****

**Estudio específico sobre datos en el deporte**

|  |  |  |
| --- | --- | --- |
| Titulación:  Máster Universitario en Big Data y Ciencia de Datos  Curso académico:  2023 – 2024 | Alumno/a:  Gil Rodrigo, Carlos  D.N.I: 70909416W  Director/a de TFM:  Reyero Diez, Raul | Convocatoria:  Primera |

02 Mayo 2024

Índice

[Resumen 5](#_Toc167053827)

[1. Introducción 6](#_Toc167053828)

[2. Objetivos 7](#_Toc167053829)

[3. Estado del Arte y Marco teórico 8](#_Toc167053830)

[4. Desarrollo del proyecto y resultados 9](#_Toc167053831)

[4.1. Metodología 9](#_Toc167053832)

[4.2. Planteamiento del problema 11](#_Toc167053833)

[4.3. Desarrollo del proyecto 11](#_Toc167053834)

[4.3.1 Selección del conjunto de datos 11](#_Toc167053835)

[4.3.2 Limpieza y transformaciones de datos 12](#_Toc167053836)

[4.3.3 Evaluación de diferentes técnicas de minería de datos 13](#_Toc167053837)

[4.3.4 Implementación de diversos algoritmos de clustering 13](#_Toc167053838)

[4.3.5 Validación de algoritmos 15](#_Toc167053839)

[4.3.6 Aplicación de algoritmos seleccionados en relación con los objetivos 16](#_Toc167053840)

[4.3.7 Análisis de resultados 17](#_Toc167053841)

[4.3.8 Visualización de resultados y conclusiones 17](#_Toc167053842)

[4.4. Herramientas utilizadas 18](#_Toc167053843)

[4.5. Resultados 18](#_Toc167053844)

[4.4.1 Resultados del proyecto 18](#_Toc167053845)

[4.4.2 Interpretación de los resultados 18](#_Toc167053846)

[4.4.3 Comparación de los resultados con investigaciones previas 18](#_Toc167053847)

[5. Conclusión y trabajos futuros 18](#_Toc167053848)

[6. Referencias 18](#_Toc167053849)

[Apéndice I 19](#_Toc167053850)

[Anexos I 20](#_Toc167053851)

Índice de ilustraciones

[Ilustración 1. Arquitectura Jupyter Cliente - Servidor. Fuente: https://www.paradigmadigital.com/dev/jupyter-data-science-aplicada/ 4](#_Toc6399225)

Índice de tablas

**No se encuentran elementos de tabla de ilustraciones.**

# Resumen

# Introducción

El punto de partida de este Trabajo de Fin de Máster se enfoca en determinar el verdadero impacto de ciertos factores en los resultados de los partidos. Si bien se ha debatido ampliamente sobre este tema, la evidencia concreta que respalde estas afirmaciones ha sido escasa, especialmente en lo que respecta a la influencia del factor de jugar como local o visitante. La discrepancia entre la percepción popular y los datos empíricos constituye el punto de partida de esta investigación.

En este contexto, se enfoca en analizar resultados históricos de la Premier League, considerada la liga de fútbol más prominente y seguida a nivel mundial. El propósito es evidenciar la verdadera importancia de los factores en cuestión. Para lograrlo, se emplean técnicas de minería de datos, las cuales permiten extraer información significativa y descubrir patrones ocultos en los datos. Este enfoque metodológico posibilita identificar tendencias, agrupaciones y correlaciones entre diversas variables clave, lo que contribuye a una comprensión más profunda y precisa de los elementos que inciden en los resultados de los encuentros futbolísticos.

A partir de los resultados obtenidos, se busca no solo dilucidar la relevancia de estos factores, sino también ofrecer una contribución significativa al entendimiento general del deporte y su relación con el análisis de datos. Como aficionado apasionado del deporte y entusiasta del mundo de la ciencia de datos, esta investigación aspira a aportar nuevas perspectivas y conocimientos al campo, enriqueciendo así el debate académico y práctico en torno a esta temática.

# Objetivos

Este Trabajo de Fin de Máster se fundamenta en una serie de objetivos generales y específicos, que establecen el propósito y la metodología de la investigación.

El objetivo principal es evidenciar la influencia de diversos factores, destacando la diferencia entre el desempeño en condición de local y visitante, en los resultados de los eventos deportivos, a través del análisis de datos y la aplicación de técnicas de minería de datos.

En lo que respecta a los objetivos específicos, cabe resaltar que estos constituyen los pilares fundamentales que dirigirán la investigación hacia la consecución de los resultados deseados, proporcionando una estructura clara y precisa para el desarrollo metodológico y el análisis de los datos. A continuación, se detallan los objetivos específicos de este trabajo:

* Recopilar y estructurar de manera exhaustiva los datos de los resultados de los partidos de la Premier League de los últimos años.
* Identificar y seleccionar las variables relevantes que podrían influir en los resultados de los partidos, tales como la localía, el rendimiento de los equipos y otros factores determinantes del desarrollo del juego.
* Realizar procesos de limpieza, procesamiento y análisis exploratorio de datos como etapas previas a la aplicación de técnicas de minería de datos con el fin de descubrir patrones y relaciones ocultas entre las variables analizadas.
* Validar los resultados obtenidos mediante un análisis riguroso y comparaciones con estudios previos realizados en el ámbito.
* Interpretar y discutir los hallazgos encontrados, proporcionando una comprensión detallada sobre la importancia relativa de los factores analizados en los resultados de los partidos de fútbol.
* Contribuir al avance del conocimiento en el campo del análisis de datos aplicado al deporte, proporcionando una base sólida para futuras investigaciones en este ámbito.

Estos objetivos se alinean con las competencias y contenidos específicos del título de máster en Big Data y Ciencia de Datos, ya que implican la aplicación de conocimientos teóricos y prácticos en el análisis de datos para abordar una problemática concreta en el contexto del deporte profesional.

# Estado del Arte y Marco teórico

# Desarrollo del proyecto y resultados

En este apartado, se proporciona una exposición completa del proceso de desarrollo del proyecto, junto con los resultados alcanzados a lo largo de su ejecución. Se abordan con detalle aspectos importantes que abarcan desde la selección de la metodología más apropiada hasta la conclusión del análisis de los datos recopilados.

Se examinan en profundidad cada uno de los pasos dados, desde el planteamiento inicial del problema hasta la aplicación de las técnicas y herramientas utilizadas durante el desarrollo del proyecto. Asimismo, se detallan los resultados obtenidos como producto de este proceso, ofreciendo una perspectiva detallada de su relevancia y significado en el contexto del estudio llevado a cabo.

## Metodología

La metodología empleada para llevar a cabo el desarrollo del proyecto aborda aspectos relacionados con la planificación y la gestión de tareas. A continuación, se describen las etapas clave del proceso, desde la recopilación inicial de datos hasta la documentación final del trabajo realizado.

1. La planificación del proyecto se llevó a cabo mediante la definición de una serie de etapas y actividades necesarias para su ejecución. Se estableció un cronograma detallado que incluía fechas límite medianamente rígidas para cada una de las etapas y tareas, con el fin de garantizar un desarrollo organizado del trabajo.
2. La primera etapa del proceso consistió en la búsqueda y recopilación de datos acordes a la temática del Trabajo de Fin de Máster y a los objetivos de investigación establecidos. La información necesaria para el análisis fue obtenida de manera cuidadosa, asegurando la selección del conjunto de datos adecuado.
3. Una vez recopilados los datos, se procedió a realizar un proceso exhaustivo de limpieza y transformación. Se identificaron y corrigieron posibles errores, duplicados e inconsistencias en los datos, y se prepararon para su posterior análisis.
4. Se llevaron a cabo pruebas con diferentes técnicas de minería de datos, considerando opciones como regresión, clasificación y agrupamiento, para explorar la información contenida en los datos recopilados. Tras una evaluación exhaustiva, se determinó que la técnica de clustering era la más adecuada para alcanzar los objetivos del proyecto, siendo seleccionada como la opción final.
5. Se implementaron y ejecutaron varios algoritmos de clustering, incluyendo KMeans, DBSCAN, Clustering Jerárquico, Mean Shift y Propagación de Afinidad, con el objetivo de encontrar resultados significativos. Se realizaron ajustes en los parámetros de los algoritmos según las características específicas de los datos y se llevó a cabo una completa evaluación de los resultados obtenidos.
6. Los resultados de los algoritmos de clustering fueron validados mediante técnicas de validación interna y externa. Se evaluó la calidad de los clusters generados utilizando métricas como el coeficiente de silueta, el coeficiente de Davies-Bouldin, el índice de Calinski-Harabasz y la suma de cuadrados dentro del grupo (SSE).
7. Se utilizaron los algoritmos de clustering que produjeron resultados satisfactorios para abordar los objetivos del proyecto, los cuales habían sido planteados como interrogantes al conjunto de datos. Estos algoritmos permitieron explorar y encontrar patrones subyacentes en los datos, facilitando así la obtención de conceptos relevantes para responder a las preguntas inicialmente planteadas.
8. Se analizaron los resultados obtenidos a partir de los clusters identificados, extrayendo conclusiones relevantes para la investigación.
9. Los resultados obtenidos se representaron mediante gráficas y visualizaciones para facilitar su interpretación.
10. A lo largo de todo el proceso, se realizó una documentación detallada de las actividades llevadas a cabo. Se registraron los pasos realizados, los resultados obtenidos y las decisiones tomadas en cada parte del proceso, garantizando la trazabilidad del trabajo realizado.

La metodología implementada facilitó el desarrollo efectivo del proyecto, abarcando desde la recolección inicial de datos hasta la documentación final del trabajo realizado. El enfoque organizado utilizado en la planificación y gestión de tareas fue fundamental para cumplir con los objetivos establecidos al inicio y obtener resultados relevantes en el análisis de los datos de la Premier League.

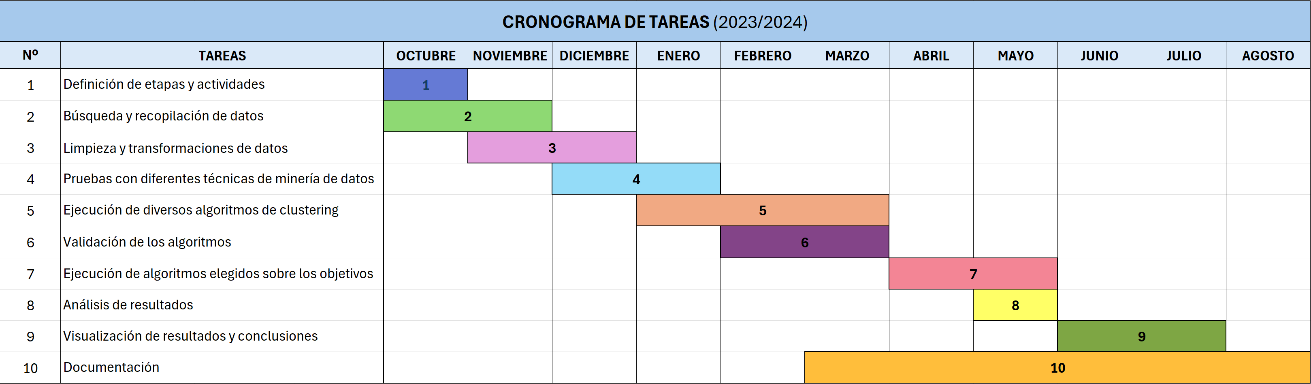


Ilustración 2. Cronograma de las tareas definidas. Elaboración propia.

## Planteamiento del problema

La investigación deportiva ha sido tradicionalmente un campo de interés tanto para aficionados como para académicos, destacando la relevancia de variables específicas en el desarrollo y desenlace de eventos deportivos, en particular en el fútbol. Dentro de este contexto, surge una interrogante recurrente sobre la influencia que ciertos factores, como la posesión del equipo, el rendimiento ofensivo y el factor de localía, pueden tener en los resultados de los partidos.

A pesar del extenso debate y la relevancia percibida de estos aspectos, la literatura científica carece en gran medida de estudios específicos que brinden una evidencia concluyente sobre su impacto real en el rendimiento deportivo. Esta brecha en el conocimiento motiva la concepción de esta investigación, que se propone abordar de manera rigurosa y analítica la relación entre estas variables y los resultados de los partidos de fútbol.

El objetivo central de este Trabajo de Fin de Máster radica en proporcionar una respuesta fundamentada a la cuestión fundamental sobre la importancia de ciertos factores en el contexto deportivo. Para lograr este propósito, se adoptará un enfoque metodológico meticuloso, basado en el análisis de datos y la aplicación de técnicas estadísticas avanzadas, con el fin de generar conclusiones sólidas y robustas.

A través de esta indagación, se aspira a contribuir al avance del conocimiento en el ámbito del análisis deportivo, llenando un vacío existente en la literatura científica y proporcionando información relevante para diversos actores del mundo del deporte, desde entrenadores y jugadores hasta analistas y gestores. Se anticipa que los hallazgos obtenidos tendrán implicaciones significativas para la comprensión y optimización del rendimiento deportivo, así como para la formulación de estrategias y tácticas efectivas en el terreno de juego.

## Desarrollo del proyecto

En este apartado, se detalla el proceso de desarrollo del proyecto, abordando diversas etapas y aspectos fundamentales que han contribuido al logro de los objetivos planteados. A lo largo de estos subapartados, se ofrece una visión completa del proceso de desarrollo del proyecto, desde la selección de datos hasta la interpretación de los resultados, ofreciendo una comprensión profunda y detallada del trabajo realizado.

### 4.3.1 Selección del conjunto de datos

La selección del conjunto de datos se llevó a cabo considerando el propósito y alcance del proyecto. Tras una exhaustiva investigación de diversas fuentes de datos relacionadas con el ámbito deportivo, se optó por un conjunto de datos que abarcara el histórico de partidos de la Premier League, la liga inglesa de fútbol reconocida por su competitividad y popularidad a nivel mundial. Este conjunto de datos fue obtenido de Kaggle, una plataforma web ampliamente utilizada por la comunidad de Data Science debido a su amplia variedad de conjuntos de datos y recursos.

El dataset, antes de ser sometido a procesos de limpieza y transformación, constaba de 114 columnas que detallaban una amplia gama de características de los partidos disputados entre los años 2010 y 2021. Este amplio conjunto de datos ofreció una sólida base informativa que facilitó la realización del análisis y la consecución efectiva de los objetivos del proyecto.

### 4.3.2 Limpieza y transformaciones de datos

La preparación del conjunto de datos se fundamentó en un riguroso proceso de limpieza y transformación, con el fin de adecuar los datos para su posterior utilización en técnicas avanzadas de minería de datos. A continuación, se detallan los tratamientos implementados durante esta importante etapa del proyecto:

* Tratamiento de missing values: El tratamiento de valores desconocidos o perdidos se realizó mediante técnicas como la eliminación de filas que contenían valores perdidos cuando su proporción era significativa y afectaba la calidad del análisis, y la estimación basada en otros valores del conjunto de datos.
* Tratamiento de outliers: El tratamiento de valores fuera de rango se llevó a cabo mediante técnicas como la eliminación de valores atípicos del análisis cuando se consideró que estos podían distorsionar los resultados de manera significativa, y la segmentación de datos de forma que los valores atípicos se agruparan por separado. Los valores fuera de rango o atípicos son datos que difieren significativamente del resto de observaciones, y pueden ser resultado de errores o simplemente representar valores que se encuentran fuera del rango normal esperado.
* Tratamiento de datos innecesarios: Este proceso abarca diversos tipos de tratamientos con el propósito común de lograr que todos los datos del modelo contribuyan significativamente al objetivo del proyecto. Todos los datos resultantes deben aportar valor, y para eliminar aquellos que no lo hacen, se implementaron los siguientes tratamientos:
* Tratamiento de datos contaminados: Se detectaron y eliminaron datos que podrían estar contaminados debido a errores de registro o problemas de calidad.
* Tratamiento de datos inconsistentes: Se solventaron discrepancias e incoherencias en los datos, incluyendo variaciones o formatos no uniformes.
* Tratamiento de datos duplicados: Se identificaron y eliminaron registros duplicados con el fin de prevenir la redundancia y asegurar la integridad del conjunto de datos.
* Tratamiento con los tipos de datos: Se revisaron y ajustaron los tipos de datos, asegurando que cada variable tuviera el tipo de dato adecuado para su posterior análisis.

Adicionalmente, se realizaron otras transformaciones importantes, tales como la normalización de texto mediante la capitalización, la eliminación y el renombramiento de características irrelevantes, y la conversión de fechas a formatos estandarizados que resultasen útiles para el análisis. Asimismo, se empleó la matriz de correlaciones para identificar las relaciones más destacadas dentro del conjunto de datos, lo que facilitó la implementación de los tratamientos y eliminaciones pertinentes.

Una vez concluidas las diversas técnicas de tratamiento de los datos del conjunto de datos seleccionado, se procedió a la evaluación de las técnicas de minería de datos que podrían ser empleadas para este proyecto.

### 4.3.3 Evaluación de diferentes técnicas de minería de datos

Desde el inicio, la evaluación de las técnicas de análisis de datos potencialmente aplicables para alcanzar los objetivos del proyecto se dividió en dos fases: la consideración de técnicas de clasificación y de técnicas de agrupamiento. La predicción no fue llevada a cabo, ya que no tenía sentido en este proyecto. El objetivo principal era analizar el histórico de los últimos años para obtener conocimientos sobre patrones y tendencias pasadas, no para proyectar datos futuros.

Después de un análisis minucioso de los objetivos y de las interrogantes iniciales relativas al conjunto de datos, y tras la evaluación de diversos algoritmos de clasificación, como la regresión logística, los K-Nearest Neighbors y los árboles de decisión, se determinó que la estrategia más adecuada consistía en la utilización de técnicas de agrupamiento, centrándose en el clustering de variables altamente correlacionadas.

### 4.3.4 Implementación de diversos algoritmos de clustering

La implementación de los diferentes algoritmos seleccionados para llevar a cabo el clustering tenía como objetivo la identificación de agrupamientos o características coincidentes, con el propósito de detectar patrones en las filas de datos. A través de la comparación de clusters, se pretendía analizar las diferencias significativas entre los elementos de estos, con un enfoque frecuente en localidad, es decir, la comparación de datos entre equipos locales y visitantes.

Se consideraron múltiples técnicas de clustering, incluyendo el biclustering, aunque se determinó que no era adecuado para este proyecto y descartó la idea. Respecto a las técnicas de clustering, se evaluaron diversas opciones como K-Means, DBSCAN, Gaussian Mixture Models (GMM), Mean Shift, Agglomerative Clustering, Spectral Clustering, Affinity Propagation y Hierarchical Clustering. Los algoritmos finalmente seleccionados y probados sobre el conjunto de datos fueron los siguientes:

* K-Means: Técnica de agrupamiento que divide un conjunto de datos en k grupos o clusters. Los datos se agrupan de tal manera que los puntos dentro del mismo cluster tengan una mayor entre sí que con los puntos de otros clusters.
* DBSCAN (Density-Based Spatial Clustering of Applications with Noise): Técnica de agrupamiento que se fundamenta en la densidad de los datos en el espacio. Este método detecta regiones de alta densidad, las cuales son interpretadas como clusters, mientras que aquellos puntos que están en zonas de baja densidad se catalogan como ruido. De esta forma, DBSCAN es capaz de identificar clusters de formas variadas y manejar de manera efectiva datos que presenten ruido o outliers.
* Hierarchical Clustering: El método de agrupamiento jerárquico organiza los datos en una estructura de árbol, fusionando repetidamente clusters cercanos hasta alcanzar un criterio establecido. Esto proporciona una representación jerárquica de los datos, donde los clusters similares se agrupan en ramas próximas y los menos similares en ramas más alejadas, facilitando explorar distintos niveles de detalle en la agrupación.
* Mean Shift: Algoritmo de agrupamiento que localiza los máximos locales de densidad para determinar los centroides de los clusters. No necesita definir previamente el número de clusters y es adecuado para conjuntos de datos no lineales o con formas irregulares de clusters. Al ser un método no paramétrico, es flexible y adecuado para diversos problemas de agrupamiento.
* Affinity Propagation: Algoritmo de agrupamiento que selecciona automáticamente representantes de los clusters y asigna puntos de datos a estos representantes. No requiere especificar el número de clusters y puede identificar tanto el número como los centroides automáticamente. Es útil cuando se desconoce el número de clústeres y se necesita una identificación precisa de los centroides.

Para evaluar estas técnicas de clustering, se formularon algunas interrogantes básicas, a partir de las cuales se obtuvieron los primeros resultados, los cuales precedieron a las validaciones de los algoritmos. Estas cuestiones abordaron aspectos tales como “Tiros respecto a tiros a puerta”, “Posesión en comparación con pases y toques de balón”, “Diferencia de goles frente a diferencia de puntos”, “Promedio de córners respecto a promedio de tiros”, “Goles locales en comparación con goles locales al descanso”, y “Promedio de goles marcados frente a promedio de goles marcados al descanso y al porcentaje de puntos ganados”. Además, se utilizó el método Elbow sobre los algoritmos para determinar el número óptimo de clusters, lo cual facilitó una agrupación más precisa y coherente de los datos analizados.

Posteriormente, tras recoger los resultados de las ejecuciones, se procedió a evaluar la precisión demostrada por cada algoritmo, descartando aquellos que no resultaban adecuados para este proyecto debido a su menor estabilidad y consistencia en la formación de clusters respecto a otros. De esta forma, se garantizó una interpretación más confiable y precisa de los resultados del clustering.

### 4.3.5 Validación de algoritmos

La validación de algoritmos se realizó teniendo en cuenta una gran cantidad de medidas de validación de clustering, tanto internas como externas. Algunas de las medidas contempladas inicialmente fueron los índices de Silhouette, Dunn, Davies-Bouldin, Calinski-Harabasz, Hubert, Jaccard, Rand, WSS (Within-Cluster Sum of Squares), SSE (Sum of Squared Errors), BSS (Between-Cluster Sum of Squares), Virtual Error y Clustering Error.

Estas medidas se clasifican en dos categorías principales: internas y externas, las cuales se diferencian en la forma en que evalúan la calidad del clustering y en la información que utilizan para dicha evaluación. Las medidas internas determinan la calidad del clustering basándose solo en la estructura intrínseca de los datos, mientras que las medidas externas comparan los clusters generados con etiquetas de referencia previamente conocidas. En el contexto de este proyecto, al carecer de etiquetas de referencia, no es posible emplear las medidas de validación externas, por lo que estas fueron descartadas.

Tras realizar un análisis de las medidas de validación interna mencionadas anteriormente, se seleccionaron cuatro de ellas para evaluar los algoritmos y comparar sus resultados:

* Silhouette Index: Mide la cohesión y separación de los clusters. Los valores cercanos a 1 indican una buena separación entre los clusters, mientras que los valores cercanos a 0 indican una superposición significativa entre los clusters. Los valores negativos sugieren que las muestras podrían haber sido asignadas al cluster incorrecto.
* Davies-Bouldin Index: Evalúa la relación entre la dispersión dentro de los clusters y la separación entre ellos. Los valores bajos indican una buena separación entre los clusters y una baja superposición. Cuanto más cercano a 0, mejor.
* Calinski-Harabasz Index: Mide la dispersión entre clusters en relación con la dispersión dentro de los clusters. Los valores más altos indican una mejor separación entre los clusters. Aunque no existe un valor específico al que deba acercarse, la comparación con otros resultados puede dar una idea relativa de la calidad del clustering.
* Sum of Squared Errors (SSE): Mide la variabilidad dentro de los clusters, evaluando la suma de las distancias al cuadrado de los puntos de datos a sus centroides de cluster respectivos. Este valor debe ser lo más bajo posible, ya que representa la variabilidad dentro de los clusters. Sin embargo, es una métrica inversa, por lo que valores más altos no son necesariamente mejores.

Los resultados de estas validaciones resultaron ser bastante significativos, particularmente en el contexto de la evaluación de tres y seis clusters, donde se observaron patrones consistentes y distintivos en la estructura de los datos. El método de validación SSE fue excluido debido a que no resultaba aplicable a todos los algoritmos de clustering contemplados para el análisis. En cuanto a las otras tres medidas de validación, se determinó que los algoritmos de clustering más estables y consistentes eran K-Means, DBSCAN, Hierarchical Clustering y Mean Shift. Se encontró que el algoritmo de Affinity Propagation mostraba resultados insatisfactorios y una notable inestabilidad, lo que llevó a su exclusión. A pesar de los resultados menos favorables de Mean Shift, se observó que aún mantenía valores lo suficientemente estables y significativos como para ser considerado útil. Por consiguiente, los cuatro algoritmos de clustering seleccionados finalmente, tras validar sus resultados, fueron: K-Means, DBSCAN, Hierarchical Clustering y Mean Shift. Estos algoritmos se emplearon para abordar cuestiones más específicas que las planteadas inicialmente, con el objetivo de alcanzar las metas del proyecto.

### 4.3.6 Aplicación de algoritmos seleccionados en relación con los objetivos

Una vez seleccionados los algoritmos de clustering más estables y significativos con el propósito de esclarecer aspectos fundamentales en los resultados de los partidos, se plantearon las siguientes cuestiones:

* Dominio en los partidos: Se analiza la relación entre el porcentaje de posesión y el número de pases y toques de balón, tanto del equipo local como del equipo visitante durante los encuentros. Este análisis permite evaluar la importancia de estas características en el dominio de un equipo sobre otro.
* Amonestaciones en los partidos: Se investiga la relación entre el número de tarjetas amarillas y rojas en función de la localidad. Esto permite examinar si los equipos visitantes tienen una mayor propensión a ser amonestados debido al estilo de juego que plantean.
* Ocasiones creadas en los partidos: Se analiza el promedio de goles anotados, tiros y tiros a puerta y córners realizados. El objetivo es observar la correlación entre estas variables para determinar cuándo un equipo crea más oportunidades.
* Equilibrio entre ataque y defensa: Se investiga la relación entre el promedio de goles anotados y encajados históricamente por los equipos, lo que permite agruparlos según su equilibrio entre ataque y defensa, distinguiendo entre equipos más ofensivos y defensivos.
* Efectividad de la posesión en cuanto a goles anotados: Se examina la relación entre la posesión del balón por parte de los equipos y los goles anotados en los partidos. Esto proporciona información sobre la importancia de la posesión y del dominio del juego en el resultado final de los encuentros.
* Agrupamiento de equipos según el histórico de resultados: Los equipos se agrupan según su éxito deportivo entre 2010 y 2021, teniendo en cuenta los equipos más victoriosos, los que han experimentado más derrotas y aquellos con un rendimiento intermedio.

[AÑADIR MÁS CUESTIONES Y MODIFICAR LAS EXISTENTES]

Como se ha podido constatar, se abordaron múltiples interrogantes relacionadas con diferentes aspectos de los partidos de fútbol, con el propósito de no solo responder a una pregunta principal, sino también a varias cuestiones vinculadas con las características de los encuentros, considerando en la mayoría de los casos el factor de la localidad.

### 4.3.7 Análisis de resultados

### 4.3.8 Visualización de resultados y conclusiones

## Herramientas utilizadas

## Resultados

### 4.4.1 Resultados del proyecto

### 4.4.2 Interpretación de los resultados

### 4.4.3 Comparación de los resultados con investigaciones previas

# Conclusión y trabajos futuros

# Referencias

# Apéndice I

# Anexos I