



UNIVERSIDAD DE LAS PALMAS DE GRAN CANARIA
Escuela de Ingeniería Informática



Análisis automático de los lanzamientos de falta en el ámbito futbolístico

Grado en Ingeniería Informática

AUTOR: Javier Torón Artiles

TUTORIZADO POR:

David Sebastián Freire Obregón

José Daniel Hernández Sosa

Julio 2023

Resumen

Este proyecto aborda la realización de análisis predictivos y estadísticos acerca de una acción técnica muy habitual en el ámbito futbolístico como son los lanzamientos de falta directa.

Su punto de origen radica en la creación de una base de datos a partir de la recopilación de un conjunto de datos no estructurados como son los vídeos de dicho lance del juego. Tras un proceso de estudio y observación de los mismos, se obtienen unos datos estructurados organizados en tablas que forman el etiquetado de los atributos de los videos. Estos detallan características como la posición de la falta en el campo, la posición del jugador, el pie de golpeo, datos sobre la barrera, por donde entra el balón en la portería, por dónde se lanza el portero, etc. La obtención de los datos tabulares es crucial para el fin que se persigue que no es otro que el análisis estadístico y la predicción del resultado final de la acción aplicando un modelo de deep learning, apoyado en algoritmos de clasificación.

Gracias a la información detallada que proporcionan las diferentes estadísticas y pronósticos recabados, se contribuye a la mejora de otros aspectos del análisis del fútbol, como la detección de patrones en los movimientos de los jugadores y la toma de decisiones técnicas y tácticas.

Abstract

This project addresses the realization of predictive and statistical analysis concerning a technical and very common action of football such as direct free kicks.

Its starting point lies in the creation of a database by collecting a set of non-structured data just as the videos of these game actions are. After the process of studying and observing these, structured data organized in tables is obtained which will become the labelling of the attributes of the videos. These detail characteristics such as the position of the foul in the field, the position of the player, the striking foot, data about the wall, the side of the goal to which the ball is directed, to where the goalkeeper launches, etc. Collecting these tabular data is crucial for achieving the purpose established which is not other than the statistical analysis and the prediction of the final outcome of the action applying a deep learning model, supported by classifying algorithms.

Thanks to the detailed information that the different statistics and forecasts gathered will provide, a contribution is made to improving other aspects of football analysis, such as the detection of patterns in the movements of players and technical and tactical decision making.

Índice general

Capítulo 1 - Introducción	1
1.1 - Contexto general del trabajo	2
1.2 - Motivación	3
1.3 - Conceptos básicos	4
1.4 - Estructura del documento	6
Capítulo 2 - Estado actual y objetivos iniciales	8
2.1 – Herramientas existentes en el mercado	8
2.2 - Objetivos	11
Capítulo 3 - Competencias específicas y aportaciones del trabajo	12
3.1 - Competencias	12
3.2 - Aportaciones	13
Capítulo 4 - Desarrollo	14
4.1 - Herramientas utilizadas:	14
4.1.1 - Easy Youtube Video Downloader Express	14
4.1.2 - ClipChamp	14
4.1.3 - Excel	15
4.1.4 - Python	15
4.1.5 - Google Colab	15
4.1.6 - Librería Scikit-learn	16
4.2 - Recopilación y descarga de vídeos	16
4.3 - Recorte y etiquetado de vídeos	16
4.3.1 - Categorías	17
4.3.1.1 - Fuente del vídeo	17
4.3.1.2 - Nombre del vídeo	18
4.3.1.3 - Lado del campo	20
4.3.1.4 - Lado de la falta	21
4.3.1.5 - Distancia de la falta	22
4.3.1.6 - Género de la competición	23
4.3.1.7 - Nombre del lanzador	24

ÍNDICE GENERAL

4.3.1.8 - Pierna con la que tira el lanzador-----	24
4.3.1.9 - Dirección a la que va el balón -----	24
4.3.1.10 - Nombre del portero-----	25
4.3.1.11 - Posición del portero -----	26
4.3.1.12 - Hacia qué lado se tira el portero -----	26
4.3.1.13 - Composición de la barrera-----	27
4.3.1.14 - Resultado final -----	27
4.3.2 - Etiquetado resultante -----	27
4.4 - Estadísticas -----	30
4.5 - Clasificadores -----	31
Capítulo 5 – Resultados obtenidos -----	37
5.1 - Estadísticas -----	37
5.1.1 - Goles-----	38
5.1.2 - Paradas -----	49
5.2 - Clasificadores -----	58
5.2.1 - Resultados iniciales-----	58
5.2.2 - Resultados obtenidos al modificar datos de entrada -----	59
5.2.3 - Función GridSearchCV -----	65
Capítulo 6 - Conclusiones y trabajo futuro -----	68
Conclusiones-----	68
El análisis estadístico -----	68
La predicción de resultados -----	70
Futuros trabajos -----	72
Bibliografía-----	75

Índice de figuras

Figura 1.1: Ejemplo de una red neuronal con diferentes capas. -----	4
Figura 2.1: Ejemplo de la probabilidad de gol calculada por Microsoft. -----	9
Figura 4.1: Ejemplo de la visión diagonal. -----	18
Figura 4.2: Ejemplo de la visión trasera. -----	19
Figura 4.3: Ejemplo de la visión delantera. -----	19
Figura 4.4: Ejemplo de la visión lateral. -----	20
Figura 4.5: Valor que se asignará a cada mitad del campo. -----	21
Figura 4.6: Valor que se asignará a los diferentes lados donde esté ubicada la falta.---	22
Figura 4.7: Valor que se asignará dependiendo si la falta es cercana o lejana. -----	23
Figura 4.8: Valor que se asignará dependiendo del lado por el que vaya el balón. -----	25
Figura 4.9: Valor que se asignará dependiendo de la posición del portero. -----	26
Figura 5.1: Porcentaje de los lanzamientos que acaban en gol en base a la dirección de la falta. -----	40
Figura 5.2: Porcentaje de los lanzamientos que acaban en gol en base al lado de la falta. -----	41
Figura 5.3: Porcentaje de los lanzamientos que acaban en gol en base a la distancia de la falta. -----	42
Figura 5.4: Porcentaje de los lanzamientos que acaban en gol en base a la ubicación y distancia de la falta. -----	44
Figura 5.5: Porcentaje de los lanzamientos de jugadores zurdos que acaban en gol en base a la ubicación y distancia de la falta. -----	46
Figura 5.6: Porcentaje de los lanzamientos de jugadores zurdos que acaban en gol en base a la ubicación y distancia de la falta. -----	48
Figura 5.7: Porcentaje de los lanzamientos que acaban en parada en base a la dirección de la falta. -----	50
Figura 5.8: Porcentaje de los lanzamientos que acaban en parada en base al lado de la falta. -----	51

ÍNDICE DE FIGURAS

Figura 5.9: Porcentaje de los lanzamientos que acaban en parada en base a la distancia de la falta. ----- 52

Figura 5.10: Porcentaje de los lanzamientos que acaban en parada en base a la ubicación y distancia de la falta. ----- 54

Figura 5.11: Porcentaje de los lanzamientos que acaban en parada en base a la colocación inicial del portero. ----- 56

Figura 5.12: Porcentaje de los lanzamientos que acaban en parada en base a la dirección en la que se lanza el portero. ----- 57

Índice de tablas

Tabla 4.1: Ejemplo del etiquetado manual de 12 videos diferentes. -----	28
Tabla 4.2: Ejemplo del etiquetado de la predicción de gol después de las modificaciones y una vez exportado a csv.-----	29
Tabla 5.1: Distribución de goles y fallos, y sus respectivos porcentajes.-----	37
Tabla 5.2: Tipos de fallos y las veces que se dio cada uno de estos con sus respectivos porcentajes. -----	38
Tabla 5.3: Goles de falta basado en la dirección de esta. -----	39
Tabla 5.4: Goles de falta basado en la ubicación de esta. -----	40
Tabla 5.5: Goles de falta basado en la distancia de esta. -----	41
Tabla 5.6: Goles de falta basado en la ubicación y distancia de esta.-----	43
Tabla 5.7: Goles de falta dependiendo de la pierna de golpeo del jugador. -----	44
Tabla 5.8: Goles de falta de jugadores zurdos basado en la ubicación y distancia de esta. -----	45
Tabla 5.9: Goles de falta de jugadores diestros basado en la ubicación y distancia de esta. -----	47
Tabla 5.10: Paradas de falta basado en la dirección de esta. -----	49
Tabla 5.11: Paradas de falta basado en la ubicación de esta.-----	50
Tabla 5.12: Paradas de falta basado en la distancia de esta.-----	51
Tabla 5.13: Paradas de falta basado en la ubicación y distancia de esta. -----	53
Tabla 5.14: Paradas de falta basado en la colocación inicial del portero dentro de la portería. -----	55
Tabla 5.15: Paradas de falta basado en la dirección en la que se lanza el portero. -----	56
Tabla 5.16: Medias y desviaciones estándar obtenidas de las predicciones de los goles y las paradas, resultados sobre 1. -----	59
Tabla 5.17: Medias y desviaciones estándar obtenidas de las predicciones de los goles, resultado final sobre 1. -----	61

ÍNDICE DE TABLAS

Tabla 5.18: Medias y desviaciones estándar obtenidas de las predicciones de las paradas, resultado final sobre 1. ----- 62

Tabla 5.19: Resultados óptimos de cada clasificador tras utilizar la función “GridSearchCV”. ----- 66

Tabla 5.20: Comparación de la media de los resultados de la predicción de goles usando los hiperparámetros devueltos por la función “GridSearchCV”. ----- 66

Tabla 5.21: Comparación de la media de los resultados de la predicción de paradas usando los hiperparámetros devueltos por la función “GridSearchCV”. ----- 67

Índice de algoritmos

Algoritmo 4.1: Obtención de la distribución de los goles y de los fallos. -----	30
Algoritmo 4.2: Separación de los datos del etiquetado de los goles en los grupos de valores y de objetivos, y la inicialización de los clasificadores. -----	33
Algoritmo 4.3: Fragmento del bucle de las 50 iteraciones para obtener las predicciones y las precisiones de cada clasificador. -----	34
Algoritmo 4.4: Fragmento de la extracción de las anomalías junto con la obtención de las medias y desviaciones estándar de las precisiones calculadas en el bucle. -----	35
Algoritmo 4.5: Fragmento que muestra el uso de la función “GridSearchCV” al clasificador MLP. -----	36

Capítulo 1 - Introducción

Poco a poco, las nuevas tecnologías se han ido implementando y utilizando en las diferentes prácticas deportivas. Dicha integración de las tecnologías en el deporte ha aportado grandes beneficios en la mejora del rendimiento de los deportistas, y parece que esta tendencia está en auge actualmente, ya que cada cierto tiempo, aparecen nuevos usos y nuevas funcionalidades que aportan beneficios a dichas prácticas. Dichos avances van desde explotar el potencial de los atletas, hasta facilitar la recuperación de estos. No únicamente beneficia el progreso del deportista como individuo, sino que, además, pueden servir para ayudar a los equipos a decidir qué metodologías, tipos de entrenamientos o tácticas son las más adecuadas o convenientes dependiendo de las diferentes situaciones o circunstancias que pueden ir surgiendo. También proporciona ciertas mejoras de las competiciones como tal, ya que pueden apoyar a la organización y al desarrollo de estas.

Este trabajo se enfocará exclusivamente en el uso de las tecnologías en el fútbol. El fútbol es un deporte conocido mundialmente, y no solamente conocido, sino que es practicado mundialmente. Es uno de los deportes con más espectadores y seguidores, y que mayor impacto tiene en la sociedad de hoy en día.

Durante los últimos años, el fútbol ha ido incorporando diferentes tecnologías para proporcionar mejoras en el desarrollo de las diferentes competiciones. Un ejemplo de una de las tecnologías más importantes implementadas recientemente es el uso del VAR. En la nomenclatura futbolística, el VAR es un sistema de videoarbitraje cuyo objetivo es ayudar al árbitro a resolver posibles dudas y conflictos con algunos lances del juego. Dicho sistema entra en juego únicamente si suceden cuatro situaciones que pueden ser decisivas en el resultado final de los partidos [1]. El VAR hace referencia a la utilización de distintas cámaras que graban las jugadas desde diferentes tomas y ángulos, para que tres árbitros, que estarán situados en un cuarto con diferentes monitores, tomen decisiones en base a las grabaciones realizadas desde los distintos ángulos.

A pesar de que el VAR es quizás el ejemplo de implementación de las tecnologías más conocido actualmente en relación con el fútbol, también existen muchas otras implementaciones de estas. Como por ejemplo, el sistema de seguimiento de jugadores,

mediante el cual se obtienen estadísticas y datos de los atletas como puede ser la distancia recorrida por un jugador. Otro caso en donde se ha hecho uso de las tecnologías es de cara a ayudar al árbitro con respecto a los denominados goles fantasmas [2]. Estos son aquellos tiros que parece que no han terminado de pasar la línea de meta pero que en realidad sí lo han hecho, o aquellos, que parece que si han sobrepasado la línea, pero en realidad no. Para estas situaciones, en cuanto el balón supere la línea de meta, el árbitro recibirá una señal en su reloj, que le indicará que el esférico ha entrado por completo en la portería, y, por ende, ha sido gol.

1.1 - Contexto general del trabajo

Como se ha comentado anteriormente, esta propuesta de TFG se enfocará en el uso de las nuevas tecnologías en el fútbol; más específicamente, en relación con los tiros de faltas directas, un lance muy recurrente en dicho deporte.

Se hará uso de un modelo de deep learning para la predicción del resultado de los libres directos utilizando los datos tabulares descriptivos de la acción. Esto quiere decir, que se intentará predecir si el lanzamiento de falta directa acabará en gol o si la parará el portero. No solamente se intentará pronosticar el resultado final de un lanzamiento de falta directa, sino que, además, se realizará un análisis estadístico de dicho lance del juego, herramientas que puede servir de apoyo a lo que se conoce como ‘Sports Analytics’.

Para llevar a cabo dicha predicción, se utilizará como referencia los datos obtenidos a raíz de observar y caracterizar múltiples lanzamientos de faltas; para así procurar determinar algún aspecto descriptivo de la acción a partir de los datos tabulares recopilados. Es decir, que basado en observaciones del pasado se intentará pronosticar resultados del futuro. Por lo tanto, se deberán analizar y estudiar lanzamientos de falta directa, para, basado en la caracterización de ciertos atributos y categorías predefinidas, poder obtener cualquier tipo de información, o patrones, que puedan servir como identificativos en los cuales basar los pronósticos.

1.2 - Motivación

El reconocimiento de acciones es una práctica habitual en la vida cotidiana de las personas de hoy en día, y en muchas ocasiones se hace inclusive inconscientemente. Los humanos realizan dicho reconocimiento de acciones de forma inconsciente debido a que se basan en sus experiencias y observaciones pasadas y en los gestos y movimientos que realiza la persona o ser vivo que está siendo examinado. Dichos atributos no son únicamente utilizados para tratar de identificar qué acción está siendo realizada, sino que además son aplicados a la hora de tratar de predecir movimientos futuros.

El pronóstico de futuros eventos es otro comportamiento el cual se realiza muy a menudo en la sociedad actual y que tiene un gran impacto en ciertos sectores y hábitos de la vida, pudiendo incluso desencadenar enormes beneficios o pérdidas económicas. Un claro ejemplo de esto es en el ámbito de las inversiones. Todo accionista querría poder predecir cuándo es el mejor momento para realizar la compra/venta de acciones. Los pronósticos se encuentran también muy presentes en el mundo del deporte, y un claro ejemplo de esto son las apuestas deportivas. A todo aquel que realiza apuestas, le encantaría poder tener la certeza de que acertará con su apuesta, para así poder obtener mayores beneficios y menos pérdidas.

La predicción de futuros acontecimientos no es utilizada únicamente con fines lucrativos. Todos hemos escuchado alguna vez a alguien intentar adivinar qué equipo será el ganador de un partido, y no solo eso, sino cuál va a ser el equipo ganador de una competición específica. Al igual que se pronostica con el resultado de encuentros deportivos, también se intenta estimar cómo acabarán ciertos lances del juego. Cuántas veces se ha escuchado en una tanda de penaltis, que un determinado jugador que va a lanzar, lo va a fallar o marcar; y no solo eso, sino, cuántas veces se ha escuchado que un determinado jugador lo va a tirar hacia la izquierda o hacia la derecha, o que lo va a tirar fuera, etc.

Esto demuestra, que el reconocimiento de acciones y la predicción de movimientos se encuentran actualmente muy presentes en una gran cantidad de situaciones de la vida cotidiana. Debido a esta gran trascendencia, que tienen las predicciones en la actualidad, se ha considerado que un trabajo en este ámbito puede resultar muy interesante y de gran utilidad.

1.3 - Conceptos básicos

A continuación, se detallan y describen algunos de los conceptos básicos más importantes que son necesarios entender para la correcta comprensión de dicho trabajo.

- Los lanzamientos de faltas directas o también conocidos como libres directos. Son aquellas faltas en las que los jugadores disparan a portería. Es decir, que no se la pasan a ningún compañero ni la intentan centrar. Por tanto, dicho trabajo se centrará única y exclusivamente en aquellos lanzamientos de faltas directas, los que sean golpes a puerta.
- Sports Analytics es el término utilizado que hace referencia a la colección y al uso de datos con los cuales realizar análisis estadísticos avanzados en el ámbito del deporte [3]. Mediante el proceso de dichos análisis estadísticos se propone utilizar modelos matemáticos [4] para investigar el rendimiento de los atletas y de los equipos, entre otros, para intentar obtener ventajas competitivas.
- Las redes neuronales artificiales (conocidas como redes neuronales) intentan simular el comportamiento del cerebro humano; y son aquellas que permiten a una computadora procesar los datos y la información de una forma similar a los humanos. Por lo tanto, hace uso de nodos o neuronas conectadas para formar una estructura que cuenta con diversas capas por las que pasarán los datos [5]. La figura 1.1 muestra un ejemplo de diferentes capas de una red neuronal.

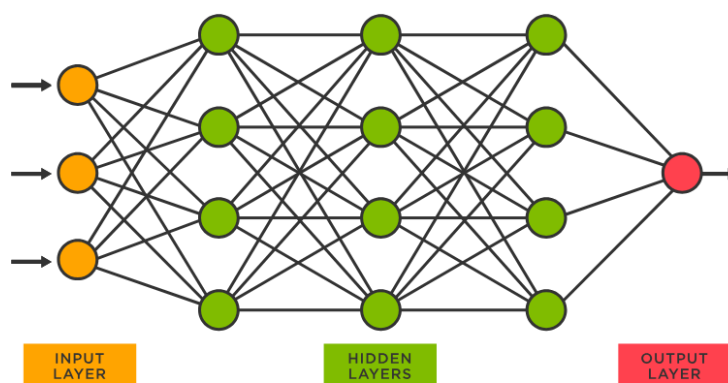


Figura 1.1: Ejemplo de una red neuronal con diferentes capas.

Imagen obtenida de: ¿Qué es una red neuronal? | TIBCO Software

- El aprendizaje automático o machine learning es un tipo de Inteligencia Artificial en donde los computadores tienen la capacidad de aprender sin la necesidad de programas con instrucciones explícitas. El machine learning se basa en el procesamiento de datos masivos para intentar identificar patrones o características para, por ejemplo, predecir valores futuros o agrupar los datos según estos patrones identificados [6].

Para realizar dicho aprendizaje, se necesitan dos fases, la fase de entrenamiento, en la cual se debe proporcionar los datos de entrada para que el modelo busque en ella los patrones; y la fase de pruebas o verificación, en donde el modelo proporciona el resultado final, la predicción de los valores futuros, por ejemplo, y se evalúa su precisión [6]. Debido a que se necesitan dos fases, antes de comenzar con el entrenamiento, se deberá dividir los datos para que en dichas fases no se utilice la misma información. Normalmente, se destina una mayor cantidad de datos para la fase de entrenamiento que para la fase de verificación, la distribución dependerá del caso en particular del cual se trate, pero habitualmente se reparte con un 70% de los datos para entrenar y el restante 30% para verificar.

- El aprendizaje automático cuenta con dos enfoques principales, el aprendizaje supervisado y el no supervisado. El aprendizaje supervisado es aquel en el cual los datos de entrada deben estar correctamente etiquetados y el resultado que proporcionará será específico y este estará ya predefinido. Para ello, será necesario que se lleve a cabo el etiquetado de los datos que va a utilizar el modelo antes de iniciar el aprendizaje. El aprendizaje no supervisado no necesita tanta intervención humana, ya que no es necesario ni el etiquetado previo de los datos, ni que el resultado esté previamente predefinido [7].
- El aprendizaje profundo o el deep learning es una red neuronal de tres o más capas de nodos, y es un subconjunto de machine learning. El aprendizaje profundo necesita menos intervención humana que el aprendizaje automático, debido a que con machine learning, lo habitual es que el humano es quien le debe indicar a la computadora la estructura y cuáles son las características identificadoras de los datos que se están estudiando (el etiquetado de los datos). Mientras que en deep learning no se necesita que los datos hayan sido procesados previamente, ya que,

los datos no tienen por qué estar estructurados, y es la propia red neuronal la que puede identificar cuáles son las características o los atributos identificativos. Es decir, que en el aprendizaje profundo los modelos pueden procesar texto o imágenes no estructuradas, a diferencia del aprendizaje automático [8].

- Un clasificador es un algoritmo que tiene como objetivo intentar categorizar o determinar una etiqueta a los datos que recibe. Por lo que un modelo de clasificación, después de la fase de entrenamiento, en la cual ha extraído las conclusiones y los patrones necesarios para realizar los pronósticos, es aquel que a partir de los nuevos datos de entrada predice la etiqueta o categoría de estos [9].

1.4 - Estructura del documento

El trabajo está dividido en capítulos, en los cuales, cada uno de estos, indaga en las diferentes áreas que abarca el trabajo. Dichos capítulos son los siguientes:

- Capítulo 1 - Introducción: Indica brevemente el contexto general del trabajo y lo que este va a abarcar, junto con la importancia y el uso de dicho ámbito en la sociedad actual. Además, describe los conceptos básicos para ayudar al entendimiento del trabajo y detalla la estructura de este.
- Capítulo 2 - Estado actual y objetivos iniciales: Ilustra cuáles son los objetivos que abarca este trabajo y aquellos establecidos inicialmente. Además, proporciona una imagen general del mercado actual relacionado con la temática del trabajo y aquellos trabajos o proyectos similares que se encuentran ya publicados.
- Capítulo 3 - Competencias específicas y aportaciones del trabajo: Indica cuales son las competencias desarrolladas y como han sido estas cubiertas, y las posibles aportaciones que puede tener este trabajo en la sociedad.
- Capítulo 4 - Desarrollo: Detalla los pasos llevados a cabo durante el trabajo, como se han realizado estos, y las herramientas y los recursos utilizados.

- Capítulo 5 - Resultados obtenidos: Muestra y describe los resultados obtenidos mediante el uso de los diferentes clasificadores y los diferentes análisis estadísticos realizados.
- Capítulo 6 - Conclusiones y trabajo futuro: Indica las conclusiones extraídas a raíz de los resultados obtenidos, y detalla posibles consideraciones para intentar mejorar dichos resultados. Proporciona, además, ideas para reforzar y expandir futuros trabajos basados en el ámbito estudiado.

Capítulo 2 - Estado actual y objetivos iniciales

2.1 – Herramientas existentes en el mercado

Tal y como se comentó en el capítulo introductorio del trabajo, actualmente se recopilan muchos datos acerca de las estadísticas del fútbol, y por ello, el Sports Analytics se encuentra muy presente en dicho deporte.

En la actualidad, se obtienen una gran cantidad de estadísticas relacionadas con el fútbol. Se recopilan estadísticas de las actuaciones de los equipos en general, pero también de las actuaciones de los propios jugadores individualmente. Por ejemplo, algunos de los datos que se recogen son los minutos jugados, goles, asistencias, tarjetas amarillas, tarjetas rojas, faltas cometidas, faltas recibidas, regates realizados con éxito, número de pases, kilómetros recorridos, etc. A raíz de estas estadísticas que se obtienen durante los diferentes partidos jugados, se pueden sacar conclusiones acerca de los rendimientos de los jugadores y de los equipos. Como por ejemplo, se puede identificar los mejores equipos ofensivos, que serían aquellos que más goles marcan, los mejores equipos defensivos, los máximos goleadores, los máximos asistentes... Además, se podría indagar aún más, y obtener la media de cuántos goles por partido marca un equipo o un jugador e incluso obtener un promedio de cuantos minutos han de pasar para que un equipo o un jugador marque un gol, por ejemplo.

Este tipo de datos puede ser de gran utilidad tanto para los entrenadores como para los propios jugadores, ya que pueden adaptar sus tácticas y formas de juego, en base a su rival, o a las diferentes circunstancias que se puedan dar.

Hoy en día, el Sport Analytics en el fútbol no solamente se basa en la recopilación de los datos de las valoraciones y el rendimiento de los jugadores y los equipos en los partidos, sino que tiene muchas otras utilidades.

Durante el transcurso de esta temporada 2022-2023, en la liga de fútbol española llamada “LaLiga”, se implantó un modelo en las retransmisiones de los partidos capaz de reconocer la probabilidad de gol que habría en la acción. Dicha probabilidad de gol ha

sido calculada por Microsoft, y se obtiene en tan sólo 30 segundos, es decir, casi en tiempo real, por lo que, los espectadores pueden disfrutar del dato al ver la segunda repetición del gol. LaLiga, es la primera liga de fútbol en tener dicha utilidad implantada [10].



Figura 2.1: Ejemplo de la probabilidad de gol calculada por Microsoft.

Imagen obtenida de: LaLiga: la primera en contar en sus retransmisiones un modelo avanzado de Probabilidad de Gol graficado casi en tiempo real con tecnología de Microsoft – Centro de noticias

Otra de las utilidades es la monitorización de los jugadores en los entrenamientos. De esta forma, se puede optimizar las sesiones de entrenamiento para mejorar el rendimiento y prevenir las lesiones de los futbolistas [11].

El análisis de los datos es utilizado, además, para las estrategias a seguir en relación con la venta de entradas y marketing, para intentar convencer a los espectadores para que acudan a los estadios a ver los partidos. Las casas de apuestas, también se basan en modelos predictivos utilizando todos los datos disponibles, para establecer las ofertas [12].

Esto demuestra que el fútbol ya cuenta con una gran implementación de lo denominado Sports Analytics. Esto va más allá de lo meramente relacionado con el estudio de las tácticas y rendimientos de jugadores y equipos, ya que abarca otros ámbitos relacionados con este deporte.

Si se busca el uso de Sports Analytics específicamente en el estudio de los libres directos, apenas hay diversidad en los datos. Puesto que, dichos datos solamente detallan el número de lanzamientos realizados y la cantidad de estos convertidos en gol durante una temporada o durante el transcurso de una competición específica. Por ejemplo, se ha encontrado que en las recientes campañas la cantidad de libres directos que acaban en gol en LaLiga ha disminuido considerablemente [13]. Estos son los únicos tipos de datos que se recogen acerca de este lance del juego.

Se ha realizado una labor de búsqueda en diferentes portales web, entre los que se encuentran: “Papers With Code”, “Kaggle”, “UCI Machine Learning Repository” y “Google Dataset Search”, para intentar encontrar bases de datos de lanzamientos de faltas directas en el fútbol. Se han encontrado diversos conjuntos de datos (o, en inglés, “datasets”) que son normalmente presentados en tablas, y artículos de investigación (o, en inglés, “papers”) relacionados con el deporte y también con el fútbol. Sin embargo, no se ha encontrado ninguno que contenga datos o análisis de los tiros de libres directos.

Con respecto a los datasets encontrados, muchos de estos, contienen datos e información acerca de los partidos y las estadísticas de estos. Un ejemplo de esto es la base de datos llamada “International football results from 1872 to 2023” [14], en donde se encuentran los resultados, y los goleadores de los partidos entre selecciones nacionales desde 1872 hasta el 2023.

Un ejemplo de una de las publicaciones encontradas se llama “SoccerNet: A Scalable Dataset for Action Spotting in Soccer Videos” [15], y hace uso de un dataset llamado “SoccerNet”. Dicha publicación, no se basa en las estadísticas de los partidos, sino que trata de identificar los diferentes eventos que suceden durante un partido de fútbol. En este caso, consideran evento a cuando se marca un gol, a cuando el árbitro le saca una tarjeta a un jugador (bien sean amarillas o rojas), y a cuando se realiza una sustitución.

Se puede apreciar que actualmente se están realizando múltiples estudios relacionados con la recopilación y el tratamiento de los datos del fútbol; pero, tal y como se ha comentado, dichos estudios no están relacionados con los lanzamientos de falta. Por lo que, a pesar de que los análisis deportivos, y especialmente los análisis futbolísticos, son una materia que está siendo ampliamente investigada hoy en día, no se han encontrado publicaciones ni bases de datos sobre el lance del juego del que trata este trabajo.

2.2 - Objetivos

Los objetivos previstos inicialmente en el documento del TFT01 son los que se listan a continuación:

- Recopilación de vídeos positivos y negativos de lanzamientos de falta directa en el deporte del fútbol
- Corte de vídeos y etiquetado manual
- Visualización de las características etiquetadas
- Aplicación de modelo de apariencia para la clasificación de movimiento
- Diseño del conjunto experimental
- Inferencia y validación

Debido a la falta de tiempo, el objetivo que trata sobre la aplicación de un modelo de apariencia para la clasificación del movimiento ha sido sustituido por un análisis tabular de los datos usando distintos clasificadores.

Por lo tanto, los objetivos principales de este trabajo son la creación de una base de datos de vídeos de lanzamientos de falta directa, junto con el correspondiente etiquetado de cada uno de los vídeos, para posteriormente, aplicando diferentes clasificadores, intentar predecir de forma correcta si el lanzamiento de falta acabará en gol o si no, o si lo detendrá el portero.

Una vez se tenga el dataset (los datos del etiquetado de todos los videos), se hará uso de diferentes clasificadores para intentar predecir si el lanzamiento acabará en gol, o si el portero parará dicho lanzamiento. Es por ello por lo que cada clasificador se ejecutará varias veces con diferentes valores de entrada. De esta forma se podrá observar cómo varía el resultado de la predicción de cada clasificador, y valorar cuánto peso presenta cada característica o atributo identificado con respecto a la precisión de la predicción.

Además de las predicciones de los resultados de las faltas, también se obtendrán las estadísticas de los datos recabados. Se obtendrán las estadísticas y los porcentajes de, por ejemplo, los goles y los fallos dependiendo de la zona del campo en donde esté situada la falta, de la pierna con la que tire el lanzador, de la dirección hacia donde fue dirigido el lanzamiento, de la distancia a la que se encuentra la falta, etc.

Capítulo 3 - Competencias específicas y aportaciones del trabajo

3.1 - Competencias

A continuación, se detallan las competencias relacionadas con este trabajo y como han sido estas cubiertas:

- CI15 - “Conocimiento y aplicación de los principios fundamentales y técnicas básicas de los sistemas inteligentes y su aplicación práctica.”

El principio fundamental en el que está basado este trabajo es en un sistema de deep learning o aprendizaje profundo; que, partiendo de la extracción de características de unos datos no estructurados, como son los videos de los lanzamientos de falta, se han obtenido unos datos estructurados, almacenados en tablas, que forman el etiquetado de dichos atributos de los videos. A partir de estos datos tabulares, se ha hecho uso de modelos de clasificación para la predicción de futuros eventos. En conclusión, tras la obtención de los datos, y empleando diversos clasificadores, se ha pronosticado el resultado de los libres directos.

- TI6 - “Capacidad de concebir sistemas, aplicaciones y servicios basados en tecnologías de red, incluyendo Internet, web, comercio electrónico, multimedia, servicios interactivos y computación móvil.”

En el fútbol actual se genera una enorme cantidad de datos a los que, aplicando los sistemas de deep learning, apoyado con análisis estadísticos, proporcionaría la posibilidad de crear aplicaciones y servicios que apoyen en la toma de decisiones de cara a mejorar el rendimiento tanto a nivel individual como colectivo. Uno de los objetivos de este trabajo, es concebir un sistema capaz de predecir los resultados de los lanzamientos de falta. Para ello, se ha hecho uso de servicios basados en tecnologías de red para la búsqueda y descarga de videos, para, tras estos haber sido recortados y etiquetados, y utilizando el lenguaje de programación Python en un entorno de desarrollo en línea, proporcionar datos que puedan servir de cara a la toma de decisiones técnicas y tácticas.

3.2 - Aportaciones

Debido a que no se han encontrado estudios de los lanzamientos de faltas directas, esto indica que no se ha investigado mucho en el análisis de dicha jugada, ni, por ende, en la temática propuesta en este trabajo; convirtiendo a este trabajo, en un trabajo innovador. Por lo tanto, este puede servir como punto de partida para los investigadores dispuestos a indagar en el análisis de estos lances del juego; ya que, no había disponible ningún dataset y mucho menos ningún tipo de predicción de resultado acerca de los lanzamientos directos en el fútbol. De esta forma, no solo servirán los resultados obtenidos en las estadísticas y en las predicciones realizadas, sino que la base de datos de los videos de los lanzamientos de falta recopilada puede ser utilizada para otros estudios y/o para su ampliación, de modo que futuros investigadores ya contarían con una extensa base de datos, y no tendrían que empezar de cero.

De igual forma, cualquier estudio que se realice a partir de ahora en esta área, tendrá disponible los resultados y el dataset obtenido en este trabajo, por lo que se podrán llevar a cabo comparaciones de resultados. Gracias a estas comparaciones, se podrían obtener conclusiones, acerca de aspectos a mejorar. Por ejemplo, si se lleva a cabo un nuevo análisis, incorporando nuevas características, se podría comprobar el peso que dichos nuevos atributos tienen en el resultado final de la predicción.

Es por ello, que dicho trabajo puede proporcionar muchas aportaciones con respecto al entorno científico, debido a las posibles investigaciones que se pueden dar a raíz de este enfoque realizado hacia el estudio de los lanzamientos de faltas directas en el fútbol.

Capítulo 4 - Desarrollo

El desarrollo de este trabajo se ha realizado mayoritariamente de forma similar a la metodología en cascada de los proyectos de software. Esto se debe a que, el proyecto inicialmente se dividió en fases, y no se comenzaba la siguiente fase, hasta que la anterior ya estuviese completada.

Las fases en las que se ha dividido el desarrollo del trabajo son:

- Recopilación y descarga de vídeos
- Recorte y etiquetado de los videos
- Obtención de estadísticas
- Uso de clasificadores para la predicción de resultados

A continuación, se explicará más en detalle, las tareas realizadas en cada una de las etapas del proyecto, y las herramientas utilizadas para su realización.

4.1 - Herramientas utilizadas:

4.1.1 - Easy Youtube Video Downloader Express

Easy Youtube Video Downloader Express es una extensión de Mozilla Firefox que proporciona la posibilidad de descargar videos de YouTube. Por lo que se ha utilizado dicha extensión para poder descargar los videos de los lanzamientos de falta.

4.1.2 - ClipChamp

ClipChamp es un editor de videos, con el que se pueden llevar a cabo diversas tareas diferentes relacionadas con la creación y edición de videos. Dicha plataforma permite hacer el recortado de los vídeos, por lo que ha sido la herramienta que se ha utilizado para realizar dicha labor.

4.1.3 - Excel

Excel es una herramienta que permite crear y modificar hojas de cálculo. En dicho programa es donde se ha almacenado en una tabla todos los datos sobre el etiquetado de los vídeos. Una vez completado el etiquetado, se ha exportado la hoja de cálculo donde se encontraba la información a un fichero con formato CSV, que es un fichero que contiene los valores separados por comas, y de ahí su nombre “Comma Separated Values”.

4.1.4 - Python

Python es un lenguaje de programación de alto nivel orientado a objetos y en la actualidad muy popular, especialmente en todo aquello relacionado con las aplicaciones web, el desarrollo de software, la ciencia de datos y el machine learning [16]. Es el lenguaje de programación que se ha utilizado en el entorno de desarrollo en línea para aplicar los diferentes modelos, y obtener las estadísticas de la información recopilada.

4.1.5 - Google Colab

Google Colab es un entorno de desarrollo en línea que permite ejecutar comandos escritos en el lenguaje de programación de Python. Google Colab es muy utilizado en relación con la inteligencia artificial, y con la ciencia de datos. Esto se debe a que se pueden utilizar múltiples librerías de Python para el análisis y la visualización de los datos, pero, sobre todo, porque al estar ubicado en los servidores de la nube de Google, se puede hacer uso de sus GPUs y TPUs [17]. Es por ello por lo que se ha utilizado dicho entorno de desarrollo en línea para la aplicación de los modelos de clasificación para la predicción de los resultados.

4.1.6 - Librería Scikit-learn

Scikit-learn es una de las librerías de código abierto de machine learning de Python más populares. Proporciona diversas herramientas para el modelado estadístico y el aprendizaje automático, entre los que se encuentran la clasificación, la regresión, la agrupación, la reducción de la dimensionalidad, la selección del modelo, y el preprocesamiento [18]. Durante el desarrollo del trabajo, se hizo uso de esta librería, y más específicamente, de los modelos de clasificación con el objetivo de intentar predecir futuros eventos.

4.2 - Recopilación y descarga de vídeos

Se ha recopilado una extensa base de datos de lanzamientos de falta. Dichos videos han sido obtenidos y descargados desde YouTube, utilizando la herramienta “Easy Youtube video Downloader”, y escogiendo la opción de descarga denominada “MP4 720P”, para que se descarguen con una resolución de “1280x720” en formato “MPEG-4”.

Se ha conseguido recopilar un total de 610 vídeos de lanzamientos de falta directa, entre los que se encuentran a partes iguales los positivos, aquellos que acaban en gol, y los negativos, aquellos que no acaban en gol. Es decir, que dicha base de datos cuenta con 305 videos de lanzamientos de faltas directas que acaban en gol, y 305, cuyo resultado final es otro distinto al gol (un total de 610).

4.3 - Recorte y etiquetado de vídeos

Una vez realizada la descarga se ha realizado el recorte de los videos para que únicamente se muestre el lanzamiento de la falta con su resultado final. Todo lo que haya pasado antes o después del resultado directo del golpeo, no es necesario preservar. Solo se estudiará el resultado directo del lanzamiento, ya que, por ejemplo, si el tiro lo detiene el portero, pero hay rebote y en el rebote se marca gol; solamente se utilizará el video hasta la parada del portero, el resto, lo que pase en este caso con el rebote, no es objeto de estudio, por lo que se omitirá del video definitivo.

Una vez finalizado el recorte de los videos, se ha procedido a realizar el etiquetado manual de las características que presentan los diferentes lanzamientos de falta.

Por cada video, se ha recopilado información acerca de diferentes factores, como son, la ubicación del libre directo, el lanzador, el golpeo, el portero y la barrera. Por lo tanto, las características se han categorizado en los siguientes campos:

- Fuente del vídeo
- Nombre del vídeo
- Lado del campo
- Lado de la falta
- Distancia de la falta
- Género de la competición
- Nombre del lanzador
- Pierna con la que tira el lanzador
- Dirección a la que va el balón
- Nombre del portero
- Posición del portero
- Hacia qué lado se tira el portero
- Composición de la barrera
- Resultado final

A continuación, se detalla lo que se ha tenido en cuenta, y cómo se ha procedido a la hora de realizar el etiquetado y a asignar los diferentes valores de las distintas categorías.

4.3.1 - Categorías

4.3.1.1 - Fuente del vídeo

Se detalla la fuente de origen de donde se ha obtenido el vídeo en cuestión.

4.3.1.2 - Nombre del vídeo

Cada uno de los vídeos recopilados ha sido denominado de la siguiente manera: “vdXXX_jXX_visión”. Donde “XXX” hace referencia al número del video en cuestión, por ejemplo, para el primer vídeo descargado se le otorgará el número “001” (vd001_jXX_visión); mientras que “XX” es el número de la falta dentro del video, por ejemplo, para la primera falta encontrada dentro del primer vídeo descargado, se le concederá el valor “01” (vd001_j01_visión).

El campo “visión” dentro del nombre, hará referencia a la posición y el ángulo de la cámara que realizó la grabación. Las diferentes posibilidades que se han establecido son la visión diagonal, trasera, delantera y lateral.

El ángulo diagonal, será denominado “diag”, y es aquel en el cual la cámara se encuentra prácticamente a la altura del centro del campo y realiza la grabación en diagonal hacia la portería sobre la que se va a lanzar la falta. La figura 4.1 muestra un ejemplo de la visión diagonal.



Figura 4.1: Ejemplo de la visión diagonal.

Imagen capturada del vídeo: Free Kick/Women's Football - YouTube

Cuando la visión sea la trasera, esta será designada como “tras”. Esta hará referencia a cuando la cámara se encuentra detrás de la portería y que cuando el lanzador chuta, se ve el balón aproximarse de frente, tal y como muestra la figura 4.2.



Figura 4.2: Ejemplo de la visión trasera.

Imagen capturada del vídeo: The Premier League's GREATEST free-kick saves! - YouTube

La denominada visión delantera, que será indicada como “del”, es cuando la grabación se realiza de frente a la portería, y detrás del lanzador del libre directo. Un ejemplo de dicho ángulo se puede apreciar en la figura 4.3.



Figura 4.3: Ejemplo de la visión delantera.

Imagen capturada del vídeo: Goles de tiros libres femeninos que conmocionaron al mundo - YouTube

La última posibilidad es la lateral, que será llamada como “lat”, es cuando la falta es grabada a su misma altura, pero desde uno de los lados del campo. La figura 4.4 muestra cómo sería dicha visión.



Figura 4.4: Ejemplo de la visión lateral.

Imagen capturada del vídeo: TOP 5 GOLES DE FALTA LaLiga Santander 2021/2022 - YouTube

Por lo tanto, si, por ejemplo, en la primera falta recortada del primer video descargado el ángulo de la cámara fuese en diagonal, su nombre final sería: “vd001_j01_diag”.

4.3.1.3 - Lado del campo

El lado del campo hace referencia a si se trata de la mitad del campo de la izquierda o de la derecha. En caso de que el lanzamiento sea en el lado izquierdo, se etiquetará con un “1” y en caso de que sea en el derecho, se usará el “2”, tal y como muestra la figura 4.5.

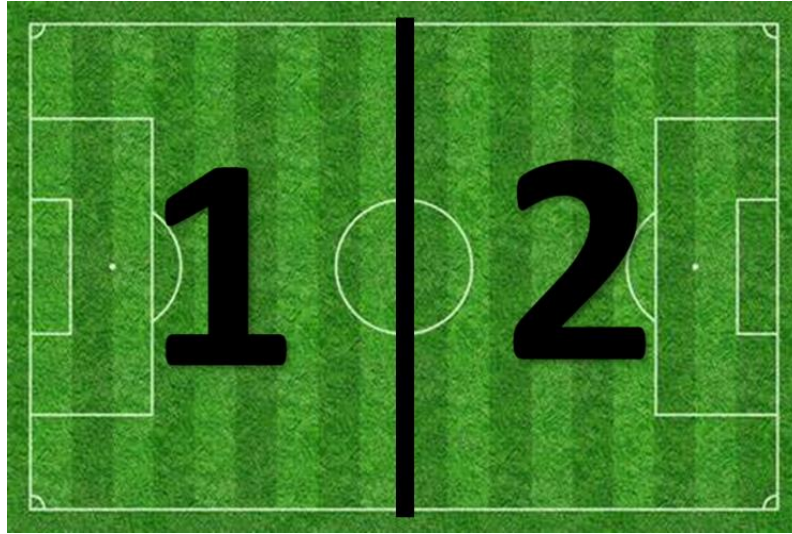


Figura 4.5: Valor que se asignará a cada mitad del campo.

Imagen modificada manualmente, obtenida de: campo de futbol png - Bing images

4.3.1.4 - Lado de la falta

En esta ocasión, se deberá indicar la ubicación de la falta con respecto a la portería a la que se va a disparar. Esto quiere decir, que se deberá indicar si la falta se encuentra a la izquierda, a la derecha o de frente a la portería. Si se encuentra a la izquierda, se le asignará un “1”; si está situada en el centro, un “2”; y si está a la derecha, un “3”.

Para facilitar la decisión de si la falta se encuentra en un lado o en el centro, se ha cogido como referencia el semicírculo que se encuentra fuera del área grande, y se han trazado dos rectas (imaginarias), que delimitan el diámetro de la medialuna (su anchura). Todo lo que se encuentre dentro de dicho diámetro se considerará centro, y lo que se encuentre a los lados de esto, se considerará o bien derecha o bien izquierda, dependiendo del lado en el que se encuentre. La figura 4.6 muestra lo descrito.

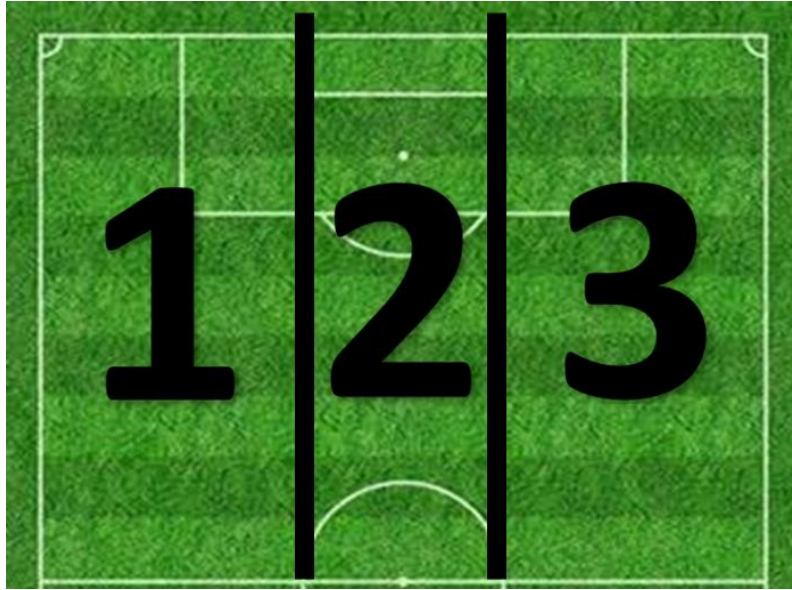


Figura 4.6: Valor que se asignará a los diferentes lados donde esté ubicada la falta.
Imagen modificada manualmente, obtenida de: campo de futbol png - Bing images

4.3.1.5 - Distancia de la falta

Esto hace referencia a cuántos metros hay desde la portería hasta la falta, es decir, si la falta se encuentra lejos o cerca de la portería. Por lo que, se ha trazado una recta horizontal que delimita lo que se considerará cerca y lo que se considerará lejos. Por lo que, todo lo que se encuentre por delante de dicha línea, se considerará cerca, y se clasificará con un “1”, y lo que se encuentre por detrás, se considerará lejos, estableciendo un “2”, tal y como muestra la figura 4.7.

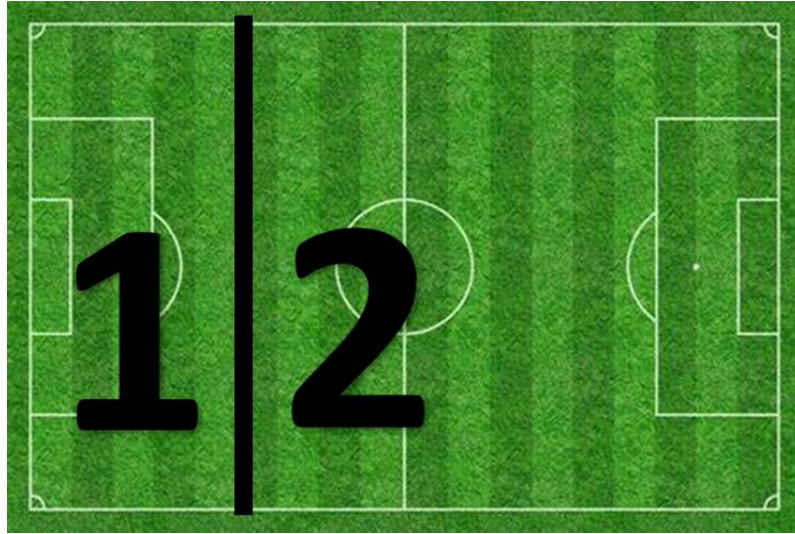


Figura 4.7: Valor que se asignará dependiendo si la falta es cercana o lejana.

Imagen modificada manualmente, obtenida de: campo de futbol png - Bing images

Para identificar de forma más precisa si la falta se considerará cerca o lejos, se hace uso de la posición de la barrera. Si la barrera se encuentra fuera del área se considerará una falta lejana, sin embargo, si esta se encuentra ubicada dentro del área, o justo en la línea, se considerará cerca.

Por lo que, para que una falta sea considerada lejana, deberá estar al menos a 25 metros de la portería. Esto se debe a que la línea horizontal que delimita el área grande está a alrededor de 16,5 metros de la portería, y en una falta, la barrera está aproximadamente a unos 9,15 metros del balón. Por lo tanto, si la barrera debe encontrarse fuera del área para que la falta sea considerada lejana, significa que el balón estará a mínimo a alrededor de 9,15 metros alejados del final del área grande, es decir, que el balón se encontrará a alrededor de 25 metros de la portería (la suma de los 9,15 metros de la barrera, y los 16,5 metros a los que se encuentra el final del área grande, dan un resultado de 25,65 metros de distancia).

4.3.1.6 - Género de la competición

Se deberá indicar si se trata de un partido de fútbol masculino, que se indicará con un “1”, o si se trata de fútbol femenino, que se etiquetará con un “2”.

4.3.1.7 - Nombre del lanzador

Se establecerá el nombre del jugador encargado de tirar el lanzamiento de la falta. En caso de que no se sepa quién es el lanzador, se utilizará la siguiente denominación: “jGXXX”. La “j” representará que se está haciendo referencia a un jugador. En donde se encuentra la “G”, se deberá especificar el género de la competición; con una “m” si es masculino, y con una “f” si es femenino. Los valores “XXX” se deberán cambiar por el número de jugador desconocido; por ejemplo, si es el primer jugador el cual se desconoce su nombre, se le pondrá “001”, y si es el segundo, “002”, y así sucesivamente. Si vuelve a aparecer un jugador, del cual no se sabe su nombre, pero que ya ha sido etiquetado siguiendo este patrón, se pondrá el mismo valor que el que se dio inicialmente, para hacer referencia a que se trata del mismo lanzador.

4.3.1.8 - Pierna con la que tira el lanzador

En caso de que el lanzador tire con la pierna izquierda se etiquetará con un “1”, y si lo hace con la derecha, con un “2”.

4.3.1.9 - Dirección a la que va el balón

Se especificará hacia dónde va dirigido el balón, más concretamente, la zona por la que el balón llega a la portería. La portería se ha dividido en seis, y dependiendo por qué zona entre el balón a la portería, se deberá etiquetar con dicho valor. Los valores establecidos son los siguientes: si va arriba a la izquierda un “1”, arriba al centro un “2”, arriba a la derecha un “3”, abajo a la izquierda un “4”, abajo al centro un “5”, y abajo a la derecha un “6”. La figura 4.8 muestra cómo se ha dividido la portería para poder establecer el valor correcto.

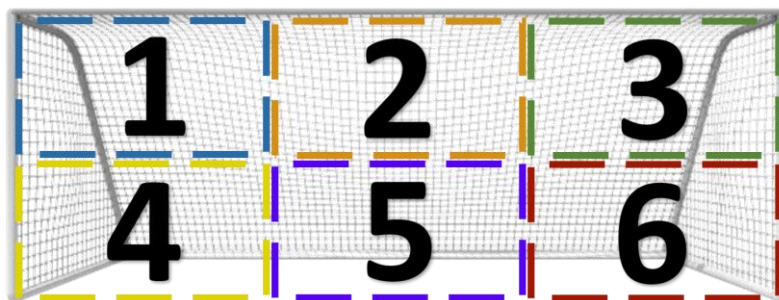


Figura 4.8: Valor que se asignará dependiendo del lado por el que vaya el balón.
 Imagen modificada manualmente, obtenida de: portería de futbol png - Bing images

En caso de que el lanzamiento no sea gol, pero se la para el portero, se le asignará el valor del recuadro hacia el cual el balón fue dirigido. Por ejemplo, si el balón va hacia arriba a la izquierda (recuadro con valor “1”), pero la para el portero, se establecerá un “1”. Lo mismo pasaría si el balón se va fuera. Por ejemplo, si el jugador tira buscando la escuadra derecha (recuadro con valor “3”), pero se va fuera, bien sea porque fue demasiado alto, o porque se va muy hacia la derecha, se va a seguir etiquetando con un “3”.

En el único caso en el que no se establecerá ninguno de estos valores, es si no se sabe a dónde va dirigido. Esto será cuando el tiro no llegue a portería, bien sea porque rebota en la barrera o porque rebota en algún otro jugador. En este caso, a estos lanzamientos se les asignará un “0” en esta categoría.

4.3.1.10 - Nombre del portero

En el caso de que se sepa quién es el portero, se indicará su nombre, y en el caso de que no se sepa se procederá de la misma forma que con el nombre del jugador, pero en esta ocasión, se utilizará una “p” en vez de una “j”. Para indicar que es un portero en vez de un jugador. Por lo que, por ejemplo, cuando aparezca el primer portero masculino que no se sabe quién es, se le asignará el valor “pm001” y cuando aparezca el segundo portero masculino que se desconoce su identidad se le asignará “pm002”; y así para los diferentes porteros que no se sepa quiénes son. Y al igual que con el jugador, si vuelve a aparecer

un portero por segunda vez, por ejemplo, que no se sabe quién es, pero ya se le ha etiquetado usando este patrón, se le asignará el mismo valor que el que se usó en la primera ocasión.

4.3.1.11 - Posición del portero

En esta categoría se deberá especificar en qué lado de la portería se encuentra el guardameta antes del lanzamiento de falta. Para ello se ha dividido la portería en tres, y se le ha asignado un valor a cada zona, tal y como se muestra en la figura 4.9. Mirando a la portería, si el portero se encuentra en la zona izquierda de la portería se le asigna un “1”, si está ubicado en el centro, un “2”, y si está en el lado derecho, un “3”.

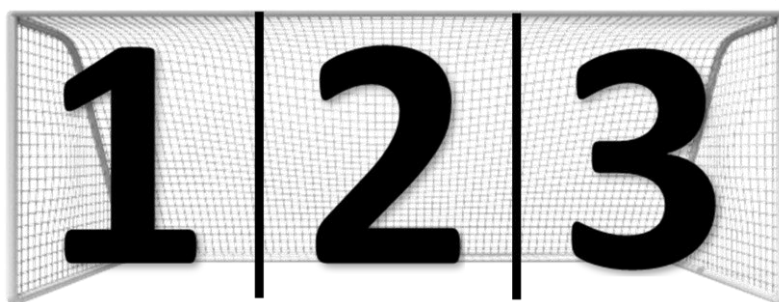


Figura 4.9: Valor que se asignará dependiendo de la posición del portero.

Imagen modificada manualmente, obtenida de: portería de fútbol png - Bing images

4.3.1.12 - Hacia qué lado se tira el portero

En este apartado, se hará referencia a la reacción del portero una vez ve el lanzamiento de la falta. Es donde se indicará si el portero se tira hacia algún lado a intentar parar el balón, o si se queda quieto, sin lanzarse. Al igual que en la categoría de la posición del portero, si se tira hacia su derecha se pondrá un “1”, si se queda quieto (si no se lanza) se establecerá un “2”, y si se tira hacia su izquierda, se asignará un “3”. La figura 4.9, también sirve como guía para decidir qué valor se especifica, dependiendo de hacia dónde se lance el guardameta.

4.3.1.13 - Composición de la barrera

Aquí, se indicará el número de jugadores que componen la barrera. Por ejemplo, si una barrera está compuesta por tres jugadores, se especificará un “3”.

En el fútbol actual se ha puesto de moda que un jugador se tumbe o se agache detrás de la barrera para intentar evitar que el lanzador marque un gol al tirar raso, y que el balón pase por debajo de la barrera cuando esta salte. Por lo que, si hay un jugador tumbado detrás de la barrera, se pondrá el símbolo “+” al lado del número que indica cuántos jugadores la componen. Si, por ejemplo, una barrera está compuesta por tres jugadores, y además hay uno tumbado por detrás, se etiquetará indicando el valor “3+”.

4.3.1.14 - Resultado final

En este apartado es donde se indicará cómo acabó la ejecución de la falta. Esto quiere decir, que se indicará si fue gol o si no. En caso de que no sea gol, hay distintos escenarios que se pueden dar, los que se contemplarán son los siguientes: si se va fuera, si le da al palo o al larguero (que se considerará palo), si le da a la barrera, si lo despeja o rebota en otro jugador que no esté en la barrera, o si se la para el portero. En caso de que sea gol, no se tendrá en cuenta como es el gol, si entra de rebote, si la toca el portero o si le da al palo y acaba entrando, etc. Sino que todos los goles se clasificarán igual, es decir que solo se tendrá un valor que simbolice el gol.

Por lo tanto, a la hora de etiquetar, si es gol se indicará un “1”, si se va fuera un “2”, si va al palo un “3”, si le da a la barrera un “4”, si lo despeja otro jugador un “5”, y si la para el portero un “6”.

4.3.2 - Etiquetado resultante

Tal y como se indicó anteriormente, se realizó el etiquetado manual de 610 videos. Dicho etiquetado se realizó originalmente en un fichero de excel y la tabla 4.1 muestra varios ejemplos del etiquetado manual que se realizó a raíz de algunos de los vídeos recopilados.

Fuente del video:	Nombre del video con la visión:	Lado del campo:	Lado de la falta:	Distancia de la falta:	Género de la competición:	Nombre del lanzador :	Pierna con la que tira el lanzador:	Dirección a la que va el balón:	Nombre del portero:	Posición del portero:	Hacia que lado se tira el portero:	Composición de la barrera:	Resultado final:
https://www.youtube.com/watch?v=ep2ktrDKO50&list=PLbdaGi5g6XB72JHEORajuc-5MgyRQleT&index=9	vd006-j01-dia	2	1	1	1	Isak	2	3	Aitor Fernandez	2	2	5	1
https://www.youtube.com/watch?v=ep2ktrDKO50&list=PLbdaGi5g6XB72JHEORajuc-5MgyRQleT&index=9	vd006-j02-tras	1	2	1	1	Joselu	2	5	Joel Robles	1	2	5+	1
https://www.youtube.com/watch?v=ep2ktrDKO50&list=PLbdaGi5g6XB72JHEORajuc-5MgyRQleT&index=9	vd006-j03-dia	2	2	1	1	Joselu	2	6	Rui Silva	3	3	4+	1
https://www.youtube.com/watch?v=AyRWH7z5jw&list=PLbdaGi5g6XB72JHEORajuc-5MgyRQleT&index=74	vd072-j01-dia	1	2	1	1	Marco Reus	2	4	Manuel Neuer	3	1	5	6
https://www.youtube.com/watch?v=QDKM7lwrq5E&list=PLbdaGi5g6XB72JHEORajuc-5MgyRQleT&index=63	vd062-j01-dia	2	2	1	1	James Maddison	2	3	Aaron Ramsdale	1	3	5+	6
https://www.youtube.com/watch?v=QDKM7lwrq5E&list=PLbdaGi5g6XB72JHEORajuc-5MgyRQleT&index=63	vd062-j02-dia	2	2	2	1	Oscar	2	1	Adrian	3	1	4	6
https://www.youtube.com/watch?v=ep2ktrDKO50&list=PLbdaGi5g6XB72JHEORajuc-5MgyRQleT&index=9	vd006-j04-dia	2	3	1	1	Unai Lopez	2	3	Ledesma	1	2	5+	1
https://www.youtube.com/watch?v=gjrs4jquLL4&list=PLbdaGi5g6XB72JHEORajuc-5MgyRQleT&index=88	vd087-j06-dia	2	2	1	1	Marcus Rashford	2	4	Hugo Lloris	1	1	4	2
https://www.youtube.com/watch?v=ep2ktrDKO50&list=PLbdaGi5g6XB72JHEORajuc-5MgyRQleT&index=9	vd006-j05-dia	2	2	1	1	Messi	1	4	Rui Silva	2	1	4+	1
https://www.youtube.com/watch?v=DElq_3KLxyY&list=PLbdaGi5g6XB72JHEORajuc-5MgyRQleT&index=104	vd102-j01-dia	1	2	1	1	Luis Suarez	2	4	Handanovic	3	2	5+	4
https://www.youtube.com/watch?v=yu-HJmFkaFI&list=PLbdaGi5g6XB72JHEORajuc-5MgyRQleT&index=127	vd126-j01-dia	1	2	1	1	Messi	1	2	Pau Lopez	2	2	3	3
https://www.youtube.com/watch?v=P_1BxnmWwCU&list=PLbdaGi5g6XB72JHEORajuc-5MgyRQleT&index=141	vd140-j01-dia	2	1	2	1	Andrea Pirlo	2	3	Joe Hart	3	2	3	3

Tabla 4.1: Ejemplo del etiquetado manual de 12 videos diferentes.

Antes de comenzar a hacer uso de los clasificadores, se ha tenido que realizar modificaciones en el fichero del etiquetado manual. Esto se debe a que: se ha tenido que cambiar los datos almacenados en algunas de las categorías mencionadas anteriormente, se ha tenido que también añadir dos columnas nuevas, e incluso se ha tenido que obviar algunas de las columnas ya existentes. Los cambios que se llevaron a cabo han sido los siguientes:

- Se han modificado los datos almacenados en el campo de composición de la barrera. Se ha tenido que eliminar el “+” que indicaba que detrás de la barrera se encontraba un jugador o bien tumbado o agachado. Si bien, debido a que dicha característica podía influir de forma determinante en el resultado final del lanzamiento de la falta, se ha añadido una nueva columna que contendrá un “1” en donde el campo de la composición de la barrera contaba con el símbolo “+”, y un “0” en aquellas que no lo tenían.
- Otra modificación que se llevó a cabo es que se ha querido pasar el campo de resultado final al formato “one hot encoding”. Esto es, que dicha columna solamente podrá tener valores binarios; o bien un “1” o bien un “0”. Esto ha

provocado que se haya reemplazado la columna de resultado final, por una columna denominada gol. En esta columna, solamente se considerará si el resultado final del libre directo fue gol, estableciendo un “1”, o si no fue gol, indicando un “0”. Es decir, que ya no se tendrán en cuenta los diferentes motivos por los cuales la falta no terminó en gol; sino solamente si fue gol o si no.

- Además, debido a que no se va a realizar una reidentificación, sino solamente se va a predecir si el resultado final es gol o si no usando los datos tabulares, los campos de: fuente del video, nombre del video, nombre del jugador y nombre del portero, se han podido obviar, y por lo tanto se han retirado.
- Lo mismo que se ha hecho para predecir si la falta acabaría en gol o no, se ha hecho para predecir si el portero pararía la falta. Por lo que, se ha creado otro fichero, pero en este caso, la columna con formato “one hot encoding”, en vez de denominarse gol, se ha denominado parada. Donde un “1” significa que el portero ha parado el lanzamiento, y un “0” que no.

Una vez realizadas las modificaciones pertinentes, los ficheros resultantes de excel se han exportado al formato csv, antes de llevar a cabo el tratamiento de la información, e intentar predecir si la falta acabaría en gol o si no, o si la pararía el portero, o si no. La tabla 4.2 muestra el etiquetado de los mismos videos que los mostrados en la tabla 4.1, pero para el caso de la predicción del gol, después de las modificaciones indicadas y una vez exportado el fichero a formato csv.

lado_campo	lado_falta	distancia	genero	pierna	direccion	posicion_portero	lado_tira_portero	composicion_barrera	jugador_tumbado_barrera	gol
2	1	1	1	2	3	2	2	5	0	1
1	2	1	1	2	5	1	2	5	1	1
2	2	1	1	2	6	3	3	4	1	1
1	2	1	1	2	4	3	1	5	0	0
2	2	1	1	2	3	1	3	5	1	0
2	2	2	1	2	1	3	1	4	0	0
2	3	1	1	2	3	1	2	5	1	1
2	2	1	1	2	4	1	1	4	0	0
2	2	1	1	1	4	2	1	4	1	1
1	2	1	1	2	4	3	2	5	1	0
1	2	1	1	1	2	2	2	3	0	0
2	1	2	1	2	3	3	2	3	0	0

Tabla 4.2: Ejemplo del etiquetado de la predicción de gol después de las modificaciones y una vez exportado a csv.

4.4 - Estadísticas

Con el etiquetado de los 610 videos ya terminado, se ha procedido a sacar las estadísticas de los datos recabados. Las estadísticas recogidas son acerca de los porcentajes de goles y paradas. La finalidad es intentar buscar patrones y características para poder optimizar el resultado final, bien sea, de cara a los porteros como a los lanzadores. Es decir, procurar conseguir conclusiones que indiquen, por ejemplo a los porteros, donde es mejor colocarse en base a las diferentes situaciones; y con respecto a los lanzadores, dependiendo de la circunstancia, a que zona de la portería deberían intentar dirigir el golpeo, por ejemplo. Los resultados obtenidos de las diferentes estadísticas se detallarán en el Capítulo 5 donde se ilustran los resultados obtenidos.

El algoritmo 4.1 muestra un fragmento del código ejecutado para la obtención de las diferentes estadísticas. En este caso, el algoritmo 4.1 muestra la parte del código en la que se pretende obtener como se han distribuido los goles y los fallos de los lanzamientos. El resultado de la ejecución de dicho algoritmo se muestra en la tabla 5.1.

```
drive.mount('/content/drive')
datosCSV = pd.read_csv("/content/drive/My
Drive/AnálisisDatasetFaltas/Etiquetado/etiquetadoDeVideos_resultadoFinal.csv")

faltasLanzadas = datosCSV

porcentajeLanzadas = (faltasLanzadas.shape[0]/datosCSV.shape[0]) * 100

faltasMarcadas = datosCSV.loc[datosCSV["resultado_final"] == 1]
porcentajeMarcadas = (faltasMarcadas.shape[0]/datosCSV.shape[0]) * 100

faltasFalladas = datosCSV.loc[datosCSV["resultado_final"] != 1]
porcentajeFalladas = (faltasFalladas.shape[0]/datosCSV.shape[0]) * 100

print("\t\tFALTAS LANZADAS")
tablaLanzamientos = pd.DataFrame({'Resultado Final': ['Goles', 'Fallos', 'Totales'],
                                'Lanzamientos': [faltasMarcadas.shape[0],
faltasFalladas.shape[0], faltasLanzadas.shape[0]],
                                'Porcentaje(%)': [format(porcentajeMarcadas, '.2f'),
format(porcentajeFalladas, '.2f'), format(porcentajeLanzadas, '.2f')]}))
display(tablaLanzamientos)
```

Algoritmo 4.1: Obtención de la distribución de los goles y de los fallos.

4.5 - Clasificadores

Se ha hecho uso del conjunto de los datos tabulares obtenidos tras el etiquetado de los videos para, aplicando diferentes modelos de clasificación, intentar predecir los resultados de las faltas. A continuación, se encuentra una lista de los diferentes clasificadores que han sido empleados de cara a realizar la predicción de los resultados de los lanzamientos de falta:

- Multi-layer Perceptron
- Logistic regression
- Nearest Neighbors (K-Nearest Neighbors)
- Support Vector Machines (Support Vector Classifier)
- Naive Bayes (GaussianNB)
- Decision Trees
- Random Forest
- Ada Boost
- Quadratic Discriminant

Antes de aplicar los modelos, se han separado los diferentes campos del etiquetado en dos grupos: el grupo de valores (“values”) y el de objetivos (“targets”). El grupo de los objetivos se ha rellenado únicamente con los datos del campo que se va a intentar predecir, mientras que, en el grupo de los valores, se han añadido los datos del resto de las columnas del etiquetado. Por ejemplo, si se requiere la predicción de si una falta acabará en gol o no, el conjunto de los objetivos contendrá los datos de la columna denominada “gol”, que es aquella que indica si es gol o si no, es decir, la que contiene un uno si la falta acaba en gol o un cero si no es gol; mientras que el conjunto de los valores contendrá el resto de las columnas del etiquetado, es decir, el resto de los atributos del lanzamiento. Dicha separación se muestra en el algoritmo 4.2, en donde, además, se muestra la inicialización de los clasificadores sin parámetros, para la utilización de los parámetros por defecto.

Una vez se han separado los datos en sus grupos correspondientes, se ha podido empezar a utilizar los clasificadores. El primer paso ha sido indicar a los diferentes modelos con que datos se realizará el entrenamiento y con qué datos se intentará pronosticar y se

comprobará la precisión de la predicción. Para ello, se ha hecho uso de la función “train_test_split”. Dicha función divide los datos de entrada en dos grupos, uno de entrenamiento y otro de prueba [19]. En este caso, dicha división se ha hecho aleatoriamente, y el grupo de entrenamiento se ha compuesto por el 75% de los datos, y el grupo de entrenamiento el restante 25%.

A la hora de entrenar y probar los modelos, para que no fuese basado únicamente en un resultado específico, se realizó un bucle de 50 iteraciones, tal y como se muestra en el algoritmo 4.3; aunque en este fragmento solo se muestra el uso de cuatro clasificadores, se utilizaron los nueve anteriormente mencionados. De esta forma, en cada iteración, se dividieron los datos originales en los grupos de entrenamiento y de prueba, por lo que cada predicción contó con unos datos de prueba diferentes, escogidos de forma aleatoria. En el algoritmo 4.3, se puede apreciar también, el uso de la función “train_test_split”, para dividir los datos en los grupos de entrenamiento y de prueba dentro del bucle. Una vez realizadas las 50 iteraciones, se obtuvieron 50 predicciones diferentes, de los nueve clasificadores utilizados, y las precisiones de cada una de estas. A continuación, se calculó la media y la desviación estándar de estos resultados. Pero antes de esto, se eliminaron los valores atípicos o extremos tras los 50 pronósticos. Para que estos no influyeran en la media. Para ello, se ha calculado el percentil 5 y el percentil 95 y se han obviados los valores que se encontraban por debajo y por encima de dichos valores, respectivamente. Esta forma de calcular las medias y las desviaciones estándar, después de eliminar los valores atípicos se muestra en el algoritmo 4.4, que al igual que en el algoritmo 4.3, solamente se observa el cálculo en cuatro de los nueve clasificadores.

Dicho procedimiento se realizó varias veces debido a que se han llevado a cabo dos tipos de predicciones diferentes. Inicialmente se ha intentado pronosticar si el lanzamiento acabaría en gol, y posteriormente, se ha intentado adivinar si el portero pararía la falta. De cada una de estas se han realizado las predicciones modificando los datos de entrada, inicialmente contaban con todos los campos, luego sin el atributo de hacia dónde se lanza el portero; luego sin el de hacia dónde va dirigido el lanzamiento; y finalmente, sin ninguna de esas dos columnas. Por lo que en total se realizaron ocho predicciones con cada clasificador, cuatro para los goles, y cuatro para las paradas.

A pesar de que el objetivo no era maximizar la precisión de las predicciones, se ha utilizado la función llamada “GridSearchCV”, para mejorar los pronósticos obtenidos. Dicha función, realiza una búsqueda exhaustiva entre diferentes hiperparámetros, y proporciona el mejor resultado y la combinación de los hiperparámetros que han conseguido que el clasificador obtenga dicho resultado [20]. Es decir que haciendo una combinación de todos los diferentes hiperparámetros indicados, dicha función es capaz de identificar cuál de todas es la mejor combinación (la que proporciona más precisión), y su resultado. En el algoritmo 4.5 se puede observar el fragmento de código en donde se hace uso de la función con el clasificador denominado "Multi-layer Perceptron”.

Una vez obtenidos los hiperparámetros idóneos de cada clasificador, se ha procedido a ejecutar cada uno de estos con dichos hiperparámetros para comparar con los resultados obtenidos inicialmente; en donde se hacía uso de los clasificadores con los parámetros por defecto, tal y como muestra el algoritmo 4.2. Para la ejecución con los nuevos hiperparámetros se ha utilizado el bucle de las 50 iteraciones, anteriormente mencionado y que se muestra en el algoritmo 4.3, para, una vez obviados los valores extremos, calcular la media y las desviaciones estándar de las precisiones de los pronósticos proporcionados por cada clasificador, lo que se puede apreciar en el algoritmo 4.4.

```
datosCSV = pd.read_csv("/content/drive/My
Drive/AnálisisDatasetFaltas/Etiquetado/etiquetadoDeVideos_gol.csv")

#Se separan los datos del etiquetado en los grupos de valores y de objetivos
values = datosCSV.iloc[:, :-1].values
targets = datosCSV.iloc[:, -1].values

#Se inicializan los clasificadores sin parametros, con los de por defecto
mlp_classifier = MLPClassifier()
lr_classifier = LogisticRegression()
knn_classifier = KNeighborsClassifier()
svc_classifier = SVC()
gaussianNb_classifier = GaussianNB()
decTree_classifier = DecisionTreeClassifier()
randForest_classifier = RandomForestClassifier()
adaBoost_classifier = AdaBoostClassifier()
quadDisc_classifier = QuadraticDiscriminantAnalysis()
```

Algoritmo 4.2: Separación de los datos del etiquetado de los goles en los grupos de valores y de objetivos, y la inicialización de los clasificadores.

```

#Se inicializan las variables en donde se almacenaran los resultados del bucle
values_training = [[] for p in range(50)]
targets_training = [[] for p in range(50)]
values_testing = [[] for p in range(50)]
targets_testing = [[] for p in range(50)]
mlp_targetsPredicted = [[] for p in range(50)]
mlp_precisionPredicted = [[] for p in range(50)]
lr_targetsPredicted = [[] for p in range(50)]
lr_precisionPredicted = [[] for p in range(50)]
knn_targetsPredicted = [[] for p in range(50)]
knn_precisionPredicted = [[] for p in range(50)]
svc_targetsPredicted = [[] for p in range(50)]
svc_precisionPredicted = [[] for p in range(50)]

#Bucle de las 50 iteraciones que obtiene las diferentes predicciones y sus precisiones
for position in range(len(values_training)):
    values_training[position], values_testing[position], targets_training[position],
    targets_testing[position] = train_test_split(values, targets)
    #MLP
    mlp_classifier.fit(values_training[position], targets_training[position])
    mlp_targetsPredicted[position] = mlp_classifier.predict(values_testing[position])
    mlp_precisionPredicted[position] = mlp_classifier.score(values_testing[position],
    targets_testing[position])
    #LR
    lr_classifier.fit(values_training[position], targets_training[position])
    lr_targetsPredicted[position] = lr_classifier.predict(values_testing[position])
    lr_precisionPredicted[position] = lr_classifier.score(values_testing[position],
    targets_testing[position])
    #KNN
    knn_classifier.fit(values_training[position], targets_training[position])
    knn_targetsPredicted[position] = knn_classifier.predict(values_testing[position])
    knn_precisionPredicted[position] = knn_classifier.score(values_testing[position],
    targets_testing[position])
    #SVC
    svc_classifier.fit(values_training[position], targets_training[position])
    svc_targetsPredicted[position] = svc_classifier.predict(values_testing[position])
    svc_precisionPredicted[position] = svc_classifier.score(values_testing[position],
    targets_testing[position])

```

Algoritmo 4.3: Fragmento del bucle de las 50 iteraciones para obtener las predicciones y las precisiones de cada clasificador.

```
mlp_sinAnomalias = pd.DataFrame({'mlp': mlp_precisionPredicted})
percentil05 = mlp_sinAnomalias.quantile(0.05)
percentil95 = mlp_sinAnomalias.quantile(0.95)
mlp_sinAnomalias = mlp_sinAnomalias.drop(mlp_sinAnomalias[(mlp_sinAnomalias.mlp
< percentil05.mlp) | (mlp_sinAnomalias.mlp > percentil95.mlp)].index)
mlp_media_todo = mlp_sinAnomalias.mean()
mlp_sd_todo = mlp_sinAnomalias.std()

lr_sinAnomalias = pd.DataFrame({'lr': lr_precisionPredicted})
percentil05 = lr_sinAnomalias.quantile(0.05)
percentil95 = lr_sinAnomalias.quantile(0.95)
lr_sinAnomalias = lr_sinAnomalias.drop(lr_sinAnomalias[(lr_sinAnomalias.lr <
percentil05.lr) | (lr_sinAnomalias.lr > percentil95.lr)].index)
lr_media_todo = lr_sinAnomalias.mean()
lr_sd_todo = lr_sinAnomalias.std()

knn_sinAnomalias = pd.DataFrame({'knn': knn_precisionPredicted})
percentil05 = knn_sinAnomalias.quantile(0.05)
percentil95 = knn_sinAnomalias.quantile(0.95)
knn_sinAnomalias = knn_sinAnomalias.drop(knn_sinAnomalias[(knn_sinAnomalias.knn
< percentil05.knn) | (knn_sinAnomalias.knn > percentil95.knn)].index)
knn_media_todo = knn_sinAnomalias.mean()
knn_sd_todo = knn_sinAnomalias.std()

svc_sinAnomalias = pd.DataFrame({'svc': svc_precisionPredicted})
percentil05 = svc_sinAnomalias.quantile(0.05)
percentil95 = svc_sinAnomalias.quantile(0.95)
svc_sinAnomalias = svc_sinAnomalias.drop(svc_sinAnomalias[(svc_sinAnomalias.svc
< percentil05.svc) | (svc_sinAnomalias.svc > percentil95.svc)].index)
svc_media_todo = svc_sinAnomalias.mean()
svc_sd_todo = svc_sinAnomalias.std()
```

Algoritmo 4.4: Fragmento de la extracción de las anomalías junto con la obtención de las medias y desviaciones estándar de las precisiones calculadas en el bucle.

```
param_grid = {
    'hidden_layer_sizes': [(20,), (50,), (100,), (200,)],
    'activation': ['identity', 'logistic', 'tanh', 'relu'],
    'solver': ['lbfgs', 'sgd', 'adam'],
    'alpha': [0.0001, 0.001, 0.001, 0.1, 0.32, 1, 3.16, 10, 32],
    'learning_rate': ['constant', 'invscaling', 'adaptive'],
    'max_iter': [100, 200, 500, 1000, 2000]
}
grid_search = GridSearchCV(mlp_classifier, param_grid=param_grid, cv=5)
grid_search.fit(values_train, targets_train)
print("-----")
print("Best parameters: ", grid_search.best_params_)
print("Best score: ", grid_search.best_score_)
print("-----")
```

Algoritmo 4.5: Fragmento que muestra el uso de la función “GridSearchCV” al clasificador MLP.

Capítulo 5 – Resultados obtenidos

En este capítulo se ilustran los resultados obtenidos tanto del análisis estadístico realizado, como del uso de los diferentes clasificadores, y el estudio del peso que tienen algunos de los atributos del etiquetado.

Cabe destacar que los resultados que se muestran en este capítulo hacen referencia únicamente al estudio de los 610 lanzamientos de falta. Por lo tanto, en este prototipo al no contar con una muestra amplia, se pueden dar casos en los cuales los resultados no sean del todo representativos; ya que, se trata de una primera aproximación al estudio del lance del juego en cuestión.

5.1 - Estadísticas

Lo primero que se ha obtenido, es la distribución de los diferentes resultados de los lanzamientos de falta. En la tabla 5.1, se puede apreciar que el 50% de los lanzamientos acabaron en gol, y el otro 50% no, es decir que, de los 610 lanzamientos, 305 acabaron en gol y los restantes 305 fueron fallados.

	Resultado Final	Lanzamientos	Porcentaje(%)
0	Goles	305	50.00
1	Fallos	305	50.00
2	Totales	610	100.00

Tabla 5.1: Distribución de goles y fallos, y sus respectivos porcentajes.

Existen cinco tipos de fallos que se han producido, tal y como se ha comentado previamente. Estos son: que el portero haga una parada, que el lanzamiento se vaya fuera, que se vaya al palo (bien sea al larguero o a los palos), que choque en la barrera o que lo despeje o rebote en algún otro jugador que no se encontrara en la barrera. La tabla 5.2

detalla de qué forma se han distribuido todos los lanzamientos fallidos, es decir, cuántas ocurrencias ha tenido cada tipo de fallo.

	Tipo de fallo	Lanzamientos	Porcentaje(%)
0	Parada	155	50.82
1	Fuera	73	23.93
2	Palo	25	8.20
3	Barrera	41	13.44
4	Despeje	11	3.61
5	Totales	305	100.00

Tabla 5.2: Tipos de fallos y las veces que se dio cada uno de estos con sus respectivos porcentajes.

Una vez se ha visto la imagen global del reparto de los desenlaces de los tiros libres, se han estudiado las estadísticas acerca de los goles y las paradas realizadas en las 610 jugadas.

5.1.1 - Goles

Se ha comprobado como varía el número de goles marcados en relación con el número de lanzamientos y las diferentes situaciones que se han dado.

La tabla 5.3 muestra los datos con respecto al total de tiros que se hace a cada lado de la portería, cuántos de estos acaban en gol, cuántos no, y sus respectivos porcentajes.

	Direccion de la falta	Lanzamientos	Goles	Fallos	Goles(%)	Fallos(%)
0	Arriba Izquierda	190	104	86	54.74	45.26
1	Arriba Centro	46	11	35	23.91	76.09
2	Arriba Derecha	184	109	75	59.24	40.76
3	Abajo Izquierda	69	39	30	56.52	43.48
4	Abajo Centro	17	5	12	29.41	70.59
5	Abajo Derecha	62	37	25	59.68	40.32
6	Totales	568	305	263	53.70	46.30

El número de faltas de las cuales se desconoce su dirección es de un total de: 42
Que suma a un total de lanzamientos de: 610

Tabla 5.3: Goles de falta basado en la dirección de esta.

Se ha obtenido que hay un mayor número de lanzamientos dirigidos a la parte superior de la portería, especialmente, a las dos escuadras, a la derecha y a la izquierda. Además, las zonas de la portería a la que menos dirigen las faltas son a la zona central, tanto la parte alta como la parte baja. Al mismo tiempo, estas son las zonas en las que menos goles se marcan por número de lanzamientos, siendo ambos porcentajes de goles por intentos inferior al 30%.

Estos resultados se pueden observar en la figura 5.1, que ilustra el porcentaje de los tiros que acaban en gol dependiendo de la zona de la portería a la que va dirigida la falta. Basado en dichos porcentajes, se podría recalcar que la mejor opción para el lanzador sería tirar la falta hacia alguno de los lados, y más especialmente hacia el lado derecho. Esto se debe a que si el jugador chuta a la parte derecha de la portería, casi el 60% de sus tiros dirigidos hacia ese lado acabarían en gol.

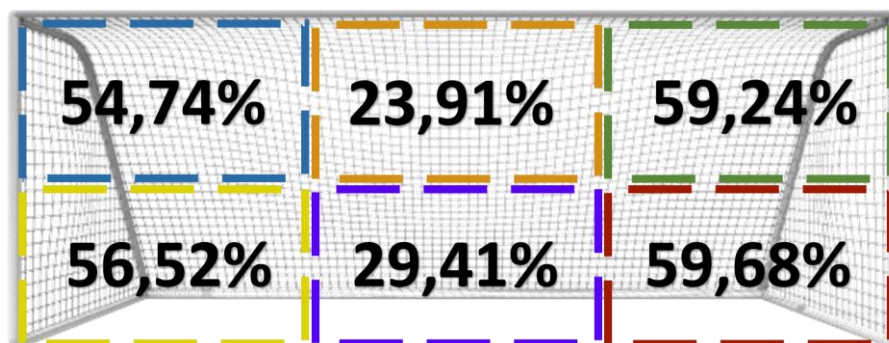


Figura 5.1: Porcentaje de los lanzamientos que acaban en gol en base a la dirección de la falta.

Imagen modificada manualmente, obtenida de: porteria de futbol png - Bing images

La tabla 5.4 refleja la información de los libres directos teniendo en cuenta donde se ubica cada uno de estos. Es decir, indica el número de tiros, goles y fallos, junto con sus porcentajes, dependiendo de si la falta estaba situada en el lado izquierdo, derecho, o central del campo.

	Lado de la falta	Lanzamientos	Goles	Fallos	Goles(%)	Fallos(%)
0	Izquierda	154	71	83	46.10	53.90
1	Centro	324	173	151	53.40	46.60
2	Derecha	132	61	71	46.21	53.79
3	Totales	610	305	305	50.00	50.00

Tabla 5.4: Goles de falta basado en la ubicación de esta.

Se ha comprobado que dependiendo de donde esté ubicada la falta existe una mayor probabilidad de gol. Esto se debe a que más de la mitad de las faltas lanzadas desde la zona central, han acabado en gol. Apenas hay diferencia entre las faltas lanzadas desde la izquierda, y aquellas lanzadas desde la derecha.

La figura 5.2 muestra visualmente el porcentaje de los goles que se han marcado dependiendo de la ubicación de la falta. Se puede apreciar lo comentado anteriormente de que apenas hay diferencia entre una falta por la derecha o por la izquierda, pero que existe una mayor probabilidad de marcar si la falta se encuentra centrada.

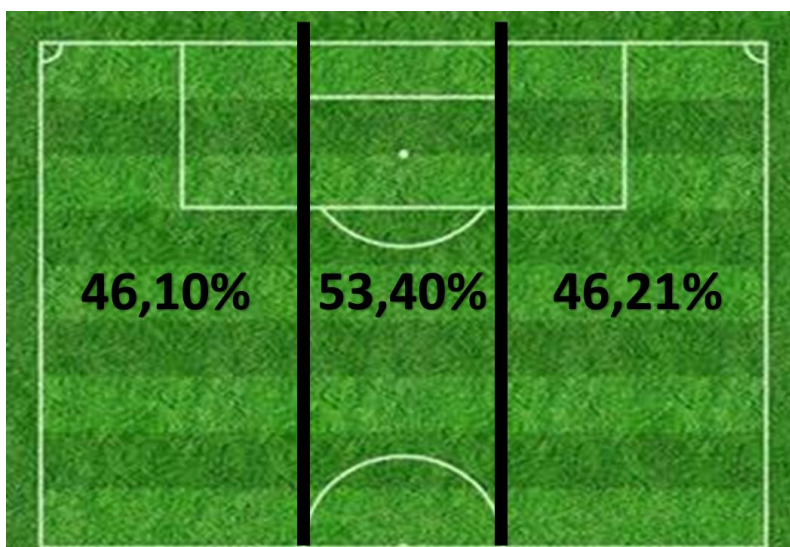


Figura 5.2: Porcentaje de los lanzamientos que acaban en gol en base al lado de la falta.

Imagen modificada manualmente, obtenida de: campo de futbol png - Bing images

La tabla 5.5 refleja las faltas marcadas y falladas, pero esta vez basado en la distancia a la que se encuentra esta. Se puede observar que las faltas cercanas tienen una mayor probabilidad de acabar en gol que las faltas lejanas.

	Distancia de la falta	Lanzamientos	Goles	Fallos	Goles(%)	Fallos(%)
0	Cerca	391	203	188	51.92	48.08
1	Lejos	219	102	117	46.58	53.42
2	Totales	610	305	305	50.00	50.00

Tabla 5.5: Goles de falta basado en la distancia de esta.

Tal y como se indicó, y se muestra en la figura 5.3, existe una mayor probabilidad de marcar las faltas cercanas en comparación con las lejanas. Ya que, de las faltas cercanas, se marcaron el 51,92% de los intentos, mientras que, de las lejanas, se marcaron el 46,58%.

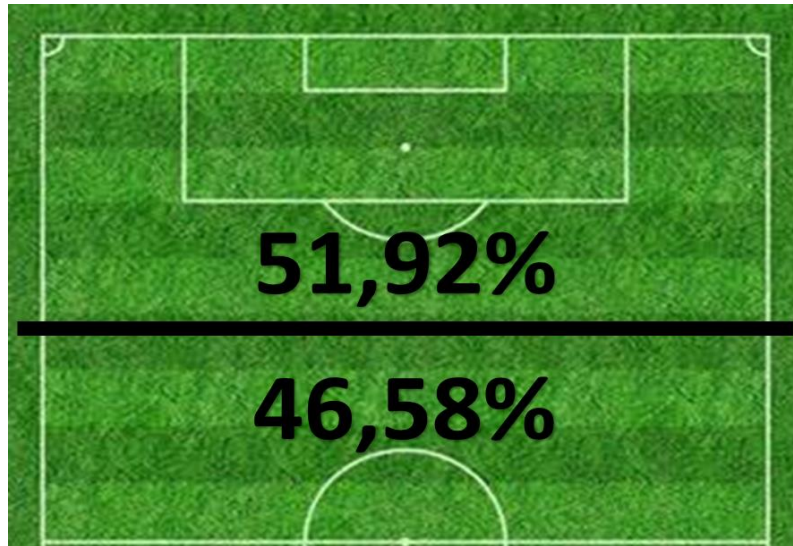


Figura 5.3: Porcentaje de los lanzamientos que acaban en gol en base a la distancia de la falta.

Imagen modificada manualmente, obtenida de: campo de futbol png - Bing images

Según los datos mostrados hasta el momento, la falta ideal, sería aquella falta cercana, que se encuentre centrada, y que el lanzador decida tirarla hacia uno de los lados, y en especial, al lado derecho de la portería.

La tabla 5.6 indica la información relativa a las faltas lanzadas teniendo en cuenta su ubicación en el campo, y su distancia con respecto a la portería.

FALTAS DESDE EL LADO IZQUIERDO:

	Distancia de la falta	Lanzamientos	Goles	Fallos	Goles(%)	Fallos(%)
0	Cerca	116	59	57	50.86	49.14
1	Lejos	38	12	26	31.58	68.42
2	Totales	154	71	83	46.10	53.90

FALTAS DESDE EL LADO CENTRAL:

	Distancia de la falta	Lanzamientos	Goles	Fallos	Goles(%)	Fallos(%)
0	Cerca	189	107	82	56.61	43.39
1	Lejos	135	66	69	48.89	51.11
2	Totales	324	173	151	53.40	46.60

FALTAS DESDE EL LADO DERECHO:

	Distancia de la falta	Lanzamientos	Goles	Fallos	Goles(%)	Fallos(%)
0	Cerca	86	37	49	43.02	56.98
1	Lejos	46	24	22	52.17	47.83
2	Totales	132	61	71	46.21	53.79

Tabla 5.6: Goles de falta basado en la ubicación y distancia de esta.

Si se examinan los porcentajes de goles obtenidos, aquellos de la figura 5.4, no existe un patrón claro, debido a que no son solamente las faltas cercanas aquellas que tienen un mayor número de goles con respecto al número de tiros realizados, ya que las faltas lejanas ubicadas en la derecha son las segundas que mayor probabilidad de gol presentan. Esto se debe a que, de todos los tiros realizados desde esa zona del campo, el 52,17% de los lanzamientos acabaron en gol. Sin embargo, la misma zona, pero por el lado izquierdo, es la zona con menor probabilidad de gol, porque tan solo el 31,58% de los tiros realizados acabaron siendo exitosos.

Lo que sí que demuestra la figura 5.4, es lo comentado anteriormente, que la ubicación de la falta ideal sería en la zona central del campo y cercanas a la portería.

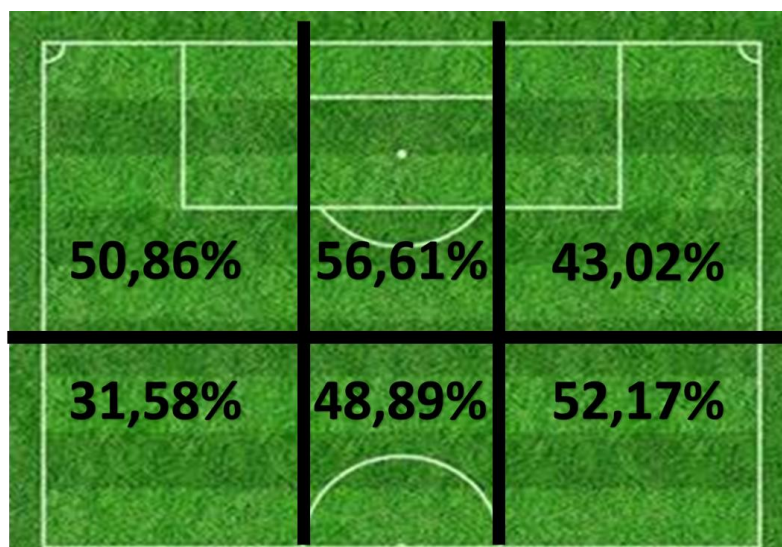


Figura 5.4: Porcentaje de los lanzamientos que acaban en gol en base a la ubicación y distancia de la falta.

Imagen modificada manualmente, obtenida de: campo de futbol png - Bing images

La tabla 5.7 distingue entre los lanzamientos realizados con la pierna derecha o con la izquierda. Ilustra como apenas existe diferencia entre ambos atributos, ya que, en ambos casos, el porcentaje de goles ronda el 50%.

Lo normal, es que un jugador golpee una falta con su pierna más hábil, y rara vez se observa lo contrario. Por ende, la tabla 5.7 demuestra, que en general, apenas hay diferencia entre la efectividad que podría tener un diestro o un zurdo a la hora de tirar una falta. Dado que tanto los jugadores diestros como zurdos tienen una probabilidad de marcar de alrededor del 50%.

	Pierna del lanzador	Lanzamientos	Goles	Fallos	Goles(%)	Fallos(%)
0	Zurda	221	110	111	49.77	50.23
1	Diestra	389	195	194	50.13	49.87
2	Totales	610	305	305	50.00	50.00

Tabla 5.7: Goles de falta dependiendo de la pierna de golpeo del jugador.

La tabla 5.8 muestra los datos de los lanzamientos de los jugadores zurdos (aquellos que tiran con la pierna izquierda) en base a la ubicación y la distancia de la falta.

ZURDOS TIRANDO FALTAS DESDE EL LADO IZQUIERDO:						
	Distancia de la falta	Lanzamientos	Goles	Fallos	Goles(%)	Fallos(%)
0	Cerca	15	5	10	33.33	66.67
1	Lejos	4	0	4	0.00	100.00
2	Totales	19	5	14	26.32	73.68

ZURDOS TIRANDO FALTAS DESDE EL LADO CENTRAL:						
	Distancia de la falta	Lanzamientos	Goles	Fallos	Goles(%)	Fallos(%)
0	Cerca	80	51	29	63.75	36.25
1	Lejos	34	13	21	38.24	61.76
2	Totales	114	64	50	56.14	43.86

ZURDOS TIRANDO FALTAS DESDE EL LADO DERECHO:						
	Distancia de la falta	Lanzamientos	Goles	Fallos	Goles(%)	Fallos(%)
0	Cerca	61	27	34	44.26	55.74
1	Lejos	27	14	13	51.85	48.15
2	Totales	88	41	47	46.59	53.41

Tabla 5.8: Goles de falta de jugadores zurdos basado en la ubicación y distancia de esta.

Al apreciar los datos obtenidos en la tabla 5.8, se pueden obtener el número de faltas que han acabado en gol, dependiendo de a qué zona del campo hace referencia. Una ilustración de esto se expone en la figura 5.5. Los porcentajes de la figura 5.5, indican dónde tienen mayor probabilidad de marcar aquellos jugadores que tiren la falta con la pierna izquierda.

Por lo tanto, se manifiesta que a los jugadores que vayan a tirar con la pierna zurda, no se le dan bien las faltas situadas en el lado izquierdo del campo, ya que únicamente marcaron

el 26,32% de todas las faltas lanzadas desde esa banda, marcaron cinco faltas cercanas (un 33%) y ninguna lejana (un 0%), de un total de las 19 faltas lanzadas desde la izquierda.

Las faltas desde la derecha tienen una mayor probabilidad de acabar en gol, un 46,59%. El dato interesante de las faltas ubicadas en la derecha es que, las lejanas tienen una mayor probabilidad de gol, un 51,85%, mientras que las cercanas tienen un 44,26%.

En la zona del medio es donde más goles marcaron, con un total de 64 tantos; y significa, que marcaron el 56,14% de todas las faltas lanzadas desde el centro. La falta favorita de los zurdos es aquella cercana que se encuentra en la zona central, ya que obtuvieron una tasa de acierto del 63,75%. Sin embargo, con respecto a los libres directos situados en el centro, el porcentaje de acierto es de tan solo el 38,24%

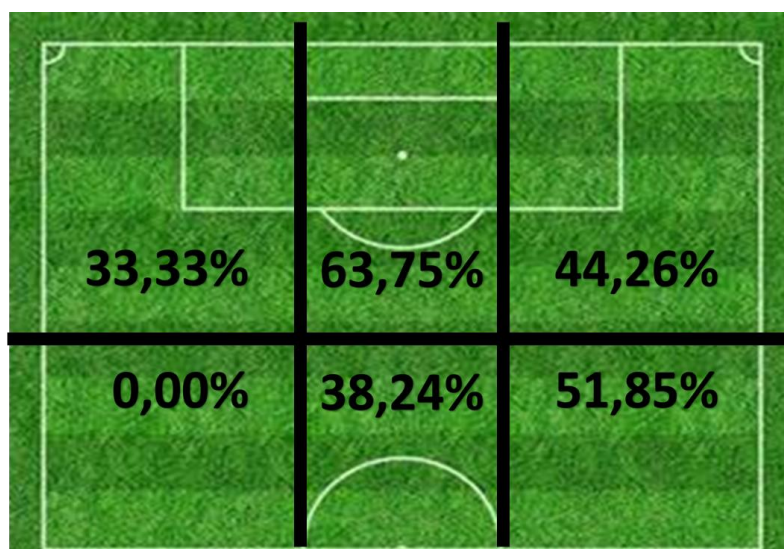


Figura 5.5: Porcentaje de los lanzamientos de jugadores zurdos que acaban en gol en base a la ubicación y distancia de la falta.

Imagen modificada manualmente, obtenida de: campo de futbol png - Bing images

En la tabla 5.9 se observan los datos de los tiros libre, pero en este caso, de los jugadores diestros (aquellos que tiran con la pierna derecha) y en base a la ubicación y la distancia de la falta. En ella se puede apreciar cuál es la zona del campo que mejor se les da a los jugadores que tiran la falta con la derecha, y al mismo tiempo, cuál es la zona del campo en la que desempeñan una peor labor en relación con los lanzamientos de falta.

DIESTROS TIRANDO FALTAS DESDE EL LADO IZQUIERDO:						
	Distancia de la falta	Lanzamientos	Goles	Fallos	Goles(%)	Fallos(%)
0	Cerca	101	54	47	53.47	46.53
1	Lejos	34	12	22	35.29	64.71
2	Totales	135	66	69	48.89	51.11

DIESTROS TIRANDO FALTAS DESDE EL LADO CENTRAL:						
	Distancia de la falta	Lanzamientos	Goles	Fallos	Goles(%)	Fallos(%)
0	Cerca	109	56	53	51.38	48.62
1	Lejos	101	53	48	52.48	47.52
2	Totales	210	109	101	51.90	48.10

DIESTROS TIRANDO FALTAS DESDE EL LADO DERECHO:						
	Distancia de la falta	Lanzamientos	Goles	Fallos	Goles(%)	Fallos(%)
0	Cerca	25	10	15	40.00	60.00
1	Lejos	19	10	9	52.63	47.37
2	Totales	44	20	24	45.45	54.55

Tabla 5.9: Goles de falta de jugadores diestros basado en la ubicación y distancia de esta.

La tabla 5.9, indica que los diestros logran el mayor número de goles con respecto al número de intentos en la zona central del campo, con un 51,90%; mientras que la banda derecha contiene los peores registros, con un 45,45% de acierto; y la banda izquierda cuenta con un porcentaje de goles del 48,89%.

Los porcentajes del número de goles en base al número de tiros se puede percibir de forma visual en la figura 5.6. En ella se puede apreciar, al igual que en el caso de los zurdos, que la zona con peores resultados es la zona más lejana de la izquierda, cuyo porcentaje de gol es del 35,29%. Sin embargo, para los diestros, la zona de mayor efectividad no es como los zurdos, en el centro y cerca, sino que es cerca, pero por el lado de la izquierda, un 53,47%.

Al igual que con los zurdos, los resultados vuelven a sorprender por el lado derecho, ya que el porcentaje de tantos de las faltas lejanas vuelve a ser superior al de las cercanas por dicho lateral, un 12,63% más.

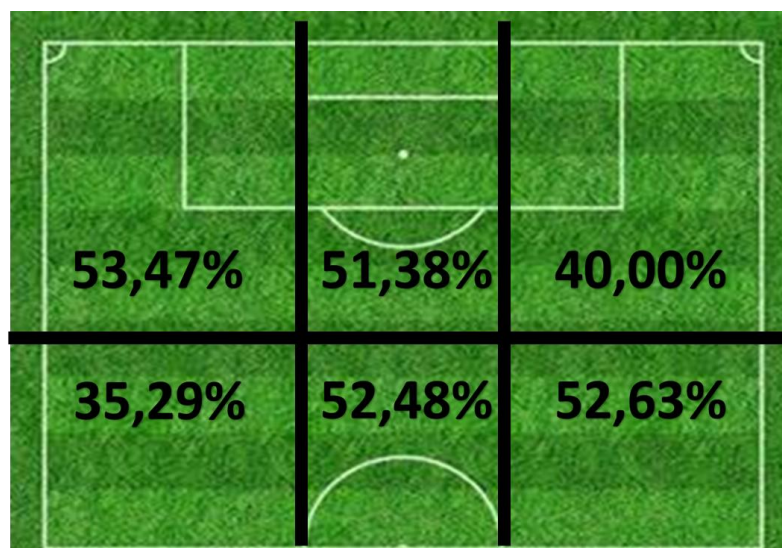


Figura 5.6: Porcentaje de los lanzamientos de jugadores zurdos que acaban en gol en base a la ubicación y distancia de la falta.

Imagen modificada manualmente, obtenida de: campo de futbol png - Bing images

Si se realiza una comparación entre los lanzamientos de los zurdos, y el de los diestros, se puede distinguir, que los diestros en general tienen mejores registros. Esto se debe a que, de las seis zonas del campo identificadas, tienen mayor porcentaje de goles, en cuatro de ellas. En todos los lanzamientos de los diestros exceptuando en una de las zonas que tienen un 35,29% de acierto, se encuentran entre el 40% y el 55%. Sin embargo, en los lanzamientos de los jugadores zurdos, existe mayor diferencia entre los porcentajes, ya que tres de sus valores se encuentran por debajo del rango 40%-55%, solamente dos se encuentran dentro de dicho rango, y hay uno, el último, que está situado por encima, un 63,75%.

Por lo tanto, si se tuviese en cuenta únicamente estos 610 tiros de falta, a raíz de dicha comparación, un entrenador o los propios jugadores, podrían tomar la decisión, de que los jugadores diestros serán los que tiren casi todas las faltas directas, exceptuando las

que sean cercanas, y estén tanto en el centro como en la derecha. Esto es, porque es en estas zonas donde los zurdos tienen mejores registros que los diestros. En las faltas cercanas del medio los zurdos tienen un porcentaje de acierto del 63,75%, mientras que los diestros del 51,38%. Y con respecto a las faltas de cerca, pero por la derecha, los zurdos han obtenido una eficacia del 44,26%, mientras que los diestros del 40,00%. Si bien, en las restantes cuatro zonas, los diestros han tenido mayores porcentajes de gol.

5.1.2 - Paradas

El otro estudio que se ha llevado a cabo es la relación entre el número de paradas con respecto al número de lanzamientos.

Al igual que con los goles, se ha comenzado con las estadísticas basadas en la dirección del lanzamiento. La tabla 5.10 muestra los datos obtenidos sobre las paradas teniendo en cuenta hacia donde ha dirigido el balón el lanzador.

	Dirección de la falta	Lanzamientos	Paradas	No Paradas	Parada(%)	No Paradas(%)
0	Arriba Izquierda	190	47	143	24.74	75.26
1	Arriba Centro	46	17	29	36.96	63.04
2	Arriba Derecha	184	41	143	22.28	77.72
3	Abajo Izquierda	69	22	47	31.88	68.12
4	Abajo Centro	17	12	5	70.59	29.41
5	Abajo Derecha	62	16	46	25.81	74.19
6	Totales	568	155	413	27.29	72.71

El número de faltas de las cuales se desconoce su dirección es de un total de: 42
Que suma a un total de lanzamientos de: 610

Tabla 5.10: Paradas de falta basado en la dirección de esta.

Los porcentajes de paradas que se muestran en la tabla 5.10, también se pueden observar en la figura 5.7. Esta indica que el mayor porcentaje de paradas fue en aquellas faltas dirigidas al medio, en especial a la parte baja del medio, en donde el porcentaje de atajadas

fue del 70,59%. También muestra que las que le resultan más complicadas a los porteros son aquellas que van a las esquinas superiores, bien sea la izquierda o la derecha, ya que lograron, respectivamente, un 24,74% y un 22,28% de paradas. Con respecto a los balones que iban dirigidos a la parte baja de la portería los datos muestran que los porteros obtuvieron mayor éxito parando aquellos que iban hacia la izquierda.

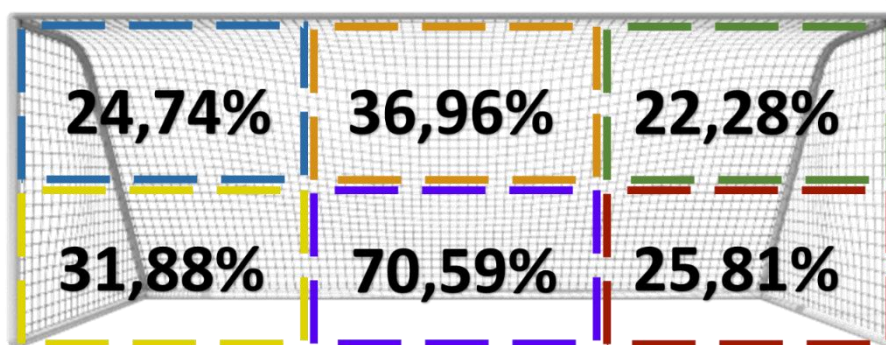


Figura 5.7: Porcentaje de los lanzamientos que acaban en parada en base a la dirección de la falta.

Imagen modificada manualmente, obtenida de: porteria de futbol png - Bing images

Lo siguiente ha sido comprobar el número de paradas con respecto a la localización de la falta, que se puede observar en la tabla 5.11. El porcentaje de las paradas realizadas según el lado en el campo se encuentra en la figura 5.8; donde indica que no existe mucha diferencia de paradas en base a donde estuviese ubicado el tiro libre, ya que las faltas que se han tirado desde el lado izquierdo y el lado central cuentan con una probabilidad de parada de alrededor del 25%, un 25,97% y un 25,62% respectivamente. Aunque las faltas desde el lado derecho hayan sido las que peores registros hayan proporcionado, fueron detenidos el 24,24% de los tiros; por lo que no presenta mucha diferencia.

	Lado de la falta	Lanzamientos	Paradas	No Paradas	Paradas(%)	No Paradas(%)
0	Izquierda	154	40	114	25.97	74.03
1	Centro	324	83	241	25.62	74.38
2	Derecha	132	32	100	24.24	75.76
3	Totales	610	155	455	25.41	74.59

Tabla 5.11: Paradas de falta basado en la ubicación de esta.

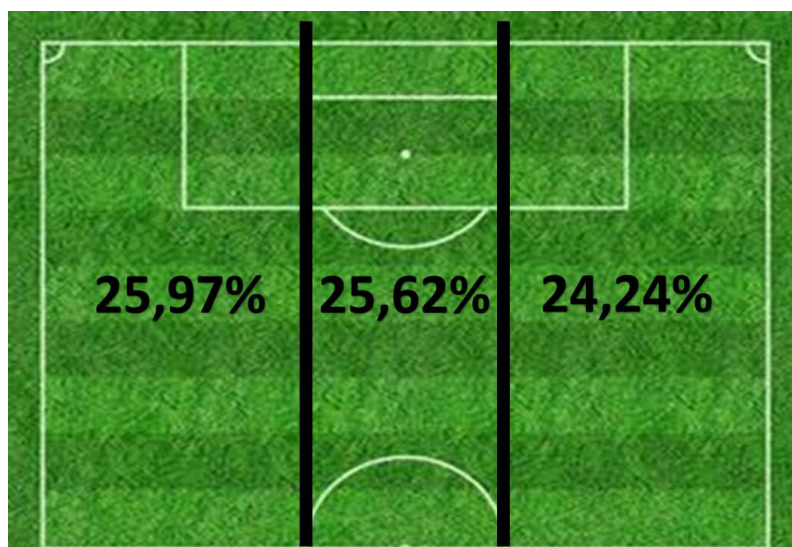


Figura 5.8: Porcentaje de los lanzamientos que acaban en parada en base al lado de la falta.

Imagen modificada manualmente, obtenida de: campo de futbol png - Bing images

Con respecto a cómo varían las paradas según la distancia, la tabla 5.12 contiene dicha información. Parece ser, que, al igual que teniendo en cuenta únicamente la ubicación de la falta, la distancia tampoco afecta significativamente en el porcentaje de paradas realizadas. Esto se puede observar en la figura 5.9. Se aprecia que, sea un lanzamiento lejano o cercano, ambos porcentajes rondan el 25%, un 25,32% para las cercanas, y un 25,57% para las lejanas. Esto indica que la distancia de la falta no provoca mucha variación en el desempeño de los guardametas, y en su efectividad.

	Distancia de la falta	Lanzamientos	Paradas	No Paradas	Paradas(%)	No Paradas(%)
0	Cerca	391	99	292	25.32	74.68
1	Lejos	219	56	163	25.57	74.43
2	Totales	610	155	455	25.41	74.59

Tabla 5.12: Paradas de falta basado en la distancia de esta.

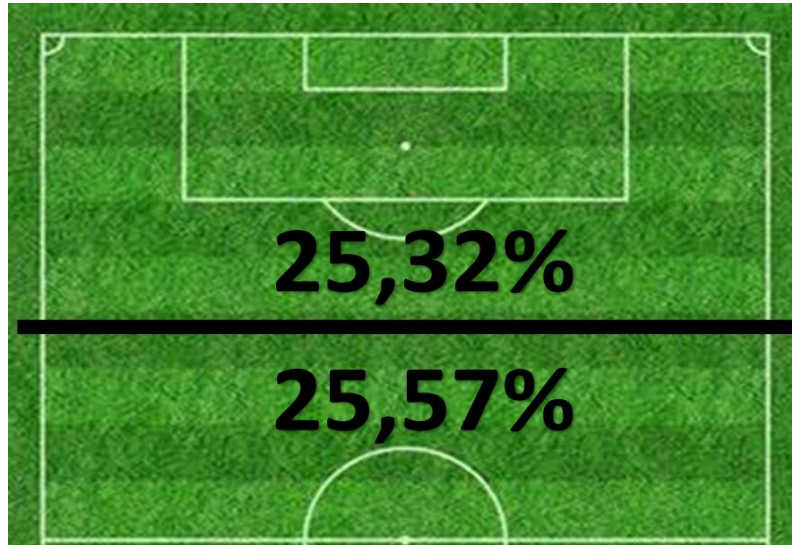


Figura 5.9: Porcentaje de los lanzamientos que acaban en parada en base a la distancia de la falta.

Imagen modificada manualmente, obtenida de: campo de futbol png - Bing images

Según los datos recogidos hasta ahora, si tenemos en cuenta lo recabado por separado, se tiene que las faltas que resultan más sencillas de detener para los porteros son aquellas faltas lejanas, situadas en el lado izquierdo del campo, y que van dirigidas hacia la zona central inferior de la portería.

La tabla 5.13, recoge los datos de las paradas realizadas teniendo en cuenta la combinación de las características de la distancia y de la ubicación de la falta.

FALTAS DESDE EL LADO IZQUIERDO:

	Distancia de la falta	Lanzamientos	Paradas	No Paradas	Paradas(%)	No Paradas(%)
0	Cerca	116	32	84	27.59	72.41
1	Lejos	38	8	30	21.05	78.95
2	Totales	154	40	114	25.97	74.03

FALTAS DESDE EL LADO CENTRAL:

	Distancia de la falta	Lanzamientos	Paradas	No Paradas	Paradas(%)	No Paradas(%)
0	Cerca	189	42	147	22.22	77.78
1	Lejos	135	41	94	30.37	69.63
2	Totales	324	83	241	25.62	74.38

FALTAS DESDE EL LADO DERECHO:

	Distancia de la falta	Lanzamientos	Paradas	No Paradas	Paradas(%)	No Paradas(%)
0	Cerca	86	25	61	29.07	70.93
1	Lejos	46	7	39	15.22	84.78
2	Totales	132	32	100	24.24	75.76

Tabla 5.13: Paradas de falta basado en la ubicación y distancia de esta.

La figura 5.10, ilustra los porcentajes obtenidos de la tabla 5.13, según la zona del campo a la que hace referencia.

Se puede apreciar, como no se dio el resultado que se esperaba al interpretar los datos de forma manera independiente. Ya que antes de realizar el último cálculo, y de tener en cuenta la distancia y la ubicación de forma conjunta, se había observado, que el mejor caso, en donde tendrían más probabilidad de atajar, eran aquellas faltas lejanas ubicadas en la izquierda. Nada más lejos de la realidad, ese tipo de falta es el segundo con menos porcentaje de paradas, con tan solo un 21,05%; solo encontrándose detrás de aquellas faltas lejanas lanzadas desde la derecha, que han proporcionado un porcentaje del 15,22%.

Los datos obtenidos reflejan un resultado que se podría considerar lógico, y es que las faltas cercanas que se encuentran centradas son las que menos probabilidad de parada presentan, con un 22,22%. Las que más son aquellas centradas de larga distancia, con un 30,37%. Lo curioso, es que en ambos laterales la probabilidad de atajar aumenta conforme

más cerca se encuentra la falta. Esto se debe a que la probabilidad de parar las faltas cercanas que se encuentran en los lados es superior a las de parar las lejanas de dichos lados; que, como se ha destacado previamente, son las que menos probabilidad de ser atajadas presentan, un 15,22% y un 21,05%. Mientras que las laterales de corta distancia tienen un 29,07% las de la derecha, y las de la izquierda un 27,59%.

Estos porcentajes obtenidos no indican que el resto de lanzamientos hayan terminado en gol, sino que no los ha detenido el portero. Por lo tanto, el hecho de que los porcentajes obtenidos en las faltas laterales lejanas sean inferiores, no significa que haya habido más goles desde ahí, solamente que no han sido parados; se han podido ir fuera o haber sido desviados por la barrera, por ejemplo.

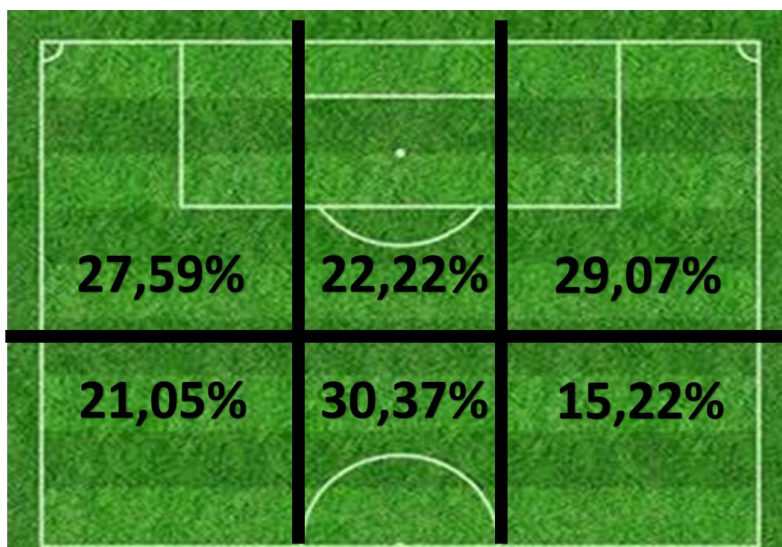


Figura 5.10: Porcentaje de los lanzamientos que acaban en parada en base a la ubicación y distancia de la falta.

Imagen modificada manualmente, obtenida de: campo de futbol png - Bing images

Una vez analizados los factores que no son controlados por el portero, se han estudiado las acciones realizadas por los diferentes porteros, como son: su colocación inicial dentro de la portería, y hacia donde se lanza para intentar detener el balón. Para ello, debido a que se trata de los actos del portero, se hará referencia a la visión del portero, esto quiere decir, que cuando se haga referencia a un lado de la portería, se hará desde la visión que

tiene el guardameta. Por lo que, si se nombra el lado izquierdo, se está haciendo referencia al lado izquierdo del portero.

La tabla 5.14 contiene los datos referentes a los lanzamientos, y las distintas posiciones en las que se puede colocar el portero de cara a intentar parar la falta. Tal y como se comentó, los lados de dicha tabla hacen referencia a los lados del portero, el lado izquierdo, es el lado izquierdo del portero.

Se puede apreciar en la tabla, que la zona en el que más se situaron los porteros es la central, pues se ubicaron ahí en 291 ocasiones de 610, es decir, en el 47,70% de las veces. Mientras que el lado en el que menos se situaron, es en el derecho, que tan sólo se ubicaron en dicha zona en 135 lanzamientos, el 22,13% de las faltas.

	Ubicacion inicial	Lanzamientos	Paradas	No Paradas	Paradas(%)	No Paradas(%)
0	Izquierda	184	47	137	25.54	74.46
1	Centro	291	84	207	28.87	71.13
2	Derecha	135	24	111	17.78	82.22
3	Totales	610	155	455	25.41	74.59

Tabla 5.14: Paradas de falta basado en la colocación inicial del portero dentro de la portería.

La figura 5.11, muestra los porcentajes de paradas que realizaron los guardametas en función de donde se colocaron inicialmente. Ilustra, que la mejor zona para colocarse es la zona central, debido a que es aquí donde obtuvieron el mayor número de paradas por número de intentos, un porcentaje de paradas del 28,87%. En las ocasiones en las que se pusieron en el lado derecho es cuando desempeñaron peor labor, pues pararon el 17,78% de las faltas, mientras que cuando estaban en el izquierdo, pararon el 25,54% de los lanzamientos.

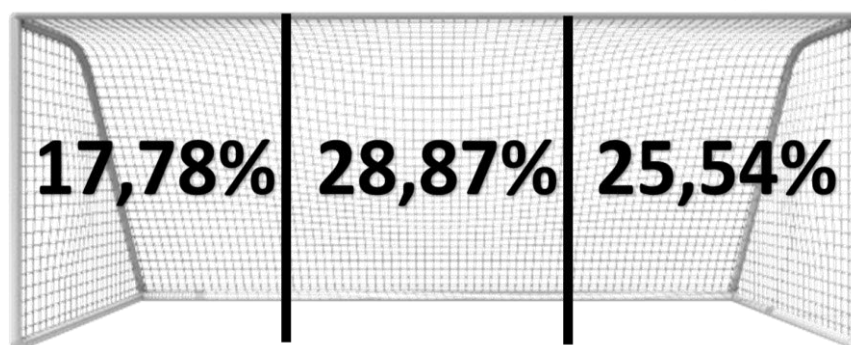


Figura 5.11: Porcentaje de los lanzamientos que acaban en parada en base a la colocación inicial del portero.

Imagen modificada manualmente, obtenida de: porteria de futbol png - Bing images

La tabla 5.15 indica las veces en las que los porteros se lanzaron a los diferentes lados para intentar detener la falta, y el número de veces que consiguieron detener el tiro, basado en dicha dirección. Al igual que sucedió en el caso anterior, en la tabla 5.14, y tal y como se había comentado anteriormente, los lados son con respecto al portero. Es decir que, si se lanza hacia la izquierda, significa que se lanza hacia el lado izquierdo del portero.

Se aprecia, que el acto más repetitivo es no lanzarse, esto significa, que o bien los porteros identificaron que por mucho que se tirasen, no iban a llegar debido a un golpeo imparables, o observaron que el balón se iba fuera y no era necesario lanzarse, o la última opción, que los porteros se encontraban bien colocados y no tuvieron la necesidad de lanzarse para detener el lanzamiento. En el 38,20% de las faltas, los porteros no se lanzaron, ya que no se tiraron en 233 faltas de las 610. Hacia el lado al que menos se lanzaron fue al izquierdo, pues solo se lanzaron en 176 ocasiones, mientras que, al derecho, se tiraron en 201 faltas.

	Lado al que se lanza	Lanzamientos	Paradas	No Paradas	Paradas(%)	No Paradas(%)
0	Hacia la Izquierda	176	48	128	27.27	72.73
1	No se lanza	233	37	196	15.88	84.12
2	Hacia la Derecha	201	70	131	34.83	65.17
3	Totales	610	155	455	25.41	74.59

Tabla 5.15: Paradas de falta basado en la dirección en la que se lanza el portero.

En la figura 5.12 se ilustran los porcentajes de parada obtenidos en base a si los porteros se lanzaron a su izquierda o a su derecha. Según los datos recabados, estos indican que a los porteros se les da mejor lanzarse hacia su derecha que hacia su izquierda, pues cuando se lanzan a la derecha, tienen una tasa de parada del 34,83%. Mientras que cuando lo hacen hacia la izquierda, han obtenido un 27,27% de éxito.

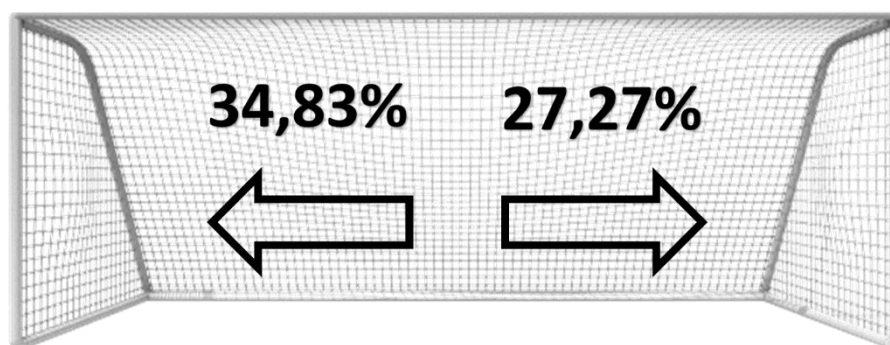


Figura 5.12: Porcentaje de los lanzamientos que acaban en parada en base a la dirección en la que se lanza el portero.

Imagen modificada manualmente, obtenida de: portería de fútbol png - Bing images

A pesar de que en la figura 5.12 no se indique el porcentaje de paradas de aquellas faltas en las que los porteros decidieron no lanzarse, en la tabla 5.15 se puede observar dicho valor. En las faltas en las que los porteros no se lanzaron, pararon 37 faltas de 233 faltas en las que no se lanzaron. Esto es un porcentaje de parada del 15,88%. Esto muestra, que el gran número de estas faltas, en las que no se lanzaron, fue porque o bien decidieron que era un lanzamiento imparable, o porque tenían la certeza de que el balón no iba a acabar dentro de la portería, es decir, que se iba fuera.

Por lo tanto, los porteros han obtenido un mayor porcentaje de paradas cuando se han lanzado hacia su derecha. Esto puede indicar que o bien el portero se coloca mejor cuando no está en la derecha, estando en el centro o en la izquierda, por lo que al lanzarse a su derecha llegan mejor al balón; o bien, que los porteros en general se lanzan mejor hacia su derecha.

5.2 - Clasificadores

Siguiendo lo comentado en el capítulo 4, donde se detalló el desarrollo del trabajo, se han utilizado diversos clasificadores para la predicción del resultado. Se ha intentado predecir, en un primer lugar, si el lanzamiento acabaría en gol; y en segundo lugar, si el lanzamiento lo detendría el portero. Para los resultados de cada predicción, se ha realizado un bucle de 50 iteraciones, a raíz del cual se han obtenido 50 predicciones con cada clasificador, se han eliminado los valores atípicos obtenidos, y se ha calculado la media y la desviación estándar de las predicciones proporcionadas por cada clasificador.

Dado que se trabaja con una muestra reducida, los resultados proporcionados por los clasificadores dependerán de la división realizada entre los datos empleados para el entrenamiento y aquellos destinados para las pruebas. Dependiendo de dicha división, se obtendrán diferentes resultados con pronósticos más o menos precisos.

5.2.1 - Resultados iniciales

En la tabla 5.16 se pueden apreciar los resultados obtenidos de los diferentes modelos, usando los parámetros por defecto. En la parte alta, se observan las medias y las desviaciones estándar obtenidas a la hora de predecir los goles, es decir, que los modelos han sido entrenados con el etiquetado cuya columna objetivo es la que indica si la falta terminó en gol o si no (“1” = gol, “0” = no gol). Mientras que la parte baja, muestra los resultados de entrenar a los modelos para predecir la parada, el etiquetado con la columna que indica si la falta la detuvo el portero o si no (“1” = parada, “0” = no parada).

Los resultados ilustran que el porcentaje de acierto en los pronósticos de las paradas es superior al de los goles. Dichos resultados se podrían considerar correctos, ya que en el dataset con el que se trabaja, la relación de faltas acabadas en gol y no gol, es del 50% cada una, por lo que el clasificador debería tener una precisión cercana al 50%. Mientras que la relación de las faltas paradas y no paradas es aproximadamente del 25% las paradas y el 75% las no paradas, por lo que la precisión debería rondar el dicho 75%.

Cada encabezado hace referencia a un clasificador diferente que se ha utilizado. Los clasificadores que se han empleado son:

- “MLP” - Multi-layer Perceptron
- “LR” - Logistic regression
- “KNN” - Nearest Neighbors (K-Nearest Neighbors)
- “SVC” - Support Vector Machines (Support Vector Classifier)
- “GAUSSIANNB” - Naive Bayes (GaussianNB)
- “DECTREE” - Decision Trees
- “RANDFOREST” - Random Forest
- “ADABOOST” - Ada Boost
- “QUADDISC” - Quadratic Discriminant

		Predicción de goles original:									
		Medida	MLP	LR	KNN	SVC	GAUSSIANNB	DECTREE	RANDFOREST	ADABOOST	QUADDISC
0	Media	0.63	0.59	0.60	0.62	0.58	0.59	0.64	0.63	0.61	
1	Desviacion estandar	0.03	0.03	0.03	0.03	0.04	0.03	0.03	0.02	0.03	

		Predicción de paradas original:									
		Medida	MLP	LR	KNN	SVC	GAUSSIANNB	DECTREE	RANDFOREST	ADABOOST	QUADDISC
0	Media	0.74	0.74	0.71	0.75	0.74	0.69	0.72	0.75	0.72	
1	Desviacion estandar	0.03	0.02	0.02	0.02	0.03	0.03	0.03	0.02	0.02	

Tabla 5.16: Medias y desviaciones estándar obtenidas de las predicciones de los goles y las paradas, resultados sobre 1.

5.2.2 - Resultados obtenidos al modificar datos de entrada

Con los diferentes clasificadores, no se ha pretendido maximizar la precisión de las predicciones, sino más bien, investigar cómo afectaría, a dicha precisión, la omisión de alguno o algunos de los atributos recabados durante el etiquetado de los videos. Se ha investigado acerca de la posible existencia de algún campo que contase con mayor peso que el resto a la hora de influir en la decisión de la predicción final. Las características que se han investigado son la dirección del lanzamiento de la falta, que hace referencia hacia qué zona de la portería tiró el lanzador; y hacia donde se lanzó el portero para intentar detener el tiro.

En las tablas 5.17 y 5.18 se detallan los resultados obtenidos al aplicar los clasificadores utilizando los hiperparámetros por defecto. El resultado que se muestra es la media de la precisión que ha tenido el clasificador a la hora de realizar las diferentes predicciones. Dicho de otra forma, la media de los porcentajes de cuántos pronósticos han sido acertados.

En dichas tablas, se ha utilizado la misma nomenclatura. Por lo que a continuación se explicará qué significan las distintas columnas que se pueden observar.

Cada columna hace referencia a la diferencia en los datos de entrada que se le han proporcionado a los clasificadores. Las diferentes columnas son:

- “Todo” - Es aquella en donde se ha hecho uso del fichero csv del etiquetado original. Es decir, con todas las columnas.
- “Sin Movimiento Portero” - Es aquella en donde, de los datos del fichero csv original, se ha quitado la columna que hacía referencia al lado hacia el cual se lanzaba el portero. Es decir, con todas las columnas excepto la denominada “lado_tira_portero”.
- “Sin Direccion” - Es aquella en donde, de los datos del fichero csv original, se ha quitado la columna que hacía referencia a la dirección hacia la que iba dirigido el lanzamiento. Es decir, con todas las columnas excepto la denominada “direccion”.
- “Sin Movimiento Portero y Sin Direccion” - Es aquella en donde, de los datos del fichero csv original, se ha quitado la columna que hacía referencia a la dirección hacia la que iba dirigido el lanzamiento, y la columna que hacía referencia al lado hacia el cual se lanzaba el portero. Es decir, con todas las columnas excepto las denominadas “direccion” y “lado_tira_portero”.

En la tabla 5.17, se muestran las medias y desviaciones estándar obtenidas por cada clasificador una vez realizadas las predicciones de los goles; y en la tabla 5.18, se muestran las medias y desviaciones estándar obtenidas por cada clasificador una vez realizadas las predicciones de las paradas.

MLP:					
	Medida	Todo	Sin Movimiento Portero	Sin Dirreccion	Sin Movimiento Portero y Sin Dirreccion
0	Media	0.63	0.60	0.61	0.58
1	Desviacion estandar	0.03	0.03	0.03	0.03
LR:					
	Medida	Todo	Sin Movimiento Portero	Sin Dirreccion	Sin Movimiento Portero y Sin Dirreccion
0	Media	0.59	0.58	0.56	0.57
1	Desviacion estandar	0.03	0.03	0.04	0.03
KNN:					
	Medida	Todo	Sin Movimiento Portero	Sin Dirreccion	Sin Movimiento Portero y Sin Dirreccion
0	Media	0.60	0.58	0.58	0.54
1	Desviacion estandar	0.03	0.03	0.03	0.03
SVC:					
	Medida	Todo	Sin Movimiento Portero	Sin Dirreccion	Sin Movimiento Portero y Sin Dirreccion
0	Media	0.62	0.58	0.60	0.59
1	Desviacion estandar	0.03	0.03	0.02	0.03
GAUSSIANNB:					
	Medida	Todo	Sin Movimiento Portero	Sin Dirreccion	Sin Movimiento Portero y Sin Dirreccion
0	Media	0.58	0.56	0.56	0.53
1	Desviacion estandar	0.04	0.03	0.03	0.03
DECTREE:					
	Medida	Todo	Sin Movimiento Portero	Sin Dirreccion	Sin Movimiento Portero y Sin Dirreccion
0	Media	0.59	0.60	0.58	0.56
1	Desviacion estandar	0.03	0.03	0.03	0.03
RANDFOREST:					
	Medida	Todo	Sin Movimiento Portero	Sin Dirreccion	Sin Movimiento Portero y Sin Dirreccion
0	Media	0.64	0.62	0.60	0.57
1	Desviacion estandar	0.03	0.03	0.03	0.02
ADABOOST:					
	Medida	Todo	Sin Movimiento Portero	Sin Dirreccion	Sin Movimiento Portero y Sin Dirreccion
0	Media	0.63	0.62	0.59	0.58
1	Desviacion estandar	0.02	0.03	0.02	0.03
QUADDISC:					
	Medida	Todo	Sin Movimiento Portero	Sin Dirreccion	Sin Movimiento Portero y Sin Dirreccion
0	Media	0.61	0.58	0.59	0.57
1	Desviacion estandar	0.03	0.03	0.03	0.03

Tabla 5.17: Medias y desviaciones estándar obtenidas de las predicciones de los goles, resultado final sobre 1.

MLP:						
	Medida	Todo	Sin Movimiento Portero	Sin Dirreccion	Sin Movimiento Portero y Sin Dirreccion	
0	Media	0.74	0.74	0.75		0.74
1	Desviacion estandar	0.03	0.02	0.03		0.03
LR:						
	Medida	Todo	Sin Movimiento Portero	Sin Dirreccion	Sin Movimiento Portero y Sin Dirreccion	
0	Media	0.74	0.74	0.75		0.75
1	Desviacion estandar	0.02	0.02	0.03		0.03
KNN:						
	Medida	Todo	Sin Movimiento Portero	Sin Dirreccion	Sin Movimiento Portero y Sin Dirreccion	
0	Media	0.71	0.70	0.72		0.70
1	Desviacion estandar	0.02	0.02	0.02		0.02
SVC:						
	Medida	Todo	Sin Movimiento Portero	Sin Dirreccion	Sin Movimiento Portero y Sin Dirreccion	
0	Media	0.75	0.75	0.75		0.75
1	Desviacion estandar	0.02	0.02	0.03		0.03
GAUSSIANNB:						
	Medida	Todo	Sin Movimiento Portero	Sin Dirreccion	Sin Movimiento Portero y Sin Dirreccion	
0	Media	0.74	0.74	0.75		0.75
1	Desviacion estandar	0.03	0.03	0.03		0.03
DECTREE:						
	Medida	Todo	Sin Movimiento Portero	Sin Dirreccion	Sin Movimiento Portero y Sin Dirreccion	
0	Media	0.69	0.68	0.69		0.69
1	Desviacion estandar	0.03	0.03	0.03		0.02
RANDFOREST:						
	Medida	Todo	Sin Movimiento Portero	Sin Dirreccion	Sin Movimiento Portero y Sin Dirreccion	
0	Media	0.72	0.71	0.70		0.70
1	Desviacion estandar	0.03	0.03	0.03		0.02
ADABOOST:						
	Medida	Todo	Sin Movimiento Portero	Sin Dirreccion	Sin Movimiento Portero y Sin Dirreccion	
0	Media	0.75	0.74	0.75		0.75
1	Desviacion estandar	0.02	0.02	0.03		0.03
QUADDISC:						
	Medida	Todo	Sin Movimiento Portero	Sin Dirreccion	Sin Movimiento Portero y Sin Dirreccion	
0	Media	0.72	0.72	0.72		0.72
1	Desviacion estandar	0.02	0.03	0.03		0.03

Tabla 5.18: Medias y desviaciones estándar obtenidas de las predicciones de las paradas, resultado final sobre 1.

En las tablas 5.17 y 5.18, se ha seguido el mismo procedimiento, pero la primera proporciona los datos de la predicción del gol, y la segunda de la predicción de la parada. A continuación, se analizará como varió el resultado de la precisión de los pronósticos conforme se iban eliminando las columnas indicadas de los datos de entrada, con respecto a cuando el clasificador contaba con todos los campos del etiquetado.

- En el primer caso, en la predicción de los goles (tabla 5.17), la predicción más alta utilizando todos los campos del etiquetado la generó el clasificador “Random Forest”, con un 64% de precisión, el peor de los clasificadores obtuvo un 58% de precisión, el “Naive Bayes (GaussianNB)”; y la media de la precisión calculada por todos los clasificadores fue del 61%.

Si se observa como varía la precisión de los clasificadores conforme se han ido eliminando columnas, se aprecia que en casi todos los casos el valor de la precisión ha disminuido con respecto a cuando se encontraban presentes todos los campos. El único caso en el que no sucedió fue utilizando el clasificador llamado “Decision Trees”, que, tras haber eliminado la columna que indica la información sobre hacia qué lado se lanzó el portero, la precisión aumentó, con un incremento del 1%.

Si se comparan los resultados obtenidos cuando únicamente se quita una de las dos columnas, o la columna del movimiento del portero, o la que indica la dirección del lanzamiento, no se puede determinar cual tiene mayor peso a la hora de realizar el pronóstico. Esto se debe a que, dependiendo del clasificador, en algunos casos al quitar una columna se obtiene menor precisión, pero en otros no.

Lo que sucede con todos los clasificadores, es que la precisión disminuye cuando se quitan ambos campos, el del movimiento del portero, y el de la dirección del lanzamiento. Llegando incluso en algunos casos a descender hasta un 7%, que es el caso de “Random Forest”, con respecto a cuando se utilizan todas las columnas.

- En el segundo caso, en la predicción de las paradas (tabla 5.18), “Ada Boost” y “Support Vector Machines” son los clasificadores más precisos utilizando todas las columnas del etiquetado, con un 75% de precisión, mientras que la peor

precisión obtenida fue del 69%, el clasificador llamado “Decision Trees”; y la media de todas las precisiones obtenidas es del 72,89%.

En esta ocasión, al obviar los datos de hacia dónde se lanzó el portero, la precisión se mantuvo igual, o disminuyó. Se mantuvo igual en cinco ocasiones, y disminuyó en las otras cuatro. Sin embargo, cuando se retiraron los datos acerca de la dirección del lanzamiento, se dieron los tres resultados posibles. En un caso disminuyó, en cuatro se mantuvo igual, y en los cuatro restantes incrementó.

Cuando se quitaron ambas columnas, la precisión se mantuvo igual a la inicial en cinco ocasiones, disminuyó en dos, y aumentó en dos.

Se puede apreciar que existe una diferencia en lo significativo que han sido ambos campos, pero, solamente, dependiendo de si se está prediciendo si el resultado final es un gol o una parada. Ya que, tienen un mayor grado de repercusión en los pronósticos de los goles.

Inicialmente, dado que la precisión de las predicciones de los goles era superior al 50%, se pensó que existían campos que tenían más peso que el resto a la hora de decantar la predicción del resultado final. Estos eran los campos de hacia dónde iba dirigida la falta, y hacia donde se lanzaba el portero, es decir, los campos investigados.

Originalmente se pensó que podían ser estos campos, debido a que no son atributos que describen el lance del juego antes de que la falta sea lanzada; sino que estos describen las decisiones tomadas por los jugadores en base a la situación concreta y a lo que en ese momento consideren que puede ser mejor. Ya que, por ejemplo, un portero no va a decidir tirarse hacia su izquierda, si el balón va dirigido hacia su derecha. Es por ello, que si en el etiquetado, los datos indican que el balón va dirigido a la derecha del portero, y este se tira hacia su izquierda (al lado contrario), se podría considerar que se está influyendo en la predicción que va a realizar el clasificador; esto se debe a que en una situación como está, o el balón entra en la portería, o se va fuera o al palo, pero ya le está indicando al modelo, que el portero no se la va a parar.

Es por esto, que se pensaba que quizás la información contenida en esas dos columnas, orientaban a los modelos a decantarse por gol o no gol (o parada o no parada).

Por lo tanto, tras analizar los resultados obtenidos, se puede determinar que estos dos campos apenas influyen en la predicción de las paradas. Esto se debe a que no se ha encontrado ningún patrón que demuestre que las columnas, individualmente o combinadas, repercutan en el resultado final.

Con respecto al pronóstico de los goles sí que afectan, y tienen un peso significativo, dado que la precisión de la predicción en uno de los casos ha llegado a disminuir un 7%. Sin embargo, los resultados de las precisiones no han llegado a disminuir hasta el 50%, por lo que deberán existir otros campos que influyen en los clasificadores a la hora de intentar predecir si el lanzamiento acabará en gol o si no.

5.2.3 - Función GridSearchCV

A pesar de que el objetivo no era maximizar la predicción, se ha utilizado la función llamada “GridSearchCV”, que tal y como se ha explicado previamente, sirve para encontrar los hiperparámetros que mejor resultado produzcan; es decir, aquellos que mejor precisión proporcionen a las predicciones realizadas por el clasificador.

Se ha aplicado dicha función a los diferentes clasificadores, y tanto para el pronóstico de goles como el de paradas. La tabla 5.19 muestra los resultados óptimos que se han obtenido en cada clasificador.

La tabla 5.19 muestra el mejor resultado obtenido con cada clasificador, y en todos los casos, mejora los datos obtenidos originalmente. Sin embargo, en los datos recogidos inicialmente, se obtuvo la media de las 50 predicciones, entre las cuales se descartaron los valores atípicos. Por lo tanto, para realizar una comparación justa, se ha vuelto a ejecutar el bucle de las 50 iteraciones, pero con los hiperparámetros calculados por la función.

Precision de las predicciones			
	Clasificador	Goles	Paradas
0	MLP	0.658720	0.781175
1	LR	0.606211	0.776828
2	KNN	0.625824	0.783325
3	SVC	0.647611	0.779025
4	GAUSSIANNB	0.610511	0.759436
5	DECTREE	0.641185	0.785619
6	RANDFOREST	0.656450	0.781223
7	ADABOOST	0.649857	0.783493
8	QUADDISC	0.619279	0.765886

Tabla 5.19: Resultados óptimos de cada clasificador tras utilizar la función “GridSearchCV”.

Las tablas 5.20 y 5.21 muestran las medias y desviaciones estándar obtenidas, después de las 50 iteraciones y tras eliminar los valores extremos, de la precisión de los pronósticos de los goles y de las paradas, respectivamente.

Predicción de goles original:										
	Medida	MLP	LR	KNN	SVC	GAUSSIANNB	DECTREE	RANDFOREST	ADABOOST	QUADDISC
0	Media	0.63	0.59	0.60	0.62	0.58	0.59	0.64	0.63	0.61
1	Desviacion estandar	0.03	0.03	0.03	0.03	0.04	0.03	0.03	0.02	0.03

Predicción de goles con los hiperparamtros optimos:										
	Medida	MLP	LR	KNN	SVC	GAUSSIANNB	DECTREE	RANDFOREST	ADABOOST	QUADDISC
0	Media	0.61	0.58	0.59	0.63	0.58	0.60	0.63	0.62	0.62
1	Desviacion estandar	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03

Tabla 5.20: Comparación de la media de los resultados de la predicción de goles usando los hiperparámetros devueltos por la función “GridSearchCV”.

Predicción de paradas original:										
	Medida	MLP	LR	KNN	SVC	GAUSSIANNB	DECTREE	RANDFOREST	ADABOOST	QUADDISC
0	Media	0.74	0.74	0.71	0.75	0.74	0.69	0.72	0.75	0.72
1	Desviacion estandar	0.03	0.02	0.02	0.02	0.03	0.03	0.03	0.02	0.02

Predicción de paradas con los hiperparamtros optimos:										
	Medida	MLP	LR	KNN	SVC	GAUSSIANNB	DECTREE	RANDFOREST	ADABOOST	QUADDISC
0	Media	0.74	0.75	0.74	0.75	0.75	0.74	0.75	0.75	0.74
1	Desviacion estandar	0.03	0.03	0.03	0.03	0.03	0.02	0.03	0.02	0.03

Tabla 5.21: Comparación de la media de los resultados de la predicción de paradas usando los hiperparámetros devueltos por la función “GridSearchCV”.

Se puede apreciar de los resultados mostrados en las tablas 5.20 y 5.21 que no en todos los casos, la media de la precisión ha aumentado. Sobre todo, con respecto a la predicción de goles, porque en relación con el pronóstico de las paradas, la precisión ha aumentado en todos los casos excepto en tres, que ha proporcionado el mismo resultado.

Mientras que, con la predicción de los goles, ha habido clasificadores que, sí que han mejorado el pronóstico, otros lo han empeorado, y otros han proporcionado la misma precisión que la original.

Capítulo 6 - Conclusiones y trabajo futuro

En este capítulo se detallarán las conclusiones expuestas a raíz de los resultados obtenidos con este prototipo. Siempre con la precaución y teniendo en cuenta, que, debido a las limitadas muestras, puede ser que los resultados no sean del todo representativos. Además, se indicarán las posibles mejoras que se podrían llevar a cabo de cara a los futuros trabajos que se realicen sobre este ámbito.

Conclusiones

El análisis estadístico

Dado que en el planteamiento inicial de este trabajo se quería intentar predecir el resultado final de los lanzamientos de falta, un requisito era obtener el mismo número de faltas fallidas que de marcadas. Como bien es sabido, en la vida real esto no es del todo así; puesto que se fallan muchas más faltas de las que se marcan, y rara vez se tiene un reparto del 50-50. Sin embargo, si se observa el reparto de los diferentes motivos de fallos, sí que tiene una semejanza con la realidad, dado que el reparto de dichos motivos parece coherente, ya que, alrededor del 50% son paradas, casi el 24% se van fuera, y los restantes 26% se dividen entre tiros a los palos, a la barrera, o despejes y/o rebotes en otro jugador.

Primero se ha llevado a cabo un análisis estadístico de los goles, y se ha obtenido que, como era de esperar, los jugadores tienen una mayor probabilidad de marcar si tiran la falta a un lado, mientras que si la tiran al centro existe menos probabilidad de gol. También se ha obtenido que los lanzamientos desde el centro son los que más se suelen marcar, y que también se suelen marcar más los cercanos que los lejanos. Por lo que, la ubicación de la falta óptima es aquella falta cercana que se encuentra en la zona del medio. Con respecto a lanzadores zurdos y diestros, apenas hay diferencias entre unos u otros, ya que la diferencia encontrada fue del 0,36%.

A partir de este punto, se ha realizado un estudio para encontrar cual sería el perfil del lanzador óptimo en cada situación. Ya que dependiendo de la zona del campo en la que esté ubicada la falta, será mejor un lanzador zurdo, o uno diestro. Se ha encontrado que,

en las faltas cercanas, si esta se encuentra en la zona central, o en la derecha, el especialista debería de ser un zurdo, mientras que, si está en la izquierda, la mejor opción es un diestro. En cuanto a las faltas lejanas, lo mejor, en los tres casos, son los jugadores que golpeen con la pierna derecha.

Con respecto a los porteros, se ha obtenido que las faltas que les son más sencillas de detener son aquellas que van dirigidas al medio, especialmente a la parte baja del centro. Con respecto a la ubicación de la falta en el campo que mejor se les da, esta, es la zona central más lejana; y le siguen, las faltas cercanas, pero laterales, tanto de la izquierda como de la derecha. Mientras que las laterales lejanas son en las que peores registros tienen. Esto no significa que el resto que no han parado han acabado en gol, sino que no han sido paradas; por lo que se han podido ir fuera.

Se ha analizado también, cuál sería la posición inicial que deberían de tomar los porteros para tener mayor probabilidad de parada. Ha resultado que, si los porteros se colocan en la zona central, es donde más probabilidad de parar tienen. Seguido de la zona izquierda del portero, y quedando su derecha como la peor posición. Además, han obtenido mejores registros cuando se lanzan hacia su derecha que hacia su izquierda.

Por lo tanto, ambas estadísticas van de la mano, ya que, el portero se podría colocar entre la zona central y la izquierda, para intentar, de esta forma, forzar que el jugador lance hacia la derecha del guardameta; debido a que, se les da mejor a los porteros lanzarse hacia su lado derecho.

Las estadísticas, en este primer acercamiento al estudio y análisis de los libres directos, podrían llegar a ser de gran utilidad para los entrenadores y jugadores de cara a intentar sacar el mayor provecho a los lanzamientos de falta, siempre y cuando se incorpore un mayor número y variedad de tiros libres. El objetivo sería no solamente ampliar la base de datos, sino abarcar un mayor número de las diferentes casuísticas y desenlaces de los lanzamientos de falta directa que se puedan dar. Por lo tanto, si a este prototipo se le añadieran un número considerable de casos, podría proporcionar resultados que si sean representativos y se podrían obtener estadísticas más ajustadas a la realidad.

Tal y como se ha comentado, un total de 610 lanzamientos no proporciona una imagen representativa, por lo que no es un base de datos lo suficientemente amplia como para

realizar un estudio estadístico completo. Si se estima que, por ejemplo, se realizan dos lanzamientos de falta directa por partido, y en “LaLiga”, hay un total de 20 equipos, cada equipo jugaría un total de 38 partidos, por lo que, habrá un total de 760 partidos en una temporada; es decir, que en una temporada se podrían lanzar un total de 1520 faltas directas; y, por ende, aumentar el número de lanzamientos estudiados sería lo óptimo de cara a mejorar los resultados obtenidos.

La predicción de resultados

A pesar de que inicialmente pueda resultar contradictorio que la precisión de los pronósticos de las paradas haya sido ampliamente superior a la de los goles, los resultados son los esperados. Y es que es normal que los resultados de la predicción de las paradas sean superiores a los de los pronósticos de los goles. Esto se debe a que, los porteros han conseguido detener el 50,82% de aquellas faltas falladas, pararon un total de 155 lanzamientos de los 610 que han sido analizados. Es decir, que los porteros han conseguido parar el 25,41% de los tiros. Debido a esta escasez de paradas es por lo que la precisión del pronóstico es tan elevada, y el ejemplo que se detalla a continuación, demuestra, de forma simplificada, el porqué de los resultados.

En este ejemplo, se ha supuesto que los clasificadores intentan predecir los resultados de los 610 lanzamientos, y a la hora de predecir las paradas, predicen que ninguna de las 610 faltas las va a parar el portero. Por ello, predicen que el resultado final será siempre 0, ya que los guardametas nunca detienen el golpeo. Esto quiere decir, que, de los 610 lanzamientos, el modelo únicamente se equivoca en 155 de estos, ya que fueron los 155 que detuvieron los porteros. Por lo tanto, al solamente confundirse en 155 lanzamientos, obtienen una tasa de fallo del 25,41%, y, por ende, una precisión del 74,59%.

Con respecto a los resultados obtenidos de las predicciones de los goles. Estos han sido superiores a lo esperado. Esto es debido a que, el 50% de los tiros acabaron en gol, y el restante 50% no. De modo que lo esperado era que el clasificador proporcionará una precisión del 50%. Sin embargo, todos los modelos consiguieron una precisión cercana al 60%. Dado estos resultados, se asumió, que algunos de los atributos etiquetados, estaban causando que los clasificadores tuvieran en su poder características que les

servían como indicadores del resultado final, es decir, que hacían decantar la balanza. Por este motivo fue por el que se decidió investigar como afectaba la ausencia de algunos de los campos del etiquetado.

Se había asumido que las características que proporcionaban un mayor peso en el resultado final del clasificador eran aquellas que entraban en juego a partir del lanzamiento de la falta, es decir, una vez pasado el instante inicial; aquellas que entran en juego en base a las acciones que realizan los jugadores. Estos atributos son: hacia dónde dirige el lanzador el tiro, y hacia qué lado se tira el portero para intentar detenerlo.

Se decidió que estos dos atributos eran los que probablemente hacían decantar a los modelos, ya que no eran características descriptivas de la falta, sino de la acción; y esto provoca que la información no sea objetiva. Esto se debe a que un jugador va a intentar tirar la falta a aquella zona en la que piense que el portero no va a llegar; y de igual forma pasa con el portero, y es que este no se va a lanzar hacia el lado contrario hacia el que va el balón dirigido, ya que, si hiciera esto, no tendría opción alguna de pararlo.

Tras ejecutar los modelos variando los datos de entrada en la predicción de los goles, se observó como con tan solo quitar una de las columnas mencionadas, sin importar cuál de las dos fuera, la mayoría de las precisiones de los pronósticos disminuyeron. Al quitar ambos atributos, se pudo incluso apreciar una mayor reducción. A pesar de que no se consiguió reducir el resultado hasta el 50%, en algunos casos si se redujo considerablemente, quedando la peor de las predicciones con un 53% de precisión.

Por lo tanto, se puede concluir con que dichas columnas ejercen un peso significativo en los modelos a la hora de predecir si el lanzamiento acabaría en gol o no. Es cierto, que estos no son los únicos dos atributos que influyen en los clasificadores de esta forma, ya que el resultado sigue sin ser el 50%.

La eliminación de estas columnas no afectó a las predicciones de las paradas, pero tal y como se comentó anteriormente, los resultados proporcionados por los clasificadores son los esperados, ya que no hay un reparto equitativo entre los lanzamientos que acabaron en parada y los que no. Por lo tanto, estos resultados no son concluyentes, ya que se debería comprobar el peso de dichos atributos, cuando se tenga el mismo número de paradas que de no paradas; tal y como sucede con los goles, que la mitad de los

lanzamientos si acaban en gol y la otra mitad no. Así que, hasta que no se tenga el mismo número de paradas que de no paradas, no se podrá identificar el peso que dichas características puedan tener en la decisión final del pronóstico de paradas.

Futuros trabajos

En esta sección se pretende detallar las posibles mejoras que se pueden incorporar a dicho proyecto, para así mejorar los resultados obtenidos de cara a futuras ampliaciones o trabajos.

Tal y como se ha comentado anteriormente, aumentar los datos no estructurados, como son el número de videos, que contiene la base de datos, sería un factor imprescindible que aportaría beneficios tanto al análisis estadístico como a la predicción de los resultados de los futuros lanzamientos.

Con respecto a las estadísticas recabadas, se podrían tener en cuenta aún más circunstancias y situaciones para incorporar nuevas métricas y de esta forma hacer que este estuviera aún más completo. Como, por ejemplo, de cara a la posición del portero, combinar en que zona de la portería se debería colocar el portero en base a la colocación de la falta; o inclusive, una más avanzada sería encontrar el lugar en donde se debería de colocar el portero, en base a la posición de la falta, y en base a si el jugador que va a lanzar es zurdo o diestro.

De esta forma, se podría obtener cual es la ubicación en la que se debería ubicar el portero, para intentar complicárselo al lanzador lo máximo posible, e intentar forzar que o bien este lance a donde peor se le da, o a donde mejor se le da al portero. Esta estadística también sería útil de cara al lanzador, para intentar dirigir la falta a donde peor se les dé a los porteros.

Además de añadir nuevas estadísticas, se podría cambiar la definición de algunos de los campos. Por ejemplo, en las faltas cercanas, cuando estas se encuentren en algunas de las bandas, pero estén casi a la altura del punto penalti, o por delante, es decir, casi como un córner, se podría realizar una distinción en este tipo de faltas cercanas laterales con respecto al resto. Lo mismo sucede con las faltas lejanas que están por detrás de la altura

del semicírculo del centro del campo. Esto es, porque en este tipo de faltas, el tipo de golpeo que se realiza es diferente, ya que el lanzador no puede tirar igual una falta que se encuentre a 25 metros, que otra que esté a 50 metros. Por ello, quizás sería beneficioso cambiar como se definen las características de algunos de los atributos, y añadir nuevas opciones.

Dicho lance del juego tiene muchos condicionantes que podrían influir en el resultado final de este, por lo que también se podrían añadir nuevos atributos para tener en cuenta en el etiquetado de los videos. Como podrían ser, por ejemplo: la superficie con la que golpea el jugador la falta, si un jugador se pone dentro de la portería para intentar tapar el lado en el que no esté colocado el portero, si algún jugador intenta obstaculizar al portero o a la barrera, la altura o la envergadura del portero, los factores medio ambientales como la lluvia o el viento, etc.

A parte de estos, también entran en juego los factores psicológicos que afectan a los deportistas. Como por ejemplo: los nervios que puedan tener tanto el lanzador como el portero y la importancia del partido de cara a esta tensión nerviosa; en que minuto del partido se encuentra, debido no solo al cansancio físico que puedan tener, sino en la importancia que puede tener marcar o detener el lanzamiento de cara al resultado final del partido; el resultado actual del partido, ya que si el lanzador va ganando podría tirar más relajado; en que campo se está jugando, si como local o como visitante, es decir, si se tiene a la afición a favor o en contra, etc. Todos estos factores, aunque sean muy complicados de medir, juegan un papel importantísimo en el resultado final de este lance del juego.

Otro aspecto que se podría tener en cuenta es la capacidad que tiene ese preciso lanzador o portero, de marcar o parar respectivamente la falta. Ya que diferentes lanzadores o porteros tienen mejores desempeños y resultados en dichas jugadas. Por lo que tener en cuenta quién es el individuo que realiza la acción también sería interesante.

La incorporación de todas estas características, y la modificación de la definición de los campos ya existentes, beneficiaría tanto a los resultados estadísticos como a los de la predicción de los futuros resultados.

De cara a mejorar los resultados de los pronósticos de los goles y las paradas, sería realmente importante realizar un reconocimiento de las acciones en los vídeos, para que así, utilizando modelos de deep learning, se puedan obtener las características de la apariencia del movimiento y obtener los aspectos descriptivos más significativos de las acciones; y así poder predecir con más precisión si la falta acabaría en gol o en parada.

Cabe reincidir en que los resultados obtenidos, tanto mediante las estadísticas como con el uso de los clasificadores, pudieran no ser todo lo preciso que se desearía. Esto se debe a que, por ejemplo, en el caso de las estadísticas, el cambio del resultado final de una única falta podría, hasta en un caso, aumentar el porcentaje un 25%. Este caso es el mostrado en la figura 5.8, que muestra que los zurdos tirando faltas desde el lado izquierdo, han tirado cuatro faltas lejanas, las han fallado todas y han obtenido un porcentaje de acierto del 0%. Si en vez de fallar las cuatro, hubiesen marcado una, el porcentaje de acierto sería del 25%. Esto muestra que para que los resultados de este primer acercamiento sean más representativos, se debería ampliar la base de datos, para así contemplar mayor variedad de circunstancias.

Lo mismo sucede con los clasificadores, otra vez, debido a la reducida base de datos, los resultados varían en base a los datos utilizados para entrenar los modelos y los utilizados para probarlos. Ya que dependiendo con que valores se entrene, se obtendrá diferentes valores de precisión en las predicciones.

En conclusión, este prototipo podría servir como una primera aproximación de cara al uso de Sports Analytics y pronósticos en los lanzamientos de falta en el fútbol. Pero para ello, sería necesario, como mínimo, que se aumentase el número de vídeos estudiados y se hiciese uso de la aplicación de modelos de apariencia para la clasificación del movimiento, y así utilizar valores continuos. Sin embargo, en este lance del juego, al igual que en el fútbol en general, existen múltiples de posibles improvisaciones de los jugadores que pueden hacer que toda estadística o predicción se vuelva ineficaz en un determinado instante, ya que siempre va a depender de la genialidad de los jugadores, y más aún de su actitud, ganas y pasión que le pongan a cada jugada del partido.

Bibliografía

- [1] “¿Qué es el VAR en el fútbol? Cómo se aplica y cuándo se utiliza el videoarbitraje”, Marca, 2018. [En línea]. Disponible en: <https://www.marca.com/futbol/2018/03/02/5a997c7f22601d4f6b8b462b.html>. [Accedido: 6-jun 2023]
- [2] “Éstas son las tres tecnologías que usa la FIFA para ayudar a los árbitros”, besoccer, 2017. [En línea]. Disponible en: <https://es.besoccer.com/noticia/estas-son-las-tres-tecnologias-que-usa-la-fifa-para-ayudar-a-los-arbitros-17-05-2017>. [Accedido: 6-jun 2023]
- [3] “Los Sports Analytics: "Una Nueva Esperanza"”, 9plus6, 2017. [En línea]. Disponible en: <https://9plus6.com/los-sports-analytics-una-nueva-esperanza/>. [Accedido: 7-jun 2023]
- [4] “How Sports Analytics Are Used Today, by Teams and Fans”, Built In, 2022. [En línea]. Disponible en: <https://builtin.com/big-data/big-data-companies-sports>. [Accedido: 7-jun 2023]
- [5] “¿Qué es una red neuronal?”, AWS, 2023. [En línea]. Disponible en: <https://aws.amazon.com/es/what-is/neural-network/>. [Accedido: 9-jun 2023]
- [6] “¿Qué es el aprendizaje automático?”, Microsoft Azure, 2023. [En línea]. Disponible en: <https://azure.microsoft.com/es-es/resources/cloud-computing-dictionary/what-is-machine-learning-platform>. [Accedido: 9-jun 2023]
- [7] “¿Qué es el aprendizaje automático?”, Oracle, 2023. [En línea]. Disponible en: <https://www.oracle.com/es/artificial-intelligence/machine-learning/what-is-machine-learning/>. [Accedido: 9-jun 2023]
- [8] “¿Qué es Deep Learning?”, IBM, 2023. [En línea]. Disponible en: <https://www.ibm.com/es-es/topics/deep-learning>. [Accedido: 9-jun 2023]

- [9] “¿Qué es un clasificador en el aprendizaje automático?”, Examinar, 2022. [En línea]. Disponible en: <https://examinar.net/que-es-un-clasificador-en-el-aprendizaje-automatico/>. [Accedido: 9-jun 2023]
- [10] “LaLiga: la primera en contar en sus retransmisiones un modelo avanzado de Probabilidad de Gol graficado casi en tiempo real con tecnología de Microsoft”, Microsoft, 2022. [En línea]. Disponible en: <https://news.microsoft.com/es-es/2022/01/21/laliga-la-primera-en-contar-en-sus-retransmisiones-un-modelo-avanzado-de-probabilidad-de-gol-graficado-casi-en-tiempo-real-con-tecnologia-de-microsoft/>. [Accedido: 12-jun 2023]
- [11] “How Data Science is Changing Soccer”, DataCamp, 2022. [En línea]. Disponible en: <https://www.datacamp.com/blog/how-data-science-is-changing-soccer>. [Accedido: 12-jun 2023]
- [12] “Sport Data Analytics: el dato que necesitas para ser N°1 en el mundo deporte”, Inesem, 2021. [En línea]. Disponible en: <https://www.inesem.es/revistadigital/gestion-empresarial/sport-data-analytics/>. [Accedido: 12-jun 2023]
- [13] “Las dos causas por las que cada vez se meten menos goles de falta en LaLiga”, FCBN, 2022. [En línea]. Disponible en: https://www.fcbarcelonanoticias.com/laliga/dos-causas-por-cada-vez-se-meten-menos-goles-falta-en-laliga_283955_102.html. [Accedido: 13-jun 2023]
- [14] “International football results from 1872 to 2023”, Kaggle, 2023. [En línea]. Disponible en: <https://www.kaggle.com/datasets/martj42/international-football-results-from-1872-to-2017>. [Accedido: 15-jun 2023]
- [15] “SoccerNet”, Papers With Code, 2023. [En línea]. Disponible en: <https://paperswithcode.com/dataset/soccernet>. [Accedido: 15-jun 2023]
- [16] “¿Qué es Python?”, AWS, 2023. [En línea]. Disponible en: <https://aws.amazon.com/es/what-is/python/>. [Accedido: 18-jun 2023]

- [17] “Te damos la bienvenida a Colaboratory”, Google Colaboratory, 2023. [En línea]. Disponible en: https://colab.research.google.com/#scrollTo=GJBs_fIRovLc. [Accedido: 18-jun 2023]
- [18] “Librería scikit-learn python, aprende todo”, Tokio School, 2023. [En línea]. Disponible en: <https://www.tokioschool.com/noticias/que-es-scikit-learn/>. [Accedido: 18-jun 2023]
- [19] “sklearn.model_selection.train_test_split”, Scikit-learn, 2023. [En línea]. Disponible en: https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.train_test_split.html. [Accedido: 20-jun 2023]
- [20] “3.2. Tuning the hyper-parameters of an estimator”, Scikit-learn, 2023. [En línea]. Disponible en: https://scikit-learn.org/stable/modules/grid_search.html#grid-search. [Accedido: 20-jun 2023]