Applied Soft Computing Journal 102 (2021) 107082

Contents lists available at ScienceDirect

Applied Soft Computing Journal

journal homepage: www.elsevier.com/locate/asoc

Review article

Computational Intelligence in the hospitality industry: A systematic literature review and a prospect of challenges

Juan Guerra-Montenegro [a,*], Javier Sanchez-Medina [a], Ibai Laña [c], David Sanchez-Rodriguez [b], Itziar Alonso-Gonzalez [b], Javier Del Ser [c,d]

[a] Innovation Center for the Information Society (CICEI), ULPGC, Spain
[b] Institute for Technological Development and Innovation in Communications (IDeTIC), ULPGC, Spain
[c] TECNALIA, 48160 Derio (Bizkaia), Spain
[d] University of the Basque Country UPV/EHU, 48013 Bilbao, Spain

ARTICLE INFO

Article history:
Received 15 January 2020
Received in revised form 3 December 2020
Accepted 2 January 2021
Available online 6 January 2021

Keywords:
Computational Intelligence
Machine Learning
Travel industry
Hospitality and tourism management
Predictive analysis
Hotel industry

ABSTRACT

This research work presents a detailed survey about *Computational Intelligence* (CI) applied to various Hotel and Travel Industry areas. Currently, the hospitality industry's interest in data science is growing exponentially because of their expected margin of profit growth. In order to provide precise state of the art content, this survey analyzes more than 160 research works from which a detailed categorization and taxonomy have been produced. We have studied the different approaches on the various forecasting methods and subareas where CI is currently being used. This research work also shows an actual distribution of these research efforts in order to enhance the understanding of the reader about this topic and to highlight unexploited research niches. A set of guidelines and recommendations for future research areas and promising applications are also presented in a final section.

Contents

* Corresponding author.
E-mail address: juanantonio.montenegro@ulpgc.es (J. Guerra-Montenegro).

https://doi.org/10.1016/j.asoc.2021.107082
1568-4946/

1. Introduction

The interest on Machine Learning (ML) has been growing exponentially over the last decades [1], especially after the recent increase in the availability of data in every field in Engineering. The benefits of using ML range from exploratory analysis (the understanding of the structure and internal relationships of data), to predictive analysis (modeling the observed processes in order to forecast their future evolution) and prescriptive analysis (generating higher level recommendations to decision makers or managers).

Before proceeding further, in order to provide a commonly accepted definition of Computational Intelligence (CI), definitions previously set in the literature have been examined in depth. Siddique and Adeli [2] stated that CI is a set of methodologies and computational approaches inspired by nature to address complex real-world problems for which traditional or mathematical modeling are not applicable for some reasons. The IEEE Computational Intelligence Society defined that these methodologies comprise Artificial Neural Networks (ANNs), fuzzy systems and evolutionary algorithms. These methods are close to how humans reason by using incomplete knowledge, producing adaptable control actions, thus making CI systems capable of learning from experiential data.

Additionally, it is necessary to define Machine Learning as the use of computer algorithms that automatically improve through experience evolving behaviors based on data [3], and is often considered as part of the Artificial Intelligence field. ML is used in a wide set of research fields, such as speech recognition, robot control or computer Vision [4]. It can be further divided in four main disciplines: supervised learning, semisupervised learning, unsupervised learning and reinforcement learning. The first kind is based on feeding an algorithm with a set of annotated examples (data instances), for it to create a model that is able to produce the annotation for any unseen data instance. On the other hand, an unsupervised learning algorithm uses unlabeled data to create a model that generates a certain output based on an input or a transformation of an input, which is useful in clustering problems or in dimensionality reduction. Semisupervised learning is a mix of the aforementioned two methodologies, where the algorithm has to deal with partially labeled data to generate a model, which usually is a mixture of both supervised and unsupervised learning algorithms. Lastly, reinforcement learning is based on the concept of "training by reward", where the algorithm observes the environment, perform actions on it, and get rewards or penalties in return, creating policies that define the best action to take in a given situation. Additionally, an emerging ML venue called online learning is getting attention in the last decade due to its inherent ability to learn from data streams (instead of datasets or Big Data), giving models the capacity to adapt to changes inside streaming data, making them resilient to the wear of time.

In the recent years, the use of CI methods applied to the hospitality industry has also seen a considerable increase, yielding positive results in diverse areas, such as tourism demand forecasting or energy consumption [1,5]. The use of these techniques not only creates larger profit margins for the industry, but also increases the quality of the offered services to tourists and visitors. Profiling and categorizing users, customizing the services to match each different customer profile is a good example of this. However, using predictive data analytics with Big Data in order to gain knowledge from real-time data instead of historical data is still a new paradigm for the hospitality industry [6].

The goal of this research work is to completely explore the field of hospitality and tourism industry to understand how CI is being applied, surveying what techniques are more frequently used to do specific tasks and proposing a new categorization based on the applications of ML on this area. This was done by classifying the twelve different research areas that were found in four big blocks: Management and Revenue Estimation; Profiling and Recommendation Systems; Tourism Demand Forecasting; and Weather Forecasting and Environmental Risks Assessment. Additionally, this research work aims to identify potential unexploited application niches, sharing with the reader the opportunities for the taking in researching new CI methodologies and applications.

More than 160 relevant articles and research works have been studied and classified to provide an actual state of the art regarding the use of CI applied to the tourism and hospitality industries, yielding 12 different areas where CI was applied. In this research work, we will provide a description of each technique used to extract and parse information and the area of knowledge where those techniques were applied, along with references to the research works that were found inside these categories.

The rest of this research work is organized as follows: In the next section, a description of the research tools and methodologies used in this research is presented. After it, a new section with a description about the more frequent ML families of techniques is introduced. In Section 4, a new categorization of ML techniques for the hospitality and tourism industry is proposed, dividing the previously studied related literature in four big blocks and ten sub-blocks in total. As a result of the examination of the current state of the art, challenges of the field in regard to the CI developments are presented in Section 6. Finally, Section 8 exposes concluding remarks drawn from this study, along with promising research areas.

2. Research methodologies and scope

For the critical literature review discussed in the following sections, we have analyzed 160 papers from different publishers and journals, alongside other research works used to support different claims, to describe different CI methodologies, and to establish diverse hospitality industry concepts and definitions. A thorough search was done in the Core Collection available from the Web of Science, defined by keywords such as "Hospitality Industry", "Tourism", "Data Stream Mining", "Machine Learning", "Computational Intelligence" and/or "Deep Learning". This was combined with a "Search by topic" heuristic, which searches for the aforementioned keywords inside article abstracts, titles and/or tagged keywords, jointly with a time window that spans from 1998 to 2020. A bibliography organizing tool was later used to compile all the references and to provide basic insights into all the reviewed research works, such as how many papers where published in a certain year. After this initial organization, all the aforementioned research works were thoroughly inspected and analyzed to find and classify which CI method is in use, and to ascertain the hospitality area where such method is applied.

After all the research works were analyzed, we discovered that a suitable categorization for CI applied to the hospitality industry was yet to be defined. This research gap allowed presenting useful insights about how research efforts are currently distributed in the hospitality industry, leading to a potential taxonomy to organize all of these efforts. This proposed taxonomy was created by assessing every reviewed research work towards finding the area where it applies a certain CI methodology. After this area analysis, all research efforts inside the hospitality area can be organized in four main blocks, two of them being "Tourism Demand Forecasting" and "Weather Forecasting and Environmental Risks Assessment". The other two blocks were denoted as "Management and Revenue Estimation" and "Profiling and Recommendation Systems" as per how they comprise different parts of the present literature. The first one is composed by six subcategories related to how different hospitality areas are managed, such as booking systems or forecasting of various matters. The latter is composed by four subcategories related to geo-tagging, sentiment analysis, behavior analysis, and recommender systems.

3. Considered computational intelligence areas and techniques

CI is a broad field that has permeated almost any knowledge and research area. There is no clear consensus on how to perform an accurate taxonomy of CI techniques and methods, with an assortment of approaches [7,8], and the focus set on different features. In this section we propose a taxonomy of the most representative families of CI learning techniques from the perspective of the hospitality industry, after surveying their extent across the consulted literature. A major division can be made among the CI methods into three main categories: modeling, optimization and simulation [9]. In this division, each category tries to clear the unknown in an equation where one element of the scheme *input data-model-output* is the unknown. However, the relevant corpus of the hospitality sector CI-related literature revolves around modeling and optimization. Our purpose is to emphasize the techniques that have been found in the studied literature body, and how, in various cases, hybrid methodologies between topics are used. Thus, the coverage of this classification is reduced to the types of techniques and methods that have been found relevant for this sector, and parting from an ontology based on the above mentioned division. Fig. 1 presents a suggested taxonomy of the different CI techniques, based on all the reviewed research works.

In this taxonomy, CI methodologies have been classified attending to their main objectives, which are modeling and optimization. ML modeling refers to the discovery of relations between data inputs and outputs. In the other hand, ML optimization is focused on finding which inputs are maximizing or minimizing the output of a ML model. Lastly, this taxonomy gives place to a classification of the studied research works into one of the categories or subcategories of CI, which is presented later after a second categorization according to the application domain is performed.

3.1. Machine learning

ML constitutes one of the largest categories of CI, including most known applications of this kind of techniques, like classification, forecasting, clustering, regression or pattern discovery. It is frequently divided into two subcategories, depending on the way in which the learning process is produced. If the process consists of finding a function that maps an input to an output, the learning is supervised, and when the process consists of finding patterns of input data without knowing the output, the learning is unsupervised. ML methods are thus aligned to the modeling category, in the division proposed by [9], as they are oriented to finding the model. In the last years, ML has been applied in different areas, demonstrating that it can be a useful tool in order to solve different kinds of problems by classifying data or by predicting situation outcomes. In the hospitality industry, this computational approach yields useful results that allow hospitality establishments to gain competitive advantages over their competitors.

3.1.1. Artificial neural networks

ANNs are inspired by the functionality of the human brain, and are composed of various types of layers which contain neurons. The first one is the input layer, then there might be one, two or more hidden layers and, lastly, a final output layer [10]. They were invented back in the 40 s, and have had popularity peaks and valleys since then. ANNs are defined mimicking the structure of the human brain with neurons and axons and dendrites connecting them. The mathematical transaction consists of a set of neurons arranged in layers, and interconnected in a particular fashion. Each neuron performs an arithmetic operation (typically a summation) among all of the incoming connections, each connection weighted by some value. Each neuron also has an activation function and the output of it is fed to the next layer or the output of the ANN.

Conventional neural networks have proven to be an effective technology for structural pattern recognition [10]. Recently, the available abundance of computing power combined with the use of GPUs instead of CPUs, for their Very Long Instruction Word instructions and the consequent suitability to process large matrices with algebraic operations, is renewing the interest of the scientific community in ANNs, in particular those included in the so called Deep Learning area (Convolutional Neural Networks (CNN), recurrent neural networks, long short term memory, and autoencoders, among others), consisting in general of an increment of the number of hidden layers inside the Neural Network.

Among the limitations for ANNs is their need of a high size of data for training, their generally long training time and a

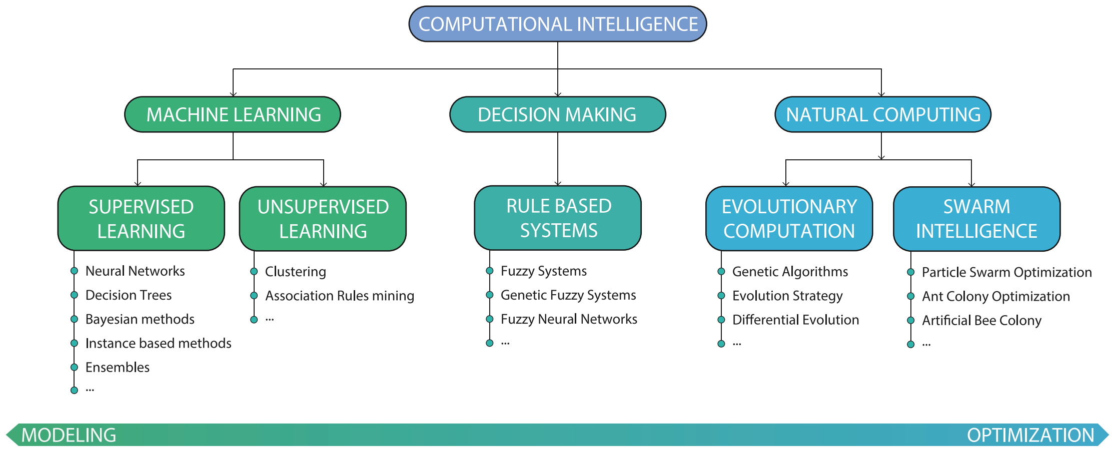

Fig. 1. Taxonomy of Machine Learning, decision making and natural computing methods.

tendency to overfit their models. There is also another concern about the interpretability of the Deep Learning, learned models, which in some contexts may weight severely against their use.

3.1.2. Decision trees

Decision trees are another case of supervised learning, and are a very common technique in sequential decision making [11]. They are fast, they produce intelligible models and can be easily tweaked to operate in an adaptive online learning setup. Essentially, a decision tree is a flowchart-like decision structure composed of nodes and leaves. Each node evaluates a condition of a particular feature of the dataset, and depending on the evaluation of such condition for each observation of the dataset, the evaluation progresses down to another connected node for a subsequent feature evaluation, or to a leaf, where the particular category that is predicted by the model is located. The learning of a decision tree model consists of the training of the node tree structure, the features to be evaluated at each node and the thresholds used to decide a each node the next one. The most interesting feature of this kind of models is its interpretability and that their structure contains, as a bonus value, interesting information on the relative importance of features. Decision trees are most frequently applied to classification problems.

3.1.3. Probabilistic and bayesian methods

According to [12], the Bayesian paradigm "interprets probability as the subjective experience of uncertainty [...] In this paradigm, the classic example of the subjective experience of uncertainty is the notion of placing a bet". Additionally, in [12] is stated that there are three main components of Bayesian statistics: Background knowledge, the information contained in the data, and *posterior inference* which is obtained by combining the first two components.

Bayesian and probabilistic inference are learning methods which, based on the observation of instances of data, update a hypothesis probability incrementally. In that sense, they can be considered in the supervised learning subset of ML, typically applied to classification. Two of the main strengths of Bayesian Networks (the most well known kind in this field), are their ability to incorporate previous knowledge of experts (accelerating the learning process), and that the outputs are not just categories, but also confidence levels. Additionally, one of the most used Bayesian Methods is Naive Bayes, which consists of a conditional probability model which assumes that the value of a particular feature is independent of the value of any other feature, given the class variable.

3.1.4. Instance based learning

Instance based learning algorithms are very similar to edited nearest neighbor algorithms, and they also are a derivation of the nearest neighbor pattern classifier [13]. These algorithms are model learning methods based on learning from the past cases, not trying to generalize models. Instead, the idea is to remember all previous cases and assimilate each new observation to the ones already learned.

Good examples of this family of methods are the k-Nearest Neighbors (KNN) [14], a non-parametric method useful in regression and classification which consists of an input of the 'k' closest training samples in the feature space, and the output is a class membership (classification) or the property value for the object (regression); and Radial Basis Function (RBF) networks [15], which is a special case of a simple ANN that uses RBF as activation functions.

3.1.5. Ensembles

Ensemble techniques are also finding their spot inside the ML panorama. As defined by [16], "an ensemble consists of a set of individually trained classifiers (such as neural networks or decision trees) whose predictions are combined when classifying novel instances". By the combination of models that may work well in some cases, (or be complemented with others if they have a poorer performance), a *better than the parts* model is built. The way of combining them may be as simple as voting, or selecting the more voted category in case of a classification problem.

3.1.6. Clustering

In [17] it is mentioned that "data clustering is the process of identifying natural groupings or clusters within multidimensional data based on some similarity measure (e.g. euclidean distance)", and that it is also an important process in ML and pattern recognition. Clustering uses a broad set of unsupervised ML techniques, where each observation is not associated with a particular dependent variable. Instead, observations are grouped by similarity measurements. Therefore, the knowledge is extracted in terms of how samples are grouped around one or several features, identifying the possible modes in the observation space. The main advantage of clustering techniques is that observations do not need to be labeled (no category has to be associated with each observation) meaning great savings in terms of data preprocessing.

3.1.7. Association rules mining

The objective of association rule mining is to extract meaningful correlations, associations, frequent patterns or even casual structures among different sets of items on data repositories [18]. The knowledge pursued in this kind of technique is the extraction of frequently associated sets of items, for example, when it comes down to items that are very frequently bought together in the supermarket, giving marketing managers opportunities in the sense of advancing what is most likely needed by a particular customer or profile of customer. The extraction of these association rules may prove computationally expensive, but they are quite frequently for maximizing the profit margin of retailers, or other service providers.

3.2. Decision making

Data-driven decision making systems are those oriented to help their users make better decisions. These include fuzzy logic systems, recommender systems, and others that fall within different categories, reason for which this type of systems are placed halfway between modeling and optimization areas. Although the output of almost any CI system can be considered as an assistance to decision making, we have focused here in rule based systems, which cannot be cataloged in any of the other proposed categories. Phillips-Wren and Hoskisson [19] expose in their research that the hospitality industry is becoming more aware of the competitive advantage of using data in decision making provides, though huge quantities of data (Big Data) make this kind of decision making difficult. Additionally, it is also stated that Big Data is getting more important for company leaders because it can be directly tied to value generation.

3.2.1. Fuzzy Systems

In 1965, [20] defined a *fuzzy set* as "a class of objects with a continuum of grades of membership". These sets were inspired by the non discrete reasoning way humans have, where a decision variable may contain vague (fuzzy) values like "low", "medium" and "high". Thanks to this fundamental conceptual change, a tremendous corpus of research has been developed in the last 50 years in many fields, particularly in control engineering and robotics. The flexibility and capability of incorporating human knowledge into this way of modeling makes fuzzy logic a very important family of ML.

3.3. Natural computing

Computation based in natural processes or organisms is widely extended and applied for optimization tasks, as well as for model tuning and adjustment. These methods are used often to explore a wide solution space, which is coded in the shape of members of a population that evolve (evolutionary computation), or interact among them (swarm intelligence), in order to solve the problem in an efficient way. This computing discipline has interesting applications inside the hospitality industry, where it has been used to optimize different areas such as resource allocation (water and energy consumption) or booking management.

3.3.1. Evolutionary computation

John H. Holland [21] set the bases of genetic algorithms, the initial seed of the vast area evolutionary computation is these days. Inspired by C. Darwin's theory of evolution of the species, Holland proposed an abstraction where the species were possible combinatorial solutions to a very complex problem, and through selection, recombination (crossover) and mutation these solutions evolve to adapt to the environment, maximizing a previously defined fitness function.

3.3.2. Swarm intelligence

In 1989, Gerardo Beni defined swarm intelligence as "the collective behavior of decentralized, self-organized systems, natural or artificial". These systems usually consist of a population of simple agents that interact with the environment and with each other. Although these agents follow simple rules with no centralized control schemes, they tend to form an intelligent global behavior. Examples of this kind of systems are bird flocks, herds, ant and bee colonies or even bacterial growth.

3.4. Computational intelligence in the hospitality industry: a literature review

The quantification of techniques used in the hospitality sector is visually depicted in Fig. 2 showing that, inside CI, ML is the most used set of methodologies in the hospitality industry. Inside this area, the most used methodologies appeared to be probabilistic/Bayesian methods and instance/linear methods because these are classic methodologies that have been present in this area for a longer time. However, ANNs are also starting to rise as a frequently used methodology in this research field, specially Deep Learning which is playing a notable role in Big Data solutions because of its capacity to harvest knowledge from complex systems [22].

Evolutionary computation is also being used in the area (in a lesser way, though) to solve different optimization problems like resource allocation and forecasting, which are just a few of the main difficulties inside the hospitality panorama. A good example of this is energy consumption, an area where hospitality establishments frequently struggle to adapt because of their need of efficient energy management methods, necessary to guarantee their performance and sustainability [23].

Lastly, fuzzy and rule-based systems are being used in a lesser degree in hospitality to forecast various kinds of problems, although they tend to be not as versatile as ML algorithms.

4. State of the art categorization

A new categorization of hospitality and tourism industry literature has been carried out in this literature review. This proposed taxonomy is based on the analysis performed in the reviewing process, which uncovered the need for a categorization of the applications of CI to the hospitality industry. Fig. 3 shows the proposed state of the art categorization, which is divided into four big blocks, with 10 subcategories in total. The four main blocks of the new categorization are:

- Management and Revenue Estimation, which comprises everything that has to do with hospitality management, resource allocation, market-related affairs and revenue management;
- Profiling and Recommendation systems, which are composed by everything that has to do with customer/tourist profiling and analysis and with recommendation systems;
- Tourism Demand Forecasting, which is a huge area by itself, and comprises anything that has to do with demand forecasting, from room allocation to seasonal occupation patterns; and finally,
- Weather Forecasting and Environmental Risks Assessment, an area that features research works about weather and climate predictions and potential environmental hazards, all applied to the hospitality and tourism industry.

Additionally, Fig. 4 shows the relation between this new categorization and the state of the art techniques, giving useful insights about methodologies and how much are they applied in a certain research area.

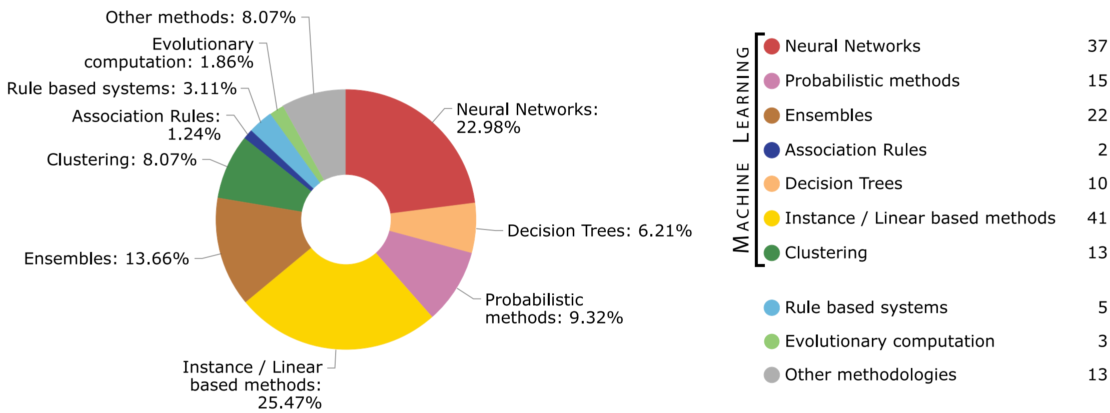

Fig. 2. Reviewed methodologies' distribution.

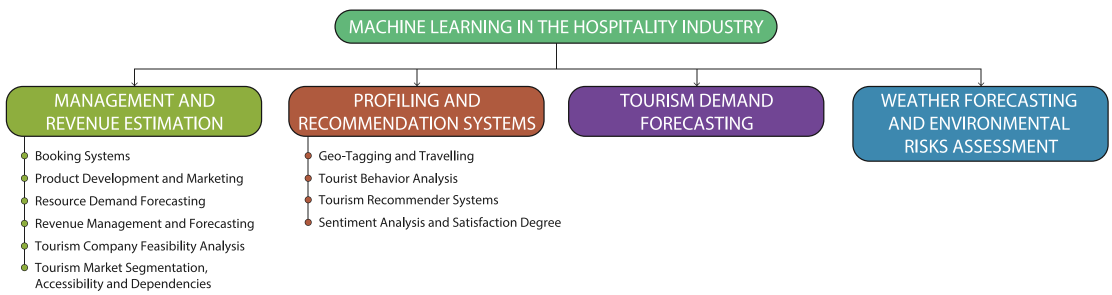

Fig. 3. Proposed state of the art categorization.

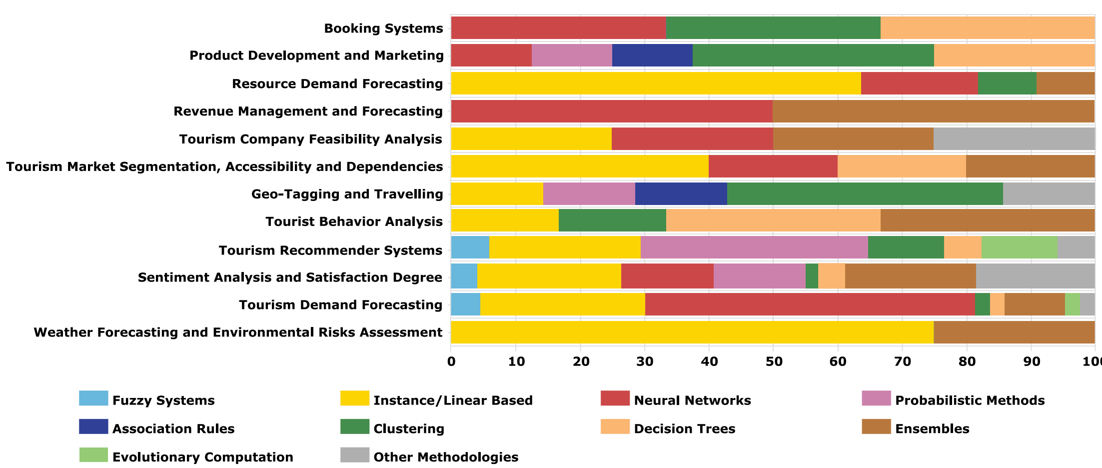

Fig. 4. Distribution of reviewed literature per area and computational intelligence technique.

Next, a summary of each block is described alongside references of the reviewed research works in order to provide a detailed state of the art description in how CI is being applied in the hospitality and tourism industry. Lastly, the reviewed literature is classified according to this taxonomy in Table 1.

4.1. Management and revenue estimation

The hotel industry is gradually starting to exploit the benefits of using ML and data mining methodologies to improve their services, or even to create new ones based on them. In this section we will discuss about several services and infrastructures where they are being applied, namely *booking systems*, easing the procedure of booking into their hotels; *product development and marketing*, where ML methods are applied in order to enhance business decisions; *resource demand forecasting*, where resources like energy must be intelligently allocated; *revenue management and forecasting*, where ML methods are used to predict and manage different aspects of money income, which is essential in a world where any unexpected event may result in a cancellation; *tourism company feasibility analysis*, which is of dire importance in order to start competing in a disputed market and, lastly, *tourism market segmentation, accessibility and dependencies*, where CI can be used to identify various parameters in market distribution and access opportunities.

4.1.1. Booking systems

In these days with such a big global market, it is very important for hospitality businesses to provide potential customers with online booking services, as stated in [26, p. 46], where is also exposed that "the software used to support customers' booking must be also a 'guide' in order to route customers preferences

Table 1
Bibliography review with the considered ML categories, crossed with the hospitality industry main areas.

Paradigm	Family	Techniques	Areas			
			Management and revenue estimation	Profiling and recommendation systems	Tourism demand forecasting	Weather forecasting and environmental risks assessment
Modeling	Supervised learning	Artificial neural networks	[24–38]	[39–50]	[1,32,51–68]	[69]
		Decision trees	[24,29,35,70,71]	[72–81]	[1,61,82,83]	[69]
		Probabilistic/Statistical/Bayesian models	[84]	[1,39,45,74,78,78,85–93, 93–95,95–97,97–112]	[33,58,63,113,114]	
		Ensembles	[35,102]	[77,115,116]	[61,82,117]	[69]
		Instance-based models	[5,24,27–29,35,42,102, 118–123,123–129]	[44,74,75,77,85,89,90,95, 96,98,130–147]	[1,33,51,57,61,65, 67,82,83,117,148–159]	[69,160–162]
	Unsupervised learning	Clustering	[5,71,163–165]	[44,97,166–170]	[117,171]	
		Association rules mining	[164]	[133,172–175]		
Optimization	Decision making	Fuzzy systems	[37]	[176,177]	[117,178]	
	Evolutionary computation	Evolutionary algorithms	[31,121,123,125,126]	[124,176,179]	[59,117,149,150, 152,154,157,171, 178,180]	[162]

since the early phase in which he/she states his/her own preliminary requirements or chooses services, making on-line booking to deal with strategic instruments in order to pursue two relevant aspects of the online market: customer loyalty and customer take over''.

Although not focused on booking systems, [165] described a clustering approach by using k-means to extract business intelligence hidden in web log data by storing it in a database format to process it by using x-means (a variation of k-means clustering), revealing a significant difference between client needs and thus demonstrating that this technique is useful for a precise description of the information needs of clients that use a tourism company's website. A year later, [26] studied the use of ANNs applied to on-line booking systems by using a *multi-layer perceptron* as an inference tool in order to approximate a suitable, fast booking solution for clients who want to book a room inside a hotel. The model they used was trained to automatically solve room allocation depending on how many rooms were unoccupied and how many clients are booking for a room. In this research work is also mentioned that a RBF may have represented a good alternative to their methodology. Lastly, [35] applied different ML methodologies to forecast flight ticket prices, discovering that AdaBoost decision trees yield a satisfactory performance against least squares regression, logistic regression, ANNs, decision trees, random forests and k-nearest neighbors.

In this area, the application of ML methodologies is finding a limitation in terms of commercial time because the time needed to train a multi-layer perceptron ANN is noticeably high. The ML methodology must be fast enough to propose a suitable booking solution for the customer in a short time period in order for it to be effective against a human operator.

There is not much work related to the use ML methodologies in this area, but it proves to be a very interesting research niche where ML forecasting could be applied in order to uncover new services that can be offered to the customer in real time, while the booking process is being done.

4.1.2. Product development and marketing

In the hospitality industry, it is not enough to have a good product, you also have to enhance it over time in order to stay ahead of your competitors. In such a hostile environment, high-return opportunities must be seized in order to take a determinant advantage. According to [71], business managers have to exploit their organizations' generated an collected daily data and transform it into useful information and knowledge in an automatic, intelligent way. With the rapid development of the mobile Internet industry, personal intelligent terminals have been used widely, which makes all kinds of information on the network growing exponentially. [100] exposes that the data mining technology, which can be used to obtain valuable information, is constantly developing. This allows to gain information from the customers in order to improve the services that a hospitality establishment is currently offering, resulting in better tourist reviews and renewed services.

One of the first examples where CI was applied is in [71], where a clustering algorithm was developed using a self-organizing map in order to discover hidden knowledge inside Big Data from a hotel duty-free shop, applying this knowledge to marketing decisions. This was also done by [24] by using different techniques such as ANNs, decision trees and nearest neighbor ML, along with other statistic learning methods.

Also, mining and analyzing customer knowledge might lead to develop new tourism products, as seen in [164], where clustering analysis was applied to generate association rules along the apriori algorithm for data mining information in order to extract useful insights into customer knowledge. In [181], an interesting twice learning approach is taken to mine tourist preferences by using C4.5 and decision trees. CI has also been applied to mine bluetooth tracking data, allowing to obtain tourist attraction patterns using association rules, as seen in [175]. Lastly, in [100], an improved version of a ML technique known as latent dirichlet allocation is used to profile tourism activity, along with the discovery of new trends.

One of the common problems found in this area is the size of the stored data and the performance of the data mining algorithms. It is also usual to have unsorted data, small datasets or, because of data protection laws, not even having any training data at all. Additionally, as seen in [175], using tracking methodologies to gain data has certain limits in terms of user collaboration, meaning that the distribution of logging devices need the collaboration of the tracked individuals. Possible new research lines might include ways to collect data from users in anonymous and innocuous ways in order to create useful datasets that could be used to improve this research area. This could yield more accurate forecasting models that would lead to a better product development and more precise marketing strategies, increasing client satisfaction and revenues.

4.1.3. Resource demand forecasting

Resources like energy and water are of dire importance in the hospitality industry because of their high strategic value. Being able to forecast their demand may result not only in preventing water or electricity shortages, but also in an increase of available resources that are not going to be used, sparing them for a more suitable occasion. ML is currently being used to forecast various kinds of resource demand, namely electricity, water, gas and traffic.

[121] mentions the suitability of SVMs to forecast electricity load, and uses Support Vector Regression (SVR) combined with genetic algorithms in order to tune up SVM parameters to increase forecasting accuracy. [123] also applied SVR to forecast electric load by optimizing the method with chaotic genetic algorithms and adding a seasonal component, creating a cyclic electric load forecasting model that yielded better forecasting results than ARIMA (AutoRegressive Integrated Moving Average) and other SVR models. In [118], is mentioned that SVMs are being widely used in these matters because of the idea of structural risk minimization. In this same research work, SVR is used to analyze various kinds of data inside the tourism economy, such as electric and water consumption, by modeling traffic demand and, additionally, monthly tourist quantity data. Also, in [126] it is said that better energy planning and administration can be provided by an accurate forecast of monthly electricity consumption, using a hybrid SVR + fruit fly algorithm optimization to forecast electricity consumption in a seasonal basis. Power grid evolution in a smart system, or Smart Grid, is also exposed in [139], where the importance of high accuracy is said to be a key factor in a power intelligence program. In this research work, another hybrid forecasting model is presented by combining SVR with the cuckoo search algorithm and singular spectrum analytics, additionally combining these two methods with a Seasonal ARIMA (SARIMA). [31] exposed an Adaptive Neuro-Fuzzy Inference Systems (ANFIS) methodology combined with genetic algorithms that proved to successfully forecast energy needs in the short term. [125] exposed a swarm intelligence approach in forecasting power load by applying this technique on hybrid SARIMA and SVR models optimized by cuckoo search and singular spectrum analytics, which yielded impressive forecasting results. In [127], a partial functional linear regression model for power prediction is used as the main methodology instead of SVR, mainly because of the adequacy of the energy system's previous day intra-day power output as a functional predictor, combined with climatic variables used as covariates. Also, [34] exposes that one of the most important components of information and communication technologies applied to water management is the recent addition of water demand forecasting methods. In their research work, they prove the adequacy of ANNs and ANFIS to forecast water demand. [5] proposed a clustering-based hybrid approach to hourly forecast electricity demand by using fuzzy C-means alongside a hybrid sensor-based model. Additionally, in [128] a hybrid SVR method is used to forecast non-stationary power demand. Lastly, [36] exposes a gas consumption forecasting method based on ANNs for residential sectors capable of forecast the demand of this resource with a 7 days time window.

As a conclusion, the widespread use of SVMs to forecast various kinds of resource demands means its feasibility as a standard method, but there might be more suitable models that have not yet been discovered. An interesting approach for resource forecasting might be the use of Data Stream Mining in order to generate adaptive models that change over time by learning from the new data they are receiving, creating models that can be valid for bigger time periods.

4.1.4. Revenue management and forecasting

In an ever-changing hospitality industry, it is not only important to forecast resources, but also the possible revenues the industry might generate. ML has been found to be a feasible revenue forecasting mechanism, allowing this industry to predict its money income by using various ML methodologies. One of the most common practices in the hospitality industry is revenue management, which is used to help establishments decide on room allocation and rating and as stated in [124], this practice is difficult but essential for creating high-quality revenue budgets.

In [27], a combined revenue management forecasting method is exposed. Based on a demand forecasting module consisting on various ML methods, namely exponential smoothing, pickup methods, moving average, Holt's methods, linear regression and ANNs; combined with an optimization module and a human decision/expert knowledge module. A hybrid revenue forecasting method was also exposed in [124], based on a combination of various ML methods such as Fuzzy Least Squares (FLS), SVR and genetic algorithms (GA) which creates their effective FLSSVRGA seasonal revenue forecasting method. Bugarsky et al. [43] used ANNs combined with RBF and scaled conjugated gradient to create a decision support system for classification of hotel guests by their additional spending in different hotel services. Lastly, Shehhi and Karathanasopoulos [38] compared various models, namely univariate SARIMA, ANFIS, Deep Learning restricted Boltzmann machines and polynomial smooth SVMs to forecast room prices inside hospitality establishments in the Gulf Cooperation Council countries, showing that ANFIS-based models yield a superior performance, closely followed by Deep Learning methods.

Revenue forecasting might be tricky because of its values' variance, making it difficult to predict incomes generated on a demand basis. A promising research line might be the use of more ML methodologies combined with tourism demand forecasting methods to generate hybrid forecasting systems capable not only of predicting revenues based on room occupation, but also on a more precise season-based approach defined by forecasting tourism demand.

4.1.5. Tourism company feasibility analysis

ML and data mining have improved forecasting in many ways, but one of the most interesting ones is forecasting the possible success or bankruptcy of a company. This can be done by studying already known failure or success cases and comparing them to an already running company. [129, p. 622] exposes that "the development of firm failure prediction models for the tourism industry benefits managers, customers, investors, and government officials by reducing loss among hotel-related businesses", and in [28] is also stated that the use of these mechanisms as early warning systems or in aiding decision makers is useful to predict bankruptcy.

Different CI methodologies have been used in this field. In [28], various methodologies for hotel bankruptcy prediction were compared in terms of overall classification, prediction accuracy and relative error cost ratios by comparing the functional characteristics of ANNs, logistic, multivariate discriminant analysis and SVM models. Also, in [129], a SVM approach is taken in order to correct imbalanced samples on a dataset composed by Chinese hotel business and tourism companies using a minority-samples generating approach based on a random percentage distance to the nearest neighbor, along with a nearest neighbor SVM. [29] stated on an article that the effort, time and cost of managing possible claims can be considerably reduced by proactively forecasting disputes in the initial public–private partnership phase, thus exposing an ensemble of various ML techniques (SVMs, ANNs and C5.0) used to classify dispute propensity in terms of overall performance measures, also exposing that SVM is the best single model technique for this task. Lastly, in [37] ANNs with ANFIS are used to forecast the success of newly launched services in tourism.

[37] states that there is little literature about forecasting the launch of new services and products, even when the interest on this subject is greatly increasing, especially when applied to tourism. This makes this area a nice niche of research where great discoveries might be made by applying CI methodologies not only to avoid bankruptcy on a tourism company, but also to improve their services and their relations with other companies and stakeholders.

4.1.6. Tourism market segmentation, accessibility and dependencies

Investing in any hospitality service might be risky if market needs and segments are not properly identified. In [70] is mentioned that in a market where consumer needs are diverse, it is necessary to homogenize them in terms of attitudes, behaviors and demands to gain competitive advantage. Market accessibility is also an important factor since there is no use in investing in a hospitality establishment if its potential market is already full of competitors. There are also cross-dependencies between different markets, and tourism is not an exception.

In the recent years, the use of ML techniques in this area has grown exponentially. In [70] the C5.0 decision tree classification algorithm was used to extract useful knowledge to identify different market segments for optimal customer management. Another pretty interesting application is found in [42], where is stated that the evaluation and assessment of an hotel's location site is of dire importance to its business prosperity (specially in the long-term), by cause of the impossibility to relocate the establishment and the extensive sunk costs; so a ML approach was developed for hotel location evaluation by using different ML methods, namely projection pursuit regression, ANNs, SVR and boosted regression in combination with a web-based GIS application. Another location planning classification method is also seen in [120], where SVMs are used along GIS tools to monitor the land cover change and land use, performing the ecological evaluation for a determined tourism area. Cross-dependencies between markets are also an important factor for forecasting refinement, and in [58] a multi-layer perceptron ANN is used along a Gaussian process regression model to significantly improve the forecasting accuracy. Lastly, in [119] the relationship between hotel price and market accessibility is investigated by the use of a hedonic pricing framework, based on a multi-level dataset containing various quality-signaling factors, using a three-level mixed-effect linear regression model.

As was stated in [33], one of the biggest pitfalls on this matter might be that there is not enough research about it, being a pretty unexploited research area where big discoveries can be made. Applying other ML methodologies than SVMs and ANNs might yield a surprising effect on forecasting market segmentation, where Gaussian process regression is used to increase the forecasting precision of a multi-layer perceptron. Future research on these lines might prove fruitful for tourism market forecasting matters.

4.2. Profiling and recommendation systems

In the hospitality industry, along with other fields, it is of paramount importance to provide good services, alternatives and installations to the clients in order not only to rise as an hospitality establishment, but also to keep that position against others inside this ever-changing industry. ML techniques have also found a nice niche inside this field because of their variety of applications and their potential in gaining insights from Big Data. In this category we are going to expose different applications of these techniques in various fields, which allows the industry to extract knowledge from vast amounts of data such as customer reviews, behavior or even geo-tagged media.

4.2.1. Geo-Tagging and traveling

Geo-Tagging allows to add location data to different file types, such as photos or social network publications. This information yields a big potential in knowledge discovery, which can be obtained by using data mining techniques and ML approaches. One of the most explored areas in this topic is the use of these techniques to discover popular tourist attractions by mining knowledge from all sorts of geo-tagged media. In [170] data clustering is combined with fast search algorithms and density peaks to discover popular tourist attractions by using geo-tagged social media Big Data. Additionally, in [139] a SVM model is used in combination with a ranking method to rank tourism attractions by applying ML over a photographic dataset.

We can also find a pretty useful application of CI to tourism in [84], where a naive Bayesian model is applied over a photograph dataset in order to create a ML model that recommends suitable advertising photos for destination marketing organization. K-Means mining is also used in [163] to gain valuable destination data, which has a high potential in the marketing area.

Finding patterns in travelers' behavior can be also achieved by means of ML. In [166] a Markov chain-based clustering method is exposed by applying this ML technique to a photograph dataset, allowing to mine travel behaviors from tourists by processing geo-tagged photographs. Additionally, in [108] sequential pattern mining is applied to gain insights about tourists' past destinations and potential future ones. Lastly, a distributed sampling association rule Mining method is defined in [172], which is focused in analyzing the holiday traveler's destination traveling behavior.

Classifying these sort of data in order to mine knowledge from it usually requires clustering classification techniques to process the data in a more effective way. As exposed in [166] insights about travel behavior have been sought by tourism managers for a long time, specially for the purposes of product development, destination management and attraction marketing. This is important because spatial clustering is the key to find attractive attractions from geo-tagged data, as explained by [170].

New research lines might be conducted in terms of creating more efficient clustering algorithms, new categorization techniques or even applying already known ML algorithms. As seen in [84] and [139], SVMs and naive Bayes techniques have been successfully used in order to gain knowledge from geo-tagged data. Additionally, the use of data stream mining techniques might increase the valid time of life of the obtained forecasting models because of its inherent adaptability.

4.2.2. Sentiment analysis and satisfaction degree

One of the biggest areas where ML is applied to tourism is sentiment analysis. According to [182], "Sentiment analysis, also called opinion mining, is the field of study that analyzes people's opinions, sentiments, evaluations, appraisals, attitudes, and emotions towards entities such as products, services, organizations, individuals, issues, events, topics, and their attributes". With these techniques it is possible to evaluate the satisfaction degree of customers by mining opinions and comments from Social Networks and review sites. This allows hospitality enterprises to enhance their services according to these reviews.

A plethora of ML techniques have been applied in this area, but the 34 per cent of the analyzed research works on this topic expose the use of SVM algorithms. In [130–132,134,135,138,141,142] and [147] SVMs are used to extract sentiment data from reviews, blogs and various platforms, and in [74,77,85,90,95,133,137,140] and [98] SVMs are used along with N-Gram kernels, various naive Bayes techniques, ANNs, association rules, maximum entropy classifiers, C4.5 decision trees and random forests, JRIP and SVR.

Deep learning methodologies have also been widely used on this area. Ma et al. [110] applied a pre-trained CNN combined with natural language processing and a recurrent neural network in order to discover how helpful are user-provided images in hotel reviews. Shoukry and Aldeek [50] analyzed the role of the Internet of Things in the increase of customer satisfaction inside the hospitality industry by also comparing a CNN with a SVM network-based Deep Learning model and an artificial neural network, resulting in the CNN having best performance than the other two modeling counterparts. Ren et al. [49] also applied Deep Learning methodologies to grasp a comprehensive understanding of the preconceptions reflected in hotel reviews by analyzing images published by clients. Cheng et al. [46] trained a deep CNN model with AirBnB reviews to predict potential guests' trust perception over a hospitality establishment. Chang et al. [47] analyzed hotel reviews and responses by using visual analytics, computational linguistics and Deep Learning to detect proactive hotel responses, using a CNN-based multi-feature fusion system. Lastly, Luo et al. [48] applied Deep Learning to model the experiences of Chinese economy hotel clients, using a bidirectional long short-term memory model combined with a conditional random field model.

Other methodologies used in this area also yield positive results, as seen with naive Bayes and Bayesian networks in [39,78,86,87,92,103] and [112]; natural language processing has also been used to extract insights into sentiment classification, as showcased in [93,107] and [109]. In [40] and [41] we can see that ANNs have been used as well in order to mine and classify opinions from customers in different environments. Other ML methodologies that have also been applied to sentiment analysis are Latent Dirichlet Allocation [101,104,183] and [111], C4.5 and classification and regression trees [76] and [80], locally weighted linear regression [146], N-Gram classification [136], contrast targeted association rule mining [173], fuzzy logic algorithms [177] and [81], gradient boosting [116], multivariate regression [144] and natural language processing [106]. A research work performed by [45] also showcases how clustering can be applied to Sentiment Analysis by using ANFIS, combined with a dimensionality reduction approach to simplify the algorithm's offline training time. Lastly, an interesting research work by [184] expose that is possible to detect fake reviews by applying an ensemble of various methods, such as KNN, logistic regression, SVMs, random forests, gradient boosting and multi-layer perceptron.

Although having a huge quantity of methods to choose from, SVMs usually perform better than any of the other methodologies when text-mining sentiment from reviews and opinions. However, Deep Learning methodologies are slowly starting to be applied in this field because of their speed, compared to other methods. Given the huge size of the datasets that are used in these areas, there is still some niche for optimization in this regard, along the application of new ensemble methods and techniques.

4.2.3. Tourist behavior analysis

Predicting customer behavior is a pretty interesting topic in terms of business. Although it seems that in the hospitality industry this is somehow not a trend, tourist behavior forecasting is starting to grow as a research area. Applying effective forecasting techniques, it is possible to predict where and what a customer is going to do, potentially increasing savings in many of the areas that a hospitality establishment covers.

In 2002, [72] stated that with the increasing competition in the hospitality industry, it is of paramount importance for an hotel's survival to procure services for the changing life styles and preferences of the customers, thus proposing the use of C5.0 decision trees to predict customer loyalty and customers' most valuable services, along with segmenting customer population and defining which segment is best suited for the hotel's services. Also, in [73], C4.5 decision trees are used alongside χ^2 statistical method to find factors that influence tourists' consumption and comprehensive evaluation. SVMs have also been used in this research field, as seen in [148], where is stated that the complexity, nonlinearity and noise of the raw tourism data may create challenges for existing CI techniques, thus proposing a SVM-based classification with two nonlinear feature projection techniques (ISOMAP and probabilistic mapping technique) for tourism data analysis. In 2017, clustering techniques are used by [167] in geotagged Flickr photos uploaded by customers in order to analyze and predict tourist behavioral patterns at specific destinations. In [115] a non-linear regression method which comprises least absolute shrinkage and selection operator and random forests is used in order to determine the total expenditure distribution of cruise tourists in Uruguay. Lastly, [102] created an indoor mapping system by developing a ML radiosity-based model using an ensemble of Bayesian Networks, KNN, multi-layer perceptron, random forests, SVMs and sequential minimal optimization, which was applied to WiFi fingerprinting to reduce the amount of time needed to create a WiFi radio map compared with the previous traditional, manual model.

Tourist behavior analysis is starting to grow as an important research area, but the complexity of the data used to train models, along their possible noise in terms of information, generates a possible research sub-area where data pretreatment is mandatory to get clean datasets that can be used to successfully train ML forecasting models. There is also the need of experimenting with other ML techniques different to decision trees or SVMs, like Deep Learning.

4.2.4. Tourism recommender systems

[176, p. 1] mentions that "applications that deliver multimedia content to and display such content on mobile devices have become increasingly common in recent years". While planning a tourism trip, it is pretty common to accept suggestions from your trip planner to enhance your visit to already known touristic places, or even to discover new ones that may be added to your travel itinerary. These factors make tourism recommender systems a growing research area where CI techniques are being applied to find topics that may be of interest for the trip-planner tourist.

In this area, the most used ML methodologies are Bayesian methods and SVMs. In [88] Bayesian networks are used along an Engel–Blackwell–Miniard model and Google Maps for a tourist attractions intelligent recommendation system. It is also stated in [75] that a recommender system that gathers information from the web might end up having duplicate data on its database, so they designed a SVM approach combined with decision trees to solve this problem. In [89] is stated that the majority of the existing tourism recommender systems use content and knowledge-based approaches, which suffer from the 'cold start' problem and need enough historical rating and extra knowledge data, thus exposing a recommendation system that categorizes the tourists using their demographic information and then makes recommendations based on demographic classes by using naive Bayes, Bayesian networks and SVMs. Additionally, in [91] is exposed that tourism services are highly context-sensitive, this being one of the reasons to develop a tourism attraction recommender system based on context. Authors in [94] also developed a context-aware recommender system based on an improved naive Bayes algorithm. In [143] is exposed that geo-tagged photos on social media reveal the trajectories of tourists and their preferences on landmarks and routings, which allows for the creation of a Road-based travel recommendation system that uses SVMs and binary logistic regression as classifiers, along other filtering methods. In [96] naive Bayes, SVMs and gradient boosting are used as methodologies for developing a feasible classifier for a hotel recommendation system. Authors in [99] developed a support system to recognize tourism places on the web pages based on naive Bayes. SVR is also used in [44], combined with ANFIS to enhance the predictive accuracy of a recommender system along various techniques for clustering. Lastly, in [145] a SVM classification technique was applied to model a grading scheme for peer-to-peer accommodation in order to avoid problems like information asymmetry and overload.

Other ML methodologies applied to tourism recommender systems include genetic algorithms [176] and [179], fuzzy logic and association rule mining [174], data clustering [168,169], latent dirichlet allocation + natural language processing [97] and decision trees + KNN [79].

With the increasing use of recommendation systems for different hospitality services, one of the most interesting sub-areas where research might be useful is in optimizing different parameters, such as dataset sizes and algorithm execution speeds, given that recommendation systems are usually executed in real-time. Additionally, creating adaptive algorithms that can recommend different things based on current trends might also be an interesting research topic.

4.3. Tourism demand forecasting

Along with customer sentiment analysis and satisfaction degree, demand forecasting is one of the biggest research areas in hospitality forecasting. Being able to predict how many customers a hospitality establishment or service is going to get at a certain time is undoubtedly useful to prepare services for a certain load or to coordinate to prevent overbooking or service overloading. With the rise of ML methodologies, forecasting these parameters is somehow easier than before, although it depends on the methodologies used for this purpose.

The use of ANNs is kind of standard, comprising almost a 54% of the used techniques, followed by SVMs which are usually combined with ANNs to improve the forecasting models. In 2005, [51] researched about the feasibility of SVMs combined with Back-Propagation Neural Networks (BPNN) to successfully forecast tourism demand. A few years later, [157] applied a novel approach in this field by combining SVR with genetic algorithms in order to find and apply optimal parameters to construct the SVR model. Additionally, [149] also applied the latter technique demonstrating that a combination of genetic algorithms and SVR, namely GA-SVR, is a feasible technique for tourism demand forecasting. Authors in [25] applied ANFIS to demonstrate its feasibility over other three models, namely fuzzy time series, Gray forecasting model and Markov residual modified model, in forecasting tourist arrivals. In [52], an empirical mode decomposition model combined with BPNNs was used to forecast tourism demand in a more precise way by decomposing raw data and by summing the predictions of both models. In 2013, authors in [30] proposed a hybrid forecasting model for customer cancellations, based on the combination of BPNNs and generalized regression neural networks. In 2014 [53] proposed an ANN forecasting model combined with time series, thus being able to predict tourism demand in a seasonal basis. [32] states that BPNNs suffer from significant drawbacks, so an ensemble of BPNN combined with Bagging was used to overcome these inherent drawbacks. In 2015, [83] compared various ML techniques, namely Seasonal ARIMA, v-SVR and a multi-layer perceptron ANN, to find the most suitable one for tourism demand forecasting, being the second one the best for these kind of predictions. Additionally, [54] used multi-layer perceptron and SVR models to deterministically generate auxiliary variables that outline different time series components and enhance forecasting performance. With this ANN trend in demand forecasting research, [55] compared three different ANN techniques: multi-layer perceptron, RBF and Elman networks, in order to compare their performance, finding that the two first techniques outperform the latter one, and also demonstrating that dimensionality is very important for long-term predictions. In the research work exposed by [56] the researchers used an extreme learning machine to calculate different variables that improve the final forecasting model, even outperforming SVRs. In 2016, [33] exposed the importance of forecast horizon on model selection by comparing performances of SVR with a RBF kernel to ANNs using a linear model as a benchmark. [59] also compared the performance of BPNNs against other ANN techniques, such as KNN and multiple linear regression by using genetic algorithms to optimize BPNN parameters, thus demonstrating that BPNNs have smaller prediction errors in terms of root mean square error. [60] also studied the feasibility of BPNNs to forecast tourism demand by using MATLAB. [57] studied the suitability of SVR, Gaussian process regression and ANNs regional predictions combined to generate more accurate forecasting models, which was specially true when the forecast horizons increase. In 2017, [61] examined various methodologies, namely ANNs, locally deep SVMs, decision jungles, decision trees and boosted decision trees, to accurately forecast hotel booking cancellations. [62] compared the performance of Deep Learning, SVMs and ANNs for tourist number forecasting, finding that Deep Learning outperforms the other two methods in accuracy. [63] also compared the feasibility of Gaussian process regression against ANNs in a multiple-input multiple-output setting, finding that as the models' memory increases the forecasting performance of Gaussian process regression, though ANNs using RBF outperform Gaussian process regression for long-term forecasting. In [64], BPNNs are once again compared against linear regression methods, proving their feasibility as a tourism demand forecasting method and a decision making tool. Additionally, [65] proposed a possible tourism demand forecasting method by means of a combined cross-view model based on BPNN and SVR algorithms. [67] also used ANNs to forecast tourist arrivals by comparing them to SVMs, the latter being outperformed by the first method in terms of root mean absolute error. In [105], a Gray-Markov model is used to predict foreign tourist income by incorporating neural networks in their forecasting model. [68] proposed a kernel-based extreme learning machine to forecast tourist arrivals, being a successful model because its better precision compared to other methods. In [66] BPNNs are used to predict tourist arrivals in Bali. Lastly, in 2018 Kamel et al. [1] investigated the performance of seven different ML methods (multi-layer perceptron, RBF, generalized regression neural networks, KNN, classification and regression trees, SVR and Gaussian process regression), showing that there are differences between these methods, but also that there is no *best* method in the obtained results, which were analyzed by mean absolute percentage error.

SVMs and SVRs have also been used in tourism demand forecasting, although it seems that these methods' forecasting accuracy is often outperformed by ANNs. However, there have been many studies about this topic. In [122] a SVMs are used with kernel logistic regression to successfully forecast cancellation rates, allowing to manage revenue in a more accurate way. [117] presented a useful tourism demand forecasting hybrid model by combining fuzzy C-Means with logarithm least-squares SVR (called LLS-SVR), additionally using genetic algorithms in order to optimally select the parameters of their hybrid model. Additionally, [152] demonstrated that their adaptive genetic algorithms combined with seasonal SVR, outperform normal SVR and BPNNs in tourist flow forecasting. In 2016, [153] used SVMs alongside kernel logistic regression to prove that is possible to generate models to predict booking cancellations with high accuracy. In [82] various regression methods, including SVR, were analyzed to discover that SVR outperforms multiple linear regression and multi-layer perceptron regression models in tourism demand forecasting. In [154] SVR is also hybridized with a seasonal component and optimized by the fruit fly optimization algorithm, yielding positive results that position this model as a feasible tourism forecast solution. [161] extracted fuzzy Takagi–Sugeno rules from trained SVMs to increase the tourism demand forecasting accuracy, and also providing understandable information for decision makers. [158] combined wavelet analysis with SVMs, creating a "WSVM" model that outperforms a normal SVM-based model. In 2017, [159] proposed a modified least square SVM to forecast passenger flow in holidays for the metro system, using an improved particle swarm optimization algorithm is used to optimize the parameters. [155] exposed a seasonal variation SVR method to forecast tourist flows, discovering that SVR is more precise than multivariate linear regression methods. Lastly, [156] developed a hybrid SVR algorithm, combined with the bat algorithm in order to forecast tourist volume by optimizing SVR parameters with it, creating a BA-SVR method that can outperform the normal SVR method.

Other ML methods have also been applied in this area, but in a minor extend. As seen before, genetic algorithms have also been used to optimize different parameters inside other ML methods, yielding more precise results than their not optimized counterparts. [180] and [178] combined genetic algorithms with fuzzy logic techniques. [150] also applied chaotic genetic algorithms to forecast tourism demand and overcome classic problems caused by this method. [171] combined clustering with fuzzy logic techniques along genetic algorithms using symbiotic evolution for fitness assignment. In [113] Gaussian process regression applied to tourism demand is compared against autoregressive moving average and SVM models by adding a sparsity component which reduces the computational complexity and increases its generalization ability. [114] applied ANOVA and stepwise regression in Chinese gaming destinations, generating a model that was useful in finding determinants of hotel occupancy rate. Lastly, as seen in [151], decision trees and their ensembles have also been tested in this forecasting field, the latter ones outperforming their single counterparts in terms of forecasting accuracy.

There is not much room in this area for new research opportunities, but new model ensembles and new hybrid models might be waiting to be discovered. Hybrid techniques have also been proven to be useful in tourism demand forecasting, sometimes outperforming ANNs and SVMs. Additionally, tourism demand is an ever-changing field where tourist flow usually changes over time, so the use of data stream mining techniques might be useful not only in terms of data storage and processing optimization, but also in models that can adapt over time.

4.4. Weather forecasting and environmental risks assessment

Weather prediction, although not directly related with tourism, is one of the most recurrent topics in ML. In fact, possibly because of its indirect relation with tourism, there are not many studies about its possible touristic applications. Being able to forecast weather and certain environmental risks is of paramount importance for customers, and might be a potential game changer against competitors. [160] exposes a method that can successfully detect harmful algal blooms in the coast by using a kernel-based SVM, which is important because, as stated in that research work, forewarnings are provided by these systems for industries like sea food industry, tourism activities and local resource and environmental managers. Another useful approach was taken by [162], in which forest fire disaster areas are forecast by using a SVM with optimized parameters using genetic algorithms. As seen in previous articles, the "genetic algorithm-SVM" approach turns out to be more effective than the classic SVM algorithm.

[69] exposed an ensemble of ML techniques, namely multiple linear regression, classification and regression trees and ANNs for predicting skier days across six ski seasons by using local, regional and national data to create the model.

Additionally, in their research work [185] online learning is applied to forecast maximum wind conditions in the Canary Islands region by using real-time data provided by different weather stations that are distributed in these islands. Climate in Canary islands tend to be clement and stable but extreme weather conditions do occur, influencing tourism and the hospitality industry, thus making this research work useful for the established canary hospitality industry.

These are only a few applications on ML over weather and environmental forecasting, meaning that there is still much room for research in this area by using new ML techniques or applying already known models over the hospitality area.

5. How is computational intelligence currently applied to the hospitality industry?

In order to gauge how CI methods are improving the hospitality industry, it is mandatory to intersect them with the area where they are applied. In Table 1, a detailed literature breakdown is provided in what refers to the kind of CI methodologies that are applied to the previously identified hospitality industry areas.

Although there could be equally valid methodologies for carrying out a work of this nature, this methodology has been used because a thorough sweep of the research area is arguably the most appropriate strategy for retaining and not leaving aside any valuable research contribution. With such huge amounts of publications in the hospitality industry topic, it could be fairly easy to oversee research works that offer valuable insights, thus creating the need for a survey that deeply analyzes the vast majority of the works present in the field.

To realize this study, a deep search to find all research works related to hospitality industry and CI was performed. Combining keywords such as "Hospitality Industry", "Data Stream Mining", "Tourism", "Online Learning", "Computational Intelligence" or "Deep Learning", along with a *filter by topic* (title, abstract and/or keywords) search heuristic and a date range from 1998 to 2020, yielded a manifold of results from the Web of Science's Core Collection. Following these findings, a thorough examination of all such retrieved works permitted to discover which contributions fell within the scope of this work.

This process yielded more than a hundred and sixty papers which directly relate to certain CI methods with an application in one or more of the main four identified areas. To visually depict these results in an understandable and informative manner, a table containing all the analyzed references was created for the reference and guidance of researchers and newcomers arriving at the crossroads between CI and hospitality industry. On one hand, it allows the CI community to know which methods have been mostly used in each area of the hospitality industry, detecting niches of opportunity around a certain CI method that remains uncharted for every area. For the research community investigating on the hospitality industry, this table helps discriminating which CI methodologies to apply for every subtopic, or even qualitatively measure to what extent CI has been applied to their specific area of interest.

For all this, the main objective of this table is to show the union between both disciplines, as well as to uncover the possible gaps and future research opportunities that are explained and discussed in the following section.

Additionally, in Table 2 a comprehensive mix of the identified hospitality areas and subareas, typical cases of use, commonly used CI methodologies, unresolved problems and future research lines is presented. This table is meant to offer a comparative analysis of both CI methods and hospitality industry related matters, depicting how CI methodologies are being applied in the hospitality industry research areas identified in this survey.

This table has been done by thoroughly analyzing each surveyed research work and identifying key information contained on it. The first thing to categorize was on which hospitality industry topic it was located, and the problem that was addressed by a certain CI-based solution. Secondly, an overview sweep was done to correctly identify the most common CI methods used in each area, alongside the unresolved problems or difficulties that the aforementioned area presents in terms of research or execution. Lastly, possible future research avenues are mentioned to help both CI experts or hospitality industry stakeholders direct future studies or applications for CI methods or hospitality use cases.

6. Challenges

The adoption of CI techniques and methods has already be fruitful within the hospitality sector, as evinced in previous sections. However, new developments and paradigms of CI can provide further improvements in this industry, opening new challenges that should be addressed in forthcoming research. Some of the main challenges related to such technology advancements are summarized below:

6.1. Actionable intelligence

Actionability of a system refers to the ability of its results to be put to action in the particular context for which they are intended. In CI research there is a tension between providing outstanding performance results and implementing applicable models and methods. Techniques such those under the umbrella of explainable Artificial Intelligence (xAI) [186,187] can help the field stakeholders put to practice the outcomes of these CI-based systems, especially those whose inner structure and learning procedure are far from transparent to non-expert audience.

A possible solution for this kind of challenge could be to create forecasting tools which provide outputs that explain the relevance of their inputs for the elicited output, increasing the trustworthiness on the model by the decision maker and eventually, the actionability of its output. Besides, considering Transfer Learning [188] techniques could be useful to extend developments to different domains and environments.

6.2. Multi-source data fusion

The variety of data sources in hospitality industry can give place to the implementation of Data Fusion models that could obtain enriched information and insights on the available data. For instance, characterizing tourists or guests can be achieved by contemplating many sources of information. Using knowledge management to manage all the data sources present in an hospitality establishment might be a good solution.

[189] defined knowledge management as the coordination and use of the organization's knowledge resources to create a competitive advantage. Traditionally, the competitiveness of a business relied on capital, land, labor and many other tangible resources but, in recent times, it has been proven that knowledge management has become an important source of competitive advantage [190]. In fact, [191] uncovered that knowledge management can be one of the most important assets for hospitality establishments and organizations because of its ability to help these organizations create and sustain competitive advantages by using different IT tools such as competency databases, decision support systems or data warehousing. Using these kind of techniques could allow hospitality establishments to enhance their services by combining different databases and using knowledge management to uncover important insights about their clients, which will yield competitive advantages over their competitors.

Besides using specialized software to overcome the difficulties found in merging data with different structures, a potential approach to solve this challenge could be to generate a *data standard* for hospitality industry databases, generalizing database structures that are easy to combine, facilitating the data merging process between databases from different services, and increasing the potential of applications relying on CI methods.

6.3. Dynamic and online learning

The growth in computational power as well as the development of more efficient methods have pushed CI towards a more online processing paradigm. Online learning and dynamic optimization schemes, as well as the introduction of concepts such as change detection and adaptation, could introduce new perspectives to the application of CI in the hospitality industry. In their research work, [192] classifies different ensemble algorithms for different data stream mining tasks into a useful taxonomy, additionally presenting found research problems and future research lines, making this work of paramount importance for future online learning developments.

Trends in the hospitality industry change, thus making ML models less accurate over time. This can provoke several losses in terms of revenue by various factors, such as erroneous tourism demand predictions or wrong resource demand forecasts. Online learning provides a solution to these problems. In this ML paradigm data is treated as data streams, making it potentially infinite. As stated by [193], "Data continuously arrives in real time, and may be changing over time. In this setting, predictive models need to operate fast, fit into limited memory, and adapt online; otherwise their accuracy will degrade over time". This is opposed to the classic batch learning paradigm, where a predictor is generated only when the entire dataset has been analyzed.

With online learning, it is possible to analyze data incrementally as they arrive, without any need for retaining them in memory. When combined with techniques for detecting and adapting to concept drift, online learning models for prediction tasks result to be not only efficient, but also resilient to the wear of time. Several approaches have hitherto gravitated on this paradigm. For instance, L. Lobo et al. [194] demonstrated that spiking neural networks [195] are notably effective in real-time stream mining, unleashing a promising research area inside online learning around spiking neural networks thanks to the low computational cost and high representational power of these models.

Using this methodology in the hospitality industry might yield competitive advantages for its establishments, because of its ability to update continuously its captured knowledge and adapt to the natural data changes resulting from the non-stationary nature of hospitality processes. When applied to resource or tourism demand, decision making could be enhanced by more accurate model predictions, potentially increasing incomes and improving the establishment's adaptability to different kinds of events.

Table 2
Found hospitality research areas, crossed with typical use cases, frequently used CI methods, known limitations/problems and possible future research lines.

Hospitality topic	Typical	Commonly used CI-based solutions	Unsolved	Future research lines
Booking systems	– Manage booking inside an establishment. – Enhance booking systems. – Increase booking system speed.	Clustering Decision trees Artificial neural networks	– Long model training time. – Slow vs. a human agent.	– Fast + reactive CI methods to generate faster systems. – Use of Online Learning to enhance adaptiveness and system speed.
Product development and marketing	– Development of new products. – Enhance marketing decisions. – Obtain customer knowledge.	Clustering Decision trees	– Size of the stored data. – Small data availability. – Anonymity of data.	– Anonymous + innocuous data collection. – Tracking methods to collect anonymous tourist data. – Online Learning could reduce the size of stored data.
Resource demand forecasting	– Electricity load forecasting. – Water consumption prediction. – Gas consumption forecasting.	Artificial neural networks Instance/Linear based methods	– System adaptability.	– Use of Online Learning to create long-term adaptable systems.
Revenue management and forecasting	– Guest expenditure. – Income prediction. – Price forecasting.	Artificial neural networks Ensembles	– Value variance. – Difficulty to predict income on a demand basis.	– Hybrid forecasting systems (revenue + tourism demand). – Seasonal prediction systems.
Tourism company feasibility analysis	– Bankruptcy prediction. – Claim and dispute management. – Forecast the success of newly launched services.	Artificial neural networks Instance/Linear based methods Ensembles	– Unexplored research niche.	– Experiment with different CI methods to increase available knowledge.
Tourism market segmentation, accessibility and dependencies	– Identify market segments. – Location assessment for new hotels. – Identify relations between hotel price and market accessibility.	Instance/Linear based methods	– Unexplored research niche.	– Experiment with different CI methods to increase available knowledge.
Geo-tagging and traveling	– Discover popular tourism attractions or establisments. – Establishment ranking. – Find patterns in traveler behavior.	Clustering	– Static predictive models. – Not enough CI methods have been applied to explore the matter.	– Explore more efficient clustering algorithms. – Apply different CI methods to assess their effectivity. – Use of Online Learning to increase a model's time of life.
Tourist behavior analysis	– Client loyalty prediction. – Find factors that could influence the client. – Analyze client expenditure distribution.	Decision trees ensembles	– Data complexity. – Data is usually "noisy".	– Find suitable data pretreatment methods to clean noisy data. – Apply Deep Learning methods to overcome excessive data complexity. – Apply Online Learning methods to create long-term, adaptive models.
Tourism recommender systems	– Recommend different attractions or establishments. – Create personalized recommendations. – Create touristic routes based on clients' preferences.	Probabilistic methods	– Big dataset sizes. – Slow speed algorithm execution.	– Parameter optimization. – Online Learning methods could help to decrease data size and to create adaptive models.
Sentiment analysis and satisfaction degree	– Extract sentiment data from reviews, blogs, social networks etc. – Analysis of images published by clients in different platforms. – Detection of fake reviews and comments.	Instance/Linear based methods Ensembles	– Huge dataset size. – Long model training time.	– Optimization of known methods. – Speed up model generation by using Deep Learning techniques. – Online Learning methods could reduce dataset size.
Tourism demand forecasting	– Prepare services for a certain load. – Prevent service overload. – Forecast tourist arrivals. – Predict booking cancellations.	Artificial neural networks	– Intensely researched area. – Trends in data usually change.	– Exploration of hybrid methods and ensembles. – Use of Online Learning to create versatile algorithms.
Weather forecasting and environmental risks assessment	– Predict extreme or certain weather conditions. – Event planification.	Instance/Linear based methods	– Predictive models usually get outdated quickly. – Not enough CI methods have been applied to explore the matter.	– More experimentation with different CI methods is needed.

6.4. Encrypted data models

One of the world's biggest concerns about Big Data is that, despite the data being anonymized, it is possible to recover bits of information that allow identification of individuals by different means. Until now, classic data encryption has been found not to be suitable for knowledge mining because of its very nature: To work with something that has been encrypted, it must first be decrypted. However, new privacy-preserving paradigms in the Artificial Intelligence realm such as Homomorphic Encryption or Federated Learning can be useful in resolving this problem, allowing stakeholders from the hospitality industry safely work and exchange customer data without any privacy-related concerns.

Homomorphic Encryption was firstly theorized by [196], and it is based on the use of determined mathematical properties present in certain encryption schemes. These properties allow to perform operations over encrypted data without decryption, presenting results that are still encrypted but can be further processed or decrypted. Later on, [197] proposed the first working homomorphic encryption system, which evaluates different low-degree polynomials over previously encrypted information by using an evaluation function. Actual developments in homomorphic encryption are presented by authors such as [198] (Fan–Vercauteren cryptography scheme) and [199] (Brakerski–Gentry–Vaikuntanathan cryptography scheme), and are based on Oded Regev's *Ring Learning With Errors* problem [200]. Particularly for the hospitality industry, the use of homomorphic encryption can be massively adopted when characterizing the customer as a whole, building CI models without compromising protected features that the customer him/herself could regard as confidential (e.g. incomes, sexual orientation, gender and other aspects alike).

However, if the focus is placed on the exchange of hospitality related information among stakeholders, reluctance arises due to the strong competitiveness existing in the sector. The combination of the information generated by the customer at different locations/over different time frames could provide enormous modeling benefits in terms of the accuracy under which models fed with such combined data could perform. However, in practice most companies are not open to sharing the information generated by the customer at their premises.

The recent advent of *Federated Learning* can change the game in this matter by establishing the technical grounds to share model-related privacy-preserving information (e.g. gradients of neural networks) rather than raw data among distributed models [201]. Contributions from all such models are centralized and processed, yielding an aggregated representation that can be delivered back to the models and combined with the locally learned knowledge for an improved performance. This paradigm is reaching maturity in the last couple of years, showing a great potential to ignite the adoption of CI models in privacy-sensitive application domains such as health and industry [202]. Without a doubt, Federated Learning will also take a major role in future deployments of CI over the hospitality industry.

6.5. Data biasing anticipation

When the phenomenon to be modeled is not stationary in terms of its statistical behavior, the pattern to be learned by a CI model may undergo changes that eventually make the model obsolete. This situation is widely referred to as Concept Drift [203], i.e. a change in the process generating the data distribution to be learned that is not explicitly reflected in the input data themselves. This issue is particularly frequent in processes generating fast data streams (e.g. electronic purchases), and often leads to a significant deterioration of the performance of the predictive models learned over time. That chance in the statistical characteristics of the phenomenon can be assimilated as a progressive data bias. The reasons for that concept drift may be diverse. Sometimes it can be because of human behavior in a decision making context. For example, the Braess paradox [204] in dynamic travel assignment explains how leaving the route choice exclusively to drivers may end up in a Nash equilibrium state, which may not be system (globally) optimal. In other words, leaving the route choice to drivers alone with their local (selfish) optimization criteria may end up in a worsening on the total performance of a particular traffic network. This paradox has no analytical solution. The development of any predictive model and dynamic route assignment trained for a particular demand distribution will naturally result in a bias in the behavior of that demand distribution and the consequent deterioration on the performance of such setup. The only practical strategy is to rely on a methodology that can cope with concept drifting in the statistical distribution of the network demand. Specifically, a dynamic methodology is needed to detect that demand bias when it happens, thereby triggering an adaptation strategy. In other words, the model learning and dynamic route assignment methodologies must be incremental and adaptive to cope with such concept drift in the demand signal of the traffic network.

A similar situation holds in the hospitality industry: when information produced by the customer changes his/her behavior, a human-induced data bias may appear as a result of the decisions taken therefrom. A clear example are recommendation systems, which may rely on assorted CI techniques, such as predictive modeling and ranking methods. For instance, a customer of a hotel can make decisions on the basis of recommended items, which are often fed back to the recommendation for its update. When a contextual change in their habits is not reflected in the data provided to the engine (e.g. a change of marital status), decisions may change radically, making the recommendation engine outdated and ultimately, recommendations useless until the model learns to grasp the new context of the user. Depending on the speed and severity of the change, there may exist a significant delay until the recommendation model provides meaningful outputs for the new concept.

Concept drift detection, characterization and adaptation techniques aim precisely at shortening the time required by the model to reflect the new data distribution. The prevalence of concept drift and data bias in the touristic domain, where human decisions are subject to a wide variety of contextual factors that are not explicitly accounted for in the collected data, makes it of utmost necessity to further study how CI models can learn from and efficiently adapt to non-stationary scenarios.

7. Directions for future research works

The increase in computational power and the development of new methodologies to obtain knowledge from all kinds of data sources, which range from small sensors to Big Data, have created new ways of analyzing huge amounts of data in less time. However, that there is still ample room for new CI applications in the hospitality industry. The use of Deep Learning methodologies provides useful insights in various forecasting fields inside hospitality, but it is necessary to create new, actionable methods that reduce the opacity of the layers inside a Deep Learning system, easing their application and understandability. This opens up a new research field where a model's performance will be important, but also its transparency will be prime in order to offer what could be called a *solid* model. New or updated methodologies which can render a good performance while being applicable to other fields should be researched to enrich the technological landscape of the hospitality industry.

Additionally, the computational power available for analyzing data increases in an exponential manner. The data increase in volume, speed and heterogeneity also requires enormous computational resources to gain insights from them, hindering their analysis via CI-based models. Moreover, the hospitality industry is in need for models that not only work in real-time, but also work in variable conditions and scenarios. This lays a fertile soil for new research works based on adaptive, real-time CI models where a reliable model works in real time, and also adapts to new trends in the data, allowing the hospitality establishments to forecast different variables such as seasonal booking spikes, possible resource outcomes or revenue estimations. There already are various methodological approaches for this purpose, such as the aforementioned family of online learning models, which provide these kind of results. However, since it is an continuously evolving field, the latest emerging research works should be progressively imported for applications in the hospitality industry known to be subject to high levels of variability and non-stationary exogenous factors affecting their data flows.

Nevertheless, beyond adaptability and actionability, data privacy also stands as a key component when developing a good model. Hospitality industry's generated data is a treasure trove in terms of data value, but also a treacherous water to sail in because there are huge amounts of data that hold personal protected information about the clients that use hospitality establishments. We currently live in a world where data leaks often occur, and where the leaked information can be accessed with little to no difficulty. As stated in one of the previous sections of this research work, homomorphic encryption applied to the data used to train different models could keep the model's performance while also keeping data private, thus creating the need of encrypted data models where CI could be applied. Additionally, the use of federated learning methodologies could also enable interesting collaborative scenarios between various stakeholders, thus producing more precise models without compromising the privacy of data collected by each party.

Lastly, there is no short answer to the question how to overcome the data bias problem. Further research is needed to understand how CI models can reliably detect and efficiently adapt to concept drifts present in data. The truth is that, unless such capabilities are ensured in CI models, undesirable effects would appear when making decisions from the output of these models, such as revenue losses or badly allocated resources. Research efforts in this direction could entail new adaptive CI methods and fast-learning models where adaptive capabilities become the key for their deployability in real-world hospitality scenarios.

In short, new research venues inside CI applied to the hospitality industry should include efforts towards new methods that allow different hospitality-related data sources to be fused together in a privacy-aware fashion, so that confidentiality is preserved and data sharing is encouraged among stakeholders. It is also of dire importance to investigate models that resiliently adapt to changes in data distribution, featuring effective mechanisms to circumvent the data bias problem often present in use cases of this industry.

8. Conclusions and outlook

This research work explores the applications of CI in all the sub-fields of hospitality and tourism industry, exposing the most frequently used techniques and methodologies and additionally finding potential unexploited research niches. Based on this exploration, a new categorization of the State of the Art has been proposed, comprising more than 160 research works regarding CI applied to hospitality. Our literature study has revealed that probabilistic/Bayesian Methods are mostly used along with instance/linear methods, with ANNs fast growing in usage in the recent years, all of them being used mostly for classification and forecasting. Decision trees and clustering methodologies are also of common use in this area. The use of ensembles is also starting to grow, specially when it comes to tune parameters for data pre-processing by using evolutionary computation algorithms. Evolutionary computation is also used to optimize parameters along various methodologies, and is commonly used in forecasting various parameters in the hospitality industry.

With the increasing computational power and the recent advancements on ML, hospitality and tourism forecasting seems to become a hot topic in the years to come. However, there is a lot of research that has been done regarding *safe* methodologies, instead of researching new CI methods and applications that could not only improve the results of the current methodologies, but even surpass them in terms of performance or usability in the long run. The use of ensemble learners already prove this, as they usually yield better results than the original, separated methods, by balancing each learner intrinsic bias and variance. Another promising research line might be the use of data stream mining to improve performance on big datasets in real-time operation, additionally allowing the resulting model to adapt over time and overcome concept drift. Finally, as it seems to be happening in every other area, Deep Learning will be used soon, specially in applications where Big Data is available, and the interpretability and transparency of the resulting models is not a critical issue.

However, limitations and assumptions of this research work must also be stated. One hundred and sixty research works in a 20 year range (from 1998 to 2018) were reviewed in order to extract current prospects in CI applied to the hospitality industry. These prospects were used not only to showcase how CI methods are being applied, but also to identify and propose a novel taxonomy based on how the hospitality industry's research efforts are currently distributed, based on the aforementioned methodologies. One of the main limitations of this research was the huge amount of literature present in this area, thus limiting the reviewed research works under the assumption that the most relevant ones in their area or subarea are representative of what is currently being done. Although being a limitation, this presents an opportunity to create new future reviews based on a deeper analysis of each subarea. Additionally, it was found out that there are many research works that do not exactly fall inside CI because of the methodologies used in them. This limitation can also be seen as an unprecedented opportunity to delve deeper in new atypical ways to discover applications or areas, which might ultimately lead to undiscovered research avenues.

On a closing note, the set of identified challenges present an unique opportunity not only as an insufficiently explored avenue for the research community working on CI, but also for the hospitality industry to become fully aware of the benefits that this research area can bring. We utterly believe that the material and prospects provided in our study is just an informed sample of the huge potential that underlies beneath the adoption of computational intelligence in this sector. It is time for the research community to join this path, stepping on grounds of evidence as those given in this survey, so that CI spreads over the whole hospitality industry in years to come.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

Research work co-funded by Agencia Canaria de Investigación, Innovación y Sociedad de la Información (ACIISI) de la Consejería de Economía, Industria, Comercio y Conocimiento, Spain and by Fondo Social Europeo (FSE) Programa Operativo Integrado de Canarias 2014–2020, Spain, Eje 3 Tema Prioritario 74 (85%). This paper was developed under project "Sistema de vigilancia meteorológica para el seguimiento de riesgos medioambientales", VIMetRi-MAC (ref. MAC/3.5b/065), funded by Programa de Cooperación Territorial INTERREG V A Spain–Portugal, MAC 2014–2020, Spain. Ibai Laña and Javier Del Ser also acknowledge funding support from the Basque Government, Spain through the EMAITEK and ELKARTEK funding programs, Basque Government, Spain. Javier Del Ser receives funding support from the Consolidated Research Group MATHMODE, Spain (IT1294-19) granted by the Department of Education of the Basque Government.

References

[1] N. Kamel, A. Atiya, N. Gayar, H. El-Shishiny, Tourism demand foreacsting using machine learning methods, ICGST Int. J. Artif. Intell. Mach. Learn. 8 (2018) 1–7.
[2] N. Siddique, H. Adeli, Computational Intelligence: Synergies of Fuzzy Logic, Neural Networks and Evolutionary Computing, John Wiley & Sons, 2013.
[3] T. Mitchell, Introduction to machine learning, Mach. Learn. 7 (1997) 2–5.
[4] T.M. Mitchell, Machine learning and data mining, Commun. ACM 42 (11) (1999) 30–36.
[5] Y. Chen, H. Tan, U. Berardi, Day-ahead prediction of hourly electric demand in non-stationary operated commercial buildings: A clustering-based hybrid approach, Energy Build. 148 (2017) 228–237.
[6] B. Marr, How big data and analytics are changing hotels and the hospitality industry, 2016, http://Forbes.com.
[7] W. Pedrycz, J.F. Peters, Computational Intelligence in Software Engineering, Vol. 16, World Scientific, 1998.
[8] Y. Wang, On abstract intelligence: Toward a unifying theory of natural, artificial, machinable, and computational intelligence, Int. J. Softw. Sci. Comput. Intell. (IJSSCI) 1 (1) (2009) 1–17.
[9] A.E. Eiben, J.E. Smith, et al., Introduction to Evolutionary Computing, Vol. 53, Springer, 2003.
[10] S.-C. Wang, Artificial neural network, in: Interdisciplinary Computing in Java Programming, Springer US, Boston, MA, 2003, pp. 81–100.
[11] H. Ishwaran, J. Rao, Decision tree: Introduction, Encyclopedia Med. Decis. Mak. (2009) 323–328.
[12] R. Schoot, D. Kaplan, J. Denissen, J. Asendorpf, F. Neyer, M. Aken, A gentle introduction to Bayesian analysis: Applications to developmental research, Child Dev. 85 (2013).
[13] D.W. Aha, D. Kibler, M.K. Albert, Instance-based learning algorithms, Mach. Learn. 6 (1) (1991) 37–66.
[14] B. Silverman, M. Jones, E. Fix, J. Hodges, An important contribution to nonparametric discriminant analysis and density estimation: Commentary on fix and hodges (1951), Int. Statist. Rev. / Rev. Int. Statist. 57 (1989) (1951).
[15] M.J.L. Orr, Introduction to Radial Basis Function Networks, Technical Report, Center for Cognitive Science, University of Edinburgh, 1996.
[16] D. Opitz, R. Maclin, Popular ensemble methods: An empirical study, J. Artif. Int. Res. 11 (1) (1999) 169–198.
[17] M. Omran, A. Engelbrecht, A. Salman, An overview of clustering methods, Intell. Data Anal. 11 (2007) 583–605.
[18] S. Kotsiantis, D. Kanellopoulos, Association rules mining: A recent overview, GESTS Int. Trans. Comput. Sci. Eng. 32 (2005) 71–82.
[19] G. Phillips-Wren, A. Hoskisson, Decision support with big data: A case study in the hospitality industry, Front. Artif. Intell. Appl. 261 (2014) 401–413.
[20] L. Zadeh, Fuzzy sets, Inf. Control 8 (3) (1965) 338–353.
[21] J.H. Holland, Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control and Artificial Intelligence, MIT Press, Cambridge, MA, USA, 1992.
[22] Q. Zhang, L.T. Yang, Z. Chen, P. Li, A survey on deep learning for big data, Inf. Fusion 42 (2018) 146–157.
[23] J.-L. Casteleiro-Roca, J.F. Gómez-González, J.L. Calvo-Rolle, E. Jove, H. Quintián, B. Gonzalez Diaz, J.A. Mendez Perez, Short-term energy demand forecast in hotels using hybrid intelligent modeling, Sensors 19 (11) (2019).
[24] V. Cho, P. Leung, Towards using knowledge discovery techniques in database marketing for the tourism industry, J. Qual. Assur. Hosp. Tour. 3 (3–4) (2002) 109–131.
[25] M.-S. Chen, L.-C. Ying, M.-C. Pan, Forecasting tourist arrivals by using the adaptive network-based fuzzy inference system, Expert Syst. Appl. 37 (2) (2010) 1185–1191.
[26] R. Bettin, F. Mason, M. Corazza, G. Fasano, An Artificial Neural Network Technique for On-Line Hotel Booking, SSRN Scholarly Paper ID 2037481, Social Science Research Network, Rochester, NY, 2011.
[27] N. Gayar, A. Hendawi, H. El-Shishiny, A proposed decision support model for hotel revenue management, Int. J. Artif. Intell. Mach. Learn. 8 (2008).
[28] S.Y. Kim, Prediction of hotel bankruptcy using support vector machine, artificial neural network, logistic regression, and multivariate discriminant analysis, Serv. Ind. J. 31 (3) (2011) 441–468.
[29] Chou Jui-Shen, Lin Chieh, Predicting disputes in public-private partnership projects: Classification and ensemble models, J. Comput. Civ. Eng. 27 (1) (2013) 51–60.
[30] H.-C. Huang, A. Y Chang, C.-C. Ho, Using artificial neural networks to establish a customer-cancellation prediction model, Prz. Elektrotech. 89 (2013).
[31] K. Kampouropoulos, F. Andrade, A. Garcia, L. Romeral, A combined methodology of adaptive neuro-fuzzy inference system and genetic algorithm for short-term energy forecasting, Adv. Electr. Comput. Eng. 14 (1) (2014) 9–14, WOS:000332062300002.
[32] Xue-Bo, L. Shi-Ting, Management of tourism resources and demand based on neural networks, in: 2014 7th International Conference on Intelligent Computation Technology and Automation, 2014, pp. 348–351.
[33] O. Claveria, E. Monte, S. Torra, Modelling cross-dependencies between Spains regional tourism markets with an extension of the Gaussian process regression model, SERIEs 7 (3) (2016) 341–357.
[34] D. Kofinas, E. Papageorgiou, C. Laspidou, N. Mellios, K. Kokkinos, Daily Multivariate Forecasting of Water Demand in a Touristic Island with the Use of Artificial Neural Network and Adaptive Neuro-Fuzzy Inference System, Ieee, New York, 2016, WOS:000386536900007.
[35] J. Lu, Machine learning modeling for time series problem: Predicting flight ticket prices, 2017, arXiv:1705.07205 [cs], arXiv:1705.07205.
[36] A. Porto, E. Irigoyen, Gas consumption prediction based on artificial neural networks for residential sectors, in: International Joint Conference SOCO'17-CISIS'17-ICEUTE'17 León, Spain, September 6–8, 2017, Proceeding, in: Advances in Intelligent Systems and Computing, Springer, Cham, 2017, pp. 102–111.
[37] G.S. Atsalakis, I.G. Atsalaki, C. Zopounidis, Forecasting the success of a new tourism service by a neuro-fuzzy technique, European J. Oper. Res. 268 (2) (2018) 716–727, WOS:000430888300022.
[38] M. Al Shehhi, A. Karathanasopoulos, Forecasting hotel room prices in selected GCC cities using deep learning, J. Hosp. Tour. Manag. 42 (2020) 40–50.
[39] W.B. Claster, H. Dinh, M. Cooper, Nave Bayes and unsupervised artificial neural nets for cancun tourism social media data analysis, in: 2010 Second World Congress on Nature and Biologically Inspired Computing (NaBIC), 2010, pp. 158–163.
[40] A. Sharma, S. Dey, A document-level sentiment analysis approach using artificial neural network and sentiment lexicons, SIGAPP Appl. Comput. Rev. 12 (4) (2012) 67–75.
[41] P. Phillips, K. Zigan, M.M. Santos Silva, R. Schegg, The interactive effects of online reviews on the determinants of swiss hotel performance: A neural network analysis, Tour. Manag. 50 (2015) 130–141.
[42] Y. Yang, J. Tang, H. Luo, R. Law, Hotel location evaluation: A combination of machine learning tools and web GIS, Int. J. Hosp. Manag. 47 (2015) 14–24.
[43] V. Bugarski, D. Mati, F. Kuli, Classification of hotel guests by predicted additional spending with ANN decision support system, in: 2017 IEEE 15th International Symposium on Intelligent Systems and Informatics (SISY), 2017, pp. 000071–000076.
[44] M. Nilashi, K. Bagherifard, M. Rahmani, V. Rafe, A recommender system for tourism industry using cluster ensemble and prediction machine learning techniques, Comput. Ind. Eng. 109 (2017) 357–368.
[45] M. Nilashi, A. Ahani, M.D. Esfahani, E. Yadegaridehkordi, S. Samad, O. Ibrahim, N.M. Sharef, E. Akbari, Preference learning for eco-friendly hotels recommendation: Amulti-criteria collaborative filtering approach, J. Cleaner Prod. 215 (2019) 767–783.
[46] X. Cheng, S. Fu, J. Sun, A. Bilgihan, F. Okumus, An investigation on online reviews in sharing economy driven hospitality platforms: A viewpoint of trust, Tour. Manag. 71 (2019) 366–377.
[47] Y.-C. Chang, C.-H. Ku, C.-H. Chen, Using deep learning and visual analytics to explore hotel reviews and responses, Tour. Manag. 80 (2020) 104129.
[48] J. Luo, S.S. Huang, R. Wang, A fine-grained sentiment analysis of online guest reviews of economy hotels in China, J. Hosp. Mark. Manag. (2020) 1–25.

[49] M. Ren, H.Q. Vu, G. Li, R. Law, Large-scale comparative analyses of hotel photo content posted by managers and customers to review platforms based on deep learning: implications for hospitality marketers, J. Hosp. Mark. Manag. (2020) 1–24.
[50] A. Shoukry, F. Aldeek, Attributes prediction from IoT consumer reviews in the hotel sectors using conventional neural network: deep learning techniques, Electron. Comm. Res. 20 (2) (2020) 223–240.
[51] P.-F. Pai, W.-C. Hong, C.-S. Lin, Forecasting tourism demand using a multifactor support vector machine model, in: Computational Intelligence and Security, in: Lecture Notes in Computer Science, Springer, Berlin, Heidelberg, 2005, pp. 512–519.
[52] C.-F. Chen, M.-C. Lai, C.-C. Yeh, Forecasting tourism demand based on empirical mode decomposition and neural network, Knowl.-Based Syst. 26 (2012) 281–287.
[53] T.M. Berenguer, J.A.M. Berenguer, M.E.B. Garca, A.P. Pol, J.J.M. Moreno, Models of artificial neural networks applied to demand forecasting in nonconsolidated tourist destinations, Methodol.: Eur. J. Res. Methods Behav. Soc. Sci. 11 (2) (2014) 35–44.
[54] S. Cankurt, A. Subasi, Developing tourism demand forecasting models using machine learning techniques with trend, seasonal, and cyclic components, Balkan J. Electr. Comput. Eng. 3 (1) (2015).
[55] O. Claveria, E. Monte, S. Torra, Tourism demand forecasting with neural network models: Different ways of treating information, Int. J. Tour. Res. 17 (5) (2015) 492–500.
[56] X. Wang, H. Zhang, X. Guo, Demand forecasting models of tourism based on ELM, in: 2015 Seventh International Conference on Measuring Technology and Mechatronics Automation, 2015, pp. 326–330.
[57] O. Claveria, E. Monte, S. Torra, Combination forecasts of tourism demand with machine learning models, Appl. Econ. Lett. 23 (6) (2016) 428–431.
[58] O. Claveria, S. Torra, E. Monte, Modelling tourism demand to spain with machine learning techniques. the impact of forecast horizon on model selection, Rev. Econ. Apl. XXIV (72) (2016).
[59] E. Noersasongko, F.T. Julfia, A. Syukur, Purwanto, R.A. Pramunendar, C. Supriyanto, A tourism arrival forecasting using genetic algorithm based neural network, Indian J. Sci. Technol. 9 (4) (2016).
[60] J. Sun, T. Chang, Prediction of rural residents' tourism demand based on back propagation neural network, Int. J. Appl. Decis. Sci. 9 (3) (2016) 320–331.
[61] N. Antonio, A.d. Almeida, L. Nunes, Predicting hotel booking cancellations to decrease uncertainty and increase revenue, Tour. Manag. Stud. 13 (2) (2017) 25–39.
[62] Y.W. Chang, C.Y. Tsai, Apply deep learning neural network to forecast number of tourists, in: 2017 31st International Conference on Advanced Information Networking and Applications Workshops (WAINA), 2017, pp. 259–264.
[63] O. Claveria, E. Monte, S. Torra, 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, Rochester, NY, 2017.
[64] R. Folgieri, T. Baldigara, M. Mamula, Artificial neural networks-based econometric models for tourism demand forecasting, in: S. Markovic, D.S. Jurdana (Eds.), 4th International Scientific Conference: Tosee - Tourism in Southern and Eastern Europe 2017, Vol. 4, Univ Rijeka, Faculty Tourism & Hospitality Management, Opatija, Opatija, 2017, pp. 169–182, WOS:000428229600013.
[65] S. Han, Y. Guo, H. Cao, Q. Feng, Y. Li, A cross-view model for tourism demand forecasting with artificial intelligence method, in: Data Science, in: Communications in Computer and Information Science, Springer, Singapore, 2017, pp. 573–582.
[66] W. Oger Vihikan, I. Ketut Gede Darma Putra, I. Putu Arya Dharmaadi, Foreign tourist arrivals forecasting using recurrent neural network backpropagation through time, Telkomnika (Telecommun. Comput. Electron. Control) 15 (2017) 1257–1264.
[67] A. Rafidah, S. Ani, Modelling Singapore tourist arrivals to Malaysia by using SVM and ANN, SCIREA J. Math. 1 (2) (2017) 210–216.
[68] S. Sun, S. Wang, Y. Wei, X. Yang, K.L. Tsui, Forecasting tourist arrivals with machine learning and internet search index, in: 2017 IEEE International Conference on Big Data (Big Data), 2017, pp. 4165–4169.
[69] M.A. King, A.S. Abrahams, C.T. Ragsdale, Ensemble methods for advanced skier days prediction, Expert Syst. Appl. 41 (4, Part 1) (2014) 1176–1188.
[70] G. Emel, C. Taşkın, Identifying segments of a domestic tourism market by means of data mining, in: Operations Research Proceedings, Vol. 2005, 2005, pp. 653–658.
[71] S.H. Ha, S.C. Park, Application of data mining tools to hotel data mart on the intranet for database marketing, Expert Syst. Appl. 15 (1) (1998) 1–31.
[72] H. Min, H. Min, A. Emam, A data mining approach to developing the profiles of hotel customers, Int. J. Contemp. Hosp. Manag. 14 (6) (2002) 274–285.
[73] Z. Guoxia, T. Jianqing, The application of data mining in tourism information, in: 2009 International Conference on Environmental Science and Information Application Technology, Vol. 3, 2009, pp. 689–692.
[74] Z. Zhang, Q. Ye, Z. Zhang, Y. Li, Sentiment classification of internet restaurant reviews written in cantonese, Expert Syst. Appl. 38 (6) (2011) 7674–7682.
[75] A. Luberg, M. Granitzer, H. Wu, P. Jrv, T. Tammet, Information retrieval and deduplication for tourism recommender sightsplanner, in: Proceedings of the 2Nd International Conference on Web Intelligence, Mining and Semantics, in: WIMS '12, ACM, New York, NY, USA, 2012, pp. 50:1–50:11.
[76] C. Zhang, J. Zhang, Analysing chinese citizens' intentions of outbound travel: a machine learning approach, Curr. Issues Tour. 17 (7) (2014) 592–609.
[77] S. Banerjee, A.Y.K. Chua, J.-J. Kim, Distinguishing between authentic and fictitious user-generated hotel reviews, in: 2015 6th International Conference on Computing, Communication and Networking Technologies (ICCCNT), 2015, pp. 1–7.
[78] S. Nakamura, M. Okada, K. Hashimoto, An investigation of effectiveness using topic information order to classify tourists reviews, in: 2015 International Conference on Computer Application Technologies, 2015, pp. 94–97.
[79] M.E.B.H. Kbaier, H. Masri, S. Krichen, A personalized hybrid tourism recommender system, in: 2017 Ieee/Acs 14th International Conference on Computer Systems and Applications (Aiccsa), Ieee, New York, 2017, pp. 244–250, WOS:000428311600042.
[80] S. Yordanova, D. Kabakchieva, Sentiment classification of hotel reviews in social media with decision tree learning, Int. J. Comput. Appl. 158 (2017).
[81] M. Nilashi, E. Yadegaridehkordi, O. Ibrahim, S. Samad, A. Ahani, L. Sanzogni, Analysis of travellers' online reviews in social networking sites using fuzzy logic approach, Int. J. Fuzzy Syst. 21 (5) (2019) 1367–1378.
[82] S. Cankurt, Tourism demand forecasting using ensembles of regression trees, in: 2016 IEEE 8th International Conference on Intelligent Systems (IS), 2016, pp. 702–708.
[83] M. Akın, A novel approach to model selection in tourism demand modeling, Tour. Manag. 48 (2015) 64–72.
[84] N. Deng, X.R. Li, Feeling a destination through the "right" photos: A machine learning model for DMOs' photo selection, Tour. Manag. 65 (2018) 267–278.
[85] Q. Ye, Z. Zhang, R. Law, Sentiment classification of online reviews to travel destinations by supervised machine learning approaches, Expert Syst. Appl. 36 (3, Part 2) (2009) 6527–6535.
[86] A. Weichselbraun, S. Gindl, A. Scharl, A context-dependent supervised learning approach to sentiment detection in large textual databases, J. Inf. Data Manag. 1 (3) (2010) 329.
[87] K. Shimada, S. Inoue, H. Maeda, T. Endo, Analyzing tourism information on Twitter for a local city, in: 2011 First ACIS International Symposium on Software and Network Engineering, 2011, pp. 61–66.
[88] F.-M. Hsu, Y.-T. Lin, T.-K. Ho, Design and implementation of an intelligent recommendation system for tourist attractions: The integration of EBM model, Bayesian network and google maps, Expert Syst. Appl. 39 (3) (2012) 3257–3264.
[89] Y. Wang, S.C.F. Chan, G. Ngai, Applicability of demographic recommender system to tourist attractions: A case study on trip advisor, in: 2012 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology, Vol. 3, 2012, pp. 97–101.
[90] Q. Xu, H. Zhao, Using deep linguistic features for finding deceptive opinion spam - semantic scholar, 2012.
[91] W. Chang, L. Ma, Personalized e-tourism attraction recommendation based on context, in: 2013 10th International Conference on Service Systems and Service Management, 2013, pp. 674–679.
[92] W. Duan, Q. Cao, Y. Yu, S. Levy, Mining online user-generated content: Using sentiment analysis technique to study hotel service quality, in: 2013 46th Hawaii International Conference on System Sciences, 2013, pp. 3119–3128.
[93] J. Sixto, A. Almeida, D. López-de Ipia, Analysing customers sentiments: An approach to opinion mining and classification of online hotel reviews, in: Natural Language Processing and Information Systems, in: Lecture Notes in Computer Science, Springer, Berlin, Heidelberg, 2013, pp. 359–362.
[94] C.S. Namahoot, M. Brückner, N. Panawong, Context-aware tourism recommender system using temporal ontology and Nave Bayes, in: Recent Advances in Information and Communication Technology 2015, in: Advances in Intelligent Systems and Computing, Springer, Cham, 2015, pp. 183–194.
[95] Y. Zhao, S. Dong, J. Yang, Effect research of aspects extraction for chinese hotel reviews based on machine learning method, Int. J. Smart Home 9 (2015) 23–34.
[96] M. Arruza, J. Pericich, M. Straka, The automated travel agent: Hotel recommendations using machine learning, 2016.
[97] A. Ebadi, An Intelligent Hybrid Multi-Criteria Hotel Recommender System Using Explicit and Implicit Feedbacks (Master's thesis), Concordia University, 2016.

[98] Q. Li, Y. Wu, S. Wang, M. Lin, X. Feng, H. Wang, Vistravel: visualizing tourism network opinion from the user generated content, J. Vis. 19 (3) (2016) 489–502.

[99] K.E. Saputro, S.S. Kusumawardani, S. Fauziati, Development of semi-supervised named entity recognition to discover new tourism places, in: 2016 2nd International Conference on Science and Technology-Computer (ICST), 2016, pp. 124–128.

[100] Y. Yuan, J. Du, J.M. Lee, Tourism activity recognition and discovery based on improved LDA model, in: 2016 4th International Conference on Cloud Computing and Intelligence Systems (CCIS), 2016, pp. 447–455.

[101] G. Ren, T. Hong, Investigating online destination images using a topic-based sentiment analysis approach, Sustain.; Basel 9 (10) (2017).

[102] Belmonte-Fernández, R. Montoliu, J. Torres-Sospedra, E. Sansano-Sansano, D. Chia-Aguilar, A radiosity-based method to avoid calibration for indoor positioning systems, Expert Syst. Appl. 105 (2018) 89–101.

[103] C.-I. Hsu, M.-L. Shih, B.-W. Huang, B.-Y. Lin, C.-N. Lin, Predicting tourism loyalty using an integrated Bayesian network mechanism, Expert Syst. Appl. 36 (9) (2009) 11760–11763.

[104] Y. Guo, S.J. Barnes, Q. Jia, Mining meaning from online ratings and reviews: Tourist satisfactionanalysis using latent dirichletallocation, Tour. Manag. 59 (2017) 467–483.

[105] H. Yi-Chung, Predicting foreign tourists for the tourism industry using soft computing-based grey-Markov models, Sustain. Basel 9 (7) (2017).

[106] C. Haruechaiyasak, A. Kongthon, P. Palingoon, C. Sangkeettrakarn, Constructing thai opinion mining resource: A case study on hotel reviews, in: Proceedings of the Eighth Workshop on Asian Language Resouces, 2010, pp. 64–71.

[107] E. García-Barriocanal, M.-A. Sicilia, N. Korfiatis, Exploring hotel service quality experience indicators in user-generated content: a case using tripadvisor data, in: MCIS 2010 Proceedings, 2010.

[108] L. Bermingham, I. Lee, Spatio-temporal sequential pattern mining for tourism sciences, Procedia Comput. Sci. 29 (2014) 379–389.

[109] G. Li, R. Law, H.Q. Vu, J. Rong, X.R. Zhao, Identifying emerging hotel preferences using emerging pattern mining technique, Tour. Manag. 46 (2015) 311–321.

[110] Y. Ma, Z. Xiang, Q. Du, W. Fan, Effects of user-provided photos on hotel review helpfulness: An analytical approach with deep leaning, Int. J. Hosp. Manag. 71 (2018) 120–131.

[111] M.J. Sanchez-Franco, G. Cepeda-Carrion, J.L. Roldan, Understanding relationship quality in hospitality services a study based on text analytics and partial least squares, Internet Res. 29 (2019) 478–503.

[112] M.J. Sanchez-Franco, A. Navarro-Garcia, F. Javier Rondan-Cataluna, A naive Bayes strategy for classifying customer satisfaction: A study based on online reviews of hospitality services, J. Bus. Res. 101 (2019) 499–506.

[113] Q. Wu, R. Law, X. Xu, A sparse Gaussian process regression model for tourism demand forecasting in hong kong, Expert Syst. Appl. 39 (5) (2012) 4769–4774.

[114] W.S.C. Lei, C.C.C. Lam, Determinants of hotel occupancy rate in a chinese gaming destination, J. Hosp. Tour. Manag. 22 (2015) 1–9.

[115] J.G. Brida, B. Lanzilotta, L. Moreno, F. Santiaque, A non-linear approximation to the distribution of total expenditure distribution of cruise tourists in Uruguay, Tour. Manag. 69 (2018) 62–68.

[116] V. Athanasiou, M. Maragoudakis, Dealing with high dimensional sentiment data using gradient boosting machines, in: L. Iliadis, I. Maglogiannis (Eds.), Artificial Intelligence Applications and Innovations, Aiai 2016, Vol. 475, Springer-Verlag Berlin, Berlin, 2016, pp. 481–489, WOS:000392413700042.

[117] P.-F. Pai, K.-C. Hung, K.-P. Lin, Tourism demand forecasting using novel hybrid system, Expert Syst. Appl. 41 (8) (2014) 3691–3702.

[118] L. Shi-Ting, X. Bo, The application of improved SVM for data analysis in tourism economy, in: 2014 7th International Conference on Intelligent Computation Technology and Automation, 2014, pp. 769–772.

[119] Y. Yang, N.J. Mueller, R.R. Croes, Market accessibility and hotel prices in the caribbean: The moderating effect of quality-signaling factors, Tour. Manag. 56 (2016) 40–51.

[120] T.H. Zhang, H.W. Ji, Y. Hu, Q. Ye, Y. Lin, Application of classification algorithm of machine learning and buffer analysis in torism regional planning, ISPRS - Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. 42 (3) (2018) 2297–2302.

[121] C.-C. Hsu, C.-H. Wu, S.-C. Chen, K.-L. Peng, Dynamically optimizing parameters in support vector regression: An application of electricity load forecasting, in: Proceedings of the 39th Annual Hawaii International Conference on System Sciences (HICSS'06), Vol. 2, 2006, pp. 30c–30c.

[122] D. Romero Morales, J. Wang, Forecasting cancellation rates for services booking revenue management using data mining, European J. Oper. Res. 202 (2) (2010) 554–562.

[123] W.-C. Hong, Y. Dong, W.Y. Zhang, L.-Y. Chen, B. K. Panigrahi, Cyclic electric load forecasting by seasonal SVR with chaotic genetic algorithm, Int. J. Electr. Power Energy Syst. 44 (1) (2013) 604–614.

[124] K.-P. Lin, P.-F. Pai, Y.-M. Lu, P.-T. Chang, Revenue forecasting using a least-squares support vector regression model in a fuzzy environment, Inform. Sci. 220 (2013) 196–209.

[125] J. Wang, S. Jin, S. Qin, H. Jiang, Swarm intelligence-based hybrid models for short-term power load prediction, Math. Probl. Eng. (2014).

[126] G. Cao, L. Wu, Support vector regression with fruit fly optimization algorithm for seasonal electricity consumption forecasting, Energy 115 (2016) 734–745.

[127] G. Wang, Y. Su, L. Shu, One-day-ahead daily power forecasting of photovoltaic systems based on partial functional linear regression models, Renew. Energy 96 (2016) 469–478.

[128] Y. Chen, H. Tan, X. Song, Day-ahead forecasting of non-stationary electric power demand in commercial buildings: Hybrid support vector regression based, 8th International Conference on Applied Energy, ICAE2016, 8-11 October 2016, Beijing, China, Energy Procedia 105 (2017) 2101–2106.

[129] H. Li, J. Sun, Forecasting business failure: The use of nearest-neighbour support vectors and correcting imbalanced samples – evidence from the chinese hotel industry, Tour. Manag. 33 (3) (2012) 622–634.

[130] W. Li, S. Xu, W. Meng, A support vector machines method for tourist satisfaction degree evaluation, in: 2009 6th International Conference on Service Systems and Service Management, 2009, pp. 883–887.

[131] H. Xia, L. Peng, SVM-based comments classification and mining of virtual community: for case of sentiment classification of hotel reviews - semantic scholar, 2009.

[132] W. Zheng, Q. Ye, Sentiment classification of chinese traveler reviews by support vector machine algorithm, in: 2009 Third International Symposium on Intelligent Information Technology Application, Vol. 3, 2009, pp. 335–338.

[133] C.-j. Lin, P.-h. Chao, Tourism-related opinion detection and tourist-attraction target identification, Comput. Linguist. Chin. Lang. Process. 15 (2010).

[134] H.X. Shi, X.J. Li, A sentiment analysis model for hotel reviews based on supervised learning, in: 2011 International Conference on Machine Learning and Cybernetics, Vol. 3, 2011, pp. 950–954.

[135] J. Yao, H. Wang, P. Yin, Sentiment feature identification from chinese online reviews, in: Advances in Information Technology and Education, in: Communications in Computer and Information Science, Springer, Berlin, Heidelberg, 2011, pp. 315–322.

[136] W. Kasper, M. Vela, Sentiment analysis for hotel reviews, in: Proceedings of the Computational Linguistics-Applications Conference, Vol. 4, 2012, pp. 45–52.

[137] Q. Lu, L. Xiao, Q. Ye, Investigating the impact of online word-of-mouth on hotel sales with panel data, in: 2012 International Conference on Management Science Engineering 19th Annual Conference Proceedings, 2012, pp. 3–9.

[138] M. Tokuhisa, H. Shahana, M. Murata, J. Murakami, An active learning based support tool for extracting hints of tourism development from blog articles, in: 2012 IIAI International Conference on Advanced Applied Informatics, 2012, pp. 103–107.

[139] K. Jiang, H. Yin, P. Wang, N. Yu, Learning from contextual information of geo-tagged web photos to rank personalized tourism attractions, Neurocomputing 119 (2013) 17–25.

[140] H.Y. Hsieh, V. Klyuev, Q. Zhao, S.H. Wu, SVR-based outlier detection and its application to hotel ranking, in: 2014 IEEE 6th International Conference on Awareness Science and Technology (ICAST), 2014, pp. 1–6.

[141] C. Chiu, N.-H. Chiu, R.-J. Sung, P.-Y. Hsieh, Opinion mining of hotel customer-generated contents in chinese weblogs, Curr. Issues Tour. 18 (5) (2015) 477–495.

[142] A. Dickinger, J.A. Mazanec, Significant word items in hotel guest reviews: A feature extraction approach, Tour. Recreat. Res. 40 (3) (2015) 353–363.

[143] Y. Sun, H. Fan, M. Bakillah, A. Zipf, Road-based travel recommendation using geo-tagged images, Comput. Environ. Urban Syst. 53 (2015) 110–122.

[144] Z. Xiang, Q. Du, Y. Ma, W. Fan, A comparative analysis of major online review platforms: Implications for social media analytics in hospitality and tourism, Tour. Manag. 58 (2017) 51–65.

[145] E. Martin-Fuentes, C. Fernandez, C. Mateu, E. Marine-Roig, Modelling a grading scheme for peer-to-peer accommodation: Stars for airbnb, Int. J. Hosp. Manag. 69 (2018) 75–83.

[146] E. Gawlik, H. Kabaria, E. Kaur, Predicting tourism trends with google insights, 2011.

[147] P. Kumar Duvvur, C.J. Romanowski, Predicting hotel rating based on user reviews, 2016.

[148] X. Xu, R. Law, T. Wu, Support vector machines with manifold learning and probabilistic space projection for tourist expenditure analysis, Int. J. Comput. Intell. Syst. 2 (1) (2009) 17–26.

[149] Z.-j. Cai, S. Lu, X.-b. Zhang, Tourism demand forecasting by support vector regression and genetic algorithm, in: 2009 2nd IEEE International Conference on Computer Science and Information Technology, 2009, pp. 144–146.

[150] W.-C. Hong, Y. Dong, L.-Y. Chen, S.-Y. Wei, SVR with hybrid chaotic genetic algorithms for tourism demand forecasting, Appl. Soft Comput. 11 (2) (2011) 1881–1890.
[151] S. Cankurt, A. Subai, Tourism demand modelling and forecasting using data mining techniques in multivariate time series: a case study in Turkey, Turk. J. Electr. Eng. Comput. Sci. 24 (5) (2016) 3388–3404.
[152] R. Chen, C.-Y. Liang, W.-C. Hong, D.-X. Gu, Forecasting holiday daily tourist flow based on seasonal support vector regression with adaptive genetic algorithm, Appl. Soft Comput. 26 (2015) 435–443.
[153] N. Antonio, A. De Almeida, L. Nunes, Using data science to predict hotel booking cancellations, in: Handbook of Research on Holistic Optimization Techniques in the Hospitality, Tourism, and Travel Industry, IGI Global, 2016, pp. 140–166.
[154] W. Lijuan, C. Guohua, Seasonal SVR with FOA algorithm for single-step and multi-step ahead forecasting in monthly inbound tourist flow, Knowl.-Based Syst. 110 (2016) 157–166.
[155] C. Liang, W. Bi, Seasonal variation analysis and SVR forecast of tourist flows during the year: A case study of huangshan mountain, in: 2017 IEEE 2nd International Conference on Big Data Analysis (ICBDA), IEEE, Beijing, China, 2017, pp. 921–927.
[156] B. Zhang, X. Huang, N. Li, R. Law, A novel hybrid model for tourist volume forecasting incorporating search engine data, Asia Pac. J. Tour. Res. 22 (3) (2017) 245–254.
[157] K.-Y. Chen, C.-H. Wang, Support vector regression with genetic algorithms in forecasting tourism demand, Tour. Manag. 28 (1) (2007) 215–226.
[158] R. Ali, A. Shabri, N. Arifin, Y. Suhaila, A wavelet support vector machine combination model for Singapore tourist arrival to Malaysia, IOP Conf. Ser.: Mater. Sci. Eng. 226 (2017) 012077.
[159] S. Liu, E. Yao, Holiday passenger flow forecasting based on the modified least-square support vector machine for the metro system, Transp. Eng. Part A: Syst. 143 (2017) 04016005.
[160] B. Gokaraju, S. S. Durbha, R. King, N. Younan, A Machine Learning Based Spatio-Temporal Data Mining Approach for Detection of Harmful Algal Blooms in the Gulf of Mexico, in: IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 4, IEEE, 2011.
[161] X. Xu, R. Law, W. Chen, L. Tang, Forecasting tourism demand by extracting fuzzy Takagi–Sugeno rules from trained SVMs, CAAI Trans. Intell. Technol. 1 (1) (2016) 30–42.
[162] F. Xiao, Forest fire disaster area prediction based on genetic algorithm and support vector machine, Adv. Mater. Res. 446–449 (2012) 3037–3041.
[163] C.-T. Lin, Y.-L. Huang, Mining tourist imagery to construct destination image position model, Expert Syst. Appl. 36 (2, Part 1) (2009) 2513–2524.
[164] S.-h. Liao, Y.-J. Chen, M.-y. Deng, Mining customer knowledge for tourism new product development and customer relationship management, Expert Syst. Appl. 37 (6) (2010) 4212–4223.
[165] A. Pitman, M. Zanker, M. Fuchs, M. Lexhagen, Web usage mining in tourism - A query term analysis and clustering approach, in: Proceedings of the International Conference in Lugano, Switzerland, 2010, pp. 393–403.
[166] H.Q. Vu, G. Li, R. Law, B.H. Ye, Exploring the travel behaviors of inbound tourists to hong kong using geotagged photos, Tour. Manag. 46 (2015) 222–232.
[167] S.J. Miah, H.Q. Vu, J. Gammack, M. McGrath, A big data analytics method for tourist behaviour analysis, Inf. Manag. 54 (6) (2017) 771–785.
[168] D. Gavalas, M. Kenteris, A web-based pervasive recommendation system for mobile tourist guides, Pers. Ubiquitous Comput. 15 (7) (2011) 759–770.
[169] K. Jiang, P. Wang, N. Yu, Contextrank: Personalized tourism recommendation by exploiting context information of geotagged web photos, in: 2011 Sixth International Conference on Image and Graphics, 2011, pp. 931–937.
[170] X. Peng, Z. Huang, A novel popular tourist attraction discovering approach based on geo-tagged social media big data, ISPRS Int. J. Geo-Inf. 6 (7) (2017) 216.
[171] J. Shahrabi, E. Hadavandi, S. Asadi, Developing a hybrid intelligent model for forecasting problems: Case study of tourism demand time series, Knowl.-Based Syst. 43 (2013) 112–122.
[172] D. Junping, Z. Min, T. Xuyan, The realization of distributed sampling association rule mining algorithm in tourism, in: 2008 7th World Congress on Intelligent Control and Automation, 2008, pp. 183–187.
[173] S. Liu, R. Law, J. Rong, G. Li, J. Hall, Analyzing changes in hotel customers expectations by trip mode, Int. J. Hosp. Manag. 34 (2013) 359–371.
[174] J.P. Lucas, N. Luz, M.N. Moreno, R. Anacleto, A. Almeida Figueiredo, C. Martins, A hybrid recommendation approach for a tourism system, Expert Syst. Appl. 40 (9) (2013) 3532–3550.
[175] M. Versichele, L. de Groote, M. Claeys Bouuaert, T. Neutens, I. Moerman, N. Van de Weghe, Pattern mining in tourist attraction visits through association rule learning on bluetooth tracking data: A case study of ghent, Belgium, Tour. Manag. 44 (2014) 67–81.
[176] R.P. Biuk-Aghai, S. Fong, Y.W. Si, Design of a recommender system for mobile tourism multimedia selection, in: 2008 2nd International Conference on Internet Multimedia Services Architecture and Applications, 2008, pp. 1–6.
[177] M. Afzaal, M. Usman, A.C.M. Fong, S. Fong, Y. Zhuang, Fuzzy aspect based opinion classification system for mining tourist reviews, Adv. Fuzzy Syst. (2016).
[178] S. Sakhuja, V. Jain, S. Kumar, C. Chandra, S.K. Ghildayal, Genetic algorithm based fuzzy time series tourism demand forecast model, Ind. Manag. Data Syst. 116 (3) (2016) 483–507.
[179] J.H. Chen, K.M. Chao, N. Shah, Hybrid recommendation system for tourism, in: 2013 IEEE 10th International Conference on E-Business Engineering, 2013, pp. 156–161.
[180] E. Hadavandi, A. Ghanbari, K. Shahanaghi, S. Abbasian-Naghneh, Tourist arrival forecasting by evolutionary fuzzy systems, Tour. Manag. 32 (5) (2011) 1196–1203.
[181] C. Zhang, J. Zhang, Mining tourist preferences with twice-learning, in: New Frontiers in Applied Data Mining, in: Lecture Notes in Computer Science, Springer, Berlin, Heidelberg, 2011, pp. 483–493.
[182] B. Liu, Sentiment analysis and opinion mining, Synth. Lect. Hum. Lang. Technol. 5 (1) (2012) 1–167.
[183] V. Taecharungroj, B. Mathayomchan, Analysing tripadvisor reviews of tourist attractions in phuket, thailand, Tour. Manag. 75 (2019) 550–568.
[184] M. Martinez-Torres, S. Toral, A machine learning approach for the identification of the deceptive reviews in the hospitality sector using unique attributes and sentiment orientation, Tour. Manag. 75 (2019) 393–403.
[185] J.J. Sanchez-Medina, J.A. Guerra-Montenegro, D. Sanchez-Rodriguez, I.G. Alonso-Gonzalez, J.L. Navarro-Mesa, Data stream mining applied to maximum wind forecasting in the canary islands, Sensors 19 (10) (2019).
[186] D. Gunning, D.W. Aha, Darpa's explainable artificial intelligence program, AI Mag. 40 (2) (2019) 44–58.
[187] A.B. Arrieta, N. Díaz-Rodríguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, S. García, S. Gil-López, D. Molina, R. Benjamins, et al., Explainable artificial intelligence (xai): Concepts, taxonomies, opportunities and challenges toward responsible ai, Inf. Fusion 58 (2020) 82–115.
[188] S.J. Pan, Q. Yang, A survey on transfer learning, IEEE Trans. Knowl. Data Eng. 22 (10) (2010) 1345–1359.
[189] P. Drucker, Managing in the Next Society, Routledge, 2012.
[190] G. Kebede, Knowledge management: An information science perspective, Int. J. Inf. Manage. 30 (5) (2010) 416–424.
[191] F. Okumus, Facilitating knowledge management through information technology in hospitality organizations, J. Hosp. Tour. Technol. 4 (1) (2013) 64–80.
[192] B. Krawczyk, L.L. Minku, J. Gama, J. Stefanowski, M. Woniak, Ensemble learning for data stream analysis: A survey, Inf. Fusion 37 (2017) 132–156.
[193] I. Žliobaitė, A. Bifet, B. Pfahringer, G. Holmes, Active learning with drifting streaming data, IEEE Trans. Neural Netw. Learn. Syst. 25 (1) (2014) 27–39.
[194] J.L. Lobo, J.D. Ser, A. Bifet, N. Kasabov, Spiking neural networks and online learning: An overview and perspectives, Neural Netw. 121 (2020) 88–100.
[195] W. Gerstner, W.M. Kistler, Spiking Neuron Models: Single Neurons, Populations, Plasticity, Cambridge university press, 2002.
[196] R.L. Rivest, L. Adleman, M.L. Dertouzos, On Data Banks and Privacy Homomorphisms, Foundations of Secure Computation, Academia Press, 1978, pp. 169–179.
[197] C. Gentry, D. Boneh, A Fully Homomorphic Encryption Scheme, Vol. 20, Stanford University Stanford, 2009.
[198] J. Fan, F. Vercauteren, Somewhat practical fully homomorphic encryption., IACR Cryptol. ePrint Arch. 2012 (2012) 144.
[199] Z. Brakerski, C. Gentry, V. Vaikuntanathan, (Leveled) fully homomorphic encryption without bootstrapping, ACM Trans. Comput. Theory 6 (3) (2014) 13:1–13:36.
[200] O. Regev, The learning with errors problem, Invited Surv. CCC 7 (2010).
[201] J. Konen, H.B. McMahan, F.X. Yu, P. Richtárik, A.T. Suresh, D. Bacon, Federated learning: Strategies for improving communication efficiency, 2016, arXiv:arXiv:1610.05492.
[202] R.C. Geyer, T. Klein, M. Nabi, Differentially private federated learning: A client level perspective, 2017, arXiv:arXiv:1712.07557.
[203] G. Widmer, M. Kubat, Learning in the presence of concept drift and hidden contexts, Mach. Learn. 23 (1) (1996) 69–101.
[204] D. Braess, Über ein Paradoxon aus der Verkehrsplanung, Unternehmensforschung 12 (1) (1968) 258–268.