el Delta.

En el ámbito más reciente del *machine learning* existe un amplio segmento de literatura donde las redes neuronales artificiales son la metodología más utilizada. Por ejemplo, una variación de la hibridación del modelo de mesoescala de quinta generación (MM5) utilizando redes neuronales es capaz de realizar predicciones sobre la velocidad del viento a corto plazo, aplicada a la velocidad del viento promediada por hora en un parque eólico con el objetivo final de predecir la producción total de energía del parque eólico (Salcedo-Sanz et al., 2009). También se ha hecho uso del elemento lineal adaptativo (ADALINE), la *backpropagation* de redes neuronales y la función de base radial para su aplicación en la predicción de la velocidad del viento con 1 hora de antelación (Li y Shi, 2010). Incluso la ingeniería de características y el aprendizaje profundo se han utilizado para predecir la velocidad del viento a corto plazo (Dalto et al., 2015).

Adicionalmente, en trabajos recientes, se proponen técnicas avanzadas de d*eep learning* basadas en redes neuronales recurrentes (Shi et al., 2018) y redes neuronales convolucionales (Huang y Kuo, 2018) para la predicción de la energía eólica y para la predicción de la concentración de partículas en suspensión con un diámetro igual o inferior a 2,5 µm, respectivamente. En ambos casos, los resultados experimentales mostraron que, en comparación con los métodos tradicionales de *machine learning*, los sistemas propuestos obtuvieron las mejores previsiones.

No obstante, y aun teniendo en cuenta los buenos resultados obtenidos en los ejemplos anteriores, no se encontraron trabajos centrados en el desarrollo de los modelos de velocidad del viento de forma incremental y adaptativa. Este parece ser un *gap* de investigación sin cubrir, ya que la velocidad del viento es un proceso muy aleatorio, tanto en el tiempo como en el espacio, e incluso con los mejores modelos de velocidad del viento a corto plazo del estado de la técnica, los parámetros ajustados para una ubicación particular pueden no funcionar bien en otras ubicaciones con diferentes distribuciones de probabilidad (Qin et al., 2011). En otras palabras, estos modelos no son fácilmente generalizables una vez entrenados.

El siguiente paso lógico parece ser desarrollar modelos de aprendizaje que sean lo suficientemente robustos y flexibles como para poder adaptarse a distribuciones de probabilidad cambiantes, lo cual coincide con la hipótesis formulada en la presente tesis.

En lo que respecta a la predicción del viento, se han diseñado muchos enfoques basados en algoritmos tradicionales de *machine learning*, utilizando técnicas como el modelado probabilístico (Wang et al., 2018), los árboles de decisión (Lahouar y Ben Hadj Slama, 2017), o el *deep learning* (Shi et al., 2018), por nombrar algunos. Sin embargo, la mayoría de estos datos meteorológicos suelen llegar de forma continua en forma de flujos de datos de alta velocidad y volumen, existiendo al menos dos razones por las que esa minería de datos tradicional no es un ajuste óptimo para modelar tales fenómenos. Cuando se aplica el paradigma tradicional de aprendizaje de modelos *offline*, separando un conjunto de datos de entrenamiento y otro de prueba (también podría utilizarse un conjunto de datos de validación cruzada), hay que recoger enormes cantidades de datos durante un largo período de tiempo. Esto se hace bajo la asunción tanto de la estabilidad estadística del fenómeno a modelar como de la disponibilidad de una potencia de cálculo infinita en términos de memoria y CPU). Dicha asunción no es muy realista en la mayoría de los casos, ya que ese enfoque no parece capaz de analizar eficazmente una cantidad creciente de datos (Krawczyk et al., 2017). No obstante, existe un enfoque alternativo y más novedoso que podemos denominar *online learning* o minería de flujos de datos (*data stream mining*), en el que el aprendizaje del modelo se realiza de forma incremental utilizando una estrategia de aprendizaje precuencial (Dawid, 1984).

3.4. Tráfico y turismo: relaciones y nuevas tecnologías.

Con el aumento del número de vehículos y la saturación de la infraestructura vial, los factores como las emisiones, los retrasos y la congestión provocadas por el tráfico se han convertido en graves problemas para los gestores y planificadores de la movilidad, ya que dichos factores generan un gran desperdicio de recursos y restricciones a la economía, afectando al desarrollo urbano y a la calidad de vida de las personas (Chen et al., 2018; Cheng et al., 2017; Ye y Yamamoto, 2018; Zhu y Zhang, 2018). Para resolver este problema, los gobiernos están aplicando diferentes enfoques, uno de los cuales es el de los Sistemas Inteligentes de Transporte (ITS). Este enfoque combina la infraestructura de transporte con la tecnología (Lin et al., 2017), utilizando diversas tecnologías de la información basadas en la comunicación para mejorar los sistemas de transporte, la orientación de los vehículos y la previsión de la demanda de viajes (Dimitrakopoulos y Demestichas, 2010), considerándose las ITS como una de las formas más eficaces de hacer frente a la congestión de vehículos (Yang et al., 2016).

También debe mencionarse que la metodología más efectiva y extensa a la hora de predecir las condiciones del tráfico son las redes neuronales debido a la gran precisión que presentan a la hora de realizar dichas predicciones gracias a su capacidad para modelar complejos patrones de tráfico no lineales (Yu et al., 2017), y se están aplicando sobre diferentes áreas con el objetivo de producir previsiones precisas sobre diversos indicadores de tráfico. Dentro de esta área existe una gran cantidad de

enfoques que se utilizan para comprender mejor las condiciones viales causantes de congestiones o atascos en las carreteras. Los últimos avances mediante redes neuronales en este campo mediante técnicas novedosas incluyen la utilización de una red neuronal difusa basada en un sistema Takagi-Sugeno combinado con un método *k-means*, un estimador de mínimos cuadrados recursivos ponderados y una función de regresión trigonométrica, proporcionando una gran capacidad de aprendizaje y mejorando así la precisión a la hora de realizar predicciones de la velocidad del tráfico (Tang et al., 2017), la creación de enfoque basado en memoria temporal jerárquica para la previsión del flujo de tráfico a corto plazo, obteniendo mejores resultados que un enfoque de memoria a largo plazo (Mackenzie et al., 2019), o el uso de un modelo de red neuronal profunda basado en la convolución para datos de tráfico periódicos con el fin de predecir la congestión vial, convirtiendo los datos de tráfico unidimensionales en una matriz de entrada bidimensional y aplicando una serie de convoluciones para tener en cuenta la coherencia local y la periodicidad (Chen et al., 2018).

Dentro de las aplicaciones más recientes puede observarse un sistema de estimación del estado del tráfico adaptativo basado en redes neuronales evolutivas, proporcionando capacidad de adaptación al modelo sin la necesidad de un proceso de reentrenamiento (Laña et al., 2019): el uso de redes neuronales generativas antagónicas para predecir el estado del tráfico (Xu et al., 2020), *deep learning* para predecir su flujo (Zheng y Huang, 2020), o incluso redes neuronales convolucionales gráficas (Guo y Yuan, 2020; Zhou et al., 2020; Zhu et al., 2022) para realizar predicciones acerca de su volumen o su situación.

No obstante, en el proceso de realización de esta tesis doctoral no se halló ningún trabajo de investigación relevante en el que se aplicaran redes neuronales basadas en la teoría de la resonancia adaptativa, las cuales no solo han demostrado ser precisas en sus predicciones, sino que también presentan una resistencia innata hacia el *concept drift* debido a su paradigma de aprendizaje precuencial y a sus capacidades adaptativas a corto plazo (Carpenter y Grossberg, 1988), siendo estos motivos de peso para realizar una investigación sobre su posible aplicación al área de los sistemas de transporte inteligente. Las redes ART pueden aprender de los cambios en los datos y actualizar lo aprendido anteriormente con nueva información, lo que mejora su capacidad para reconocer patrones y hacer predicciones sobre grandes flujos de datos que cambian con el tiempo.

4. Metodología.

Para cumplir los objetivos específicos designados al comienzo del desarrollo de la presente tesis doctoral, se ha desarrollado una metodología principal consistente en la aplicación de *Data Stream Mining* para crear modelos predictivos perdurables en el tiempo, de gran precisión y con resistencia frente al *concept drift*.

Esta metodología engloba, a su vez, las dos metodologías utilizadas tanto para la elaboración de un modelo predictivo adaptativo para predecir la velocidad máxima del viento en cualquier punto de las Islas Canarias, como para la novedosa aplicación de una aproximación basada en redes neuronales resonantes para realizar predicciones sobre eventos de tráfico no deseados en una localización de gran valor turístico. A continuación, se describen dichas metodologías con detalle.

4.1. Desarrollo de un modelo de regresión lineal incremental y adaptativo.

Para lograr el segundo objetivo específico, consistente en predecir la velocidad máxima del viento en 68 estaciones climatológicas ubicadas por todo el archipiélago canario con un horizonte de predicción de 60 minutos, se desarrolló una metodología basada en *Data Stream Mining*. Para ello se utilizó como base un modelo de regresión lineal combinado con descenso por gradiente, (del inglés *gradient descent*), modificado de manera que pudiese operar bajo el paradigma incremental y adaptativo inherente a un flujo de datos. Esta aproximación se realizó debido a su robustez y amplia utilización gracias a que el comportamiento de este modelo es fácilmente comprensible y comparable a otras metodologías, por lo que las modificaciones aplicadas al mismo para convertirlo en un modelo *online* e incremental son también de fácil entendimiento.

El modelo de regresión lineal adaptativo se entrena gradualmente mientras los nuevos datos van llegando, haciendo que los parámetros de la ecuación del modelo vayan cambiando respecto a los datos analizados con anterioridad (Sánchez-Medina et al., 2019). A la hora de implementar una estrategia de olvido, o *forgetting strategy*, se recurrió a controlar la dimensión del vector de coeficientes de la regresión lineal, así como a la alteración adaptativa del ratio de aprendizaje del *gradient descent*: si tenemos un número de datos previamente analizados y detectamos que la precisión del modelo de regresión lineal se degrada, reducimos el número de observaciones pasadas eliminando el dato más antiguo analizado. De manera inversa, si el modelo observa que llega a un momento de estabilidad, añade las últimas observaciones al conjunto de datos analizados.

Todo lo mencionado anteriormente permite analizar, mediante un modelo de regresión lineal y *gradient descent*, un número potencialmente infinito de datos y realizar predicciones acerca de la velocidad máxima del viento en un lugar determinado, independientemente de la estación meteorológica seleccionada. El modelo también se adapta gradualmente al *concept drift* mediante la eliminación de datos analizados con anterioridad, los cuales pueden degradar el modelo predictivo con el paso del tiempo, y

mediante la adición de datos analizados en el momento presente con el objetivo de incrementar la estabilidad del modelo (Sánchez-Medina et al., 2019).

En esta investigación se ha desarrollado una metodología de minería de flujos de datos para predecir la velocidad máxima del viento (VMAX10m) en 68 estaciones meteorológicas de las Islas Canarias, con un horizonte temporal de 60 min. Como técnica de aprendizaje de parámetros se ha utilizado un método de regresión clásico, la regresión lineal con descenso de gradiente, aunque con modificaciones para que pueda operar de forma adaptativa e incremental.

El aprendizaje adaptativo que se propone a continuación se basa en el paradigma "precuencial" (Dawid, 1984). En lugar de utilizar el enfoque clásico de *machine learning* (*offline*) consistente en dos conjuntos independientes de prueba y entrenamiento, en una configuración precuencial el modelo se evalúa primero a medida que llegan nuevas observaciones, y luego vuelve a ser entrenado con esas nuevas observaciones.

En esta primera fase de la investigación se eligió la regresión lineal como modelo base para la predicción de la velocidad máxima del viento por dos razones. En primer lugar, la combinación de la regresión lineal con un modelo de aprendizaje basado en el descenso de gradiente es muy conocida, así como su robustez. Su comportamiento puede ser fácilmente comprendido y comparado con otras metodologías. Por tanto, las modificaciones aplicadas a esta metodología para hacerla incremental y adaptable pueden ser entendidas por cualquier lector medio con formación básica en *machine learning*. Debido a la falta de literatura sobre el uso del *Data Stream Mining* para esta aplicación, es necesario un estudio de referencia utilizando una metodología muy clásica como la regresión lineal y el descenso de gradiente.

La segunda razón es más técnica. La ausencia de recurrencia de estos algoritmos los hace adecuados para el aprendizaje basado en la observación y, por tanto, fácilmente adaptables al entrenamiento *online*.

Sin embargo, es evidente que la predicción de la velocidad del viento no es un fenómeno lineal, y confirmamos esa hipótesis en un análisis del coeficiente de determinación que hemos desarrollado como parte de los experimentos asociados a esta investigación. La adaptabilidad de la metodología propuesta hace frente en cierta medida a la falta de linealidad del fenómeno.

El modelo aprende gradualmente a medida que llegan nuevos lotes de datos, por lo que los parámetros de la ecuación de regresión lineal se modifican sobre las modificaciones de los lotes de datos anteriores. La estrategia de olvido o *forgetting strategy* se implementa en dos elementos diferentes. En primer lugar, se controla el tamaño de la ventana de instancias anteriores del VMAX10m (la dimensión del vector de coeficientes de regresión lineal). En segundo lugar, se modifica de forma adaptativa la tasa de aprendizaje del algoritmo de ascenso gradual. Consideremos primero la regresión lineal como marco de modelización para la predicción del viento como combinación lineal de instancias previas de la variable VMAX10m en cada estación meteorológica.

A continuación, se expondrá de manera teórica el modelo de regresión lineal utilizado, así como las estrategias seguidas para dotarlo de capacidades adaptativas y acumulativas.

4.1.1. Regresión lineal.

Supongamos que existe un conjunto de datos $\{x_i, x_{i1}, \dots, x_{ip},\}_{i=1}^{n}$, donde y_i significa la i-ésima observación de la variable dependiente (a predecir) y la respectiva i-ésima ocurrencia de todas las p variables independientes x_i. Supongamos que también existe un vector Θ que contiene p+1 valores. Entonces, un modelo de regresión lineal puede expresarse como:

$$\mathrm{y}_i = \theta_0 1 + \theta_1 x_{i1} + \theta_2 x_{i2} + \cdots + \theta_p x_{ip} + \epsilon_i, \quad i = 1, \dots, n, \tag{2}$$

donde ϵ es la perturbación o el término de error o, en otras palabras, todo lo que no puede ser explicado por el propio modelo de regresión lineal. En formulación matricial: $Y = X\theta + \epsilon$.

El proceso de aprendizaje de un modelo de regresión lineal consiste en la minimización de ϵ. La elección convencional para la función de coste es el error medio cuadrático, que puede formularse de la siguiente manera:

$$\mathrm{H}_\theta = X^T \Theta, \quad X = 1, x_1, x_2, \dots, x_n \tag{3}$$

$$J(\Theta) = \frac{1}{2m} \sum (H_\Theta - Y)^2 = \frac{1}{2m} \sum_{i=1}^{m} (e_i)^2, \tag{4}$$

donde m simboliza el número de muestras de entrenamiento.

El *gradient descent* es un método clásico para la minimización de una función de coste, un algoritmo de optimización iterativo de primer orden comúnmente utilizado para ajustar los parámetros θ en una función de regresión lineal minimizando J. Esencialmente, calcula la dirección más pronunciada, o gradiente, de la función de coste y ajusta el modelo proporcionalmente a ella. En este trabajo se utiliza el *gradient descent* para actualizar los parámetros θ como se indica en la ecuación (5).

$$\theta_j := \theta_j - \alpha \frac{\delta}{\delta \theta_j} J(\Theta) \tag{5}$$

En otras palabras, cada parámetro θ_i se actualiza de forma inversamente proporcional a la derivada parcial de la función J (coste) con respecto a cada θ_i. En esta ecuación, α es la tasa de aprendizaje que calibra la dimensión del paso dado en la dirección del gradiente en la superficie de la función de coste J. En pocas palabras, α pondera cómo la inclinación de la función de coste J para cada dimensión i de Θ se traduce en un cambio en el respectivo θ_i. Este parámetro α suele ser un valor fijo. En la metodología propuesta, este parámetro α se incrementa o disminuye dependiendo de la evolución de la función de coste en cada ejecución de la rutina de *gradient descent*.

4.1.2. Estrategia de aprendizaje adaptativo basada en Data Stream Mining.

La primera parte de la metodología de aprendizaje adaptativo propuesta consiste en tener un número variable de parámetros θ o instancias anteriores consideradas en el modelo de regresión lineal. Cuando se detecta indirectamente el *concept drift* debido a la degradación observada en el rendimiento del modelo actual, se reduce gradualmente la ventana de observaciones anteriores considerada para el aprendizaje del modelo de regresión lineal. Simplemente se elimina el elemento más antiguo de esta ventana, desactivando ese último parámetro del vector Θ cuando el coste (J) se incrementa en una cantidad estadísticamente significativa (superior al 5%). Por otro lado, cuando se detecta estabilidad por la mejora del rendimiento del modelo a medida que llegan nuevas observaciones, se vuelven a añadir, uno a uno, más parámetros o instancias pasadas de VMAX10m (con el respectivo θ_i inicializado a 0).

Las variables independientes o explicativas consideradas en el modelo entrenado de regresión lineal son los valores pasados de la respuesta o variable dependiente a predecir (viento máximo, VMAX10m) hasta una ventana máxima de valores previos NMax.

Esto significa que en la ecuación (2), la matriz de variables explicativas X se define como sigue para cada estación:

$$X = \left\{1, VMAX10m_{t-H}, VMAX10m_{t-(H+p)}, VMAX10m_{t-(H+2p)}, \dots, VMAX10m_{t-(H+NMAX\times p)}\right\}, \quad (6)$$

donde H es el horizonte de predicción, p es el periodo de tiempo de muestreo y $NMax$ es el número máximo de épocas anteriores consideradas.

La segunda parte de la estrategia de aprendizaje adaptativo propuesta consiste en reducir el parámetro de la tasa de aprendizaje α en la ecuación (5) cuando hay un incremento de la función de coste J. En cada ejecución de la rutina de *gradient descent*, el parámetro α se reduce en un factor muy pequeño. Asimismo, cuando se reduce el coste J, α se incrementa gradualmente.

La Figura 5 muestra la curva de aprendizaje correspondiente al primer lote de datos de la estación meteorológica con ID "C619Y". Se puede observar claramente cómo diferentes valores de α significan diferentes comportamientos de los algoritmos de descenso de gradiente. Valores pequeños de α significan una convergencia lenta, incluso demasiado lenta para converger de manera oportuna. Y valores mayores pueden significar saltos demasiado grandes en el espacio J, llegando a impedir la convergencia en algunos casos. El Algoritmo 1 describe la propuesta de regresión lineal incremental adaptativa basada en el *gradient descent*.

Para reducir el tamaño de la ventana W, primero se calcula $\overline{J0}$ promediando un número (W_J) de valores J anteriores. Para ser un poco más conservador, el J actual solo se considera significativamente mayor que el J_0 promediado si es un 5% mayor.

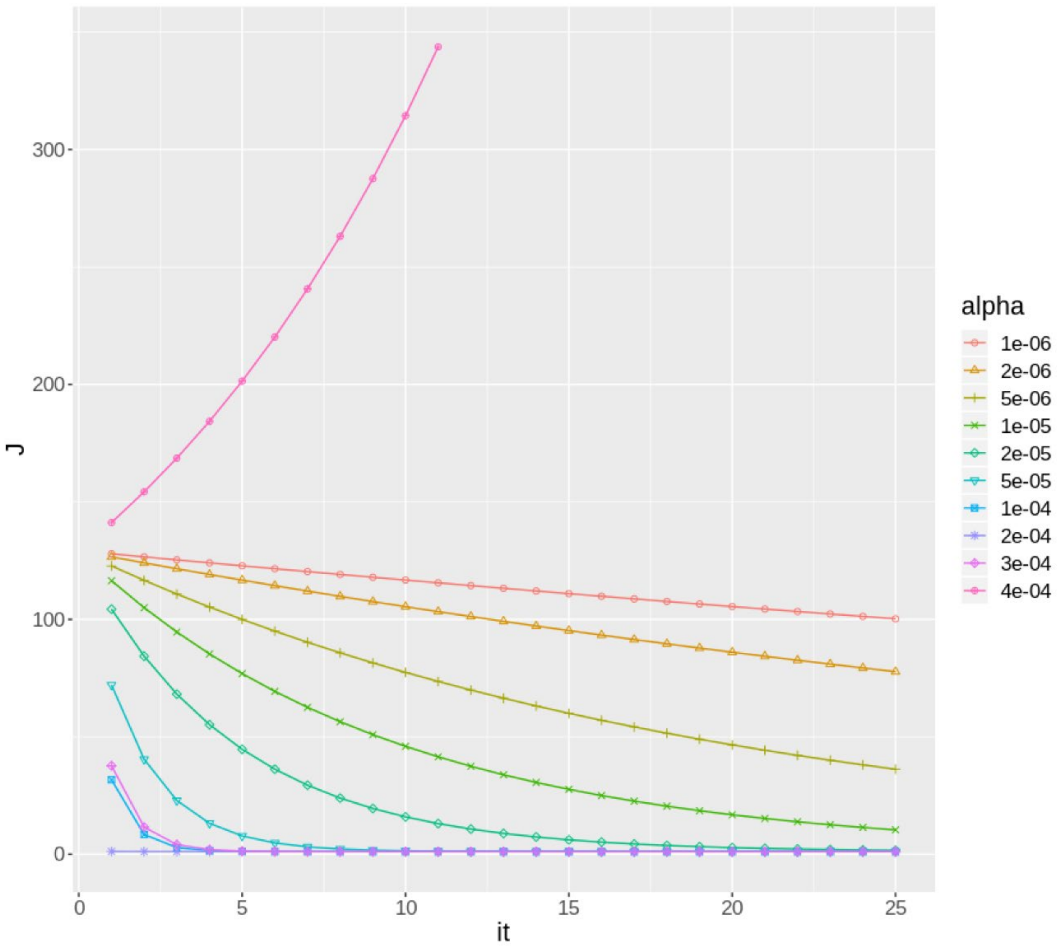


Figura 5 - *Curva de aprendizaje del gradient descent aplicada a la estación con ID C619Y en el primer lote de datos. (Sánchez-Medina et al., 2019).*

W (window size) is initialized max_W;
while *TRUE* **do**
  Data: A new batch of data is loaded (X_b, Y_b)
  Result: Θ is updated (pruned by zero padding, if W is reduced)
  initialization;
  $J = Calculate\_Cost(X_b, Y_b, \Theta, \alpha)$ for the batch is calculated;
  if $J > \overline{J_0} \times 1.05$ **then**
    if $W > min\_W$ **then**
      $W = W - 1$
  else
    if $W < max\_W$ **then**
      $W = W + 1$
  while *Convergence condition is not reached* **do**
    $\Theta = Update\_\Theta\_Gradient\_Descent(\Theta, X_b, Y_b, \alpha)$
  if $J > J_{-1}$ **then**
    $\alpha = \alpha \times (1 - f_\alpha)$
  else
    $\alpha = \alpha \times (1 + f_\alpha)$

Algoritmo 1 – Sistema de aprendizaje adaptivo para la regresión lineal mediante gradient descent. (Sánchez-Medina et al., 2019).

En cuanto a la tasa de aprendizaje α, el método propuesto consiste en reducirla en un pequeño factor $(1 - f\alpha)$ si el coste final J al final del algoritmo de *gradient descent* es mayor que el J obtenido en el lote anterior y, del mismo modo, incrementarla en una pequeña proporción $(1 + f\alpha)$ si el coste se ha reducido o se mantiene igual, siendo $f\alpha$ un pequeño número entre 0 y 1. De esta manera, cuando hay un pequeño deterioro del rendimiento del modelo, el algoritmo de *gradient descent* dará pasos ligeramente más pequeños, y cuando parezca que el modelo está mejorando lote tras lote, se permitirá que α crezca ligeramente, dirigirse hacia la convergencia durante la ejecución del *gradient descent*.

La condición de convergencia de la rutina de *gradient descent* es que, si se alcanza el final del lote o el coste no puede ser más pequeño, entonces se habrán alcanzado el coste medio de la observación anterior o el número máximo de iteraciones.

4.1.3. Estrategia acumulativa.

Como forma de comparar nuestros resultados con la estrategia basada en el flujo de datos se ha diseñado un enfoque fijo alternativo. Los mecanismos de olvido y adaptación se suprimen manteniendo fija la ventana de instancias pasadas de la variable VMAX10m y el valor α de la tasa de aprendizaje ($f\alpha = 0$).

En otras palabras, esta estrategia no considera el *concept drift*, sino que realiza un entrenamiento y aprendizaje incrementales sobre el modelo de regresión lineal afinando los Θ parámetros utilizando todas las nuevas observaciones, las cuales llegan de manera continua e ininterrumpida.

4.2. Desarrollo de un modelo de predicción basado en la Teoría de la Resonancia Adaptativa.

Se crea un modelo adaptativo capaz de realizar predicciones de tráfico fiables con una antelación de 60 minutos en zonas turísticas para dar así respuesta al tercer objetivo específico de esta tesis doctoral. Debido a la ausencia de un catálogo abierto de datos de tráfico en el archipiélago canario, se decidió escoger la ciudad de Madrid como campo de pruebas debido no solo a su elevada significancia turística y a la alta densidad de tráfico dentro de la misma, sino también a su gran infraestructura de captura de datos de tráfico, la cual es accesible de manera sencilla y abierta por cualquier persona en cualquier lugar del mundo sin ningún tipo de registro o trámite previo (Ayuntamiento de Madrid, 2022).

El lenguaje de programación escogido para desarrollar este modelo predictivo fue *Python* debido a la gran potencia y flexibilidad que ofrece en términos de escalabilidad y modularidad, seleccionando Anaconda como gestor de paquetes, *Python 3.8* como distribución para el desarrollo y *Pycharm* como entorno de programación. En primer lugar, se desarrolló un módulo de preprocesamiento de datos para filtrar la gran cantidad de información pura que ofrecen los diferentes *datasets* de tráfico de Madrid.

Dada la alta densidad de tráfico dentro de la autopista M-30, se decidió escoger el parámetro "ocupación" (referente al porcentaje de ocupación de un carril) con el objetivo de predecir posibles atascos con el margen de maniobra suficiente como para organizar cualquier actuación preventiva.

Una vez completado este módulo, se implementó una aproximación basada en la Teoría de Resonancia Adaptativa, siendo el *Laterally Primed Adaptive Resonance Theory*, o *LAPART*, el modelo de aprendizaje escogido. Esto fue debido a que su propia estructura utiliza el aprendizaje incremental y adaptativo para generar modelos predictivos basados en el comportamiento neuronal y memorístico del cerebro humano, reteniendo los datos más relevantes de una observación e identificando casos previamente vistos resistiendo el *concept drift* presente en los datos de tráfico y mostrándose resiliente al paso del tiempo (Healy et al., 1993).

Las redes neuronales ART fueron propuestas por primera vez por Carpenter y Grossberg (1988), y replican cómo el cerebro humano procesa la información cognitiva (Grossberg, 1976a, 1976b). Este tipo de redes neuronales han sido mejoradas y adaptadas a lo largo de los años en aplicaciones tecnológicas con diferentes enfoques, y son capaces de realizar aprendizaje supervisado y no supervisado, predicciones y reconocimiento de patrones. De hecho, este tipo de redes neuronales están siendo utilizadas en la actualidad para resolver problemas que surgen en una amplia variedad de áreas, como los robots autónomos (Murugan et al., 2021), la previsión de la carga eléctrica dentro de *smart grids* (da Silva et al., 2021), la susceptibilidad a las avalanchas de nieve (Yariyan et al., 2022), el análisis de la tierra o el suelo (Hu et al., 2021), la mejora de los datos sanitarios (Shobha y Savarimuthu, 2021), la elaboración de perfiles basados en roles para asegurar los sistemas de bases de datos (Brahma y Panigrahi, 2021), o incluso el reconocimiento de características del comportamiento de los peces (Yang et al., 2021), lo que demuestra que esta metodología muestra una capacidad excepcional para adaptarse a diferentes estructuras y patrones de datos.

Las redes ART cuentan con un proceso de coincidencia de patrones que utiliza la memoria de código interna activa para comparar una entrada externa con entradas previamente almacenadas o conocidas, lo que da lugar a una coincidencia que desencadena una resonancia. Esta resonancia permite al sistema ART aprender una entrada o buscar paralelamente en su memoria una coincidencia mejor. Si el sistema ART considera que la entrada analizada contiene suficiente información nueva, puede actualizar el patrón de resonancia con esa nueva información, pero si la cantidad de información nueva es suficientemente relevante, el sistema puede generar un nuevo patrón para la entrada analizada. Además, las redes ART poseen un sistema de aprendizaje rápido capaz de reaccionar ante entradas poco frecuentes que pueden requerir un recuerdo preciso inmediato. Todas estas cualidades muestran la capacidad de ART para aprender de los cambios en los datos y adaptarse a la nueva información, lo que la convierte en una metodología ideal para analizar datos en línea que presentan cambios en su estructura interna a lo largo del tiempo (*concept drift*).

A continuación, se realiza una descripción en detalle sobre las redes ART y sobre el funcionamiento de un sistema LAPART.

4.2.1. Algoritmos ART.

Es importante señalar que ha habido muchos algoritmos que se han construido a partir del concepto original de ART. El primer grupo de algoritmos ART originales fueron ART1 (Carpenter y Grossberg, 1988; Grossberg, 1987) ART2 y ART2-A (Carpenter, Grossberg, y Rosen, 1991a; Carpenter y Grossberg, 1987) y ART3 (Carpenter y Grossberg, 1990), que admitían entradas binarias, entradas continuas, un tiempo de ejecución extremadamente rápido y la regulación de la actividad sináptica mediante neurotransmisores rudimentarios, respectivamente.

De estos trabajos de investigación se derivan muchas variaciones, como ARTMAP (Carpenter, Grossberg, y Reynolds, 1991), que consisten en un acoplamiento de dos unidades modificadas de ART1 o ART2 en un esquema de aprendizaje supervisado, las activaciones gaussianas (Williamson, 1996), las cuales utilizan el cálculo basado en la teoría de la probabilidad y las funciones de activación gaussianas, logrando que los algoritmos basados en ART sean menos sensibles al ruido, y las ART de lógica difusa (Carpenter, Grossberg, y Rosen, 1991b) mediante la inclusión de un sistema de lógica difusa dentro de la estructura de reconocimiento de patrones de ART para mejorar su generalización. Por último, también se ha creado un sistema basado en el aprendizaje y la interacción de dos módulos ART, creando así el algoritmo ART preparado lateralmente (LAPART) (Healy et al., 1993), el cual se explica en la siguiente subsección.

4.2.2. Algoritmo LAPART.

Este algoritmo se clasifica dentro del grupo de algoritmos ART de lógica difusa debido a su estructura interna. LAPART es capaz de aprender asociaciones para realizar predicciones mediante la unión de dos algoritmos ART difusos, lo que proporciona al sistema la capacidad de converger rápidamente hacia una buena solución debido a su estabilidad nativa. Se entrena de forma precoz e incremental analizando clases de patrones, haciendo una predicción de la siguiente clase utilizando un patrón de una clase anterior que se ha aprendido secuencialmente. Esto hace que LAPART sea capaz de adaptarse tras su entrenamiento inicial, lo que le permite aprender de los nuevos patrones observados y lo hace adecuado para las predicciones de datos no estacionarios.

LAPART está compuesto por dos módulos ART1 (designados como ART1-A y ART1-B), e infiere la pertenencia a la clase del siguiente patrón de entrada utilizando secuencias de clases aprendidas. Se presenta una secuencia de patrones binarios I_k a los campos de entrada de ambos sistemas ART1 (primero a ART1-B y luego a ART1-A). Cuando la secuencia mencionada se presenta a ART1-A, el patrón I_{k+1} se presenta de manera simultánea a ART1-B, lo que hace que las clases se emparejen de forma $A \rightarrow B$, como puede observarse en la Figura 6. Ambos módulos ART1 tienen un nodo de vigilancia que comprueba si las plantillas existentes coinciden suficientemente con la

entrada actual, lo que da a LAPART la capacidad de elegir entre crear una nueva plantilla para una entrada o clasificar esa entrada con una plantilla ya conocida.

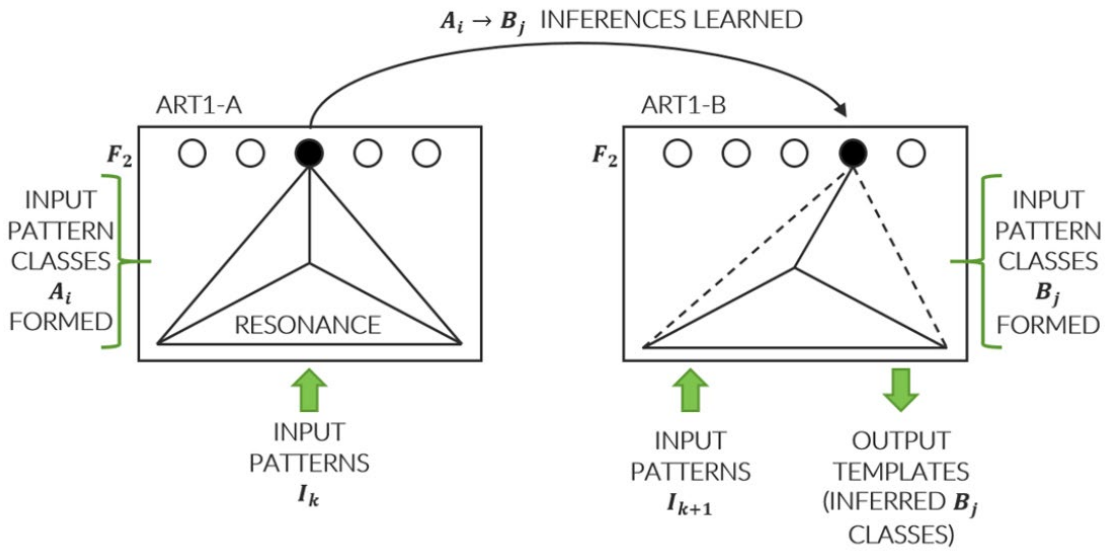


Figura 6 - Estructura del sistema LAPART, basado en el diseño original de (Healy et al., 1993).

Esto hace que LAPART se comporte como un sistema de aprendizaje supervisado, compuesto por dos sistemas no supervisados capaces de interactuar entre sí. Ambos sistemas supervisan mutuamente las relaciones de inferencia clase a clase $A_i \rightarrow B_j$ y las representaciones internas de clase. Este comportamiento genera dos posibles resultados, los cuales se describen a continuación.

El primer resultado posible consiste en que ART1-A no tiene una representación de clase para I_k, por lo que trata esta entrada como un nuevo evento. Debido a que ART1-A no contiene una plantilla resonante para I_k, un nodo F_2 se reserva para almacenar una plantilla basada en la entrada recibida. Después de un retraso controlado por el ART1-A, el ART1-B lee la entrada I_{k+1}, activando tanto su proceso de aprendizaje no supervisado como el aprendizaje sináptico de una conexión *feedforward* entre el nodo F_2 previamente reservado y su nodo de clasificación resonante B. Esta acción genera una fuerte asociación entre una clase A_i y una clase B_j, que contienen I_k e I_{k+1}, respectivamente.

El segundo resultado posible sucede cuando una representación de clase A_i ya existe en ART1-A, lo que produce una resonancia con la entrada actual I_k. Entonces, utilizando una conexión *feedforward* previamente aprendida, ART1-A prepara el nodo de clasificación de ART1-B para su clase asociada B_j. A continuación, el ART1-B lee una plantilla de clase B_j previamente almacenada y, al mismo tiempo, lee la entrada I_{k+1}. Esto permite "forzar" al ART1-B a representar su entrada basándose en la clase preparada seleccionada por el ART1-A antes de considerar cualquier posible plantilla que haya sido filtrada previamente a través de su filtro adaptativo $F_1 \rightarrow F_2$. Sin embargo, ART1-B no puede aceptar, a través de un reinicio mediante su nodo de vigilancia, la clase de la plantilla preparada.

5. Aplicaciones.

A continuación, se expondrán los resultados obtenidos mediante la aplicación de las metodologías previamente mencionadas en los apartados anteriores, dando así respuesta al segundo y tercer objetivos específicos. En primer lugar, se realizará un análisis de los resultados generados por modelo predictivo de regresión lineal adaptado para funcionar de manera online gracias a metodologías de *Data Stream Mining* aplicadas a la predicción de la velocidad máxima del viento en el archipiélago canario y, finalmente, se mostrarán los resultados obtenidos mediante la aplicación de un sistema LAPART basado en la teoría de la resonancia adaptativa a la hora de predecir el grado de ocupación de un carril en la autopista M-30 de Madrid.

5.1. *Data Stream Mining* aplicado a la predicción de la velocidad máxima del viento en las Islas Canarias.

En esta sección se detallará el uso de estrategias de *machine learning* adaptativas e incrementales para predecir la velocidad media del viento máximo (VMAX10m) en un horizonte de 60 minutos, de forma que sea lo suficientemente fiable y robusta como para hacer frente a la diversidad de la posición geográfica de cada estación en la región. Para ello, se ha abordado la tarea de modelización predictiva a través de una metodología basada en el *Data Stream Mining*, lo que significa que los modelos a desarrollar se entrenan de forma incremental y pueden adaptarse a la inestabilidad estocástica del proceso a modelizar, en este caso, la velocidad máxima del viento.

Esta parte del desarrollo de la presente tesis se enmarca en el ViMetRi-MAC ("Sistema de vigilancia meteorológica para el seguimiento de riesgos medioambientales", financiado por el Programa de Cooperación Territorial. INTERREG V A España-Portugal. MAC 2014-2020). Está clasificado en el eje prioritario 3, cuyo objetivo es mejorar la capacidad de respuesta ante posibles riesgos naturales que afecten a los archipiélagos macaronésicos del Espacio Atlántico Norte, incluyendo Madeira, Azores, Cabo Verde y Canarias, con énfasis en la adaptación al cambio climático y la prevención y gestión de riesgos. El principal objetivo de ViMetRi-MAC es promover el desarrollo de sinergias público-privadas para hacer frente a los riesgos vinculados a los fenómenos meteorológicos potencialmente causantes de catástrofes.

Con la metodología propuesta y los resultados obtenidos utilizando estaciones meteorológicas bastante diversas desde el punto de vista geográfico (Figura 7 y Tabla 3), pretendemos demostrar que es posible un enfoque diferente a través de la minería de flujos de datos, pero a costa de revisar y ajustar los métodos existentes para que funcionen en estas configuraciones en línea.

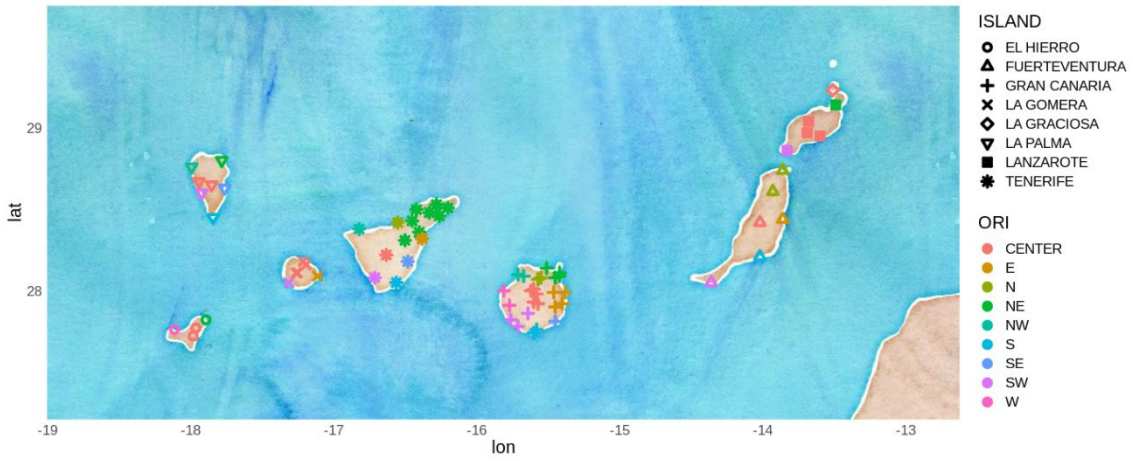


Figura 7 - Estaciones meteorológicas de la AEMET en las Islas Canarias. (Sánchez-Medina et al., 2019).

Tabla 3 - Estaciones meteorológicas de la AEMET en las Islas Canarias (Identificador, localización e isla).

IDEMA	UBI	ISLA	IDEMA	UBI	ISLA
C018J	TIAS-LAS VEGAS	LANZAROTE	C457I	VICTORIA-DEPÓSITO MARRERO	TENERIFE
C019V	YAIZA-PLAYA BLANCA	LANZAROTE	C458A	TACORONTE-A S.E.A.	TENERIFE
C029O	LANZAROTE/AEROPUERTO	LANZAROTE	C459Z	PUERTO DE LA CRUZ	TENERIFE
C038N	HARÍA-CEMENTERIO	LANZAROTE	C469N	SILOS-DEPURADORA	TENERIFE
C048W	TINAJO-LOS DOLORES	LANZAROTE	C611E	SAN MATEO (CORRAL DE LOS JUNCOS)	GRAN CANARIA
C117A	PUNTAGORDA	LA PALMA	C612F	TEJEDA-CRUZ DE TEJEDA	GRAN CANARIA
C117Z	TIJARAFE-MIRADOR TIME	LA PALMA	C614H	TEJEDA CASCO	GRAN CANARIA
C126A	EL PASO-C.F.	LA PALMA	C619X	AGAETE-CASCO	GRAN CANARIA
C129V	FUENCALIENTE-SALINAS	LA PALMA	C619Y	LA ALDEA DE SAN NICOLAS	GRAN CANARIA
C129Z	TAZACORTE	LA PALMA	C623I	SAN BARTOLOME TIRAJANA (CUEVAS DEL PINAR)	GRAN CANARIA
C139E	LA PALMA/AEROPUERTO	LA PALMA	C625O	SAN BARTOLOME TIRAJANA-LOMO PEDRO ALFONSO	GRAN CANARIA
C148F	SAUCES-S. ANDRÉS-BALSA ADEYAHAME	LA PALMA	C628B	SAN NICOLAS T.-TASARTE/COPARLITA	GRAN CANARIA
C229J	PÁJARA-PUERTO MORRO JABLE	FUERTEVENTURA	C629Q	MOGAN (PUERTO RICO)	GRAN CANARIA
C239N	TUINEJE-PUERTO GRAN TARAJAL	FUERTEVENTURA	C629X	PUERTO DE MOGÁN	GRAN CANARIA
C248E	ANTIGUA-EL CARBÓN	FUERTEVENTURA	C635B	SAN BARTOLOME TIRAJANA-H.LAS TIRAJANAS	GRAN CANARIA
C249I	FUERTEVENTURA/AEROPUERTO	FUERTEVENTURA	C639M	SAN BARTOLOME TIRAJANA-C. INSULAR TURISMO	GRAN CANARIA
C258K	LA OLIVA (CARRETERA DEL COTILLO)	FUERTEVENTURA	C639U	SAN BARTOLOME TIRAJANA (EL MATORRAL)	GRAN CANARIA
C259X	LA OLIVA-PUERTO DE CORRALEJO	FUERTEVENTURA	C648C	AGÜIMES-EL MILANO	GRAN CANARIA

IDEMA	UBI	ISLA	IDEMA	UBI	ISLA
C314Z	VALLEHERMOSO-ALTO IGUALERO	LA GOMERA	C648N	TELDE-CENTRO FORESTAL DORAMAS	GRAN CANARIA
C317B	AGULO-JUEGO BOLAS	LA GOMERA	C649I	LAS PALMAS DE GRAN CANARIA/GANDO	GRAN CANARIA
C319W	VALLEHERMOSO-DAMA	LA GOMERA	C649R	TELDE-MELENARA	GRAN CANARIA
C328W	HERMIGUA-DEPÓSITO AYUNTAMIENTO	LA GOMERA	C656V	TEROR-OSORIO	GRAN CANARIA
C329Z	SAN SEBASTIÁN DE LA GOMERA	LA GOMERA	C658X	LAS PALMAS G.C.-TAFIRA/ZURBARÁN	GRAN CANARIA
C406G	CAÑADAS PARADOR	TENERIFE	C659H	LAS PALMAS G.C. SAN CRISTÓBAL	GRAN CANARIA
C419X	ADEJE-CALDERA B	TENERIFE	C659M	LAS PALMAS DE GC. PLAZA DE LA FERIA	GRAN CANARIA
C428T	ARICO-DEPURADORA LA DEGOLLADA	TENERIFE	C665T	VALLESECO	GRAN CANARIA
C429I	TENERIFE/SUR	TENERIFE	C668V	AGAETE - SUERTE ALTA	GRAN CANARIA
C430E	IZAÑA	TENERIFE	C669B	ARUCAS-BAÑADEROS	GRAN CANARIA
C438N	CANDELARIA-DEPOSITO CUEVECITAS	TENERIFE	C689E	MASPALOMAS	GRAN CANARIA
C439J	TENERIFE-GÜIMAR	TENERIFE	C839X	TEGUISE LA GRACIOSA-HELIPUERTO	LA GRACIOSA
C446G	LAS MERCEDES-LLANO LOS LOROS	TENERIFE	C916Q	PINAR-DEPÓSITO	EL HIERRO
C447A	TENERIFE/LOS RODEOS	TENERIFE	C925F	SAN ANDRÉS-DEPÓSITO CABILDO	EL HIERRO
C449C	SANTA CRUZ DE TENERIFE	TENERIFE	C929I	EL HIERRO/AEROPUERTO	EL HIERRO
C449F	ANAGA-COL. REP. ARGENTINA	TENERIFE	C939T	SABINOSA-BALNEARIO	EL HIERRO

5.1.1. Conjunto de datos y análisis exploratorio.

El conjunto de datos utilizado para aplicar la metodología propuesta se ha obtenido a través del proyecto ViMetRi-MAC, proporcionado por la agencia pública española AEMET (Agencia Estatal de Meteorología). Se trata de un conjunto de datos patentado, y no es posible redistribuirlo. Incluye las variables medidas por 68 estaciones meteorológicas repartidas por las Islas Canarias, como se muestra en la Figura 7. Cada estación está equipada con una de las tres plataformas equivalentes siguientes, con firmware personalizado por la AEMET:

- Datalogger DLx-MET
- Vaisala HydroMet System MAWS301
- SEAC EMA55

Las observaciones utilizadas para este trabajo fueron seleccionadas en un periodo de tiempo comprendido entre las 12:10:00 del 26 de abril de 2018 y las 11:50:00 del 13 de diciembre de 2018, hora local, con una tasa de muestreo de 10 minutos. En total, se utilizaron 61.057.225 observaciones para el presente estudio. Los datos utilizados proceden de un conjunto de datos propios de la AEMET, facilitados dentro del proyecto ViMetRi-MAC. Los datos se descargaron de sus servidores en archivos por día estudiado. Cada estación meteorológica estaba equipada para recopilar diversas variables, como la velocidad máxima del viento, la velocidad media del viento, la temperatura, la humedad, la precipitación, la presión atmosférica, etc.

La Tabla 3 muestra la ubicación (UBI), el código de identificación (IDEMA) y la isla para cada una de las estaciones meteorológicas. Las Islas Canarias (España) son un grupo de ocho pequeñas islas (Gran Canaria, Tenerife, La Palma, La Gomera, El Hierro, Fuerteventura, Lanzarote y La Graciosa) situadas frente al sur de Marruecos, en el recuadro delimitador 29°29'08,4"N, 13°22'18,8"W y 27°43'21,7"N, 18°11'34,8"W. La población de las islas es de 2.127.685 habitantes, de los cuales el 42,5% se encuentra en Gran Canaria y el 39,8% en Tenerife, según el ISTAC. La principal actividad económica de las islas es el turismo. En general, hay 24.368 empresas establecidas en Gran Canaria, 27.881 en Tenerife y 12.135 en las demás islas.

La variable utilizada para la predicción del viento fue VMAX10m, que es la velocidad máxima del viento (m/s). En la Figura 8 se representa tanto VMAX10m como VV10m, que es la velocidad media del viento, ambas suavizadas a lo largo del periodo muestreado mediante un modelo aditivo generalizado (GAM) (Hastie y Tibshirani, 2017), proporcionado por la librería ggplot2 (Wickham, 2016).

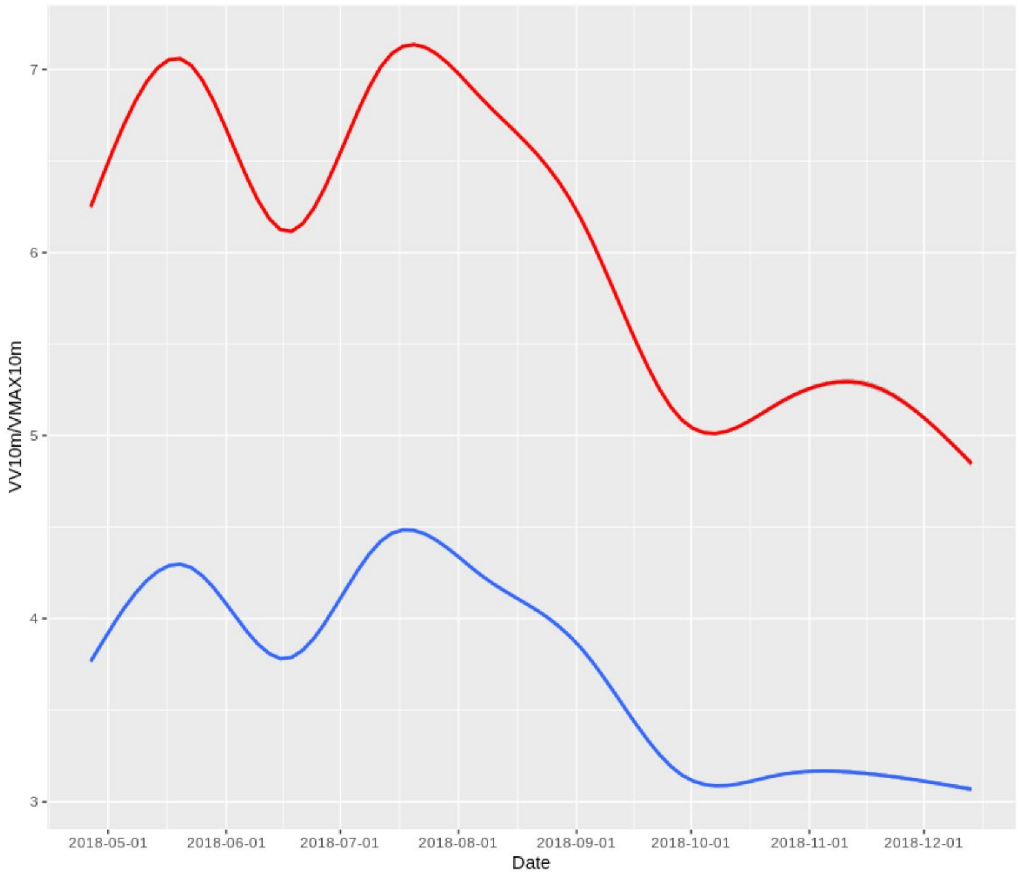


Figura 8 - *Evolución suavizada (mediante GAM) del parámetro VMAX10m (velocidad máxima del viento en rojo, parte superior) y VV10m (velocidad media del viento en azul, parte inferior) durante el periodo muestreado. (Sánchez-Medina et al., 2019).*

En los estudios preliminares realizados con este conjunto de datos, la variable explicativa más importante para la velocidad máxima del viento actual fue, con diferencia, la velocidad máxima del viento anterior. En otras palabras, VMAX10m presentaba un nivel de autocorrelación muy elevado, lo que motivó que se utilizara sólo esa variable y las instancias anteriores de la misma como variables independientes del

modelo propuesto. Sin embargo, no se puede descartar añadir otras variables al modelo para mejorar su eficiencia en una futura ampliación de esta investigación.

En la Figura 9 se puede observar que existe una clara componente norte para los vientos en Canarias, conocida por la comunidad científica como "viento comercial" por sus implicaciones históricas en el comercio español y portugués con Centro y Sudamérica durante y después del siglo XVI. Esa cifra considera sólo el mes más ventoso de 2018, que fue julio, para cada isla. Esta imagen se comparte para mostrar que, incluso cuando hay un predominio de los vientos del norte en julio, las rosas de los vientos para las ocho islas son muy diferentes. Según Bechtel (2016) y Rodríguez et al. (2010), este archipiélago presenta numerosos microclimas. Por lo tanto, un modelo entrenado para predecir el viento tiene que ser generalizable.

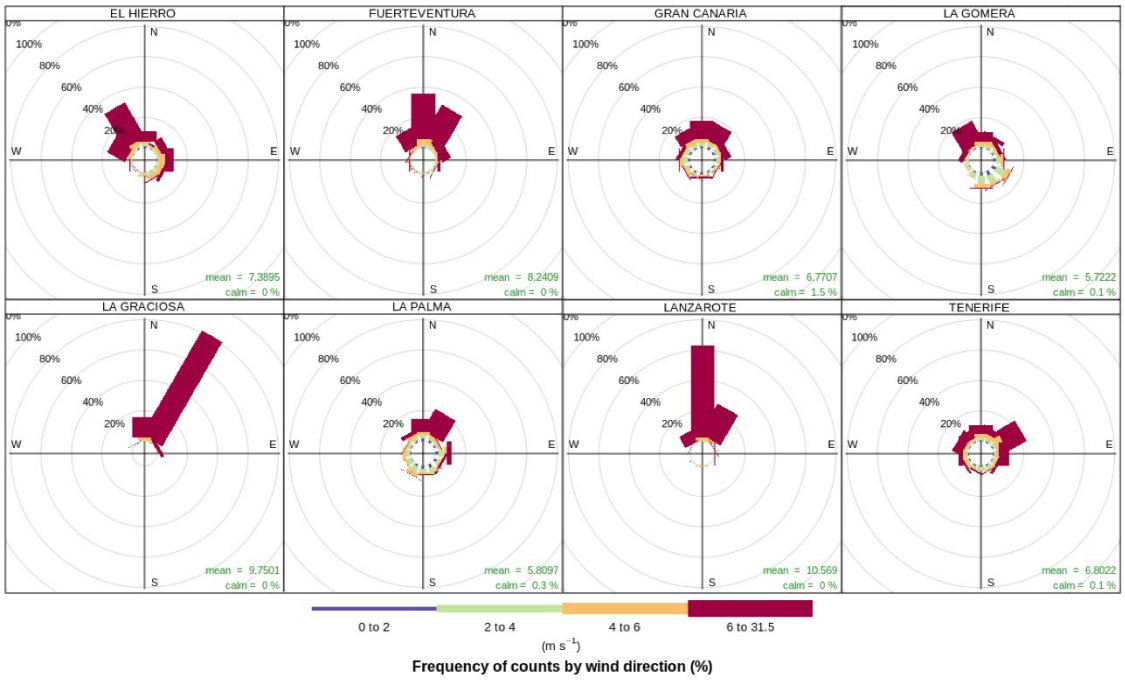


Figura 9 - *Rosas del Viento para cada isla en julio, basadas en el parámetro VMAX10m. (Sánchez-Medina et al., 2019).*

5.1.2. Experimentación y resultados.

Como se ha explicado anteriormente, para la aplicación de esta metodología las observaciones utilizadas se muestrearon desde las 12:10:00 del 26 de abril de 2018 hasta las 11:50:00 del 13 de diciembre de 2018, hora local de Canarias, a una tasa de muestreo de 10 min, recogiendo 61.057.225 observaciones en total.

El tamaño W inicial se fijó en 20, lo que significa 20 valores previos de VMAX10m, que van desde 60, 70, 80 hasta 250 min antes, muestreados cada 10 min. El tamaño mínimo de la ventana fue de 3 y el máximo de 20 desfases temporales. Esto significa que muestras de VMAX10m con un desfase desde 30 min hasta 4 h y 10 min son utilizadas para predecir 60 min en el futuro. Cuando se detecta indirectamente *concept drift* se reduce el número de instancias anteriores y, cuando se observa estabilidad, W se amplía hasta un máximo de 20.

Station	Accumulative Cost	Adaptive Cost	delta	Station	Accumulative Cost	Adaptive Cost	delta
C018J	1.61375426020819	0.736620985421285	0.877133274786904	C457I	0.839705539779162	0.519550664657003	0.320154875122159
C019V	1.878355459473	1.01303685263121	0.865318606841788	C458A	1.16721483639045	0.667056432945642	0.500158403444811
C029O	2.50941790757332	1.23698814891607	1.27242975865725	C459Z	1.14510393711185	0.639670118067816	0.50543381904403
C038N	1.6331396226027	0.876875005595046	0.756264617007655	C469N	1.06005736966017	0.463144903303695	0.596912466356476
C048W	1.39279973545229	0.747621980109606	0.645177755342681	C611E	1.1832105146369	0.616083409840493	0.567127104796405
C117A	1.54275551383038	0.735285484320889	0.807470029509494	C612F	2.0988036555081	1.02573109071779	1.07307256479031
C117Z	1.34254433286014	0.799856345613295	0.542687987246849	C614H	3.00788066682113	1.33211336366056	1.67576730316058
C126A	3.28691675415722	1.43777989451821	1.84913685963902	C619X	1.83691716397217	0.979284250906172	0.857632913065993
C129V	2.97587563821671	1.40869273463594	1.56718290358077	C619Y	4.51038917410384	1.83361119967339	2.67677797443044
C129Z	1.09389967799579	0.620063005940674	0.473836672055114	C623I	1.22384431336633	0.747689715802761	0.476154597563565
C139E	2.23076263362239	0.986847168490131	1.24391546513226	C625O	1.40545043888253	0.792201283241206	0.613249155641323
C148F	1.15697866506076	0.661473674246127	0.495504990814633	C628B	2.92609613736441	1.57479499310418	1.35130114426022
C229J	3.49354149961246	1.77829930850775	1.71524219110471	C629Q	1.54083631923994	0.959745332186988	0.581090987052952
C239N	2.22253349438313	1.22092061784035	1.00161287654278	C629X	1.97924349651942	1.12466274132171	0.854580755197713
C248E	1.78556572649646	1.03262788786312	0.752937838633341	C635B	1.19069575550491	0.540410911843024	0.650284843661886
C249I	1.56478851085226	0.855590520304028	0.70919799054823	C639M	1.32863094114035	0.842925657800837	0.485705283339513
C258K	1.27419055605137	0.764010723672497	0.510179832378873	C639U	3.04580285290737	1.41544625876787	1.6303565941395
C259X	2.09660185298836	1.13836172505819	0.958240127930171	C648C	4.13211219391727	1.8057457200552	2.32636647386207
C314Z	2.9793735169288	1.19209886834214	1.78727464858666	C648N	1.35417714223871	0.777017197619144	0.577159944619565
C317B	1.02669830811662	0.568331747240371	0.458366560876246	C649I	1.80636051812577	0.944458266105895	0.861902252019871
C319W	1.27614758138065	0.662885364452939	0.613262216927711	C649R	0.863437663222217	0.499340316398164	0.364097346824053
C328W	1.13551933116851	0.653581401411116	0.481937929757392	C656V	0.980764958569705	0.580205613946806	0.400559344622899
C329Z	2.84704442590661	1.31655473100597	1.53048969490064	C658X	0.837395750299986	0.473549820950731	0.363845929349255
C406G	2.42911402348815	1.35786350054872	1.07125052293942	C659H	1.4746354709119	0.722298737657487	0.75233673325441
C419X	1.00568138700447	0.629712650652642	0.375968736351827	C659M	0.753042011034233	0.414396106515115	0.338645904519118
C428T	1.93778372027019	1.09981540249901	0.837968317771183	C665T	1.04222267810124	0.565936488496188	0.476286189605051
C429I	3.01265857667598	1.48922711777086	1.52343145890512	C668V	3.34551228758179	1.59652483194243	1.74898745563936
C430E	2.96684267021323	1.2151632274649	1.75167944274833	C669B	0.825501955450797	0.463793594538346	0.361708360912451
C438N	1.97564374310652	1.0466977497186	0.928945993387927	C689E	2.59381140157255	1.32712549004557	1.26668591152698
C439J	2.59678646845612	1.15851895881336	1.43826750964277	C839X	1.15833659518073	0.539804913221394	0.618531681959335
C446G	2.92318389498232	1.50032505350965	1.42285884147267	C916Q	1.90490035765184	1.05282194159438	0.852078416057467
C447A	1.79757396262073	0.833812733663265	0.963761228957466	C925F	2.01412184115637	1.03763382712108	0.976488014035291
C449C	2.03496185728843	1.08563458557613	0.949327271712296	C929I	2.03940278393262	0.913065798943753	1.12633698498887
C449F	1.47025192622584	0.886884159674865	0.583367766550978	C939T	1.72684428239572	0.902405443496627	0.824438838899097

Figura 10 - *Coste medio acumulativo y adaptativo ((m/s)²) para las estaciones meteorológicas (valores estimados frente a valores observados para las dos estrategias), y diferencia entre ambos. (Sánchez-Medina et al., 2019).*

El valor inicial de α (la tasa de aprendizaje) se fijó en $1e^{-4}$. El tamaño del lote fue de 100 observaciones. El número máximo de iteraciones del algoritmo de *gradient descent* (si la convergencia no ha sido alcanzada previamente) se fijó en 1000. $\overline{J0}$ se calcula promediando 10 valores J anteriores ($W_J = 10$). Por último, el valor $f\alpha$ es de 0,01 (1 por ciento) para aumentar o reducir el valor de α durante la ejecución del *gradient descent*.

A continuación, se presentan los resultados de la aplicación de la metodología adaptativa basada en el flujo de datos frente al método lineal "acumulativo", habiendo sido explicadas ambas estrategias en la sección anterior. En la tabla representada en la Figura 10, se enumeran los valores de coste medio (error cuadrático medio, MSE) obtenidos con ambas metodologías en todo el conjunto de datos. La estrategia adaptativa fue siempre superior a la acumulativa, como puede observarse en la columna "delta" (Coste medio acumulativo-Coste medio adaptativo).

En la Figura 11 puede observarse la misma evidencia de forma más visual. Los valores extremos de los costes obtenidos para la estrategia acumulativa en comparación con la estrategia adaptativa son otra prueba de gran importancia a señalar en este análisis. Esto parece confirmar la hipótesis de que la metodología adaptativa es capaz de hacer frente al *concept drift* mucho más rápidamente que la metodología acumulativa, la cual sufre de una inercia mucho mayor debido a los datos anteriores en su modelo.

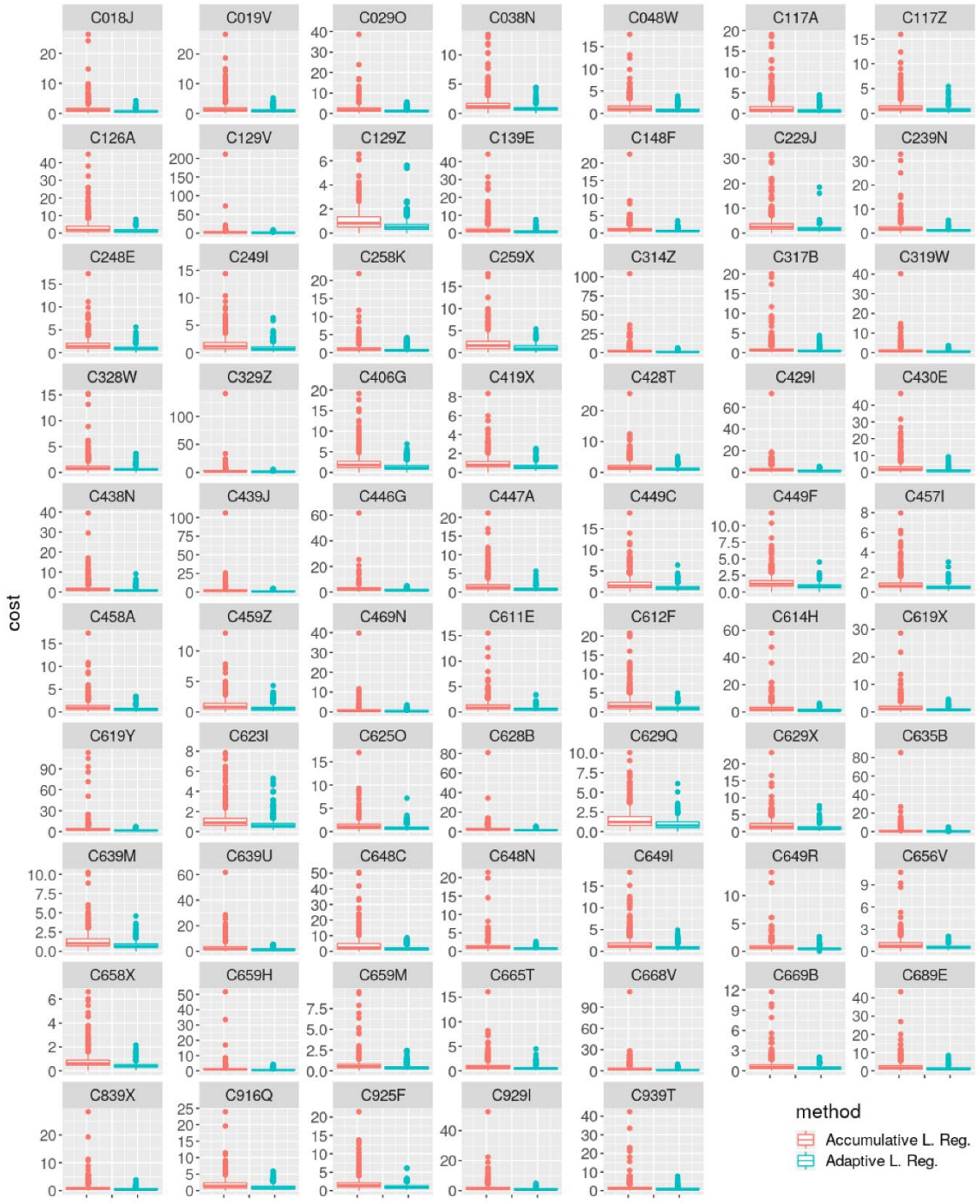


Figura 11 - Coste (Error Cuadrático Medio, MSE), $(m/s)^2$ representado en box-plots, comparando ambas metodologías para cada estación. (Sánchez-Medina et al., 2019).

Para una comprensión más detallada de la metodología propuesta, los resultados de los mejores y peores casos se muestran en la Figura 12 y la Figura 13, respectivamente. En la parte inferior de los gráficos puede verse la evolución del valor del coste (MSE) a lo largo del tiempo para ambas metodologías. La curva púrpura (superior) de la Figura 12 y la Figura 13 es el valor del coste resultante para la estrategia acumulativa, y en la curva negra punteada (inferior) puede observarse la evolución del valor del coste para la estrategia adaptativa.

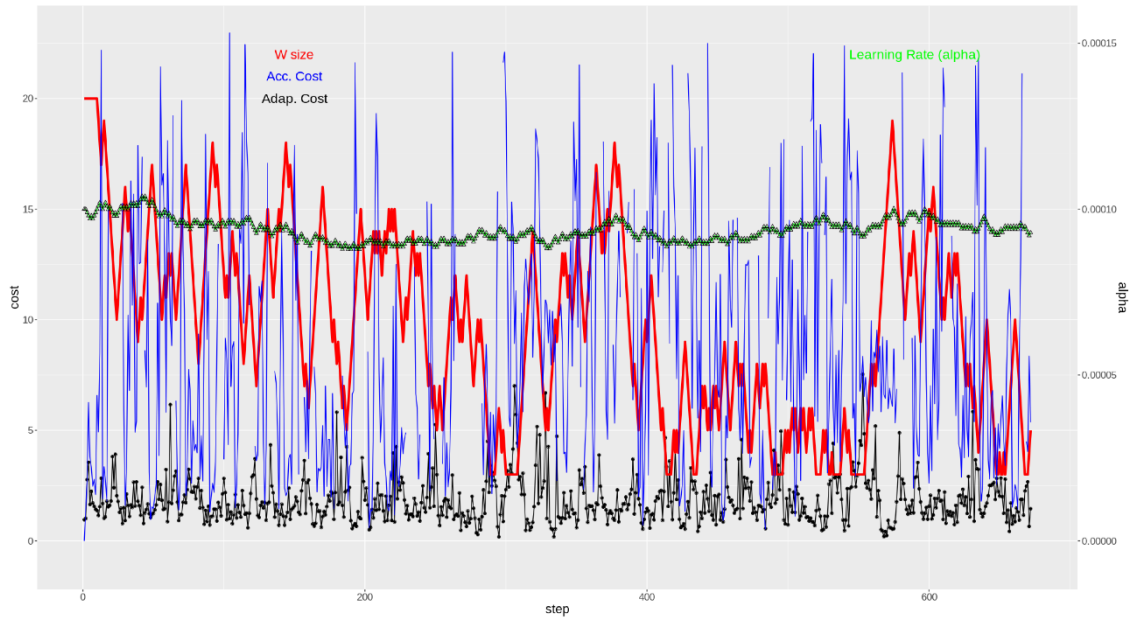


Figura 12 - *Estación meteorológica ID C619Y: Caso de mejor rendimiento para el método adaptativo. En la parte inferior, coste (MSE, (m/s)²) obtenido con la estrategia adaptativa (negro) y con la estrategia acumulativa (azul). En la parte superior, tamaño de la ventana (rojo) y tasa de aprendizaje (verde), ambos para la estrategia adaptativa. (Sánchez-Medina et al., 2019).*

En la parte superior de las dos figuras (Figura 12 y Figura 13), hay dos curvas relativas a la metodología de la estrategia adaptativa. Aquí se representa la evolución de la ventana W de instancias previas de tamaño VMAX10m (en rojo, más gruesa), pasando de 3 instancias previas a un máximo de 20. Por último, se muestra cómo evoluciona el valor α a lo largo del tiempo (línea verde formada por triángulos).

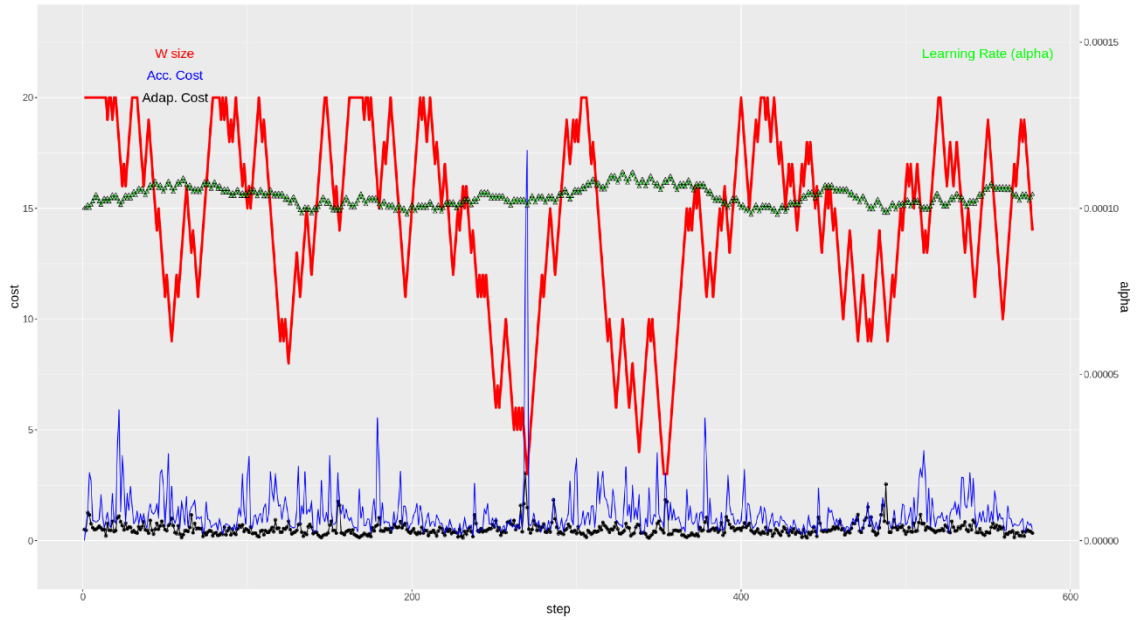


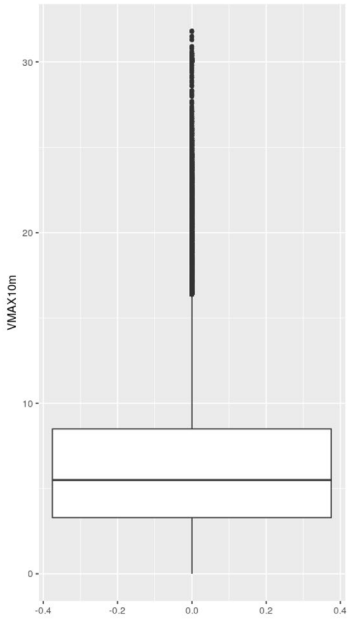
Figura 13 - *Estación meteorológica ID C457I: Caso de peor rendimiento para el método adaptativo. En la parte inferior, coste (MSE, (m/s)²) obtenido con la estrategia adaptativa (negro) y con la estrategia acumulativa (azul). En la parte superior, tamaño de la ventana (rojo) y tasa de aprendizaje (verde), ambos para la estrategia adaptativa. (Sánchez-Medina et al., 2019).*

Debe observarse que, en el gráfico de la Figura 12, los valores extremos de la representación de los costes de la estrategia acumulativa se han recortado para obtener un gráfico más claro. Véase el gráfico de caja de la estación meteorológica "C619Y" en la Figura 11 para ver el rango real de los cálculos de costes extremos para esa estación.

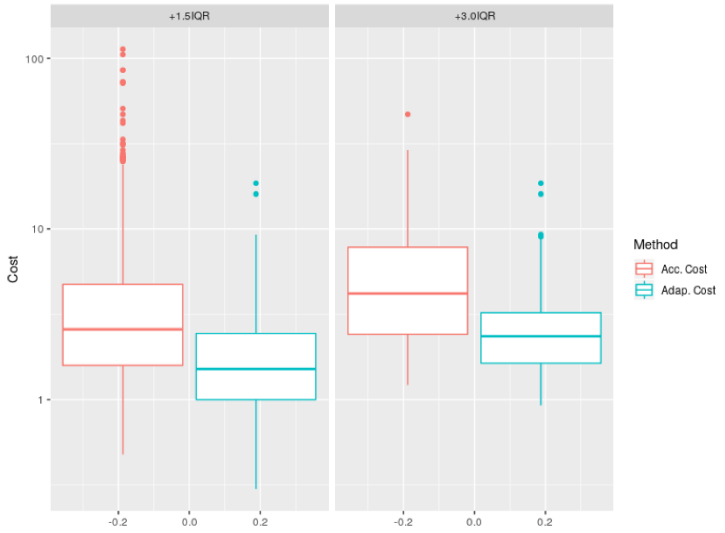
En ambos casos extremos, queda claro cómo la estrategia adaptativa contiene el error cuadrático medio de la función de regresión.

En la Figura 14a se representa la distribución de los valores de VMAX10m para todo el periodo estudiado. En dicha figura se han analizado por separado las situaciones de mayor velocidad de viento, concretamente los episodios de VMAX10m por encima del tercer cuartil más 1,5 veces el rango intercuantil y aquellos por encima del tercer cuartil más 3 veces el rango intercuantil ($Q_3 + 1{,}5 \times IQR$ y $Q_3 + 3 \times IQR$). Para los datos estudiados, $Q_3 = 8{,}5\, m/s$ y $IQR = 5{,}2\, m/s$. Por lo tanto, los dos umbrales establecidos para este análisis fueron VMAX10m por encima de 16,3 m/s (36,5 mph, 58,7 Km/h) y VMAX10m por encima de 24,1 m/s (53,91 mph, 86,76 km/h).

En la Figura 14b, se muestra la representación, en forma de *box-plot,* del coste obtenido en todas las estaciones meteorológicas durante los episodios de velocidad VMAX10m atípica. En la Figura 15, se representan los episodios atípicos VMAX10m considerados para todo el periodo estudiado. Los episodios de velocidad máxima del viento por encima del primer umbral representan el 6,39% de todas las observaciones, mientras que los episodios de velocidad máxima del viento por encima de 24,1 m/s suponen el 0,21%.



(**a**) VMAX10m (m/s) values distribution.



(**b**) Outliers cost (MSE $(\mathrm{m/s})^2$) representation for each weather station, $VMax10m > Q_3 + 1.5 \times IQR$ (16.3 m/s , 36.5 mph , 58.7 Km/h), and $VMax10m > Q_3 + 3.0 \times IQR$ (24.1 m/s, 53.91 mph, 86.76 km/h).

Figura 14 – Análisis de outliers (VMAX10m > 16.3 m/s). (Sánchez-Medina et al., 2019).

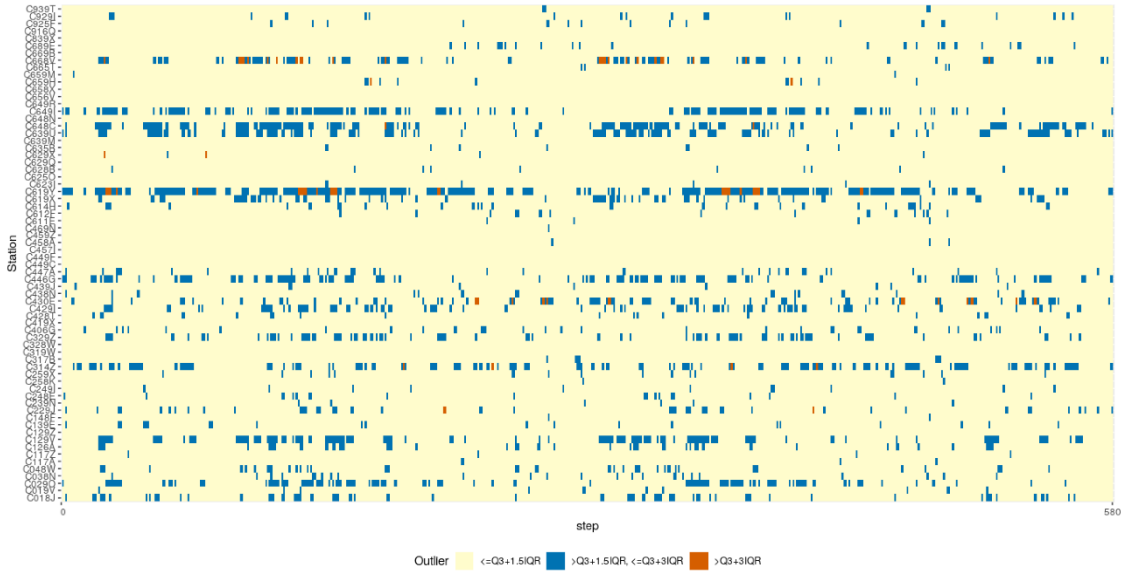


Figura 15 - Representación del coste de los valores atípicos (MSE (m/s)^2) para cada estación meteorológica, VMax10m > Q_3+1,5×IQR (16,3 m/s, 36,5 mph, 58,7 Km/h), y VMax10m > Q_3+3,0×IQR (24,1 m/s, 53,91 mph, 86,76 Km/h). (Sánchez-Medina et al., 2019).

La Figura 16 muestra la representación, en forma de *box-plot*, del coeficiente de determinación de la ecuación (7) calculado a lo largo del periodo de estudio para ambos modelos. La r^2 es una medida estadística que ofrece información sobre la bondad del ajuste de un modelo. Indica qué parte de la varianza de la variable de respuesta (VMAX10m en nuestro caso) puede ser explicada por el modelo, o lo bien que el modelo de regresión se aproxima a las nuevas observaciones.

$$r^2 = \frac{\sum(\hat{y}_i - \overline{y})^2}{\sum(y_i - \overline{y})^2} \quad (7)$$

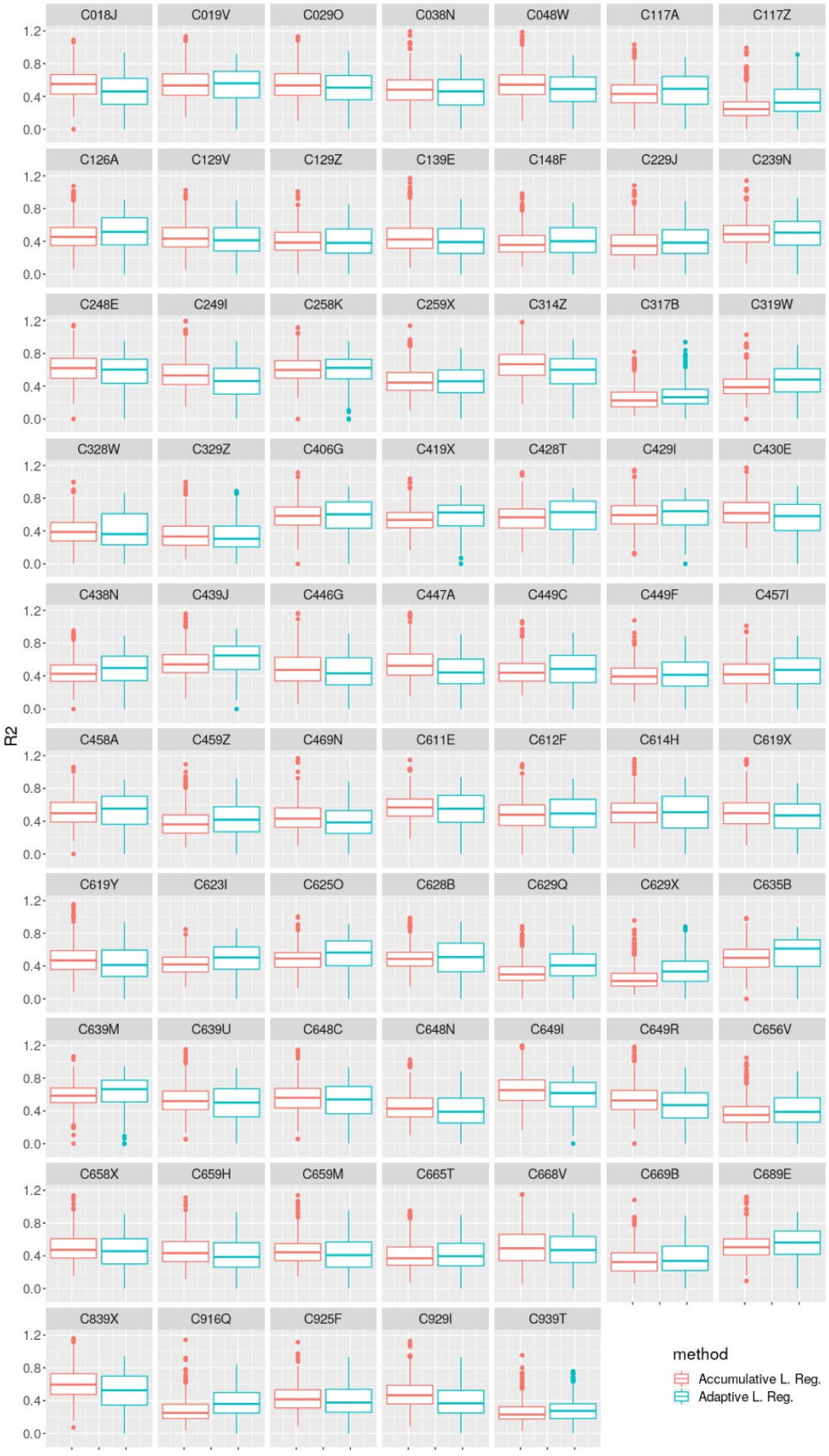


Figura 16 – Comparación en box-plots del coeficiente de determinación r^2 entre ambas estrategias para cada estación meteorológica. (Sánchez-Medina et al., 2019).

5.1.3. Discusión de resultados.

En la sección anterior se han mostrado los resultados de la aplicación de las dos estrategias propuestas, la acumulativa y la adaptativa. Mediante la inspección del coste medio, calculado como el MSE de la VMAX10m estimada frente a la VMAX10m real a lo largo de todo el periodo de estudio, visible en la Figura 10, se puede observar que para cada caso (cada estación meteorológica), el coste obtenido es siempre mayor (peor) cuando se utiliza la estrategia acumulativa. Esto significa que, aun reconociendo que los resultados son más significativos para algunas estaciones que para otras, en todos los casos, el modelo de regresión adaptativa supera a la estrategia acumulativa.

La Figura 11 muestra la distribución de los costes a lo largo de todo el periodo de estudio. Esta figura ha sido añadida con motivo de mostrar la distribución de los valores atípicos o extremos del coste. Aquí, además de mostrar el mejor rendimiento de la estrategia adaptativa para cada estación, queda patente que hay picos o costes mucho mayores para la estrategia acumulativa que para la adaptativa. Esto parece estar relacionado con los episodios de *concept drift*, en los que el concepto previamente aprendido fue considerado de manera excesiva en el modelo, afectando negativamente a la predicción del concepto actual. La estrategia adaptativa parece más reactiva a esos cambios de concepto.

Si se observan las Figuras 12 y 13 con atención, el *concept drift* puede detectarse indirectamente al observar la rápida reducción del tamaño de la ventana (en rojo). En estos episodios resulta evidente que el modelo construido con la estrategia acumulativa (en azul) tiene un gran pico en el coste calculado. Sin embargo, la estrategia adaptativa tiene un comportamiento mucho más contenido, aparentemente debido a su adaptación al nuevo concepto a modelar.

En las Figuras 14a y 14b, y en la Figura 15, se muestra el comportamiento de las dos estrategias, comparadas en la parte superior del rango VMAX10m. En concreto, se seleccionaron los eventos de VMAX10m por encima de dos niveles de umbral definidos ($Q_3 + 1{,}5 \times IQR$ y $Q_3 + 3 \times IQR$), en los que había velocidades de viento máximas superiores a 16,3 m/s y 24,1 m/s. Todo parece apoyar la hipótesis de que en las condiciones (poco frecuentes) de velocidad del viento extrema en las que se centra esta investigación, de nuevo la estrategia adaptativa supera a la acumulativa (Figura 14b). Observando la Figura 15, sería posible realizar un análisis más detallado estableciendo un umbral diferente para cada estación, ya que evidentemente hay vientos fuertes más frecuentes en diferentes estaciones meteorológicas. Sin embargo, se ha entendido que, gracias a los resultados obtenidos con este umbral global, la conclusión de que la estrategia adaptativa también supera a la acumulativa en los regímenes de alta velocidad del viento queda demostrada.

Por último, de la Figura 16 se deduce una conclusión muy importante que puede servir de motivación para futuras investigaciones. Un coeficiente de determinación superior a 0,6 se toma convencionalmente como umbral para considerar que un modelo explica una cantidad suficiente de la varianza de las variables independientes. Sin embargo, en los experimentos realizados, ninguna de las estrategias acumulativas o

adaptativas puede considerarse suficientemente ajustada para todas las estaciones meteorológicas y, adicionalmente, ninguna de las estrategias supera a la otra en ese aspecto. La conclusión es que, incluso cuando hay pruebas que muestran una clara mejora del rendimiento utilizando una estrategia adaptativa, el sesgo de un modelo de regresión lineal impide que los modelos obtenidos expliquen una parte suficiente de la varianza observada. Esto proporciona una razón para ampliar la presente investigación y considerar la adaptación de otros modelos de base no lineal para que funcionen de forma adaptativa.

5.2. Una aproximación basada en la resonancia neural para tratar flujos de datos no estacionarios de tráfico.

Canarias aun no posee una red de tráfico debidamente sensorizada que arroje datos suficientes como para generar un modelo predictivo. No obstante, y gracias a las capacidades adaptativas de las redes ART, es posible generar este tipo de modelo en base a datos de otra localización y extrapolarlo al entorno de tráfico canario una vez se cuente con suficientes datos. Por ello, se ha decidido utilizar los datos generados por la red de tráfico del Ayuntamiento de Madrid, concretamente los datos capturados por la red de sensores de la autopista M-30, la cual rodea a la ciudad y provee de diferentes accesos a puntos de gran interés turístico.

Para realizar dicho modelo se ha decidido utilizar un módulo ART de preparación lateral (LAPART) para aumentar la precisión de la predicción debido a su capacidad inherente de converger rápidamente hacia una solución fiable. Adicionalmente, los resultados obtenidos se han comparado con aquellos generados por las técnicas de descenso por gradiente estocástico (SGD), árboles adaptivos de Hoeffding y bosques aleatorios adaptivos con objetivo de comparar su rendimiento y su precisión.

Las tres principales contribuciones de esta parte de la investigación realizada pueden resumirse en:

- La propuesta de una nueva aplicación de una red neuronal ART para predecir con precisión la congestión del tráfico en función de dos o más parámetros de la carretera, como la velocidad media o la intensidad del tráfico. Este enfoque converge rápidamente hacia una solución y presenta adaptabilidad a la deriva conceptual.
- La utilización de datos de tráfico reales, preprocesados previamente para simular flujos de datos, con el objetivo de comparar diferentes modelos de referencia con una red neuronal ART.
- Una explicación detallada sobre cómo se ha realizado el preprocesado de los datos para facilitar la comprensión de la metodología propuesta.

En los epígrafes siguientes se explica de forma detallada todas las fases del procedimiento seguido en la obtención de los resultados.

5.2.1. Descripción del conjunto de datos: la autopista M-30 de Madrid.

En la actualidad, la autopista M-30 sigue siendo la carretera más transitada de España. Esta autopista posee una longitud de 32,5 kilómetros, rodeando los distritos centrales de la ciudad en los que viven más de 800.000 personas y siendo el anillo más interior de un sistema vial orbital de cuatro anillos compuesto por las autopistas M-30, M-40, M-45 y M-50. La M-30 está equipada con más de 400 sensores de tráfico que emiten diferentes datos, que se almacenan en un repositorio oficial de datos abiertos (Ayuntamiento de Madrid, 2022).

5.2.1.1. Características del conjunto de datos.

Los datos de tráfico se capturan en "puntos de control" distribuidos a lo largo de la M-30. Dichos datos se almacenan en archivos CSV con una frecuencia de muestreo de quince minutos y divididos en meses. La Tabla 4 muestra un extracto de cómo se presenta la estructura de los archivos CSV, y a continuación se define el significado de cada una de las nueve características que componen cada instancia:

- *Id:* Identificador del punto de control.
- *Fecha:* Fecha de medición de los datos, en formato *'aaaa-mm-dd hh:mm:ss*'.
- *Tipo de elemento:* Define si el sensor pertenece a la autopista M-30 (M30) o a otra vía interurbana (URB).
- *Intensidad:* Número de vehículos por hora.
- *Ocupación:* Define el porcentaje de ocupación del punto de control por parte de los vehículos.
- *Carga:* Parámetro de carga de la autopista en función de la intensidad, la ocupación de los carriles y las características de la infraestructura.
- *Velocidad media:* Velocidad media de los vehículos detectados en el periodo de integración.
- *Error:* Código de control de la validez de los datos del punto de medición.
- *Período de integración:* Número de muestras recibidas antes de la integración.

Tabla 4 - Breve extracto de los datos almacenados en los archivos CSV de tráfico de la autopista M-30 de Madrid.

ID	Fecha	Tipo_Elem	Intensidad	Ocupación	Carga	Vmed	Error	Periodo_Integración
…	…	…	…	…	…	…	…	…
1001	2019-01-03 15:15:00	M30	2532	8	0	62	N	5
1001	2019-01-03 15:30:00	M30	2628	6	0	64	N	5
1001	2019-01-03 15:45:00	M30	2580	7	0	63	N	5
1001	2019-01-03 16:00:00	M30	2136	7	0	60	N	5
…	…	…	…	…	…	…	…	…

El parámetro "*Ocupación*" fue escogido para realizar predicciones debido a su importancia en relación con los posibles atascos. Para predecirlo, se desecharon los

parámetros “*Tipo_Elem*” y “*Periodo_Integración*” por su falta de relación con los datos principales. El parámetro “*Carga*” también ha sido descartado, ya que no existe documentación oficial que permita formular la función utilizada para relacionar la intensidad, la ocupación y la velocidad media a partir de las definiciones de datos presentadas por la *Dirección General de Gestión y Vigilancia de la Circulación de Madrid*. El parámetro "*ID*" se ha usado para seleccionar el sensor que se utilizará para el entrenamiento y la previsión, y el parámetro "*Fecha*" se ha utilizado también para crear flujos de datos por orden o llegada.

5.2.1.2. Preprocesado del conjunto de datos.

Tal como se ha mencionado previamente, los datos brutos captados por los sensores de la autopista necesitan ser preprocesados con el objetivo de ser utilizados para entrenar los diferentes modelos. Para ello, se utilizó *Python 3.8* para crear un módulo de preprocesamiento de datos capaz de manejar los datos brutos, procesarlos y entregar el resultado a los algoritmos de aprendizaje para que sean entrenados con ellos. Debido a que la cantidad de datos es de un gran tamaño, se utilizó el módulo *Numpy* dentro de *Python* para facilitar el proceso de tratamiento de *arrays* de gran tamaño. Adicionalmente, también se ha utilizado el módulo *Pandas* para manejar de manera más eficiente y sencilla estos grandes conjuntos de datos, almacenándolos óptimamente dentro de *dataframes*. La Figura 17 ilustra el proceso de preprocesamiento.

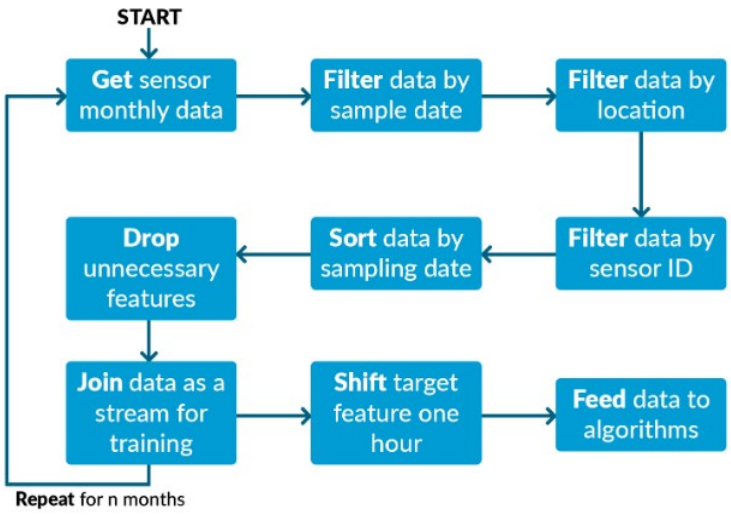


Figura 17 - Pipeline de preprocesamiento de datos. (Elaboración propia).

Supongamos que queremos entrenar nuestros modelos de forma incremental, con un mes de datos disponibles. Estos datos deben ser entregados a los algoritmos como un flujo de datos para que los modelos aprendan de forma incremental. Dado que los datos del M-30 se almacenan en archivos CSV de manera mensual, es imprescindible preprocesar dichos datos para convertirlos en un flujo de datos continuo y estable.

En primer lugar, el módulo de preprocesamiento realiza un triple filtrado de los datos. Comienza por obtener el primer archivo CSV y lo carga en un *dataframe* de Pandas. Como es posible solicitar un rango de tiempo fijo dentro de un mes, el módulo filtra los datos basándose en un intervalo de tiempo definido por el usuario, resolviendo automáticamente cualquier problema relacionado con los años bisiestos y los números de los días. A continuación, el módulo de preprocesamiento filtra los datos por ubicación, dejando fuera los sensores que no pertenecen a la autopista M-30. Los datos se filtran una última vez seleccionando únicamente el sensor definido por el usuario mediante su campo de identificación. A continuación, los datos triplemente filtrados previamente se ordenan por fecha de muestreo para garantizar su coherencia temporal. Seguidamente, se eliminan las características innecesarias de los datos, como los campos "*ID", "Tipo_Elem", "Carga", "Error" y "Periodo_Integración"*. Por último, los datos mensuales preprocesados se unen en un *dataframe* común para ser procesados como un flujo de datos. Todo este proceso debe repetirse para un número "n" de meses.

Una vez que todos los datos mensuales se han preprocesado en un único *dataframe*, los datos de la característica sobre la que se pretende realizar predicciones (previamente definida por el usuario) se desplazan cuatro filas hacia atrás. Dado que los datos tienen una frecuencia de muestreo de quince minutos, esto hará que el modelo aprenda a predecir el valor de la característica objetivo con una hora de antelación respecto a los datos reales recibidos. Por último, los datos finales preprocesados se configuran como un flujo de datos y se entregan a los algoritmos para que se entrenen en línea y de forma precuencial.

5.2.2. Experimentos y resultados.

Para evaluar la idoneidad de la Teoría de la Resonancia Adaptativa a la hora de predecir condiciones de tráfico evolutivas, el algoritmo LAPART ha sido implementado en *Python 3.8* junto con el uso de varias bibliotecas, que se detallan a continuación:

- *Pandas:* Biblioteca de análisis de datos de código abierto (McKinney, 2010).
- *Numpy:* Biblioteca de cálculo numérico de código abierto (Harris et al., 2020).
- *CSV:* Biblioteca nativa de *Python*, utilizada para leer y escribir archivos de valores separados por comas (CSV).
- *SYS:* Biblioteca nativa de *Python*, utilizada para solicitar al usuario a través de la consola la personalización del flujo de trabajo de análisis de datos.
- *Matplotlib:* Biblioteca completa para crear diferentes tipos de visualizaciones gráficas en *Python* (Hunter, 2007).
- *Sklearn: Sci-kit Learn* es una biblioteca de código abierto de herramientas sencillas y eficientes para el análisis predictivo de datos (Pedregosa et al., 2011).
- *Skmultiflow: Sci-kit Multiflow* es un paquete de *machine learning* de código abierto para gestionar flujos de datos (Montiel et al., 2018).

La implementación general consta de tres módulos, siendo el primero el módulo LAPART, que consta de submódulos de entrenamiento y de prueba. Se programó un módulo de preprocesamiento de datos para que el conjunto de datos de la autopista M-30 fuese filtrado y estructurado correctamente, siendo finalmente convertido en un flujo de datos para su correcto análisis mediante el modelo de aprendizaje. Una vez preprocesados, estos datos se introducen en los algoritmos LAPART y de evaluación comparativa con el objetivo de entrenar modelos que prevean la ocupación de carriles con un margen de una hora, basándose en la velocidad e intensidad medias de los vehículos. Adicionalmente, este módulo muestra algunas opciones de personalización para definir las fechas de entrenamiento y prueba, junto con la posibilidad de personalizar qué característica se predecirá y cuáles serán utilizadas como fuente de entrenamiento.

Por último, se creó un módulo de gestión de modelos de regresión para garantizar que dichos modelos fuesen entrenados y probados de manera correcta. Se utilizaron tres algoritmos de aprendizaje *online* para evaluar el rendimiento de LAPART, a saber:

- Descenso por gradiente estocástico (SGD).
- Bosques aleatorios adaptativos (ARF).
- Árboles adaptativos de Hoeffding (HAT).

Estos algoritmos de regresión se extrajeron de *Scikit Learn* y *Scikit Multiflow*, y se entrenan con los mismos flujos de datos preprocesados con los que se entrena LAPART.

Se utilizó un rango de fechas previamente definido para delimitar la cantidad de datos disponibles para los algoritmos con el fin de simular flujos de datos. Cada flujo de datos contiene una fecha de muestreo y tres características [a saber, la ocupación, la intensidad y la velocidad media (Vmed)], una de las cuales se elige para ser el objetivo de predicción durante el proceso de prueba y entrenamiento precuencial.

El intervalo de fechas se estableció aleatoriamente entre el 1 de enero de 2019 y el 31 de enero de 2019, proporcionando un flujo de datos compuesto por un mes entero de mediciones, con una tasa de muestreo de 15 minutos. También se eligieron aleatoriamente tres sensores situados en zonas totalmente diferentes de la autopista con el objetivo de demostrar la robustez de este modelo. Dado que la autopista M-30 es famosa por sus atascos, se escogió como característica a predecir la ocupación, que representa el porcentaje de ocupación del tráfico en la ubicación del sensor. Además, dicha característica se desplazó cuatro filas hacia arriba dentro del flujo de datos para que los algoritmos pudieran predecir con una hora de antelación los datos recibidos en ese momento.

5.2.2.1. Evaluación comparativa con modelos de referencia.

Para mostrar la robustez del modelo LAPART se seleccionaron aleatoriamente tres sensores no consecutivos (con números de identificación 6717, 3494 y 3558). Para cada sensor, se utilizaron las métricas de error absoluto medio (MAE), error cuadrático medio (MSE) y raíz del error cuadrático medio (RMSE) con el objetivo de comparar el rendimiento de LAPART con SGD, ARF y HAT. También se midió la precisión, así como la capacidad de adaptación frente al *concept drift*. Con el motivo de facilitar la lectura y el análisis, sólo se comparará frente al modelo LAPART el modelo de referencia con mejor puntuación de los tres.

5.2.2.2. Sensor 6717.

El sensor 6717 está situado en la parte derecha de la autopista M-30, al este de Madrid (Figura 18). Esta zona cuenta con una salida hacia los barrios de Chamartín y Tetuán del centro de Madrid, así como otra hacia el barrio del Pinar del Rey. En sus proximidades se encuentra también la estación de tren de Chamartín, uno de los centros neurálgicos de la ciudad en el que se realizan miles de desplazamientos diarios.

Figura 18 - Ubicación del Sensor 6717 dentro de la M-30*, al este de Madrid. (Elaboración propia a partir de Google Maps).*

Los resultados de este experimento se reflejan en la Tabla 5. Puede observarse claramente que LAPART es superior en todas las métricas analizadas. Supera completamente a SGD con significación estadística. También supera a ARF en estas métricas, aunque la diferencia es estadísticamente significativa sólo para la métrica MSE. Por último, LAPART también supera a los árboles adaptativos de Hoeffding, con significación en las métricas de MSE y RMSE.

Tabla 5 - Resultados del sensor 6717.

Modelo	MAE	MSE	RMSE
LAPART	**0.014744**	**0.000826**	**0.028754**
SGD	0.076274	0.020857	0.144420
ARF	0.021525	0.002175	0.046641
HAT	0.054240	0.007708	0.087798

En las Figuras 19 y 20 se exponen los gráficos de precisión de LAPART y ARF. Aunque el modelo ARF muestra una gran flexibilidad para adaptarse a la deriva del

concepto, LAPART es superior en este aspecto y en su capacidad para predecir correctamente los valores futuros de ocupación.

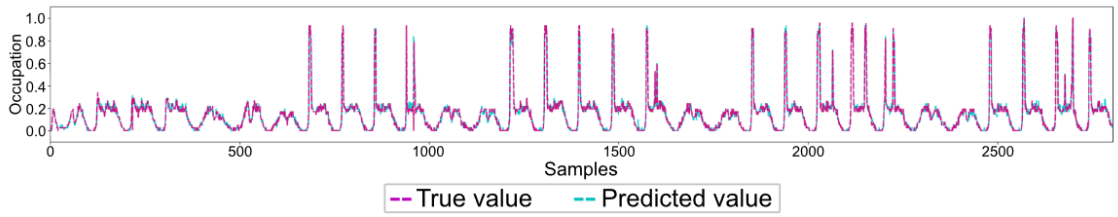


Figura 19 - Gráfica de precisión de LAPART (Sensor 6717). (Elaboración propia).

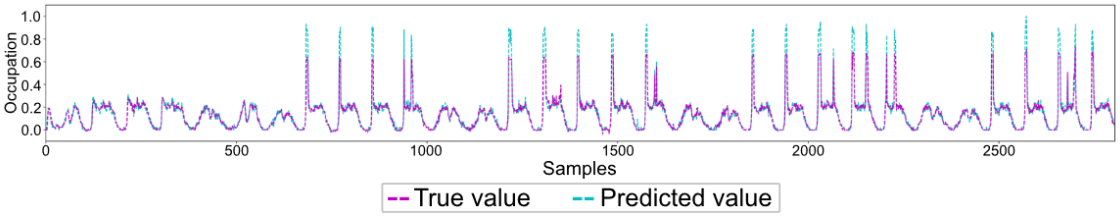


Figura 20 - Gráfica de precisión de Adaptive Random Forests (Sensor 6717). (Elaboración propia).

5.2.2.3. Sensor 3494.

Este sensor está situado en la parte superior de la autopista M-30, al norte del centro de Madrid, como puede observarse en la Figura 21. Esta zona presenta una intersección con los barrios de Mirasierra, La Paz y El Pilar. El barrio de El Pilar cuenta con uno de los centros comerciales más importantes de Madrid (La Vaguada), y La Paz tiene uno de los hospitales más concurridos, el Hospital Carlos III. La autopista M-30 pasa por estos barrios, lo que permite a los conductores tomar varias salidas en el cruce sensorizado.

Figura 21 - Ubicación del Sensor 3494 dentro de la M-30*, al norte de Madrid. (Elaboración propia a partir de Google Maps).*

Las métricas obtenidas tras el entrenamiento de los modelos se muestran en la Tabla 6. En este caso, LAPART también supera a todos los modelos de referencia en cada una de las métricas, con la excepción del MAE para el modelo ARF. En cuanto a la

significación estadística, LAPART supera claramente a SGD y HAT en las métricas MSE y RMSE.

Tabla 6 - Resultados del sensor 3494.

Modelo	MAE	MSE	RMSE
LAPART	0.013052	**0.000365**	**0.019115**
SGD	0.036032	0.004297	0.065552
ARF	**0.011518**	0.001221	0.034947
HAT	0.044219	0.006998	0.083655

Las Figuras 22 y 23 muestran los gráficos de precisión para los modelos LAPART y ARF, respectivamente. Se puede observar que para los valores más altos el ARF no es tan bueno en la predicción como el LAPART. En cuanto a la adaptabilidad de la deriva conceptual, LAPART consigue adaptarse mejor a los cambios de datos que ARF, aunque ambos muestran una gran adaptabilidad.

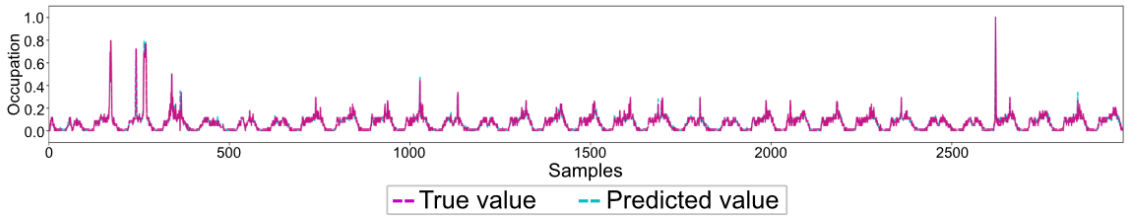


Figura 22 - Gráfica de precisión de LAPART (Sensor 3494). (Elaboración propia).

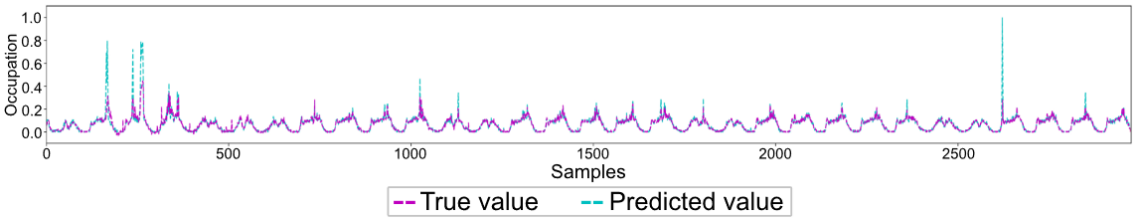


Figura 23 - Gráfica de precisión de Adaptive Random Forests (Sensor 3494). (Elaboración propia).

5.2.2.4. Sensor 3558.

El último sensor analizado se encuentra en la parte occidental de la autopista M-30, y se encuentra marcado con un círculo rojo en la Figura 24. Esta zona de la M-30 discurre paralela al río Manzanares, y pasa por delante del monumento histórico de Puerta de Hierro. En concreto, el cruce donde se encuentra el sensor da acceso al barrio de Valdemarín, donde se encuentra el hipódromo de la Zarzuela, y al barrio residencial de Aravaca.

Figura 24 - Ubicación del Sensor 3558 dentro de la M-30*, al norte de Madrid. (Elaboración propia a partir de Google Maps).*

Como se observa en la Tabla 7, las métricas obtenidas tras el entrenamiento de los modelos de regresión muestran que LAPART supera a los demás algoritmos con la excepción del modelo ARF, donde el MAE es mayor, aunque no es estadísticamente significativo. En cuanto a los resultados estadísticamente significativos, LAPART supera a los algoritmos SGD y HAT tanto en su MSE como en su RMSE.

Tabla 7 - Resultados del sensor 3558.

Modelo	MAE	MSE	RMSE
LAPART	0.016907	**0.000505**	**0.022484**
SGD	0.044800	0.005648	0.075153
ARF	**0.014467**	0.001893	0.043512
HAT	0.039429	0.005207	0.072162

Por último, las Figuras 25 y 26 muestran la precisión de los modelos LAPART y ARF, respectivamente. Ambos gráficos confirman que LAPART se adapta mejor a los cambios repentinos de los datos que ARF, también presentando este último características adaptativas, aunque a una velocidad más lenta.

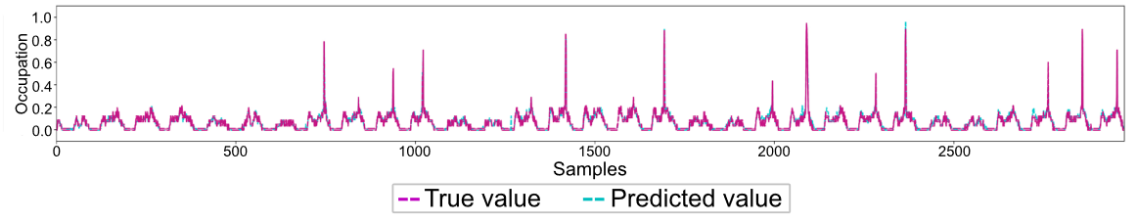


Figura 25 - Gráfica de precisión de LAPART (Sensor 3558). (Elaboración propia).

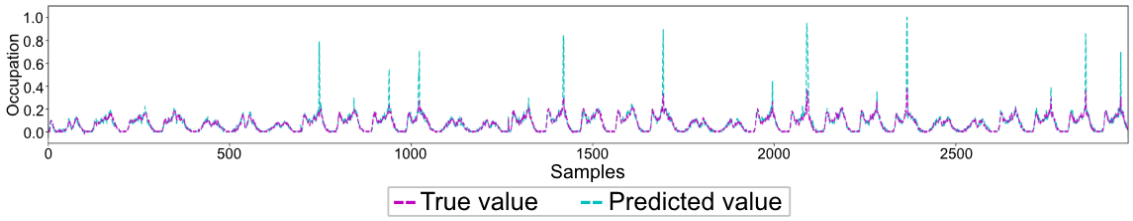


Figura 26 - Gráfica de precisión de Adaptive Random Forests (Sensor 3558). (Elaboración propia).

5.2.3. Discusión de resultados.

Este trabajo de investigación se ha centrado en la aplicación del paradigma de la ART al análisis y la predicción de flujos de datos en evolución. Más concretamente, se aplicó una red neuronal ART con preparación lateral (LAPART) para predecir la ocupación del tráfico basándose en dos características de este, a saber, la intensidad y la velocidad media. Para evaluar la idoneidad de este enfoque de resonancia adaptativa se ha utilizado un conjunto de datos del mundo real, extraído de un repositorio abierto que almacena los datos capturados por los sensores presentes en la autopista M-30 de la ciudad de Madrid, la carretera más transitada de España.

Se utilizó un mes de datos de tráfico para entrenar tanto a LAPART como a los modelos de referencia para asegurar su capacidad de adaptación frente a la deriva conceptual. Este conjunto de datos fue previamente preprocesado para seleccionar un rango de fechas, eliminar características no útiles y crear un flujo de datos consistente para ser proporcionado a los algoritmos de forma *online*. Este flujo de datos también fue preprocesado para que los algoritmos fueran capaces de generar predicciones con una hora de antelación, lo que resulta muy útil para anticiparse a los atascos y a posibles condiciones de tráfico adversas.

Se utilizaron tres sensores en diferentes ubicaciones dentro de la autopista con el objetivo de mostrar la robustez del modelo LAPART utilizando las métricas de error medio absoluto, el error medio cuadrado y la raíz del error medio cuadrado como métricas para evaluar el rendimiento de LAPART frente a modelos ya conocidos y de buen rendimiento como el descenso por gradiente estocástico, los bosques aleatorios adaptativos y los árboles adaptativos de Hoeffding. Tanto LAPART como los demás algoritmos tuvieron que hacer frente al *concept drift*, el cual degrada la precisión del modelo con el tiempo.

Los resultados experimentales mostraron que LAPART supera a los algoritmos anteriormente mencionados, con resultados estadísticamente relevantes en la mayoría de las métricas. Para los tres sensores analizados, el modelo propuesto mostró una buena capacidad de adaptación a la deriva conceptual, adaptándose rápidamente a los cambios repentinos en los datos y generando predicciones estables y precisas a pesar de estos cambios. Además, LAPART también mostró una mayor precisión en la predicción de la ocupación del carril que los modelos de referencia, obteniendo una

mayor precisión que los bosques aleatorios adaptativos en la mayoría de las ocasiones. Sin embargo, es importante señalar que este enfoque basado en ART dio un buen rendimiento en el conjunto de datos analizado, pero podría utilizarse fácilmente con otros conjuntos de datos para corroborar su rendimiento.

Como conclusión, LAPART no sólo produce mejores predicciones que los otros modelos, sino que también se adapta mejor al *concept drift*. Esto demuestra que tanto la teoría de la resonancia adaptativa como el modelo específico LAPART son soluciones adecuadas para la previsión y el análisis del tráfico debido a su inherente capacidad de adaptación y a su estructura interna de aprendizaje *online*.

6. Conclusiones.

El objetivo de esta tesis ha sido contribuir al conocimiento dentro del campo de aplicaciones de la *Computational Intelligence* sobre el sector turístico y hotelero. Para ello, diversas investigaciones han sido realizadas, las cuales han aportado resultados de gran valor, tanto teóricos como prácticos. Dichos resultados han mostrado que las metodologías *Computational Intelligence* basadas en *Data Stream Mining* no solo son más precisas que las metodologías más clásicas, sino que además resisten mejor al paso del tiempo y a los cambios provocados por el *concept drift* dentro de los datos. A continuación, se expondrán las contribuciones teóricas realizadas por esta tesis, así como sus contribuciones prácticas, limitaciones y, por último, posibles líneas de investigación que puedan ayudar a este sector a aumentar no solo su rendimiento, sino también sus beneficios.

6.1. Contribuciones teóricas.

En el ámbito teórico cabe destacar, en primer lugar, el extenso y preciso estudio del estado del arte realizado respecto a las aplicaciones de metodologías de la *Computational Intelligence* utilizadas sobre las diversas subáreas que comprenden el turismo y el sector hotelero, proporcionando adicionalmente un resumen de retos actuales que deben solventarse y de posibles líneas de investigación futuras. Tras haber analizado de manera exhaustiva más de 180 contribuciones científicas aplicadas en este sector, se pudo observar que no existía una clasificación precisa para dichas contribuciones respecto al subárea donde se aplicaban, lo que proporcionó una importante oportunidad para realizar dicha clasificación, proponiéndose así una taxonomía precisa en base a todos los estudios analizados. Una vez realizada esta clasificación, se identificaron las técnicas y metodologías de *Computational Intelligence* más utilizadas, resultando ser aquellas relacionadas con métodos probabilísticos/bayesianos y con modelos lineales o basados en instancias, observándose también un muy notable incremento en la utilización de metodologías basadas en ANN. Los árboles de decisión y las metodologías de *clustering* también son de uso común en este ámbito. Así mismo, el uso de *ensembles* está empezando a crecer, especialmente cuando se trata de ajustar los parámetros para el preprocesamiento de

datos utilizando algoritmos de computación evolutiva. Este tipo de computación también se usa para optimizar los parámetros a lo largo de varias metodologías, y se utiliza habitualmente en la previsión de varios parámetros en la industria turística y hotelera.

Con el aumento de la potencia de cálculo y los recientes avances en *machine learning*, la previsión dentro de este sector parece que se convertirá en un tema candente en los próximos años. No obstante, si algo ha quedado patente dentro de esta área es que se ha realizado una investigación excesiva en torno a metodologías de eficacia ya demostrada en lugar de estudiar metodologías innovadoras o de reciente descubrimiento, las cuales podrían arrojar mejores resultados a los ya conocidos. Adicionalmente, debe mencionarse que la utilización de metodologías de *Computational Intelligence* basadas en *online learning* dentro de esta área es más bien escasa, lo cual abre las puertas a un nicho de investigación en el que puedan generarse modelos predictivos resistentes al paso del tiempo y a los cambios de tendencia dentro de los datos. Por último, como parece estar ocurriendo en todas las demás áreas, el *deep learning* cada vez será más utilizado, especialmente en aplicaciones en las que se disponga de *Big Data* y la interpretabilidad y transparencia de los modelos resultantes no sea una cuestión crítica. Adicionalmente, un conjunto de retos identificados gracias a este estudio se presenta en la sección 6.3, y presenta una oportunidad única no sólo como una vía insuficientemente explorada para la comunidad investigadora que trabaja en *Computational Intelligence*, sino también para que el sector turístico y hotelero sea plenamente consciente de los beneficios que puede aportar esta área de investigación. Se cree firmemente que el material y las perspectivas que ofrece este estudio son sólo una muestra del enorme potencial que subyace bajo la adopción de la inteligencia computacional en este sector.

En segundo lugar, esta tesis ha contribuido al ámbito científico mediante la definición teórica de un modelo basado en *Data Stream Mining* a partir de la modernización realizada sobre un paradigma clásico del *machine learning*: la regresión lineal, todo ello aplicado con el objetivo de predecir la velocidad máxima del viento en destinos turísticos del archipiélago canario. Esta modernización se realizó mediante la reformulación del original, adaptando su funcionamiento con el objetivo de generar modelos que fuesen generados incrementalmente a partir de flujos de datos, en lugar de ser modelos finales creados a partir del entrenamiento mediante *Big Data* o grandes conjuntos de datos. Esto ha mostrado que dichos modelos incrementales son capaces de adaptarse a los cambios dentro de la distribución estadística estocástica de la variable que pretende modelarse, superando así el *concept drift* y mostrando resiliencia al paso del tiempo.

Como tercera contribución teórica, se ha demostrado que la utilización de metodologías basadas en la teoría de la resonancia adaptativa en el campo de la predicción de tráfico puede no solo competir con otras metodologías actuales, sino también ofrecer mejores resultados estadísticamente significativos. La capacidad de adaptación nativa de las redes neuronales resonantes hace que los modelos predictivos

generados sean capaces de realizar predicciones sorprendentemente precisas con un margen de tiempo relativamente generoso de 60 minutos, además de mostrar mayores capacidades de adaptación frente al *concept drift*, así como una mayor velocidad a la hora de sobreponerse a dicho fenómeno.

6.2. Contribuciones prácticas.

Respecto a las contribuciones prácticas realizadas por esta tesis, debe destacarse la gran aplicabilidad de estas al entorno turístico y hotelero. El estudio sobre el estado del arte realizado ha mostrado un *gap* investigaciones realizadas mediante metodologías basadas en *Data Stream Mining* aplicadas sobre el área turística y hotelera, lo que coincide con el objetivo principal de este trabajo de investigación. Es por ello por lo que, además de realizar las contribuciones teóricas previamente mencionadas, se ha puesto especial énfasis en hacer aplicables dichas contribuciones mediante la utilización de conjuntos de datos extraídos del mundo real, en lugar de probar los modelos mediante datos sintéticamente generados.

Como primera aproximación a una aplicación relevante para el turismo, se escogieron datos climáticos reales procedentes de todas las estaciones climatológicas de la AEMET ubicadas en el archipiélago canario con el objetivo de realizar un modelo predictivo basado en *Data Stream Mining* capaz de realizar predicciones sobre la velocidad máxima del viento con 60 minutos de margen. Los resultados obtenidos parecen confirmar cómo la estrategia de aprendizaje adaptativo propuesta logra sobreponerse al *concept drift* de una manera notablemente superior a los resultados obtenidos mediante la estrategia acumulativa. Adicionalmente, el modelo aprendido parece poseer altas posibilidades de generalización, ya que ha demostrado funcionar bien para la red geográficamente dispersa de estaciones meteorológica. Esto permitiría su aplicación en diferentes localizaciones con motivo de organizar anticipadamente diferentes tipos de eventos o traslados con independencia de la ubicación de los datos utilizados, dotando a esta contribución de un gran valor estratégico y facilitando su aplicación práctica. El aprendizaje por transferencia es fundamental para este tipo de tareas de modelización (Qin et al., 2011), de modo que los modelos de velocidad del viento pueden aplicarse a diferentes escenarios una vez entrenados.

Respecto a la segunda contribución práctica, se decidió probar un enfoque innovador de redes neuronales sobre un conjunto de datos extraído también del mundo real. El objetivo inicial era probar la viabilidad de una red neuronal basada en la teoría de la resonancia adaptativa a la hora de predecir condiciones de tráfico adversas y de adaptarse a la estructura cambiante de los datos con rapidez, todo ello con un margen de 60 minutos de antelación, para así lograr que el modelo predictivo no solo fuera preciso y adaptable, sino también rápido y fiable. Debido a la ausencia de conjuntos de datos no privativos sobre el tráfico en Canarias se decidió buscar una alternativa abierta basada en una ciudad con un alto componente turístico, siendo seleccionada para ello la ciudad de Madrid y, concretamente, la autopista M-30 debido a sus características viales caóticas y de difícil predicción.

Se utilizó un mes de datos de tráfico para entrenar tanto a LAPART como a los modelos de referencia para asegurar su capacidad de adaptación frente a la deriva conceptual. Este conjunto de datos fue previamente preprocesado para seleccionar un rango de fechas, eliminar características no útiles y crear un flujo de datos consistente para ser proporcionado a los algoritmos de forma *online*. Este flujo de datos también fue preprocesado para que los algoritmos fueran capaces de generar predicciones con una hora de antelación, lo que resulta muy útil para anticiparse a los atascos y a posibles condiciones de tráfico adversas.

Tres sensores ubicados en diferentes localizaciones dentro de la autopista M-30 fueron seleccionados aleatoriamente con el objetivo de mostrar la robustez del modelo LAPART utilizando las métricas de error medio absoluto, el error medio cuadrado y la raíz del error medio cuadrado como métricas para evaluar el rendimiento de LAPART frente a modelos ya conocidos y de buen rendimiento como el descenso por gradiente estocástico, los bosques aleatorios adaptativos y los árboles adaptativos de Hoeffding. Tanto LAPART como los demás algoritmos tuvieron que hacer frente al *concept drift*, el cual degrada la precisión del modelo con el tiempo.

Los resultados experimentales mostraron que LAPART supera a los algoritmos anteriormente mencionados, con resultados estadísticamente relevantes en la mayoría de las métricas. Para los tres sensores analizados, el modelo propuesto mostró una buena capacidad de adaptación a la deriva conceptual, adaptándose rápidamente a los cambios repentinos en los datos y generando predicciones estables y precisas a pesar de estos cambios. Además, LAPART también mostró una mayor precisión en la predicción de la ocupación del carril que los modelos de referencia, obteniendo una mayor precisión que los bosques aleatorios adaptativos en la mayoría de las ocasiones. Sin embargo, es importante señalar que este enfoque basado en ART dio un buen rendimiento en el conjunto de datos analizado, pero podría utilizarse fácilmente con otros conjuntos de datos para probar su rendimiento más a fondo.

Como conclusión, LAPART no solo produce mejores predicciones que los otros modelos, sino que también se adapta mejor al *concept drift*. Esto demuestra que tanto la teoría de la resonancia adaptativa como el modelo específico LAPART son soluciones adecuadas para la previsión y el análisis del tráfico debido a su inherente capacidad de adaptación y a su estructura interna de aprendizaje *online*. El hecho de poder predecir las condiciones de tráfico de manera precisa con un modelo altamente resistente a la obsolescencia provocada por el *concept drift* lo dota de un gran valor estratégico a la hora de planear itinerarios o eventos de todo tipo dentro del ámbito turístico.

6.3. Retos y desafíos: Inteligencia Computacional y Sector Turístico.

La adopción de técnicas y métodos de *Computational Intelligence* ya ha sido fructífera en el sector turístico y hotelero, tal como se ha puesto de manifiesto en los apartados anteriores. Sin embargo, los nuevos desarrollos y paradigmas de la *Computational Intelligence* pueden aportar nuevas mejoras a esta industria, abriendo

nuevos retos que deberán ser abordados en próximas investigaciones. A continuación, se resumen algunos de los principales retos y desafíos relacionados con estos avances tecnológicos:

6.3.1. Inteligencia práctica.

La aplicabilidad o practicidad de un sistema se refiere a la capacidad de sus resultados para ser puestos en acción en el contexto particular para el que están destinados. En la investigación sobre *Computational Intelligence* existe disparidad entre la obtención de resultados extraordinarios y la aplicación de modelos y métodos aplicables. Técnicas como las que se engloban bajo el paraguas de la Inteligencia Artificial explicable (xAI) (Barredo Arrieta et al., 2020; Gunning y Aha, 2019) pueden ayudar a los interesados en el campo a poner en práctica los resultados de estos sistemas basados en la *Computational Intelligence*, especialmente aquellos cuya estructura interna y procedimiento de aprendizaje están lejos de ser transparentes para el público no experto.

Una posible solución para este tipo de desafío podría ser la creación de herramientas de previsión que proporcionen resultados que expliquen la relevancia de sus entradas para el resultado obtenido, aumentando la confianza en el modelo por parte del responsable de la toma de decisiones y, finalmente, la capacidad de acción de su resultado. Además, considerar las técnicas de aprendizaje de transferencia (*transfer learning*) (Pan y Yang, 2010) podría ser útil para extender los desarrollos a diferentes dominios y entornos.

6.3.2. Fusión de datos de múltiples fuentes.

La variedad de fuentes de datos en la industria turística y hotelera puede dar lugar a la implementación de modelos de fusión de datos que podrían obtener información enriquecida y perspectivas sobre los datos disponibles. Por ejemplo, la caracterización de los turistas o los huéspedes puede lograrse contemplando muchas fuentes de información. Utilizar la gestión del conocimiento para gestionar todas las fuentes de datos presentes en un establecimiento hotelero podría ser una buena solución.

La gestión del conocimiento puede definirse como la coordinación y el uso de los recursos de conocimiento de la organización para crear una ventaja competitiva (Drucker, 2002). Tradicionalmente, la competitividad de una empresa se basaba en el capital, el terreno, la mano de obra y muchos otros recursos tangibles, pero, en los últimos tiempos se ha demostrado que la gestión del conocimiento se ha convertido en una importante fuente de ventaja competitiva (Kebede, 2010). De hecho, se ha descubierto que la gestión del conocimiento es uno de los activos más importantes para los establecimientos y organizaciones hoteleras y turísticas por su capacidad para ayudar a estas organizaciones a crear y mantener ventajas competitivas mediante el uso de diferentes herramientas informáticas como las bases de datos de habilidades, los

sistemas de apoyo a la toma de decisiones o los almacenes de datos (Okumus, 2013). El uso de este tipo de técnicas puede permitir a los establecimientos del sector mejorar sus servicios combinando diferentes bases de datos, utilizando la gestión del conocimiento para descubrir información importante sobre sus clientes y obteniendo ventajas relevantes sobre sus competidores.

Además de utilizar programas informáticos especializados para superar las dificultades encontradas en la fusión de datos con estructuras diferentes, un enfoque potencial para resolver este reto podría ser generar un estándar de datos para las bases de datos del sector hotelero y turístico, facilitando el proceso de fusión de datos entre bases de datos de diferentes servicios mediante la generalización de su estructura (lo que las haría fácilmente combinables) y aumentando el potencial de las aplicaciones que dependen de los métodos de *Computational Intelligence*.

6.3.3. Aprendizaje dinámico y *online*.

El aumento de la potencia de cálculo y el desarrollo de métodos más eficaces han empujado a la *Computational Intelligence* hacia un paradigma de procesamiento más dinámico. Los esquemas de aprendizaje *online* y de optimización dinámica, así como la introducción de conceptos como la detección de cambios y la adaptación, podrían introducir nuevas perspectivas en la aplicación de la *Computational Intelligence* en el sector hotelero. En su trabajo de investigación, Krawczyk et al. (2017) clasifican diferentes algoritmos de *ensemble* para diferentes tareas de minería de flujos de datos en una taxonomía útil, presentando además los problemas de investigación encontrados durante dicho estudio y las líneas de investigación futuras, lo que hace que este trabajo sea de suma importancia para futuros desarrollos de *online learning*.

Las tendencias del sector turístico y hotelero cambian, lo que hace que los modelos de *machine learning* pierdan precisión con el tiempo. Esto puede provocar varias pérdidas en términos de ingresos por diversos factores, como predicciones erróneas de la demanda turística o de la demanda de recursos. El *online learning* ofrece una solución a estos problemas. En este paradigma de *machine learning* los datos se tratan como flujos de datos, lo que los hace potencialmente infinitos. Como afirman Žliobaitė et al. (2014), los datos llegan continuamente en tiempo real y pueden cambiar con el tiempo. En este entorno, los modelos predictivos deben operar rápidamente, ajustarse a una memoria limitada y adaptarse en línea; de lo contrario, su precisión se degradará con el tiempo. Esto se opone al paradigma clásico del aprendizaje por lotes, en el que un modelo predictivo se genera solo cuando se ha analizado todo el conjunto de datos.

Gracias al *online learning* es posible analizar los datos de forma incremental a medida que llegan, sin necesidad de retenerlos en la memoria. Cuando se combinan con técnicas de detección y adaptación a la deriva conceptual, los modelos de *online learning* para tareas de predicción resultan ser no solo eficientes, sino también resistentes al desgaste del tiempo. Varios enfoques se han basado en este paradigma.

Por ejemplo, Lobo et al. (2020) demostraron que las redes neuronales en espiga (*spiking neural networks*) (Gerstner y Kistler, 2002) son notablemente eficaces en el uso de *Data Stream Mining* en tiempo real, desencadenando así una prometedora área de investigación dentro del *online learning* en torno a este tipo de red neuronal gracias al bajo coste computacional y a la alta capacidad de representación de estos modelos.

El uso de esta metodología en el sector hotelero podría suponer ventajas competitivas para sus establecimientos debido a su capacidad para actualizar continuamente el conocimiento analizado y adaptarse a los cambios naturales de los datos derivados de la naturaleza no estacionaria de los procesos hoteleros y turísticos. Si se aplica a la demanda de recursos o de turismo, la toma de decisiones podría mejorar gracias a la mayor precisión de las predicciones del modelo, lo que podría aumentar los ingresos y mejorar la adaptabilidad del establecimiento a distintos tipos de eventos.

6.3.4. Modelos de datos encriptados.

Una de las mayores preocupaciones del mundo del *Big Data* es que, a pesar de que los datos estén anonimizados, es posible recuperar fragmentos de información que permiten identificar a los individuos por diferentes medios. Hasta ahora, el cifrado clásico de los datos no resultaba adecuado para la minería del conocimiento debido a su propia naturaleza: para trabajar con algo que ha sido encriptado, primero hay que desencriptarlo. Sin embargo, los nuevos paradigmas de preservación de la privacidad en el ámbito de la inteligencia artificial, como el cifrado homomórfico o el aprendizaje federado, pueden ser útiles para resolver este problema, permitiendo a las partes interesadas del sector hotelero y turístico trabajar e intercambiar con seguridad los datos de los clientes sin ninguna preocupación relacionada con la privacidad.

El cifrado homomórfico (Rivest y Dertouzos, 1978) se basa en el uso de determinadas propiedades matemáticas presentes en ciertos esquemas de cifrado. Estas propiedades permiten realizar operaciones sobre los datos cifrados sin descifrarlos, presentando resultados que, en efecto, siguen estando cifrados pero que pueden ser procesados o descifrados posteriormente. Años después, Gentry (2009) propuso el primer sistema de cifrado homomórfico funcional, el cual utiliza una función de evaluación, basada en polinomios de bajo grado, sobre la información previamente cifrada. Los desarrollos actuales de la encriptación homomórfica son presentados por autores como Fan y Vercauteren (2012) (esquema criptográfico Fan-Vercauteren) y Brakerski et al. (2014) (esquema criptográfico Brakerski-Gentry-Vaikuntanathan), y se basan en el problema de aprendizaje en anillo con errores de Oded Regev (Regev, 2010). Especialmente para la industria turística y hotelera, el uso de la encriptación homomórfica puede adoptarse de forma masiva a la hora de caracterizar al cliente en su conjunto, construyendo modelos de *Computational Intelligence* sin comprometer características protegidas que el propio cliente podría considerar confidenciales (por ejemplo, ingresos, orientación sexual, género y otros aspectos similares).

Sin embargo, si se pone el foco en el intercambio de información entre las partes interesadas, surgen reticencias debido a la fuerte competitividad existente en el sector hotelero y turístico. La combinación de la información generada por el cliente en distintos lugares/en distintos plazos podría aportar enormes ventajas en cuanto a la precisión con la que podrían funcionar los modelos alimentados con esos datos combinados. Sin embargo, en la práctica la mayoría de las empresas no están dispuestas a compartir la información generada por el cliente en sus instalaciones.

El reciente advenimiento del aprendizaje federado puede cambiar el paradigma en torno a este asunto al establecer las bases técnicas para compartir información relacionada con un modelo que preserva la privacidad (por ejemplo, los gradientes de las redes neuronales) en lugar de los datos brutos entre los modelos distribuidos (Konečný et al., 2017). Las contribuciones de todos estos modelos se centralizan y se procesan, dando lugar a una representación agregada que puede devolverse a los modelos y combinarse con el conocimiento aprendido localmente para mejorar el rendimiento. Este paradigma ha estado madurando durante los últimos dos años, mostrando un gran potencial para lanzar la adopción de modelos de *Computational Intelligence* en dominios de aplicación sensibles a la privacidad, como la salud y la industria (Geyer et al., 2017). Por último, es importante enunciar que, con toda seguridad, el aprendizaje federado también desempeñará un papel importante en los futuros despliegues de la *Computational Intelligence* dentro del sector turístico y hotelero.

6.3.5. Anticipación a sesgos en los datos.

Cuando el fenómeno a modelar no es estacionario en cuanto a su comportamiento estadístico, el patrón a aprender por un modelo basado en *Computational Intelligence* puede sufrir cambios que eventualmente hagan que el modelo quede obsoleto. Esta situación se denomina ampliamente *concept drift* (Widmer y Kubat, 1996), es decir, un cambio en el proceso que genera la distribución de datos a aprender, el cual no se refleja explícitamente en los propios datos de entrada. Este problema es especialmente frecuente en los procesos que generan flujos de datos rápidos (por ejemplo, las compras electrónicas), y suele provocar un deterioro significativo del rendimiento de los modelos predictivos con el paso del tiempo. Ese "azar" en las características estadísticas del fenómeno puede deberse a un sesgo progresivo de los datos. Las razones detrás de este *concept drift* pueden ser diversas. A veces puede deberse al comportamiento humano en un contexto de toma de decisiones. Por ejemplo, la paradoja de Braess (Braess, 1968) en la asignación dinámica de viajes explica cómo dejar la elección de la ruta exclusivamente en manos de los conductores puede acabar en un estado de equilibrio de Nash, que puede no ser el óptimo del sistema (globalmente). En otras palabras, dejar la elección de la ruta sólo a los conductores con sus criterios de optimización locales (egoístas) puede terminar en un empeoramiento del rendimiento total de una red de tráfico particular. Es importante mencionar que esta paradoja no tiene solución analítica. El desarrollo de cualquier

modelo predictivo y de asignación dinámica de rutas entrenado para una determinada distribución de la demanda dará lugar, naturalmente, a un sesgo en el comportamiento de esa distribución de la demanda y al consiguiente deterioro del rendimiento de dicha configuración. La única estrategia práctica es apoyarse en una metodología que pueda hacer frente al *concept drift* en la distribución estadística de la demanda de la red. En concreto, se necesita una metodología dinámica que detecte esa desviación de la demanda cuando se produzca, desencadenando así una estrategia de adaptación. En otras palabras, las metodologías de aprendizaje de modelos y asignación dinámica de rutas deben ser incrementales y adaptativas para hacer frente al *concept drift* presente en la señal de demanda de la red de tráfico.

Una situación similar se da en el sector turístico y hotelero: cuando la información producida por el cliente cambia su comportamiento, puede aparecer un sesgo de datos inducido por el ser humano como resultado de las decisiones tomadas a partir de ella. Un ejemplo claro son los sistemas de recomendación, que pueden basarse en diversas técnicas de *Computational Intelligence*, como los modelos predictivos y los métodos de clasificación. Por ejemplo, un cliente de un hotel puede tomar decisiones sobre la base de los artículos recomendados que, a menudo, se retroalimentan sobre la recomendación para su actualización. Cuando un cambio contextual en sus hábitos no se refleja en los datos proporcionados (por ejemplo, un cambio de estado civil), las decisiones pueden variar radicalmente, lo que hace que el motor de recomendación quede obsoleto y, en última instancia, provocar que las recomendaciones realizadas sean inútiles hasta que el modelo aprenda a captar el nuevo contexto del usuario. Dependiendo de la velocidad y la gravedad del cambio, puede existir un retraso notable hasta que el modelo de recomendación proporcione resultados significativos para el nuevo concepto.

Las técnicas de detección, caracterización y adaptación del *concept drift* tienen como objetivo acortar el tiempo que necesita el modelo para reflejar la nueva distribución de datos. El sesgo de los datos, así como la prevalencia del *concept drift,* en un ámbito turístico donde las decisiones humanas están sujetas a una amplia variedad de factores contextuales que no se tienen en cuenta explícitamente en los datos recogidos, hace que sea muy necesario seguir estudiando cómo los modelos de *Computational Intelligence* pueden aprender de los escenarios no estacionarios y adaptarse a ellos de forma eficiente.

6.4. Limitaciones y futuras líneas de investigación.

Pese a la extensión que esta tesis ha logrado abarcar, es necesario resaltar las posibles limitaciones que hayan podido surgir a la hora de haber realizado esta investigación. En primer lugar, y aún habiéndose realizado un extenso estudio del estado del arte respecto a la aplicación de metodologías *Computational Intelligence* sobre el área turística y hotelera, es necesario destacar que los 180 artículos analizados, aun siendo una muestra relevante, estaba focalizada solo al ámbito turístico, con lo cual otras metodologías *Computational Intelligence* que pudieran ser extrapoladas a este

ámbito no han sido contempladas. Existen multitud de artículos científicos dedicados a este tipo de aplicaciones, pero se decidió analizar los más relevantes dentro del ámbito turístico con motivo de representar de manera fiel y directa la distribución de los esfuerzos de investigación más actuales dentro de esta área. Debe tenerse en cuenta que esta *review* fue realizada utilizando los artículos existentes en la *Web of Science* hasta el año 2020, debiéndose actualizar e incluso pudiéndose ampliar a los artículos recogidos en *Scopus*.

El aumento de la potencia de cálculo y el desarrollo de nuevas metodologías para obtener conocimientos de todo tipo de fuentes de datos, que van desde los pequeños sensores hasta el *Big Data*, han creado nuevas formas de analizar enormes cantidades de datos en menos tiempo. Sin embargo, aún existe un amplio margen para realizar nuevas aplicaciones de *Computational Intelligence* en el sector de turístico y hotelero. La utilización de metodologías de *deep learning* proporciona conocimientos útiles en varios campos del sector, pero es necesario crear nuevos métodos procesables que reduzcan la opacidad de las capas dentro de un sistema de este tipo con el objetivo de facilitar su aplicación y comprensibilidad. Esto abre un nuevo campo de investigación en el que el rendimiento de un modelo será importante, pero su transparencia será primordial a la hora de ofrecer lo que podría denominarse como "un modelo sólido". Deberían investigarse nuevas metodologías o actualizar las existentes para ofrecer un buen rendimiento, además de ser aplicables a otros campos para enriquecer el panorama tecnológico del sector hotelero y turístico.

Además, la potencia de cálculo disponible para analizar los datos aumenta de forma exponencial. El aumento del volumen, la velocidad y la heterogeneidad de los datos también requiere de enormes recursos informáticos para obtener información, lo que dificulta su análisis mediante modelos basados en *Computational Intelligence*. Además, este sector necesita modelos que no sólo funcionen en tiempo real, sino también en condiciones y escenarios variables. Esto abre un territorio inexplorado y fértil para nuevos trabajos de investigación basados en modelos de *Computational Intelligence* adaptativos, en los que un modelo fiable funciona en tiempo real y también se adapta a las nuevas tendencias de los datos, permitiendo a los establecimientos hoteleros y turísticos prever diferentes variables, como los picos de reservas estacionales, los posibles resultados de la distribución o reserva de recursos o las estimaciones de ingresos. Ya existen diversas aproximaciones metodológicas para este fin, como la mencionada familia de modelos de *online learning*, que proporcionan este tipo de resultados. Sin embargo, dado que se trata de un campo en continua evolución, los últimos trabajos de investigación deberían integrarse progresivamente para ser aplicados en el sector hotelero, cuyos datos están sujetos a altos niveles de variabilidad y a factores exógenos no estacionarios.

Sin embargo, más allá de la adaptabilidad y la capacidad de acción, la privacidad de los datos también es un componente clave a la hora de desarrollar un buen modelo. Los datos generados por el sector de turístico y hotelero son un tesoro en términos de valor, pero también resultan un mar traicionero en el que navegar debido a las enormes

cantidades de datos que contienen información personal protegida sobre los clientes. Actualmente vivimos en un mundo en el que las filtraciones de datos se producen con frecuencia, y en el que se puede acceder a dicha información filtrada sin apenas dificultad. Tal y como se ha expuesto en uno de los apartados anteriores de este trabajo de investigación, la encriptación homomórfica aplicada a los datos utilizados para entrenar los diferentes modelos podría mantener el rendimiento del modelo a la vez que conservar la privacidad de los datos, creando así la necesidad de modelos de datos encriptados donde aplicar la *Computational Intelligence*. Además, el uso de metodologías de aprendizaje federado también podría permitir interesantes escenarios de colaboración entre varias partes interesadas, produciendo así modelos más precisos sin comprometer la privacidad de los datos recogidos por cada parte.

Por último, no hay una respuesta breve a la pregunta de cómo superar el problema del sesgo de los datos. Es necesario seguir investigando para entender cómo los modelos de *Computational Intelligence* pueden detectar de forma fiable y adaptarse eficazmente al *concept drift* presente en los datos. Lo cierto es que, a menos que se garanticen tales capacidades en los modelos de *Computational Intelligence*, aparecerían efectos indeseables a la hora de tomar decisiones a partir de los resultados de estos modelos, tales como pérdidas de ingresos o recursos mal asignados. Los esfuerzos de investigación en esta dirección podrían implicar nuevos métodos de *Computational Intelligence* adaptativos y modelos de aprendizaje rápido en los que las capacidades de adaptación se convierten en la clave para su despliegue en escenarios del mundo real.

En resumen, las nuevas vías de investigación dentro de la *Computational Intelligence* aplicada al sector turístico y hotelero deberían redoblar esfuerzos en la creación de nuevos métodos que permitan fusionar diferentes fuentes de datos relacionados con el sector hotelero de una manera respetuosa con la privacidad, de modo que se preserve la confidencialidad y se fomente el intercambio de datos entre las partes interesadas. También es de suma importancia investigar modelos que se adapten de forma resistente a los cambios en la distribución de los datos, con mecanismos eficaces para sortear el problema del sesgo de los datos que suele estar presente en los casos de uso de esta industria.

En cuanto a futuras investigaciones respecto al modelo de regresión lineal adaptado a *Data Stream mining*, el reto es comparar esta estrategia con algunas otras metodologías clásicas de previsión de series temporales (por ejemplo, de la familia de los modelos ARIMA, las máquinas de vectores de soporte (SVM) o los métodos basados en el filtrado de Kalman) que tendrán que ser ajustadas como en el presente análisis con regresión lineal y descenso de gradiente, con el fin de operar en una configuración *online* para una comparación justa. Además, dentro del ámbito del *deep learning*, otras técnicas de *machine learning* de última generación como las redes neuronales recurrentes deberían ser un buen punto de comparación con la metodología propuesta. No obstante, las modificaciones algorítmicas necesarias para hacer a estas redes operativas en una configuración *online* de forma de aprendizaje incremental resultarán aún más difíciles debido a la naturaleza convolucional de la mayoría de ellas. También

sería posible ampliar las variables explicativas utilizadas, teniendo en cuenta, por ejemplo, la altitud, la nubosidad o la frecuencia de los vientos fuertes en cada estación meteorológica concreta, fusionando otras señales relacionadas de manera directa o indirecta con la meteorología procedentes de otro tipo de sensores, como la intensidad de la señal recibida (RSS).

Otra posible línea de actuación dentro de los márgenes de este modelo lineal modificado sería utilizar un modelo base diferente para la modelización predictiva. Como se muestra en el análisis r^2 realizado, la velocidad máxima del viento no es un fenómeno lineal modelable. Se ha utilizado un modelo de regresión lineal con una estrategia de descenso de gradiente principalmente como punto de partida, debido al extenso conocimiento que hay de su enfoque y a su sencilla comprensibilidad. El modelado flexible de la metodología propuesta, que utiliza la estrategia incremental y adaptable, compensa la linealidad de estos modelos. Sin embargo, existen otros modelos de regresión, como la regresión basada en el vector soporte, los cuales pueden hacer frente a la no linealidad del fenómeno modelado. Los autores planean ampliar esta investigación ajustando las rutinas de aprendizaje de estas metodologías para que sean adaptables e incrementales.

Por último, y respecto al modelo LAPART utilizado para realizar predicciones sobre el estado del tráfico en zonas turísticas de alta densidad, puede definirse una posible línea de actuación relativa a la utilización de una variante de las redes neuronales ART conocida como ART-2 con el objetivo de corroborar si consigue presentar similares resultados a la hora tanto de realizar predicciones de gran precisión como de adaptarse a los cambios provocados por la naturaleza no estacionaria de los datos. Dicha variante podría implementarse siguiendo la metodología descrita en esta tesis, pero modificando el modelo LAPART original mediante la sustitución de las dos redes ART-1 por redes ART-2 o ART-3. Esto generaría modelos adaptivos LAPART-2 y LAPART-3 respectivamente, los cuales podrían mostrar resultados interesantes utilizando el mismo conjunto de datos usado en la presente aproximación.

Adicionalmente, es necesario realizar más pruebas utilizando conjuntos de datos de otras vías de alta densidad con motivo de demostrar la capacidad de generalización de este modelo.

6.5. Conclusiones y recomendaciones para el sector del turismo.

En base a los resultados obtenidos en los trabajos de investigación desarrollados dentro del marco de esta tesis, se han realizado un conjunto de sugerencias y posibles líneas de actuación para mejorar los servicios proporcionados por la oferta turística del archipiélago canario con el objetivo de no solo mejorar la propia calidad de dichos servicios, sino también de facilitar la toma de decisiones y la planificación de eventos relacionados con dicha área.

6.5.1. Inversión en nuevas investigaciones basadas en *Data Stream Mining*.

En primer lugar, se han aportado pruebas que favorecen la inversión en tiempo y dinero en nuevas técnicas de *online learning* aplicadas al sector turístico canario. Queda patente que, si bien no existen muchos trabajos de investigación al respecto, el *corpus* existente demuestra que los resultados obtenidos por estas nuevas tecnologías son notablemente favorables frente a técnicas más clásicas, siendo capaces de dotar a los establecimientos turísticos o *stakeholders* de una ventaja estratégica potencialmente decisiva frente a sus competidores debido a varios factores. Por un lado, el generar modelos predictivos precisos capaces de realizar recomendaciones en diversas áreas como la predicción de la demanda energética o turística con el potencial aumento de beneficios que ello supondría, y por otro la resistencia de dichos modelos frente al paso del tiempo, haciéndolos más fiables y duraderos a pesar de los posibles cambios en las costumbres de los clientes o en las tendencias turísticas.

Es por ello por lo que, como primera recomendación, se aconseja encarecidamente el realizar nuevas investigaciones basadas en la minería en flujo de datos con el objetivo de generar sistemas de predicción capaces de mejorar la toma de decisiones, mejorando de manera potencial la percepción y la calidad de vida del turista durante su viaje en el archipiélago canario.

6.5.2. Convenios con organismos públicos para la utilización de datos.

También ha quedado demostrado en esta tesis que la predicción meteorológica precisa es posible mediante metodologías basadas en *Data Stream Mining*, permitiendo que puedan planificarse o cancelarse eventos ante condiciones meteorológicas adversas con suficiente tiempo de reacción para evitar pérdidas materiales o, lo que es más importante, posibles daños personales. El archipiélago canario presenta una condición muy particular en cuanto a microclimas se refiere (Bechtel, 2016), pero se ha demostrado que pese a ello pueden generarse predicciones de bastante precisión con un margen de una hora, siendo dicho margen potencialmente ampliable si técnicas más complejas fuesen aplicadas.

No obstante, la gran mayoría de los datos generados suele ser de carácter privado, dificultando el desarrollo de nuevas técnicas o soluciones para problemas relacionados con la predicción meteorológica. Es por ello por lo que, como segunda recomendación, se hace hincapié en la colaboración entre organismos privados y públicos para generar sistemas de predicción climatológica para así poder planificar eventos con mayores garantías de seguridad, lo cual aumentará la calidad de la visita realizada por el turista al archipiélago al poder disfrutar de eventos más seguros y mejor planificados.

6.5.3. Inversión en infraestructura vial.

También ha quedado demostrado en esta tesis que el tráfico puede impactar en la calidad de los servicios prestados a viajeros y turistas ya que, por ejemplo, la organización de un trayecto puede verse afectada de manera notable por la elección de

vías de alta congestión para las cuales podría existir una alternativa. La creación de sistemas capaces de predecir en tiempo real los posibles atascos con un generoso margen de actuación podría beneficiar enormemente a la toma de decisiones relacionada con la creación de trayectos turísticos o de transporte desde o hacia instalaciones turísticas.

También se ha demostrado que, a través de la utilización de metodologías basadas en *Data Stream Mining* y redes neuronales, pueden obtenerse resultados que como mínimo igualan en precisión y velocidad a las soluciones más frecuentes en este campo. No obstante, en Canarias aún no existen vías sensorizadas que generen datos como para poder crear modelos predictivos basados en estas metodologías, aunque la capacidad adaptativa de dichos modelos, aunque hayan sido entrenados mediante otros datos, permite su despliegue en diferentes infraestructuras. Es por ello por lo que la tercera y última recomendación de esta tesis englobaría dos actuaciones, siendo la primera y más urgente la elaboración de un plan de sensorización de las infraestructuras viales del archipiélago, con un mayor enfoque en las autovías principales de cada isla, con el objetivo de generar datos suficientes para crear modelos predictivos en tiempo real capaces de ayudar en la toma de decisiones a múltiples niveles y en múltiples sectores, entre ellos el turístico. Como segunda actuación, se debería crear un plan de investigación con el objetivo de probar diversos modelos predictivos basados en *Data Stream Mining* a fin de descubrir cual genera mejores predicciones, estableciendo así un *framework* capaz de facilitar la toma de decisiones mencionada anteriormente y de mejorar la calidad de la circulación en las autovías del archipiélago, lo que a su vez aumentaría la calidad de la estancia no solo de los turistas, sino también de los habitantes de cada isla.

7. Referencias.

Afzaal, M., Usman, M., Fong, A. C. M., Fong, S., & Zhuang, Y. (2016). Fuzzy Aspect Based Opinion Classification System for Mining Tourist Reviews. *Advances in Fuzzy Systems, 2016.* https://doi.org/10.1155/2016/6965725

Aha, D. W., Kibler, D., & Albert, M. K. (1991). Instance-based learning algorithms. *Machine Learning*, *6*(1), 37-66. https://doi.org/10.1007/BF00153759

Akın, M. (2015). A novel approach to model selection in tourism demand modeling. *Tourism Management*, *48*, 64-72. https://doi.org/10.1016/j.tourman.2014.11.004

Al Shehhi, M., & Karathanasopoulos, A. (2020). Forecasting hotel room prices in selected GCC cities using deep learning. *Journal of Hospitality and Tourism Management, 42*, 40-50. https://doi.org/10.1016/j.jhtm.2019.11.003

Albalate, D., & Bel, G. (2010). Tourism and urban public transport: Holding demand pressure under supply constraints. *Tourism Management, 31*(3), 425-433. https://doi.org/10.1016/j.tourman.2009.04.011

Ali, R., Shabri, A., Arifin, N., & Suhaila, Y. (2017). A Wavelet Support Vector Machine Combination Model for Singapore Tourist Arrival to Malaysia. *IOP Conference Series: Materials Science and Engineering, 226*. https://doi.org/10.1088/1757-899X/226/1/012077

Antonio, N., De Almeida, A., & Nunes, L. (2017). Predicting hotel booking cancellations to decrease uncertainty and increase revenue. *Tourism & Management Studies, 13*(2), 25-39.

Antonio, N., De Almeida, A., & Nunes, L. (2016). Using Data Science to Predict Hotel Booking Cancellations. En P.Vassant, & K.M. (Ed.), *Handbook of Research on Holistic Optimization Techniques in the Hospitality, Tourism, and Travel Industry* (pp. 140-166). IGI Global. https://doi.org/10.4018/978-1-5225-1054-3.ch006

Anuradha, J. (2015). A brief introduction on Big Data 5Vs characteristics and Hadoop technology. *Procedia computer science, 48*, 319-324.

Arruza, M., Pericich, J., & Straka, M. (2016). The Automated Travel Agent: Hotel Recommendations Using Machine Learning.

Athanasiou, V., & Maragoudakis, M. (2016). Dealing with High Dimensional Sentiment Data Using Gradient Boosting Machines. En L. Iliadis & I. Maglogiannis (Eds.), *Artificial Intelligence Applications and Innovations*, 2016 (Vol. 475, pp. 481-489). Springer-Verlag Berlin.

Atsalakis, G. S., Atsalaki, I. G., & Zopounidis, C. (2018). Forecasting the success of a new tourism service by a neuro-fuzzy technique. *European Journal of Operational Research, 268*(2), 716-727. https://doi.org/10.1016/j.ejor.2018.01.044

Ayuntamiento de Madrid, A. de M. (2022). Repositorio público de datos de la autopista M-30 de Madrid. Repositorio público de datos de la autopista M-30 de Madrid. https://datos.madrid.es/portal/site/egob/menuitem.9e1e2f6404558187cf35cf3584f1a5a0/?vgnextoid=374512b9ace9f310VgnVCM100000171f5a0aRCRD&vgnextchannel=374512b9ace9f310VgnVCM100000171f5a0aRCRD&vgnextfmt=default. (Revisado el 30 de mayo de 2022).

Banerjee, S., Chua, A. Y. K., & Kim, J.-J. (2015). Distinguishing between authentic and fictitious user-generated hotel reviews. *6th International Conference on Computing, Communication and Networking Technologies (ICCCNT), 2015,* 1-7. https://doi.org/10.1109/ICCCNT.2015.7395179

Barredo-Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R., & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion, 58,* 82-115. https://doi.org/10.1016/j.inffus.2019.12.012

Bechtel, B. (2016). The Climate of the Canary Islands by Annual Cycle Parameters. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLI-B8,* 243-250. https://doi.org/10.5194/isprs-archives-XLI-B8-243-2016

Belmonte-Fernández, Ó., Montoliu, R., Torres-Sospedra, J., Sansano-Sansano, E., & Chia-Aguilar, D. (2018). A radiosity-based method to avoid calibration for indoor positioning systems. *Expert Systems with Applications, 105,* 89-101. https://doi.org/10.1016/j.eswa.2018.03.054

Beni, G., & Wang, J. (1993). Swarm Intelligence in Cellular Robotic Systems. En P. Dario, G. Sandini, & P. Aebischer (Eds.), *Robots and Biological Systems: Towards a New Bionics?* (pp. 703-712). Springer. https://doi.org/10.1007/978-3-642-58069-7_38

Berenguer, T. M., Berenguer, J. A. M., García, M. E. B., Pol, A. P., & Moreno, J. J. M. (2014). Models of artificial neural networks applied to demand forecasting in nonconsolidated tourist destinations. Methodology: European Journal of Research Methods for the *Behavioral and Social Sciences, 11*(2), 35-44. http://dx.doi.org.bibproxy.ulpgc.es/10.1027/1614-2241/a000088

Bermingham, L., & Lee, I. (2014). Spatio-temporal Sequential Pattern Mining for Tourism Sciences. *Procedia Computer Science, 29,* 379-389. https://doi.org/10.1016/j.procs.2014.05.034

Bernard, M. (2016). How Big Data And Analytics Are Changing Hotels And The Hospitality Industry. *Forbes, Enero 2016.* https://www.forbes.com/sites/bernardmarr/2016/01/26/how-big-data-and-analytics-changing-hotels-and-the-hospitality-industry/

Bettin, R., Mason, F., Corazza, M., & Fasano, G. (2011). An Artificial Neural Network Technique for On-Line Hotel Booking. *Social Science Research Network, 2012*. https://papers.ssrn.com/abstract=2037481

Beven, J. (2006). Tropical cyclone report: Tropical Storm Delta, 22–28 November 2005. *National Hurricane Center, 12.*

Bifet, A., & Kirkby, R. (2009). DATA STREAM MINING A Practical Approach.

Biuk-Aghai, R. P., Fong, S., & Si, Y. W. (2008). Design of a recommender system for mobile tourism multimedia selection. *2008 2nd International Conference on Internet Multimedia*

Services Architecture and Applications, 2008 1-6. https://doi.org/10.1109/IMSAA.2008.4753931

Bossanyi, E. (1985). Short-term wind prediction using Kalman filters. *Wind Engineering, 9*(1), 1-8.

Bradley, J., Barbier, J., & Handler, D. (2013). Embracing the Internet of everything to capture your share of $14.4 trillion. *White Paper,* Cisco, 318. https://www.cisco.com/c/dam/en_us/about/ac79/docs/innov/IoE_Economy.pdf

Braess, D. (1968). Über ein Paradoxon aus der Verkehrsplanung. *Unternehmensforschung, 12*(1), 258-268. https://doi.org/10.1007/BF01918335

Brahma, A., & Panigrahi, S. (2021). Role-Based Profiling Using Fuzzy Adaptive Resonance Theory for Securing Database Systems. *International Journal of Applied Metaheuristic Computing (IJAMC), 12*(2), 36-48. https://doi.org/10.4018/IJAMC.2021040103

Brakerski, Z., Gentry, C., & Vaikuntanathan, V. (2014). (Leveled) fully homomorphic encryption without bootstrapping. *ACM Transactions on Computation Theory (TOCT), 6*(3), 1-36.

Brida, J. G., Lanzilotta, B., Moreno, L., & Santiñaque, F. (2018). A non-linear approximation to the distribution of total expenditure distribution of cruise tourists in Uruguay. *Tourism Management, 69,* 62-68. https://doi.org/10.1016/j.tourman.2018.05.006

Briguglio, L., & Briguglio, M. (2005). Sustainable tourism in small islands: The case of Malta. *Sustainable tourism in islands and small states: Case studies*, 226-317.

Budiono, O. (2009). Customer Satisfaction in Public Bus Transport: A study of travelers' perception in Indonesia.

Bugarski, V., Matić, D., & Kulić, F. (2017). Classification of hotel guests by predicted additional spending with ANN decision support system. *2017 IEEE 15th International*

Symposium on Intelligent Systems and Informatics (SISY), 2017, 000071-000076. https://doi.org/10.1109/SISY.2017.8080528

Cai, Z., Lu, S., & Zhang, X. (2009). Tourism demand forecasting by support vector regression and genetic algorithm. *2009 2nd IEEE International Conference on Computer Science and Information Technology, 2009,* 144-146. https://doi.org/10.1109/ICCSIT.2009.5234447

Cankurt, S. (2016). Tourism demand forecasting using ensembles of regression trees. 2016 *IEEE 8th International Conference on Intelligent Systems (IS), 2016,* 702-708. https://doi.org/10.1109/IS.2016.7737388

Cankurt, S., & Subasi, A. (2015). Developing tourism demand forecasting models using machine learning techniques with trend, seasonal, and cyclic components. Balkan *Journal of Electrical and Computer Engineering, 3(*1). https://doi.org/10.17694/bajece.96106

Cankurt, S., & Subaşi, A. (2016). Tourism demand modelling and forecasting using data mining techniques in multivariate time series: A case study in Turkey. *Turkish Journal of Electrical Engineering & Computer Sciences, 24*(5), 3388-3404.

Cao, G., & Wu, L. (2016). Support vector regression with fruit fly optimization algorithm for seasonal electricity consumption forecasting. *Energy, 115*, 734-745. https://doi.org/10.1016/j.energy.2016.09.065

Carpenter, G. A., & Grossberg, S. (1987). ART 2: Self-organization of stable category recognition codes for analog input patterns. *Applied Optics, 26*(23), 4919-4930. https://doi.org/10.1364/AO.26.004919

Carpenter, G. A., & Grossberg, S. (1990). ART 3: Hierarchical search using chemical transmitters in self-organizing pattern recognition architectures. *Neural Networks, 3*(2), 129-152. https://doi.org/10.1016/0893-6080(90)90085-Y

Carpenter, G. A., Grossberg, S., & Reynolds, J. H. (1991). ARTMAP: Supervised real-time learning and classification of nonstationary data by a self-organizing neural network.

IEEE Conference on Neural Networks for Ocean Engineering, 1991, 341-342. https://doi.org/10.1109/ICNN.1991.163370

Carpenter, G. A., Grossberg, S., & Rosen, D. B. (1991a). ART 2-A: An adaptive resonance algorithm for rapid category learning and recognition. *Neural Networks, 4*(4), 493-504. https://doi.org/10.1016/0893-6080(91)90045-7

Carpenter, G. A., Grossberg, S., & Rosen, D. B. (1991b). Fuzzy ART: Fast stable learning and categorization of analog patterns by an adaptive resonance system. *Neural Networks, 4*(6), 759-771. https://doi.org/10.1016/0893-6080(91)90056-B

Carpenter, G., & Grossberg, S. (1988). Adaptive Resonance Theory (art). En *The Handbook of Brain Theory And Neural Networks* (pp. 79-82). MIT Press.

Carpinone, A., Giorgio, M., Langella, R., & Testa, A. (2015). Markov chain modeling for very-short-term wind power forecasting. *Electric Power Systems Research, 122*, 152-158.

Cassola, F., & Burlando, M. (2012). Wind speed and wind energy forecast through Kalman filtering of Numerical Weather Prediction model output. *Applied Energy, 99*, 154-166.

Casteleiro-Roca, J. -L., Gómez-González, J. F., Calvo-Rolle, J. L., Jove, E., Quintián, H., Gonzalez Diaz, B., & Mendez Perez, J. A. (2019). Short-Term Energy Demand Forecast in Hotels Using Hybrid Intelligent Modeling. *Sensors, 19*(11), 2485. https://doi.org/10.3390/s19112485

Chang, W., & Ma, L. (2013). Personalized e-tourism attraction recommendation based on context. *2013 10th International Conference on Service Systems and Service Management, 2013*, 674-679. https://doi.org/10.1109/ICSSSM.2013.6602591

Chang, Y. W., & Tsai, C. Y. (2017). Apply Deep Learning Neural Network to Forecast Number of Tourists. *2017 31st International Conference on Advanced Information Networking and Applications Workshops (WAINA), 2017,* 259-264. https://doi.org/10.1109/WAINA.2017.125

Chang, Y.-C., Ku, C.-H., & Chen, C.-H. (2020). Using deep learning and visual analytics to explore hotel reviews and responses. *Tourism Management, 80,* 104129. https://doi.org/10.1016/j.tourman.2020.104129

Chen, C.-F., Lai, M.-C., & Yeh, C.-C. (2012). Forecasting tourism demand based on empirical mode decomposition and neural network. *Knowledge-Based Systems, 26*, 281-287. https://doi.org/10.1016/j.knosys.2011.09.002

Chen, J. H., Chao, K. M., & Shah, N. (2013). Hybrid Recommendation System for Tourism. *2013 IEEE 10th International Conference on e-Business Engineering, 2013,* 156-161. https://doi.org/10.1109/ICEBE.2013.24

Chen, J., Li, D., Zhang, G., & Zhang, X. (2018). Localized space-time autoregressive parameters estimation for traffic flow prediction in urban road networks. *Applied Sciences, 8(*2). Scopus. https://doi.org/10.3390/app8020277

Chen, K.-Y., & Wang, C.-H. (2007). Support vector regression with genetic algorithms in forecasting tourism demand. *Tourism Management, 28*(1), 215-226. https://doi.org/10.1016/j.tourman.2005.12.018

Chen, M., Yu, X., & Liu, Y. (2018). PCNN: Deep Convolutional Networks for Short-Term Traffic Congestion Prediction. *IEEE Transactions on Intelligent Transportation Systems, 19*(11), 3550-3559. https://doi.org/10.1109/TITS.2018.2835523

Chen, M.-S., Ying, L.-C., & Pan, M.-C. (2010). Forecasting tourist arrivals by using the adaptive network-based fuzzy inference system. *Expert Systems with Applications, 37*(2), 1185-1191. https://doi.org/10.1016/j.eswa.2009.06.032

Chen, R., Liang, C.-Y., Hong, W.-C., & Gu, D.-X. (2015). Forecasting holiday daily tourist flow based on seasonal support vector regression with adaptive genetic algorithm. *Applied Soft Computing, 26*, 435-443. https://doi.org/10.1016/j.asoc.2014.10.022

Chen, Y., Tan, H., & Berardi, U. (2017). Day-ahead prediction of hourly electric demand in non-stationary operated commercial buildings: A clustering-based hybrid approach. *Energy and Buildings, 148*, 228-237. https://doi.org/10.1016/j.enbuild.2017.05.003

Chen, Y., Tan, H., & Song, X. (2017). Day-ahead Forecasting of Non-stationary Electric Power Demand in Commercial Buildings: Hybrid Support Vector Regression Based. *Energy Procedia, 105,* 2101-2106. https://doi.org/10.1016/j.egypro.2017.03.590

Cheng, A., Jiang, X., Li, Y., Zhang, C., & Zhu, H. (2017). Multiple sources and multiple measures based traffic flow prediction using the chaos theory and support vector regression method. *Physica A: Statistical Mechanics and Its Applications, 466*, 422-434. https://doi.org/10.1016/j.physa.2016.09.041

Cheng, X., Fu, S., Sun, J., Bilgihan, A., & Okumus, F. (2019). An investigation on online reviews in sharing economy driven hospitality platforms: A viewpoint of trust. *Tourism Management, 71,* 366-377. https://doi.org/10.1016/j.tourman.2018.10.020

Chiu, C., Chiu, N.-H., Sung, R.-J., & Hsieh, P.-Y. (2015). Opinion mining of hotel customer-generated contents in Chinese weblogs. *Current Issues in Tourism, 18*(5), 477-495. https://doi.org/10.1080/13683500.2013.841656

Cho, V., & Leung, P. (2002). Towards Using Knowledge Discovery Techniques in Database Marketing for the Tourism Industry. *Journal of Quality Assurance in Hospitality & Tourism, 3(*3-4), 109-131. https://doi.org/10.1300/J162v03n03_07

Chou Jui-Sheng & Lin Chieh. (2013). Predicting Disputes in Public-Private Partnership Projects: Classification and Ensemble Models. *Journal of Computing in Civil Engineering, 27*(1), 51-60. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000197

Claster, W. B., Dinh, H., & Cooper, M. (2010). Naïve Bayes and unsupervised artificial neural nets for Cancun tourism social media data analysis. *2010 Second World Congress on Nature and Biologically Inspired Computing (NaBIC), 2010,* 158-163. https://doi.org/10.1109/NABIC.2010.5716370

Claveria, O., Monte, E., & Torra, S. (2015). Tourism Demand Forecasting with Neural Network Models: Different Ways of Treating Information. *International Journal of Tourism Research, 17(*5), 492-500. https://doi.org/10.1002/jtr.2016

Claveria, O., Monte, E., & Torra, S. (2016a). Combination forecasts of tourism demand with machine learning models. *Applied Economics Letters, 23*(6), 428-431. https://doi.org/10.1080/13504851.2015.1078441

Claveria, O., Monte, E., & Torra, S. (2016b). Modelling cross-dependencies between Spain's regional tourism markets with an extension of the Gaussian process regression model. *SERIEs, 7*(3), 341-357. https://doi.org/10.1007/s13209-016-0144-7

Claveria, O., Monte, E., & Torra, S. (2017). Regional Tourism Demand Forecasting with Machine Learning Models: Gaussian Process Regression vs. Neural Network Models in a Multiple-Input Multiple-Output Setting (SSRN Scholarly Paper ID 2945556). Social Science Research Network. https://papers.ssrn.com/abstract=2945556

Claveria, O., Torra, S., & Monte, E. (2016). Modelling Tourism Demand to Spain With Machine Learning Techniques. The Impact of Forecast Horizon On Model Selection. *Revista de Economía Aplicada, 24*(72), 109-132.

Commission (EC), E. (2010). Europe 2020: A Strategy for Smart, Sustainable and Inclusive Growth.

da Silva, M. A., Abreu, T., Santos-Júnior, C. R., & Minussi, C. R. (2021). Load forecasting for smart grid based on continuous-learning neural network. *Electric Power Systems Research, 201,* 107545. https://doi.org/10.1016/j.epsr.2021.107545

Dalto, M., Matuško, J., & Vašak, M. (2015). Deep neural networks for ultra-short-term wind forecasting. *2015 IEEE international conference on industrial technology (ICIT), 2015,* 1657-1663.

Dawid, A. P. (1984). Present Position and Potential Developments: Some Personal Views Statistical Theory the Prequential Approach. *Journal of the Royal Statistical Society: Series A (General), 147*(2), 278-290. https://doi.org/10.2307/2981683

Demchenko, Y., Grosso, P., De Laat, C., & Membrey, P. (2013). Addressing big data issues in scientific data infrastructure. *2013 International conference on collaboration technologies and systems (CTS), 2013,* 48-55.

Deng, N., & Li, X. (Robert). (2018). Feeling a destination through the "right" photos: A machine learning model for DMOs' photo selection. *Tourism Management, 65*, 267-278. https://doi.org/10.1016/j.tourman.2017.09.010

Dickinger, A., & Mazanec, J. A. (2015). Significant word items in hotel guest reviews: A feature extraction approach. *Tourism Recreation Research, 40*(3), 353-363. https://doi.org/10.1080/02508281.2015.1079964

Dimitrakopoulos, G., & Demestichas, P. (2010). Intelligent Transportation Systems. *IEEE Vehicular Technology Magazine, 5*(1), 77-84. https://doi.org/10.1109/MVT.2009.935537

Drucker, P. F. (2002). *Managing in the next society (1st ed).* St. Martin's Press.

Duan, W., Cao, Q., Yu, Y., & Levy, S. (2013). Mining Online User-Generated Content: Using Sentiment Analysis Technique to Study Hotel Service Quality. *2013 46th Hawaii International Conference on System Sciences, 2013,* 3119-3128. https://doi.org/10.1109/HICSS.2013.400

Duval, D. D. T. (2007). *Tourism and Transport: Modes, Networks and Flows*. Channel View Publications.

Ebadi, A. (2016). An intelligent hybrid multi-criteria hotel recommender system using explicit and implicit feedbacks [Masters, Concordia University]. https://spectrum.library.concordia.ca/981086/

Eiben, A. E., & Smith, J. E. (2015). *Introduction to Evolutionary Computing.* Springer. https://doi.org/10.1007/978-3-662-44874-8

Emel, G., & Taşkın, Ç. (2005). Identifying Segments of a Domestic Tourism Market by Means of Data Mining. *Operations Research Proceedings, 2005,* 653-658. https://doi.org/10.1007/3-540-32539-5_102

Fan, J., & Vercauteren, F. (2012). Somewhat practical fully homomorphic encryption. *Cryptology ePrint Archive, 2012*, 144.

Fellesson, M., & Friman, M. (2012). Perceived Satisfaction with Public Transport Service in Nine European Cities. *Journal of the Transportation Research Forum, 47*. https://doi.org/10.5399/osu/jtrf.47.3.2126

Folgieri, R., Baldigara, T., & Mamula, M. (2017). Artificial Neural Networks-Based Econometric Models for Tourism Demand Forecasting. En S. Markovic & D. S. Jurdana (Eds.), *4th International Scientific Conference: Tosee—Tourism in Southern and Eastern Europe 2017* (Vol. 4, pp. 169-182). Univ Rijeka, Faculty Tourism & Hospitality Management, Opatija. https://www.bib.irb.hr/917314

Gama, J., & Gaber, M. M. (2007). *Learning from data streams: Processing techniques in sensor networks.* Springer.

García-Barriocanal, E., Sicilia, M.-A., & Korfiatis, N. (2010). Exploring Hotel Service Quality Experience Indicators In *User-Generated Content: A Case Using TripAdvisor Data*. MCIS 2010 Proceedings. https://aisel.aisnet.org/mcis2010/33

Gavalas, D., & Kenteris, M. (2011). A web-based pervasive recommendation system for mobile tourist guides. *Personal and Ubiquitous Computing, 15*(7), 759-770. https://doi.org/10.1007/s00779-011-0389-x

Gawlik, E., Kabaria, H., & Kaur, E. (2011). Predicting tourism trends with Google Insights, 2011.

Gayar, N., Hendawi, A., & El-Shishiny, H. (2008). *A proposed Decision Support Model for Hotel Revenue Management.*

Gentry, C. (2009). *A fully homomorphic encryption scheme*. Stanford university.

Gerstner, W., & Kistler, W. M. (2002). *Spiking Neuron Models: Single Neurons, Populations, Plasticity*. Cambridge University Press.

Geyer, R. C., Klein, T., & Nabi, M. (2017). Differentially private federated learning: A client level perspective. *31st Conference on Neural Information Processing Systems (NIPS 2017), 2017*.

Gokaraju, B., S. Durbha, S., King, R., & Younan, N. H. (2011). A Machine Learning Based Spatio-Temporal Data Mining Approach for Detection of Harmful Algal Blooms in the Gulf of Mexico. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 4*(3), 710-720. https://doi.org/10.1109/JSTARS.2010.2103927

Grossberg, S. (1976a). Adaptive pattern classification and universal recoding: I. Parallel development and coding of neural feature detectors. *Biological Cybernetics, 23*(3), 121-134. https://doi.org/10.1007/BF00344744

Grossberg, S. (1976b). Adaptive pattern classification and universal recoding: II. Feedback, expectation, olfaction, illusions. *Biological Cybernetics, 23*(4), 187-202. https://doi.org/10.1007/BF00340335

Grossberg, S. (1987). Competitive learning: From interactive activation to adaptive resonance. *Cognitive Science, 11*(1), 23-63. https://doi.org/10.1016/S0364-0213(87)80025-3

Guerra-Montenegro, J., Sanchez-Medina, J., Laña, I., Sanchez-Rodriguez, D., Alonso-Gonzalez, I., & Ser, J. D. (2021). Computational Intelligence in the hospitality industry: A systematic literature review and a prospect of challenges. *Applied Soft Computing, 102*, 107082. https://doi.org/10.1016/j.asoc.2021.107082

Gunning, D., & Aha, D. (2019). DARPA's Explainable Artificial Intelligence (XAI) Program. *AI Magazine, 40(*2), 44-58. https://doi.org/10.1609/aimag.v40i2.2850

Guo, G., & Yuan, W. (2020). Short-term traffic speed forecasting based on graph attention temporal convolutional networks. *Neurocomputing, 410*, 387-393. https://doi.org/10.1016/j.neucom.2020.06.001

Guo, Y., Barnes, S. J., & Jia, Q. (2017). Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent dirichlet allocation. *Tourism Management, 59*, 467-483. https://doi.org/10.1016/j.tourman.2016.09.009

Guoxia, Z., & Jianqing, T. (2009). The Application of Data Mining in Tourism Information. *2009 International Conference on Environmental Science and Information Application Technology, 3*, 689-692. https://doi.org/10.1109/ESIAT.2009.193

Ha, S. H., & Park, S. C. (1998). Application of data mining tools to hotel data mart on the Intranet for database marketing. *Expert Systems with Applications, 15*(1), 1-31. https://doi.org/10.1016/S0957-4174(98)00008-6

Hadavandi, E., Ghanbari, A., Shahanaghi, K., & Abbasian-Naghneh, S. (2011). Tourist arrival forecasting by evolutionary fuzzy systems. *Tourism Management, 32*(5), 1196-1203. https://doi.org/10.1016/j.tourman.2010.09.015

Han, S., Guo, Y., Cao, H., Feng, Q., & Li, Y. (2017). A Cross-View Model for Tourism Demand Forecasting with Artificial Intelligence Method. *Data Science, 2017,* 573-582. https://doi.org/10.1007/978-981-10-6385-5_48

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., del Río, J. F., Wiebe, M., Peterson, P., … Oliphant, T. E. (2020). Array programming with NumPy. *Nature, 585*(7825), 357-362. https://doi.org/10.1038/s41586-020-2649-2

Haruechaiyasak, C., Kongthon, A., Palingoon, P., & Sangkeettrakarn, C. (2010). Constructing Thai Opinion Mining Resource: A Case Study on Hotel Reviews. *Proceedings of the Eighth Workshop on Asian Language Resources, 2010,* 64-71.

Hastie, T. J., & Tibshirani, R. J. (2017). *Generalized additive models.* Routledge.

Healy, M. J., Caudell, T. P., & Smith, S. G. (1993). A neural architecture for pattern sequence verification through inferencing. *IEEE Transactions on Neural Networks, 4*(1), 9-20. https://doi.org/10.1109/72.182691

Holland, J. H. (1992). *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control and Artificial Intelligence*. MIT Press.

Hong, W.-C., Dong, Y., Chen, L.-Y., & Wei, S.-Y. (2011). SVR with hybrid chaotic genetic algorithms for tourism demand forecasting. *Applied Soft Computing, 11*(2), 1881-1890. https://doi.org/10.1016/j.asoc.2010.06.003

Hong, W.-C., Dong, Y., Zhang, W. Y., Chen, L.-Y., & K. Panigrahi, B. (2013). Cyclic electric load forecasting by seasonal SVR with chaotic genetic algorithm. *International Journal of Electrical Power & Energy Systems, 44*(1), 604-614. https://doi.org/10.1016/j.ijepes.2012.08.010

Hsieh, H. Y., Klyuev, V., Zhao, Q., & Wu, S. H. (2014). SVR-based outlier detection and its application to hotel ranking. *2014 IEEE 6th International Conference on Awareness Science and Technology (iCAST), 2014* 1-6. https://doi.org/10.1109/ICAwST.2014.6981842

Hsu, C.-C., Wu, C.-H., Chen, S.-C., & Peng, K.-L. (2006). Dynamically Optimizing Parameters in Support Vector Regression: An Application of Electricity Load Forecasting. *Proceedings of the 39th Annual Hawaii International Conference on System Sciences (HICSS'06), 2*, 30c-30c. https://doi.org/10.1109/HICSS.2006.132

Hsu, C.-I., Shih, M.-L., Huang, B.-W., Lin, B.-Y., & Lin, C.-N. (2009). Predicting tourism loyalty using an integrated Bayesian network mechanism. *Expert Systems with Applications, 36(*9), 11760-11763. https://doi.org/10.1016/j.eswa.2009.04.010

Hsu, F.-M., Lin, Y.-T., & Ho, T.-K. (2012). Design and implementation of an intelligent recommendation system for tourist attractions: The integration of EBM model, Bayesian network and Google Maps. *Expert Systems with Applications, 39*(3), 3257-3264. https://doi.org/10.1016/j.eswa.2011.09.013

Hu, J., Xie, C., Xu, L., Qi, X., Zhu, S., Zhu, H., Dong, J., Cheng, P., & Zhou, Z. (2021). Direct Analysis of Soil Composition for Source Apportionment by Laser Ablation Single-Particle Aerosol Mass Spectrometry. *Environmental Science & Technology, 55*(14), 9721-9729. https://doi.org/10.1021/acs.est.0c07983

Huang, C.-J., & Kuo, P.-H. (2018). A deep CNN-LSTM model for particulate matter (PM2. 5) forecasting in smart cities. *Sensors, 18*(7), 2220.

Huang, H.-C., Y Chang, A., & Ho, C.-C. (2013). Using Artificial Neural Networks to Establish a Customer-cancellation Prediction Model. *Przeglad Elektrotechniczny, 89*, 178-180.

Hunter, J. D. (2007). Matplotlib: A 2D Graphics Environment. *Computing in Science Engineering, 9*(3), 90-95. https://doi.org/10.1109/MCSE.2007.55

Ishwaran, H., & Rao, J. S. (2009). Decision Tree: Introduction. En Kattan (Ed.), *Encyclopedia of Medical Decision Making* (pp. 323-328). Sage Publications.

Ispas, A., & Saragea, R.-A. (2011). Evaluating the image of tourism destinations. The case of the autonomous community of the Canary Islands. *Revista de Turism-Studii si Cercetari in Turism,* 12, 6-12.

Jiang, K., Wang, P., & Yu, N. (2011). ContextRank: Personalized Tourism Recommendation by Exploiting Context Information of Geotagged Web Photos. *2011 Sixth International Conference on Image and Grphics, 2011*, 931-937. https://doi.org/10.1109/ICIG.2011.48

Jiang, K., Yin, H., Wang, P., & Yu, N. (2013). Learning from contextual information of geo-tagged web photos to rank personalized tourism attractions. *Neurocomputing, 119*, 17-25. https://doi.org/10.1016/j.neucom.2012.02.049

Jiang, Y., Song, Z., & Kusiak, A. (2013). Very short-term wind speed forecasting with Bayesian structural break model. *Renewable energy, 50*, 637-647.

Jo, M.-W., Kim, H., & Shin, H.-J. (2016). Understanding traffic congestion to improve tourist satisfaction in local tourism. *International Journal of Tourism and Hospitality Research, 30*, 85. https://doi.org/10.21298/IJTHR.2016.04.30.4.85

Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science, 349*(6245), 255-260. https://doi.org/10.1126/science.aaa8415

Junping, D., Min, Z., & Xuyan, T. (2008). The realization of distributed sampling association rule mining algorithm in tourism. *2008 7th World Congress on Intelligent Control and Automation, 2008*, 183-187. https://doi.org/10.1109/WCICA.2008.4592921

Kamel, N., Atiya, A., Gayar, N., & El-Shishiny, H. (2018). Tourism Demand Foreacsting using Machine Learning Methods. *ICGST International Journal on Artificial Intelligence and Machine Learning, 8*, 1-7.

Kampouropoulos, K., Andrade, F., Garcia, A., & Romeral, L. (2014). A Combined Methodology of Adaptive Neuro-Fuzzy Inference System and Genetic Algorithm for Short-term Energy Forecasting. *Advances in Electrical and Computer Engineering, 14*(1), 9-14. https://doi.org/10.4316/AECE.2014.01002

Kasper, W., & Vela, M. (2012). Sentiment Analysis for Hotel Reviews. *Proceedings of the Computational Linguistics-Applications Conference, 2012*, 45-52.

Kbaier, M. E. B. H., Masri, H., & Krichen, S. (2017). A personalized hybrid tourism recommender system. *2017 IEEE 14th International Conference on Computer Systems and Applications (AICCSA), 2017,* 244-250.

Kebede, G. (2010). Knowledge management: An information science perspective. International *Journal of Information Management, 30*(5), 416-424. https://doi.org/10.1016/j.ijinfomgt.2010.02.004

Kim, S. Y. (2011). Prediction of hotel bankruptcy using support vector machine, artificial neural network, logistic regression, and multivariate discriminant analysis. *The Service Industries Journal, 31*(3), 441-468. https://doi.org/10.1080/02642060802712848

King, M. A., Abrahams, A. S., & Ragsdale, C. T. (2014). Ensemble methods for advanced skier days prediction. *Expert Systems with Applications, 41*(4, Part 1), 1176-1188. https://doi.org/10.1016/j.eswa.2013.08.002

Kofinas, D., Papageorgiou, E., Laspidou, C., Mellios, N., & Kokkinos, K. (2016). Daily Multivariate Forecasting of Water Demand in a Touristic Island with the Use of Artificial Neural Network and Adaptive Neuro-Fuzzy Inference System. *2016 International Workshop on Cyber-physical Systems for Smart Water Networks (CySWater), 2016*, 37-42. https://doi.org/10.1109/CySWater.2016.7469061

Konečný, J., McMahan, H. B., Yu, F. X., Richtárik, P., Suresh, A. T., & Bacon, D. (2017). Federated Learning: Strategies for Improving Communication Efficiency. *NIPS Workshop on Private Multi-Party Machine Learning, 2017.*

Kotsiantis, S., & Kanellopoulos, D. (2005). *Association Rules Mining: A Recent Overview.*

Krawczyk, B., Minku, L. L., Gama, J., Stefanowski, J., & Woźniak, M. (2017). Ensemble learning for data stream analysis: A survey. *Information Fusion, 37,* 132-156. https://doi.org/10.1016/j.inffus.2017.02.004

Kumar Duvvur, P., & Romanowski, C. J. (2016). *Predicting Hotel Rating Based on User Reviews.*

Lahouar, A., & Ben Hadj Slama, J. (2017). Hour-ahead wind power forecast based on random forests. *Renewable Energy, 109,* 529-541. https://doi.org/10.1016/j.renene.2017.03.064

Laña, I., Lobo, J. L., Capecci, E., Del Ser, J., & Kasabov, N. (2019). Adaptive long-term traffic state estimation with evolving spiking neural networks. *Transportation Research Part C: Emerging Technologies, 101*, 126-144. https://doi.org/10.1016/j.trc.2019.02.011

Lei, W. S. (Clara), & Lam, C. C. (Cindia). (2015). Determinants of hotel occupancy rate in a Chinese gaming destination. *Journal of Hospitality and Tourism Management, 22*, 1-9. https://doi.org/10.1016/j.jhtm.2014.12.003

Le-Klähn, D.-T., & Hall, C. M. (2015). Tourist use of public transport at destinations – a review. *Current Issues in Tourism, 18*(8), 785-803. https://doi.org/10.1080/13683500.2014.948812

León, F. M., Serdán, J. M. F., & Beirán, I. S. A. (2006). La inusual y anómala tormenta tropical «Delta». *Ambienta: La revista del Ministerio de Medio Ambiente, 52*, 60-65.

Li, G., Law, R., Vu, H. Q., Rong, J., & Zhao, X. (Roy). (2015). Identifying emerging hotel preferences using Emerging Pattern Mining technique. *Tourism Management, 46,* 311-321. https://doi.org/10.1016/j.tourman.2014.06.015

Li, G., & Shi, J. (2010). On comparing three artificial neural networks for wind speed forecasting. *Applied Energy, 87*(7), 2313-2320.

Li, H., & Sun, J. (2012). Forecasting business failure: The use of nearest-neighbour support vectors and correcting imbalanced samples – Evidence from the Chinese hotel industry. *Tourism Management, 33*(3), 622-634. https://doi.org/10.1016/j.tourman.2011.07.004

Li, Q., Wu, Y., Wang, S., Lin, M., Feng, X., & Wang, H. (2016). VisTravel: Visualizing tourism network opinion from the user generated content. *Journal of Visualization, 19*(3), 489-502. https://doi.org/10.1007/s12650-015-0330-x

Li, W., Xu, S., & Meng, W. (2009). A support vector machines method for tourist satisfaction degree evaluation. *2009 6th International Conference on Service Systems and Service Management, 2009,* 883-887. https://doi.org/10.1109/ICSSSM.2009.5175007

Liang, C., & Bi, W. (2017). Seasonal variation analysis and SVR forecast of tourist flows during the year: A case study of Huangshan mountain. *2017 IEEE 2nd International Conference on Big Data Analysis (ICBDA), 2017,* 921-927. https://doi.org/10.1109/ICBDA.2017.8078773

Liao, S., Chen, Y.-J., & Deng, M. (2010). Mining customer knowledge for tourism new product development and customer relationship management. *Expert Systems with Applications, 37*(6), 4212-4223. https://doi.org/10.1016/j.eswa.2009.11.081

Lijuan, W., & Guohua, C. (2016). Seasonal SVR with FOA algorithm for single-step and multi-step ahead forecasting in monthly inbound tourist flow. *Knowledge-Based Systems, 110,* 157-166. https://doi.org/10.1016/j.knosys.2016.07.023

Lin, C., & Chao, P. (2010). Tourism-Related Opinion Detection and Tourist-Attraction Target Identification. *Computational Linguistics and Chinese Language Processing, 15.*

Lin, C.-T., & Huang, Y.-L. (2009). Mining tourist imagery to construct destination image position model. *Expert Systems with Applications, 36(*2, Part 1), 2513-2524. https://doi.org/10.1016/j.eswa.2008.01.074

Lin, K.-P., Pai, P.-F., Lu, Y.-M., & Chang, P.-T. (2013). Revenue forecasting using a least-squares support vector regression model in a fuzzy environment. *Information Sciences, 220*, 196-209. https://doi.org/10.1016/j.ins.2011.09.003

Lin, Y., Wang, P., & Ma, M. (2017). Intelligent Transportation System(ITS): Concept, Challenge and Opportunity. *2017 IEEE 3rd international conference on big data security on cloud (bigdatasecurity), IEEE international conference on high performance and smart computing (hpsc), and IEEE international conference on intelligent data and security (ids), 2017,* 167-172. https://doi.org/10.1109/BigDataSecurity.2017.50

Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies, 5*(1), 1-167.

Liu, S., Law, R., Rong, J., Li, G., & Hall, J. (2013). Analyzing changes in hotel customers' expectations by trip mode. *International Journal of Hospitality Management, 34,* 359-371. https://doi.org/10.1016/j.ijhm.2012.11.011

Liu, S., & Yao, E. (2017). Holiday Passenger Flow Forecasting Based on the Modified Least-Square Support Vector Machine for the Metro System. *Transportation Engineering, Part A: Systems, 143,* 04016005. https://doi.org/10.1061/JTEPBS.0000010

Lobo, J. L., Del Ser, J., Bifet, A., & Kasabov, N. (2020). Spiking Neural Networks and online learning: An overview and perspectives. Neural Networks, 121, 88-100. https://doi.org/10.1016/j.neunet.2019.09.004

Lu, J. (2017). Machine learning modeling for time series problem: Predicting flight ticket prices.

Lu, Q., Xiao, L., & Ye, Q. (2012). Investigating the impact of online word-of-mouth on hotel sales with panel data. *2012 International Conference on Management Science Engineering 19th Annual Conference Proceedings, 2012* 3-9. https://doi.org/10.1109/ICMSE.2012.6414153

Luberg, A., Granitzer, M., Wu, H., Järv, P., & Tammet, T. (2012). Information Retrieval and Deduplication for Tourism Recommender Sightsplanner. *Proceedings of the 2nd International Conference on Web Intelligence, Mining and Semantics, 50*(1), 50-11. https://doi.org/10.1145/2254129.2254191

Lucas, J. P., Luz, N., Moreno, M. N., Anacleto, R., Almeida Figueiredo, A., & Martins, C. (2013). A hybrid recommendation approach for a tourism system. *Expert Systems with Applications, 40*(9), 3532-3550. https://doi.org/10.1016/j.eswa.2012.12.061

Luo, J., Huang, S. (Sam), & Wang, R. (2020). A fine-grained sentiment analysis of online guest reviews of economy hotels in China. *Journal of Hospitality Marketing & Management, 30*(1), 1-25. https://doi.org/10.1080/19368623.2020.1772163

Ma, Y., Xiang, Z., Du, Q., & Fan, W. (2018). Effects of user-provided photos on hotel review helpfulness: An analytical approach with deep leaning. *International Journal of Hospitality Management, 71,* 120-131. https://doi.org/10.1016/j.ijhm.2017.12.008

Mackenzie, J., Roddick, J. F., & Zito, R. (2019). An Evaluation of HTM and LSTM for Short-Term Arterial Traffic Flow Prediction. *IEEE Transactions on Intelligent Transportation Systems, 20*(5), 1847-1857. https://doi.org/10.1109/TITS.2018.2843349

Marrero, C., Jorba, O., Cuevas, E., & Baldasano, J. (2009). Sensitivity study of surface wind flow of a limited area model simulating the extratropical storm Delta affecting the Canary Islands. *Advances in Science and Research, 2*(1), 151-157.

Martinez-Torres, M. R., & Toral, S. L. (2019). A machine learning approach for the identification of the deceptive reviews in the hospitality sector using unique attributes and sentiment orientation. Tourism Management, 75, 393-403. https://doi.org/10.1016/j.tourman.2019.06.003

Martin-Fuentes, E., Fernandez, C., Mateu, C., & Marine-Roig, E. (2018). Modelling a grading scheme for peer-to-peer accommodation: Stars for Airbnb. *International Journal of Hospitality Management, 69*, 75-83. https://doi.org/10.1016/j.ijhm.2017.10.016

McKinney, W. (2010). Data Structures for Statistical Computing in Python. *Proceedings of the 9th Python in Science Conference (SCIPY 2010), 2010,* 56-61. https://doi.org/10.25080/Majora-92bf1922-00a

Miah, S. J., Vu, H. Q., Gammack, J., & McGrath, M. (2017). A Big Data Analytics Method for Tourist Behaviour Analysis. *Information & Management, 54*(6), 771-785. https://doi.org/10.1016/j.im.2016.11.011

Min, H., Min, H., & Emam, A. (2002). A data mining approach to developing the profiles of hotel customers. *International Journal of Contemporary Hospitality Management, 14*(6), 274-285. https://doi.org/10.1108/09596110210436814

Mitchell, T. (1997). Introduction to machine Learning. En *Machine Learning* (pp 2-5).

Mitchell, T. M. (1997). *Machine Learning (1.a ed.).* McGraw-Hill, Inc.

Mitchell, T. M. (1999). Machine learning and data mining. *Communications of the ACM, 42*(11), 30-36. https://doi.org/10.1145/319382.319388

Mohandes, M. A., Halawani, T. O., Rehman, S., & Hussain, A. A. (2004). Support vector machines for wind speed prediction. *Renewable energy, 29*(6), 939-947.

Montiel, J., Read, J., Bifet, A., & Abdessalem, T. (2018). Scikit-Multiflow: A Multi-output Streaming Framework. *Journal of Machine Learning Research, 19*(72), 1-5.

Murugan, S., Chinnadurai, M., & Manikandan, S. (2021). Tour Planning Design for Mobile Robots Using Pruned Adaptive Resonance Theory Networks. *Computers, Materials & Continua, 70(*1), 181. https://doi.org/10.32604/cmc.2022.016152

Nakamura, S., Okada, M., & Hashimoto, K. (2015). An Investigation of Effectiveness Using Topic Information Order to Classify Tourists Reviews. *2015 International Conference on Computer Application Technologies, 2015,* 94-97. https://doi.org/10.1109/CCATS.2015.32

Namahoot, C. S., Brückner, M., & Panawong, N. (2015). Context-Aware Tourism Recommender System Using Temporal Ontology and Naïve Bayes. En *Recent Advances in Information and Communication Technology 2015* (pp. 183-194). Springer, Cham. https://doi.org/10.1007/978-3-319-19024-2_19

Nilashi, M., Ahani, A., Esfahani, M. D., Yadegaridehkordi, E., Samad, S., Ibrahim, O., Sharef, N. M., & Akbari, E. (2019). Preference learning for eco-friendly hotels recommendation: A multi-criteria collaborative filtering approach. *Journal of Cleaner Production, 215*, 767-783. https://doi.org/10.1016/j.jclepro.2019.01.012

Nilashi, M., Bagherifard, K., Rahmani, M., & Rafe, V. (2017). A recommender system for tourism industry using cluster ensemble and prediction machine learning techniques. *Computers & Industrial Engineering, 109,* 357-368. https://doi.org/10.1016/j.cie.2017.05.016

Nilashi, M., Yadegaridehkordi, E., Ibrahim, O., Samad, S., Ahani, A., & Sanzogni, L. (2019). Analysis of Travellers' Online Reviews in Social Networking Sites Using Fuzzy Logic Approach. *International Journal of Fuzzy Systems, 21*(5), 1367-1378. https://doi.org/10.1007/s40815-019-00630-0

Noersasongko, E., Julfia, F. T., Syukur, A., Purwanto, Pramunendar, R. A., & Supriyanto, C. (2016). A Tourism Arrival Forecasting using Genetic Algorithm based Neural Network. *Indian Journal of Science and Technology, 9*(4). https://doi.org/10.17485/ijst/2016/v9i4/78722

Oger Vihikan, W., Ketut Gede Darma Putra, I., & Putu Arya Dharmaadi, I. (2017). Foreign Tourist Arrivals Forecasting Using Recurrent Neural Network Backpropagation through Time. *Telkomnika (Telecommunication Computing Electronics and Control), 15,* 1257-1264. https://doi.org/10.12928/TELKOMNIKA.v15i3.5993

Okumus, F. (2013). Facilitating knowledge management through information technology in hospitality organizations. *Journal of Hospitality and Tourism Technology, 4*(1), 64-80. https://doi.org/10.1108/17579881311302356

Omran, M. G. H., Engelbrecht, A. P., & Salman, A. (2007). An overview of clustering methods. *Intelligent Data Analysis, 11*(6), 583-605.

Opitz, D., & Maclin, R. (1999). Popular ensemble methods: An empirical study. *Journal of Artificial Intelligence Research, 11*(1), 169-198.

Orr, M. J. L. (1996). *Introduction to Radial Basis Function Networks.*

Page, S. (2007). *Tourism Management: Managing for Change.* Routledge.

Pai, P.-F., Hong, W.-C., & Lin, C.-S. (2005). Forecasting Tourism Demand Using a Multifactor Support Vector Machine Model. *Computational Intelligence and Security*, 512-519. https://doi.org/10.1007/11596448_75

Pai, P.-F., Hung, K.-C., & Lin, K.-P. (2014). Tourism demand forecasting using novel hybrid system. *Expert Systems with Applications, 41*(8), 3691-3702. https://doi.org/10.1016/j.eswa.2013.12.007

Pan, S. J., & Yang, Q. (2010). A Survey on Transfer Learning. *IEEE Transactions on Knowledge and Data Engineering, 22*(10), 1345-1359. https://doi.org/10.1109/TKDE.2009.191

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research, 12*(85), 2825-2830.

Pedrycz, W., & Peters, J. F. (1998). *Computational Intelligence in Software Engineering (Vol. 16).* WORLD SCIENTIFIC.

Peng, X., & Huang, Z. (2017). A Novel Popular Tourist Attraction Discovering Approach Based on Geo-Tagged Social Media Big Data. *ISPRS International Journal of Geo-Information, 6*(7), 216. https://doi.org/10.3390/ijgi6070216

Phillips, P., Zigan, K., Santos Silva, M. M., & Schegg, R. (2015). The interactive effects of online reviews on the determinants of Swiss hotel performance: A neural network analysis. *Tourism Management, 50,* 130-141. https://doi.org/10.1016/j.tourman.2015.01.028

Phillips-Wren, G., & Hoskisson, A. (2014). Decision Support with Big Data: A Case Study in the Hospitality Industry. *Frontiers in Artificial Intelligence and Applications, 261*(DSS 2.0), 401-413. https://doi.org/10.3233/978-1-61499-399-5-401

Pitman, A., Zanker, M., Fuchs, M., & Lexhagen, M. (2010). Web Usage Mining in Tourism—A Query Term Analysis and Clustering Approach. *Information and Communication Technologies in Tourism 2010, 2010,* 393-403. https://doi.org/10.1007/978-3-211-99407-8_33

Porto, A., & Irigoyen, E. (2017). Gas Consumption Prediction Based on Artificial Neural Networks for Residential Sectors. *International Joint Conference SOCO'17-CISIS'17-ICEUTE'17 León, Spain, September 6–8, 2017, Proceeding, 2017,* 102-111. https://doi.org/10.1007/978-3-319-67180-2_10

Qin, Z., Li, W., & Xiong, X. (2011). Estimating wind speed probability distribution using kernel density method. *Electric Power Systems Research, 81*(12), 2139-2146.

Rafidah, A., & Ani, S. (2017). Modelling Singapore Tourist Arrivals to Malaysia by Using SVM and ANN. *SCIREA Journal of Mathematics, 1*(2), 210-216.

Regev, O. (2010). The Learning with Errors Problem (Invited survey). *2010 IEEE 25th Annual Conference on Computational Complexity, 2010,* 191-204.

Ren, G., & Hong, T. (2017). Investigating Online Destination Images Using a Topic-Based Sentiment Analysis Approach. *Sustainability; Basel, 9*(10). http://dx.doi.org/10.3390/su9101765

Ren, M., Vu, H. Q., Li, G., & Law, R. (2020). Large-scale comparative analyses of hotel photo content posted by managers and customers to review platforms based on deep learning: Implications for hospitality marketers. *Journal of Hospitality Marketing & Management, 30*(1), 1-24. https://doi.org/10.1080/19368623.2020.1765226

Rivest, R. L., & Dertouzos, M. (1978). *ON DATA BANKS AND PRIVACY HOMOMORPHISMS.*

Romero-Morales, D., & Wang, J. (2010). Forecasting cancellation rates for services booking revenue management using data mining. *European Journal of Operational Research, 202*(2), 554-562. https://doi.org/10.1016/j.ejor.2009.06.006

Sakhuja, S., Jain, V., Kumar, S., Chandra, C., & Ghildayal, S. K. (2016). Genetic algorithm based fuzzy time series tourism demand forecast model. *Industrial Management & Data Systems, 116*(3), 483-507. https://doi.org/10.1108/IMDS-05-2015-0165

Salcedo-Sanz, S., Pastor-Sánchez, A., Del Ser, J., Prieto, L., & Geem, Z.-W. (2015). A coral reefs optimization algorithm with harmony search operators for accurate wind speed prediction. *Renewable Energy, 75*, 93-101.

Salcedo-Sanz, S., Pérez-Bellido, Á. M., Ortiz-García, E. G., Portilla-Figueras, A., Prieto, L., & Paredes, D. (2009). Hybridizing the fifth generation mesoscale model with artificial neural networks for short-term wind speed prediction. *Renewable Energy, 34*(6), 1451-1457.

Sanchez-Franco, M. J., Cepeda-Carrion, G., & Roldan, J. L. (2019). Understanding relationship quality in hospitality services A study based on text analytics and partial least squares. *Internet Research, 29*(3), 478-503. https://doi.org/10.1108/IntR-12-2017-0531

Sanchez-Franco, M. J., Navarro-Garcia, A., & Javier Rondan-Cataluna, F. (2019). A naive Bayes strategy for classifying customer satisfaction: A study based on online reviews of hospitality services. *Journal of Business Research, 101,* 499-506. https://doi.org/10.1016/j.jbusres.2018.12.051

Sánchez-Medina, J. J., Guerra-Montenegro, J. A., Sánchez-Rodríguez, D., Alonso-González, I. G., & Navarro-Mesa, J. L. (2019). Data Stream Mining Applied to Maximum Wind Forecasting in the Canary Islands. *Sensors, 19*(10), 2388. https://doi.org/10.3390/s19102388

Saputro, K. E., Kusumawardani, S. S., & Fauziati, S. (2016). Development of Semi-Supervised Named Entity Recognition to Discover New Tourism Places. *2016 2nd International Conference on Science and Technology-Computer (ICST), 2016*, 124-128. https://doi.org10.1109/ICSTC.2016.7877360.

Schlimmer, J. C., & Granger, R. H. (1986). Incremental learning from noisy data. *Machine learning, 1*(3), 317-354.

Scott, D., & Lemieux, C. (2010). Weather and Climate Information for Tourism. P*rocedia Environmental Sciences, 1,* 146-183. https://doi.org/10.1016/j.proenv.2010.09.011

Seco, A., González, P., Ramírez, F., García, R., Prieto, E., Yagüe, C., & Fernández, J. (2009). GPS monitoring of the tropical storm delta along the Canary Islands track, November 28-29, 2005. *Pure and applied geophysics, 166*(8), 1519-1531.

Shahrabi, J., Hadavandi, E., & Asadi, S. (2013). Developing a hybrid intelligent model for forecasting problems: Case study of tourism demand time series. *Knowledge-Based Systems, 43,* 112-122. https://doi.org/10.1016/j.knosys.2013.01.014

Sharma, A., & Dey, S. (2012). A Document-level Sentiment Analysis Approach Using Artificial Neural Network and Sentiment Lexicons. *ACM SIGAPP Applied Computing Review, 12*(4), 67-75. https://doi.org/10.1145/2432546.2432552

Shi, H. X., & Li, X. J. (2011). A sentiment analysis model for hotel reviews based on supervised learning. *2011 International Conference on Machine Learning and Cybernetics, 3,* 950-954. https://doi.org/10.1109/ICMLC.2011.6016866

Shi, Z., Liang, H., & Dinavahi, V. (2018). Direct Interval Forecast of Uncertain Wind Power Based on Recurrent Neural Networks. *IEEE Transactions on Sustainable Energy, 9*(3), 1177-1187. https://doi.org/10.1109/TSTE.2017.2774195

Shimada, K., Inoue, S., Maeda, H., & Endo, T. (2011). Analyzing Tourism Information on Twitter for a Local City. *2011 First ACIS International Symposium on Software and Network Engineering, 2011*, 61-66. https://doi.org/10.1109/SSNE.2011.27

Shi-Ting, L., & Bo, X. (2014). The Application of Improved SVM for Data Analysis in Tourism Economy. *2014 7th International Conference on Intelligent Computation Technology and Automation, 2014*, 769-772. https://doi.org/10.1109/ICICTA.2014.186

Shobha, K., & Savarimuthu, N. (2021). Clustering based imputation algorithm using unsupervised neural network for enhancing the quality of healthcare data. *Journal of Ambient Intelligence and Humanized Computing, 12*(2), 1771-1781. https://doi.org/10.1007/s12652-020-02250-1

Shoukry, A., & Aldeek, F. (2020). Attributes prediction from IoT consumer reviews in the hotel sectors using conventional neural network: Deep learning techniques. *Electronic Commerce Research, 20*(2), 223-240. https://doi.org/10.1007/s10660-019-09373-4

Siddique, N. H. (2013). *Computational intelligence synergies of fuzzy logic, neural networks and evolutionary computing / Nazmul Siddique, Hojjat Adeli. (1st edition).* John Wiley & Sons Inc.

Silverman, B. W., & Jones, M. C. (1989). E. Fix and J.L. Hodges (1951): An Important Contribution to Nonparametric Discriminant Analysis and Density Estimation: Commentary on Fix and Hodges (1951). *International Statistical Review / Revue Internationale de Statistique, 57*(3), 233-238. https://doi.org/10.2307/1403796

Sixto, J., Almeida, A., & López-de-Ipiña, D. (2013). Analysing Customers Sentiments: An Approach to Opinion Mining and Classification of Online Hotel Reviews. *Natural Language Processing and Information Systems,* 359-362. https://doi.org/10.1007/978-3-642-38824-8_38

Sun, J., & Chang, T. (2016). Prediction of rural residents' tourism demand based on back propagation neural network. *International Journal of Applied Decision Sciences, 9*(3), 320-331. https://doi.org/10.1504/IJADS.2016.081095

Sun, S., Wang, S., Wei, Y., Yang, X., & Tsui, K. L. (2017). Forecasting tourist arrivals with machine learning and internet search index. *2017 IEEE International Conference on Big Data (Big Data), 2017*, 4165-4169. https://doi.org/10.1109/BigData.2017.8258439

Sun, Y., Fan, H., Bakillah, M., & Zipf, A. (2015). Road-based travel recommendation using geo-tagged images. *Computers, Environment and Urban Systems, 53,* 110-122. https://doi.org/10.1016/j.compenvurbsys.2013.07.006

Taecharungroj, V., & Mathayomchan, B. (2019). Analysing TripAdvisor reviews of tourist attractions in Phuket, Thailand. *Tourism Management, 75*, 550-568. https://doi.org/10.1016/j.tourman.2019.06.020

Tang, J., Liu, F., Zou, Y., Zhang, W., & Wang, Y. (2017). An Improved Fuzzy Neural Network for Traffic Speed Prediction Considering Periodic Characteristic. *IEEE Transactions on Intelligent Transportation Systems, 18*(9), 2340-2350. https://doi.org/10.1109/TITS.2016.2643005

Tokuhisa, M., Shahana, H., Murata, M., & Murakami, J. (2012). An Active Learning Based Support Tool for Extracting Hints of Tourism Development from Blog Articles. *2012 IIAI International Conference on Advanced Applied Informatics, 2012,* 103-107. https://doi.org/10.1109/IIAI-AAI.2012.29

Torres, J. L., Garcia, A., De Blas, M., & De Francisco, A. (2005). Forecast of hourly average wind speed with ARMA models in Navarre (Spain). *Solar energy, 79*(1), 65-77.

van de Schoot, R., Kaplan, D., Denissen, J., Asendorpf, J. B., Neyer, F. J., & van Aken, M. A. G. (2014). A Gentle Introduction to Bayesian Analysis: Applications to Developmental Research. *Child Development, 85*(3), 842-860. https://doi.org/10.1111/cdev.12169

Versichele, M., de Groote, L., Claeys Bouuaert, M., Neutens, T., Moerman, I., & Van de Weghe, N. (2014). Pattern mining in tourist attraction visits through association rule

learning on Bluetooth tracking data: A case study of Ghent, Belgium. *Tourism Management, 44*, 67-81. https://doi.org/10.1016/j.tourman.2014.02.009

Vu, H. Q., Li, G., Law, R., & Ye, B. H. (2015). Exploring the travel behaviors of inbound tourists to Hong Kong using geotagged photos. *Tourism Management, 46,* 222-232. https://doi.org/10.1016/j.tourman.2014.07.003

Wang, G., Su, Y., & Shu, L. (2016). One-day-ahead daily power forecasting of photovoltaic systems based on partial functional linear regression models. *Renewable Energy, 96*, 469-478. https://doi.org/10.1016/j.renene.2016.04.089

Wang, J., Jin, S., Qin, S., & Jiang, H. (2014). Swarm Intelligence-Based Hybrid Models for Short-Term Power Load Prediction. *Mathematical Problems in Engineering, 2014.* https://doi.org/10.1155/2014/712417

Wang, S.-C. (2003). Artificial Neural Network. En S.-C. Wang (Ed.), *Interdisciplinary Computing in Java Programming* (pp. 81-100). Springer US.

Wang, X., Zhang, H., & Guo, X. (2015). Demand Forecasting Models of Tourism Based on ELM. *2015 Seventh International Conference on Measuring Technology and Mechatronics Automation, 2015*, 326-330. https://doi.org/10.1109/ICMTMA.2015.84

Wang, Y. (2012). On Abstract Intelligence: Toward a Unifying Theory of Natural, Artificial, Machinable, and Computational Intelligence. En *Software and Intelligent Sciences: New Transdisciplinary Findings* (pp. 18). IGI Global.

Wang, Y., Chan, S. C. F., & Ngai, G. (2012). Applicability of Demographic Recommender System to Tourist Attractions: A Case Study on Trip Advisor. *2012 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology, 3*, 97-101. https://doi.org/10.1109/WI-IAT.2012.133

Wang, Z., Wang, W., Liu, C., Wang, Z., & Hou, Y. (2018). Probabilistic Forecast for Multiple Wind Farms Based on Regular Vine Copulas. *IEEE Transactions on Power Systems, 33*(1), 578-589. https://doi.org/10.1109/TPWRS.2017.2690297

Weichselbraun, A., Gindl, S., & Scharl, A. (2010). A Context-Dependent Supervised Learning Approach to Sentiment Detection in Large Textual Databases. *Journal of Information and Data Management, 1*(3), 329.

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis.* Springer.

Widmer, G., & Kubat, M. (1996). Learning in the presence of concept drift and hidden contexts. *Machine Learning, 23*(1), 69-101. https://doi.org/10.1007/BF00116900

Williamson, J. R. (1996). Gaussian ARTMAP: A Neural Network for Fast Incremental Learning of Noisy Multidimensional Maps. Neural Networks: *The Official Journal of the International Neural Network Society, 9(*5), 881-897. https://doi.org/10.1016/0893-6080(95)00115-8

Wu, Q., Law, R., & Xu, X. (2012). A sparse Gaussian process regression model for tourism demand forecasting in Hong Kong. *Expert Systems with Applications, 39*(5), 4769-4774. https://doi.org/10.1016/j.eswa.2011.09.159

Xia, H., & Peng, L. (2009). SVM-Based Comments Classification and Mining of Virtual Community: For Case of Sentiment Classification of Hotel Reviews. *Proceedings of the International Symposium on Intelligent Information Systems and Applications (ISA'09), 2009,* 507-511.

Xiang, Z., Du, Q., Ma, Y., & Fan, W. (2017). A comparative analysis of major online review platforms: Implications for social media analytics in hospitality and tourism. *Tourism Management, 58,* 51-65. https://doi.org/10.1016/j.tourman.2016.10.001

Xiao, F. (2012). Forest Fire Disaster Area Prediction Based on Genetic Algorithm and Support Vector Machine. *Advanced Materials Research, 446-449,* 3037-3041. https://doi.org/10.4028/www.scientific.net/AMR.446-449.3037

Xu, D., Peng, P., Wei, C., He, D., & Xuan, Q. (2020). Road traffic network state prediction based on a generative adversarial network. *IET Intelligent Transport Systems, 14.* https://doi.org/10.1049/iet-its.2019.0552

Xu, Q., & Zhao, H. (2012). Using Deep Linguistic Features for Finding Deceptive Opinion Spam. *Proceedings of COLING 2012, 2012,* 1341-1350.

Xu, X., Law, R., Chen, W., & Tang, L. (2016). Forecasting tourism demand by extracting fuzzy Takagi–Sugeno rules from trained SVMs. *CAAI Transactions on Intelligence Technology, 1*(1), 30-42. https://doi.org/10.1016/j.trit.2016.03.004

Xu, X., Law, R., & Wu, T. (2009). Support Vector Machines with Manifold Learning and Probabilistic Space Projection for Tourist Expenditure Analysis. *International Journal of Computational Intelligence Systems, 2*(1), 17-26. https://doi.org/10.1080/18756891.2009.9727636

Xue-Bo, & Shi-Ting, L. (2014). Management of Tourism Resources and Demand Based on Neural Networks. *2014 7th International Conference on Intelligent Computation Technology and Automation, 2014,* 348-351. https://doi.org/10.1109/ICICTA.2014.91

Yang, L., Shen, Q., & Li, Z. (2016). Comparing travel mode and trip chain choices between holidays and weekdays. *Transportation Research Part A: Policy and Practice, 91,* 273-285. https://doi.org/10.1016/j.tra.2016.07.001

Yang, P., Takahashi, H., Murase, M., & Itoh, M. (2021). Zebrafish behavior feature recognition using three-dimensional tracking and machine learning. *Scientific Reports, 11*(1), 13492. https://doi.org/10.1038/s41598-021-92854-0

Yang, Y., Mueller, N. J., & Croes, R. R. (2016). Market accessibility and hotel prices in the Caribbean: The moderating effect of quality-signaling factors. *Tourism Management, 56,* 40-51. https://doi.org/10.1016/j.tourman.2016.03.021

Yang, Y., Tang, J., Luo, H., & Law, R. (2015). Hotel location evaluation: A combination of machine learning tools and web GIS. *International Journal of Hospitality Management, 47,* 14-24. https://doi.org/10.1016/j.ijhm.2015.02.008

Yao, J., Wang, H., & Yin, P. (2011). Sentiment Feature Identification from Chinese Online Reviews. En *Advances in Information Technology and Education* (pp. 315-322). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-22418-8_44

Yariyan, P., Omidvar, E., Minaei, F., Ali Abbaspour, R., & Tiefenbacher, J. P. (2022). An optimization on machine learning algorithms for mapping snow avalanche susceptibility. *Natural Hazards, 111*(1), 79-114. https://doi.org/10.1007/s11069-021-05045-5

Ye, L., & Yamamoto, T. (2018). Modeling connected and autonomous vehicles in heterogeneous traffic flow. *Physica A: Statistical Mechanics and Its Applications, 490,* 269-277. https://doi.org/10.1016/j.physa.2017.08.015

Ye, Q., Zhang, Z., & Law, R. (2009). Sentiment classification of online reviews to travel destinations by supervised machine learning approaches. *Expert Systems with Applications, 36*(3, Part 2), 6527-6535. https://doi.org/10.1016/j.eswa.2008.07.035

Yi-Chung, H. (2017). Predicting Foreign Tourists for the Tourism Industry Using Soft Computing-Based Grey-Markov Models. *Sustainability; Basel, 9*(7). http://dx.doi.org.bibproxy.ulpgc.es/10.3390/su9071228

Yordanova, S., & Kabakchieva, D. (2017). Sentiment Classification of Hotel Reviews in Social Media with Decision Tree Learning. *International Journal of Computer Applications, 158*. https://doi.org/10.5120/ijca2017912806

Yu, H., Wu, Z., Wang, S., Wang, Y., & Ma, X. (2017). Spatiotemporal recurrent convolutional networks for traffic prediction in transportation networks. *Sensors, 17*(7), 1501.

Yuan, Y., Du, J., & Lee, J. M. (2016). Tourism activity recognition and discovery based on improved LDA model. *2016 4th International Conference on Cloud Computing and Intelligence Systems (CCIS), 2016,* 447-455. https://doi.org/10.1109/CCIS.2016.7790300

Zadeh, L. A. (1965). Fuzzy sets. *Information and Control, 8*(3), 338-353. https://doi.org/10.1016/S0019-9958(65)90241-X

Zhang, B., Huang, X., Li, N., & Law, R. (2017). A novel hybrid model for tourist volume forecasting incorporating search engine data. *Asia Pacific Journal of Tourism Research, 22*(3), 245-254. https://doi.org/10.1080/10941665.2016.1232742

Zhang, C., & Zhang, J. (2011). Mining Tourist Preferences with Twice-Learning. *New Frontiers in Applied Data Mining,* 483-493. https://doi.org/10.1007/978-3-642-28320-8_41

Zhang, C., & Zhang, J. (2014). Analysing Chinese citizens' intentions of outbound travel: A machine learning approach. *Current Issues in Tourism, 17*(7), 592-609. https://doi.org/10.1080/13683500.2013.768606

Zhang, Q., Yang, L. T., Chen, Z., & Li, P. (2018). A survey on deep learning for big data. *Information Fusion, 42,* 146-157. https://doi.org/10.1016/j.inffus.2017.10.006

Zhang, T. H., Ji, H. W., Hu, Y., Ye, Q., & Lin, Y. (2018). Application of Classification Algorithm of Machine Learning and Buffer Analysis in Torism Regional Planning. ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information *Sciences, 42*(3), 2297-2302. https://doi.org/10.5194/isprs-archives-XLII-3-2297-2018

Zhang, Z., Ye, Q., Zhang, Z., & Li, Y. (2011). Sentiment classification of Internet restaurant reviews written in Cantonese. *Expert Systems with Applications, 38*(6), 7674-7682. https://doi.org/10.1016/j.eswa.2010.12.147

Zhao, Y., Dong, S., & Yang, J. (2015). Effect research of aspects extraction for Chinese hotel reviews based on machine learning method. *International Journal of Smart Home, 9,* 23-34. https://doi.org/10.14257/ijsh.2015.9.3.03

Zheng, J., & Huang, M. (2020). Traffic Flow Forecast Through Time Series Analysis Based on Deep Learning. *IEEE Access, 8,* 82562-82570. https://doi.org/10.1109/ACCESS.2020.2990738

Zheng, W., & Ye, Q. (2009). Sentiment Classification of Chinese Traveler Reviews by Support Vector Machine Algorithm. *2009 Third International Symposium on Intelligent Information Technology Application, 3,* 335-338. https://doi.org/10.1109/IITA.2009.457

Zhou, F., Yang, Q., Zhang, K., Trajcevski, G., Zhong, T., & Khokhar, A. (2020). Reinforced Spatiotemporal Attentive Graph Neural Networks for Traffic Forecasting. *IEEE Internet of Things Journal, 7*(7), 6414-6428. https://doi.org/10.1109/JIOT.2020.2974494

Zhu, K., Zhang, S., Li, J., Zhou, D., Dai, H., & Hu, Z. (2022). Spatiotemporal multi-graph convolutional networks with synthetic data for traffic volume forecasting. *Expert Systems with Applications, 187*, 115992. https://doi.org/10.1016/j.eswa.2021.115992

Zhu, W.-X., & Zhang, H. M. (2018). Analysis of mixed traffic flow with human-driving and autonomous cars based on car-following model. *Physica A: Statistical Mechanics and Its Applications, 496,* 274-285. https://doi.org/10.1016/j.physa.2017.12.103

Žliobaitė, I., Bifet, A., Pfahringer, B., & Holmes, G. (2014). Active Learning With Drifting Streaming Data. *IEEE Transactions on Neural Networks and Learning Systems, 25*(1), 27-39. https://doi.org/10.1109/TNNLS.2012.2236570