



Universidad del Desarrollo

Facultad de Ingeniería

Instituto Data Science

PREDICCIÓN DE RESULTADOS DE PARTIDOS DE LA LIGA PROFESIONAL DE
FUTBOL CHILENO USANDO ALGORITMOS DE MACHINE LEARNING

POR: RICARDO OVANDO FUENTEALBA

Proyecto de grado presentado a la Facultad de Ingeniería de la Universidad del
Desarrollo para optar al grado académico de Magíster en Data Science

PROFESOR GUÍA:

Sra. Loreto Bravo

Marzo 2025

SANTIAGO

AGRADECIMIENTO

Agradezco a la profesora guía Loreto Bravo por su disposición y paciencia durante el desarrollo de este informe. Además agradezco el apoyo y comprensión incondicional de mi familia que fueron de gran ayuda durante el desarrollo de este trabajo.-

TABLA DE CONTENIDO

TABLA DE CONTENIDO

RESUMEN.....	4
1. INTRODUCCIÓN.....	6
2. REVISIÓN DE LITERATURA.....	10
3. HIPÓTESIS Y OBJETIVOS.....	15
Hipótesis Principal.....	15
Hipótesis Secundarias.....	15
Objetivo Principal.....	15
Objetivos Específicos.....	15
4. DATOS Y METODOLOGÍA.....	16
4.1. DATOS.....	16
4.2. METODOLOGÍA.....	26
5. RESULTADOS.....	26
Experimento 1.....	32
Experimento 2.....	38
Experimento 3.....	45
Experimento 4.....	51
Experimento 5.....	59
6. CONCLUSIONES.....	69
BIBLIOGRAFÍA.....	72

Resumen

Este proyecto se centra en el desarrollo y evaluación de modelos de aprendizaje automático para predecir resultados de partidos de fútbol (Victoria Local, Empate, Victoria Visitante). A partir de datos históricos de ligas de fútbol, se construyó un conjunto de características (incluyendo estadísticas de temporadas previas, datos acumulados de rendimiento y contexto) con el objetivo de entrenar y comparar diferentes algoritmos de clasificación: Random Forest, XGBoost, CatBoost y Regresión Logística.

El enfoque metodológico incluyó:

1. **Preprocesamiento de Datos:**

Se aplicaron técnicas para manejar el desbalance de clases (SMOTE, ADASYN) y métodos de normalización robusta. Esto buscaba mejorar la detección de resultados menos frecuentes (empates y victorias visitantes).

2. **Optimización y Validación:**

Se emplearon búsquedas de hiperparámetros (RandomizedSearchCV), validación cruzada estratificada y ampliación de muestras con ruido para intentar incrementar la capacidad predictiva de los modelos. Además, se exploró la reducción de dimensionalidad mediante PCA, analizando el impacto en las métricas de desempeño.

3. **Comparación de Algoritmos:**

Los resultados mostraron que, si bien los modelos ensemble (Random Forest, XGBoost, CatBoost) superaron a la Regresión Logística en general, ninguno logró un salto significativo en la predicción de empates. La clase “Victoria Local” fue la más fácil de predecir, reflejando patrones más marcados, mientras

que la clase “Empate” permaneció consistentemente difícil de identificar.

4. Replicación en Diferentes Ligas:

El mismo procedimiento se aplicó a datos de la Premier League de Inglaterra con resultados similares, lo que evidencia que el reto no se limita a una sola competencia. Esto sugiere que la dificultad es intrínseca al dominio y las variables empleadas, más que a un problema de un conjunto de datos específico.

Conclusiones:

El proyecto demuestra la complejidad de predecir resultados de fútbol a partir de datos cuantitativos limitados. A pesar de esfuerzos en balanceo de datos, optimización de hiperparámetros y reducción de dimensionalidad, la clasificación del empate sigue siendo un desafío sustancial. La conclusión central es que, para mejorar el desempeño, se requieren nuevas fuentes de información más ricas (variables contextuales, tácticas, calidad individual de jugadores, condiciones externas) y/o técnicas más sofisticadas que capten la complejidad inherente al juego.

De esta forma, el proyecto aporta una visión clara de las limitaciones actuales y ofrece una ruta hacia futuras líneas de investigación, enfatizando la necesidad de enriquecer los datos y adaptar las técnicas de modelado para abordar la complejidad del fenómeno “empate” en el fútbol.

1. Introducción

La predicción de resultados en eventos deportivos, particularmente en el fútbol, se ha convertido en un área de interés creciente tanto para la investigación académica como para la industria del deporte. En el caso de la liga profesional de fútbol chileno, existen múltiples factores que justifican la necesidad de desarrollar modelos predictivos capaces de anticipar resultados con alta precisión y una interpretabilidad clara que permita a los equipos técnicos comprender y aplicar los resultados.

Este trabajo de magíster se centra en la predicción de resultados de partidos de la liga profesional de fútbol chileno mediante la aplicación de algoritmos de machine learning. El principal objetivo es desarrollar un modelo predictivo capaz de anticipar los resultados de los encuentros, utilizando variables relevantes como estadísticas de jugadores, rendimientos anteriores y condiciones del partido. Asimismo, se busca integrar características nuevas que puedan mejorar la precisión y robustez del modelo.

Uno de los principales problemas en la predicción de resultados deportivos es la calidad de los datos y la variabilidad inherente a los eventos futbolísticos. La falta de modelos robustos y específicos para la liga chilena representa una brecha que este proyecto busca llenar, ofreciendo una solución que no solo prediga resultados, sino que también permita entender qué variables afectan más el resultado de un partido.

Este proyecto de magíster se centra en la predicción de resultados de la liga profesional de fútbol chileno precisamente con el objetivo de responder a las necesidades específicas de los actores del fútbol nacional. Si bien la predicción deportiva ha sido abordada en diversos contextos, la aplicación de modelos predictivos robustos y, crucialmente, interpretables, para la liga chilena presenta una oportunidad única. Un modelo preciso no solo podría satisfacer el interés académico y de las apuestas deportivas, sino que

fundamentalmente, entregaría a los cuerpos técnicos de los clubes chilenos una herramienta valiosa para la toma de decisiones informada.

La capacidad de anticipar la probabilidad de diferentes resultados en futuros encuentros permitiría a los equipos técnicos chilenos optimizar su planificación estratégica. Por ejemplo, comprender las fortalezas y debilidades de los próximos rivales, así como la probabilidad de diferentes marcadores, podría influir en la selección de la formación inicial, la elaboración de planes tácticos específicos y la asignación de recursos durante la semana de entrenamiento. Más aún, la interpretabilidad del modelo, es decir, la capacidad de entender qué variables están impulsando las predicciones, podría ofrecer a los técnicos insights valiosos sobre los factores que más inciden en los resultados dentro del contexto particular de la liga chilena, facilitando la identificación de áreas de mejora tanto en el propio equipo como en el análisis de los contrincantes.

La relevancia de este estudio radica en la posibilidad de mejorar el análisis estratégico de los equipos, optimizar las apuestas deportivas y contribuir a la investigación académica en el campo del análisis de datos deportivos. Además, ofrece un enfoque innovador para explorar el uso de técnicas avanzadas en un contexto deportivo local, con aplicaciones potenciales a nivel global.

En este documento se describen las técnicas de recopilación y procesamiento de datos, los algoritmos seleccionados y la metodología utilizada para entrenar y evaluar el modelo predictivo. Se discuten también los resultados obtenidos y sus implicaciones para el fútbol chileno y el campo del análisis deportivo.

En resumen, este proyecto busca cerrar una brecha en el ámbito de la predicción deportiva aplicada al contexto chileno, proporcionando un modelo preciso e interpretable que podría tener un impacto positivo tanto en la toma de decisiones de los clubes como en el desarrollo del deporte en el país. Además, este proyecto podría servir

como base para futuras investigaciones en otros deportes o ligas locales, estableciendo un estándar de excelencia en el uso de tecnología y ciencia de datos en el deporte chileno.

2. Revisión de Literatura

La predicción de resultados en el fútbol ha sido un campo de investigación activo que busca aplicar métodos estadísticos y de machine learning (ML) para pronosticar los desenlaces de los partidos. Esta revisión explora la literatura relevante, con el objetivo de identificar las metodologías predominantes, los modelos predictivos alternativos, lo crucial de las variables predictivas (características), el impacto de la temporalidad de los datos y las limitaciones más frecuentes encontradas en estos estudios. Para este propósito, se han implementado los modelos más prometedores de la literatura para analizar su rendimiento bajo un protocolo unificado (2).

Metodologías y Modelos Predictivos en el Fútbol

Se ha empleado una variedad de técnicas estadísticas y de machine learning para abordar la predicción de resultados de fútbol:

Modelos Estadísticos Basados en Puntuación: Los modelos de Poisson, incluyendo el modelo de Poisson bivariado y el modelo de doble Poisson (9), se utilizan para modelar la cantidad de goles que cada equipo podría marcar (2). Estos modelos permiten estimar la probabilidad de diferentes marcadores y, por ende, del resultado del encuentro (victoria, empate o derrota) (2). Asimismo, los sistemas de calificación Elo y otras variaciones como pi-ratings y los ratings de Berrar evalúan la capacidad relativa de los equipos basándose en su desempeño histórico. La computación de pi-ratings enfatiza el resultado del partido (ganar, empatar o perder) en mayor grado que el margen de goles. Una versión modificada de pi-ratings también incorporó un factor de forma del equipo que busca identificar el rendimiento superior o inferior continuo. Cuatro características de calificación (dos para el equipo local y dos para el visitante) se utilizaron como entradas para la Red Bayesiana Híbrida (7). Estos ratings pueden usarse directamente para la predicción o como features en modelos más complejos. Hubáček et al. realizaron una revisión experimental para comparar la efectividad de varios modelos basados en puntuación (2). También se propusieron embeddings relacionales de equipos (3). El

algoritmo PageRank también se ha utilizado para estimar la fortaleza de los equipos (2) (7).

Modelos de Regresión: La regresión logística es una técnica estadística popular para modelar la probabilidad de resultados categóricos, como la victoria del equipo local, el empate o la victoria del equipo visitante (7). Se han utilizado diversas variables como inputs para estos modelos, incluyendo estadísticas históricas de los equipos, calificaciones (como las FIFA ratings) (10) e incluso información subjetiva (11). Se ha utilizado la regresión logística regularizada para estimar la contribución de los jugadores.

Otras Técnicas de Machine Learning: Se han explorado otros algoritmos como las Máquinas de Vectores de Soporte (SVM) (7) para la clasificación de resultados de partidos.

La Importancia de las Variables Predictivas (Features)

La selección de variables relevantes es fundamental para el éxito de cualquier modelo predictivo en el fútbol (4*). La literatura resalta la importancia de considerar:

Datos Históricos del Rendimiento de los Equipos: Esto incluye resultados de partidos anteriores, número de goles marcados y recibidos (5*), estadísticas de ataque y defensa (5*), posesión del balón, disciplina (7*), etc. La creación de características históricas es una práctica común (7*).

Sistemas de Calificación y Rankings: El uso de calificaciones como Elo o FIFA ratings puede proporcionar una medida de la fuerza relativa de los equipos (10*). Los pi-ratings y los ratings de Berrar también son utilizados.

Factores del Partido: La localía o condición de visitante es una variable predictiva importante en muchos estudios (10*) (6*). El historial de enfrentamientos directos entre los equipos también se considera relevante (11*).

Features Derivadas e Ingeniería de Features: La creación de nuevas variables a partir de las existentes, como la diferencia de goles promedio o el porcentaje de victorias recientes, puede mejorar la capacidad predictiva de los modelos (6*). Se ha propuesto dividir las características en promedios en casa y fuera (6*). También se ha sugerido añadir atributos de jugadores, tiros, córners, fuerza de ataque y rendimiento de los entrenadores (6*). Se han utilizado métodos de selección de características basados en la experiencia de expertos (7*). Algunos estudios han utilizado conjuntos de características que incluían variables seleccionadas por expertos (12*).

La Ventana de Tiempo de los Datos

La cantidad y la actualidad de los datos históricos utilizados para entrenar los modelos son factores críticos (10):

Datos Históricos Extensos: Utilizar un amplio historial de partidos permite a los modelos aprender patrones a largo plazo en el rendimiento de los equipos (10).

Relevancia Temporal: Las temporadas más antiguas pueden volverse menos relevantes debido a cambios en los equipos y el rendimiento de los jugadores. Es importante considerar la actualidad de los datos.(2)(7)(11)

División Temporal de Datos: Al dividir los datos en conjuntos de entrenamiento y prueba, se debe mantener el orden cronológico (2) para asegurar que los modelos se evalúen en datos futuros que no se utilizaron para el entrenamiento (2). Los conjuntos de datos deben estar ordenados cronológicamente (2).

Limitaciones Recurrentes en la Literatura

A pesar de los avances en las técnicas de predicción, la literatura señala varias limitaciones persistentes:

Naturaleza Estocástica del Fútbol: La imprevisibilidad inherente al deporte hace que la predicción con alta precisión sea un desafío (13). Se considera poco probable alcanzar una precisión del 100% con métodos estadísticos.

Dificultad en la Predicción de Empates: La predicción del empate es consistentemente más difícil que la predicción de victorias (8).

Calidad y Disponibilidad de Datos: La falta de conjuntos de datos estandarizados y la dificultad para cuantificar factores cualitativos relevantes (como la moral del equipo o factores tácticos) limitan la capacidad predictiva de los modelos. Muchos conjuntos de datos no son accesibles públicamente, lo que dificulta la replicación de la investigación. La mayoría de los sitios web presentan los resultados e información sobre fútbol como codificados en HTML.(2)(7)(8)

Sesgo en los Datos Históricos: Los datos históricos pueden contener sesgos que afecten el rendimiento de los modelos en el futuro.(14)(15)(2)

Baja Precisión Predictiva: Se ha destacado la baja precisión predictiva como un problema general, lo que subraya la necesidad de más investigación para lograr resultados confiables. La precisión obtenida en algunos estudios, incluso con técnicas avanzadas, puede ser limitada (2)(7)(14)(10)

Relación de la Literatura con el Trabajo Realizado

La literatura revisada informa el trabajo realizado de las siguientes maneras:

Las metodologías y modelos explorados en el trabajo (como regresión logística y métodos de ensamble basados en árboles) son consistentes con los enfoques comunes en la literatura para la predicción de resultados de fútbol (2) (1). La decisión de probar estos modelos se basa en su aplicación previa y los resultados reportados en otros estudios (1).

La importancia de las variables como datos históricos de rendimiento (5), sistemas de calificación (10) y factores del partido (10) (6), que probablemente se consideran en el trabajo realizado, se subraya consistentemente en la literatura (5). La ingeniería de features es un tema recurrente en las direcciones futuras de investigación (6).

Los desafíos encontrados en el trabajo, como la dificultad para predecir ciertas clases de resultados (posiblemente el empate) (8), son un reflejo de las limitaciones recurrentes

identificadas en la literatura (8). Varios estudios señalan que la predicción del empate es particularmente desafiante (8).

Las estrategias de preprocesamiento de datos y evaluación de modelos utilizadas en el trabajo (como la división temporal de datos (2) y la validación cruzada) son prácticas estándar en el campo de la predicción deportiva con machine learning, tal como se describe en la literatura (2). La necesidad de utilizar conjuntos de datos relevantes y de buena calidad es enfatizada.

Las líneas futuras de investigación sugeridas en la literatura, como la incorporación de fuentes de datos más ricas y contextuales, y la exploración de modelos más sofisticados (3), ofrecen posibles direcciones para mejorar el trabajo realizado. La necesidad de abordar la dificultad específica de predecir empates (8) y de aumentar la precisión general son temas recurrentes. Se sugiere desarrollar software capaz de recopilar todas las características y organizarlas. El mantenimiento automatizado de la base de datos del sistema de predicción podría ser necesario para garantizar que el sistema se actualice constantemente. Se podría realizar más investigación para comprender las características que afectan los resultados de los partidos de fútbol.

En resumen, esta revisión de literatura, generada, proporciona una visión general de las metodologías, características, consideraciones temporales y limitaciones comunes en la predicción de resultados de fútbol utilizando métodos estadísticos y machine learning. Además, contextualiza el trabajo realizado dentro de este campo de investigación, resaltando la coherencia entre los enfoques adoptados y los desafíos encontrados con lo reportado en la literatura existente. La persistente dificultad en la predicción del empate y la necesidad de mejorar la precisión general emergen como desafíos centrales en este dominio (8)

3. Hipótesis y Objetivos

Hipótesis Principal:

La implementación de un modelo de aprendizaje automático basado en características seleccionadas de la liga profesional de fútbol chileno permitirá predecir los resultados de los partidos con una precisión superior a los métodos tradicionales

Hipótesis Secundarias:

1. El uso de características relevantes como datos históricos mejorará significativamente la precisión del modelo predictivo.
2. La correcta selección de técnicas de preprocesamiento y estrategias de división de datos históricos aumentará la capacidad del modelo para generalizar correctamente a nuevos datos.
3. La comparación de diferentes algoritmos de aprendizaje automático permitirá identificar el modelo más preciso y robusto para la predicción de resultados en la liga chilena de fútbol.
4. Las pruebas de validación y optimización mejorarán el rendimiento global del modelo, aumentando su capacidad de predicción y su utilidad para análisis estratégicos en el fútbol profesional.

Objetivo Principal:

construir un modelo para predecir resultados en la liga chilena de futbol

Objetivos Específicos

1. Implementar un modelo de aprendizaje automático para la predicción de resultados de la liga chilena de fútbol.

2. Seleccionar e implementar características clave relevantes para mejorar la precisión del modelo predictivo.
3. Comparar modelos de predicción en términos de precisión y capacidad de interpretación.
4. Realizar pruebas de validación para optimizar el rendimiento del modelo.

4. Datos y Metodología

4.1. Datos

Análisis Exploratorio de los datos:

Volumen de Datos

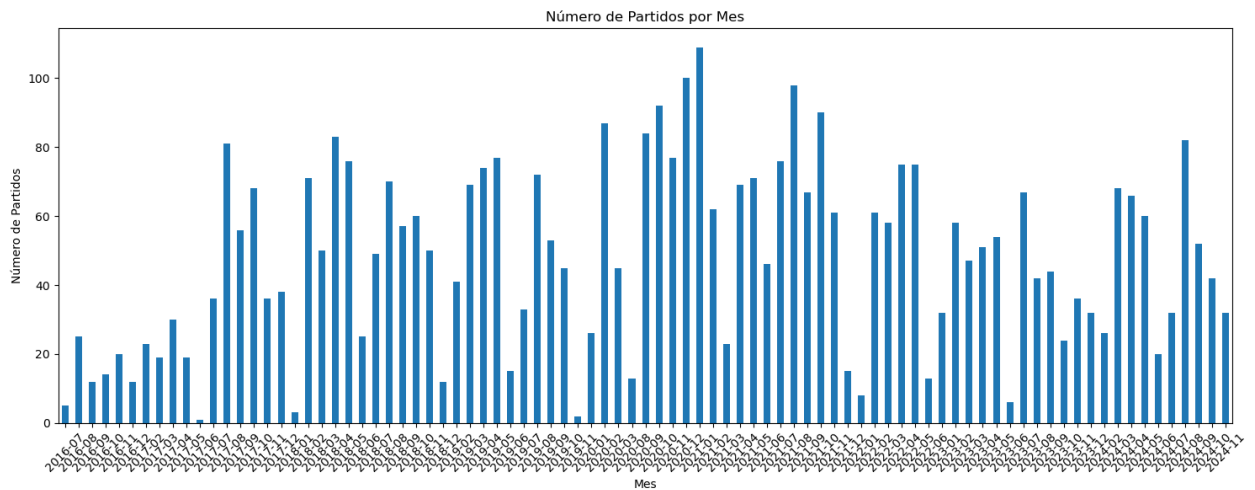
- Se analizaron 2,548 partidos en total
- Cada partido tiene 109 características rastreadas
- El conjunto de datos ocupa 2.07 MB
- No se encontraron entradas duplicadas

Período de Cobertura

- Inicio: 31 de mayo de 2017
- Fin: 10 de Noviembre de 2024
- Duración total: **2720 días** o aproximadamente **7.45 años**.

Problemas encontrados:

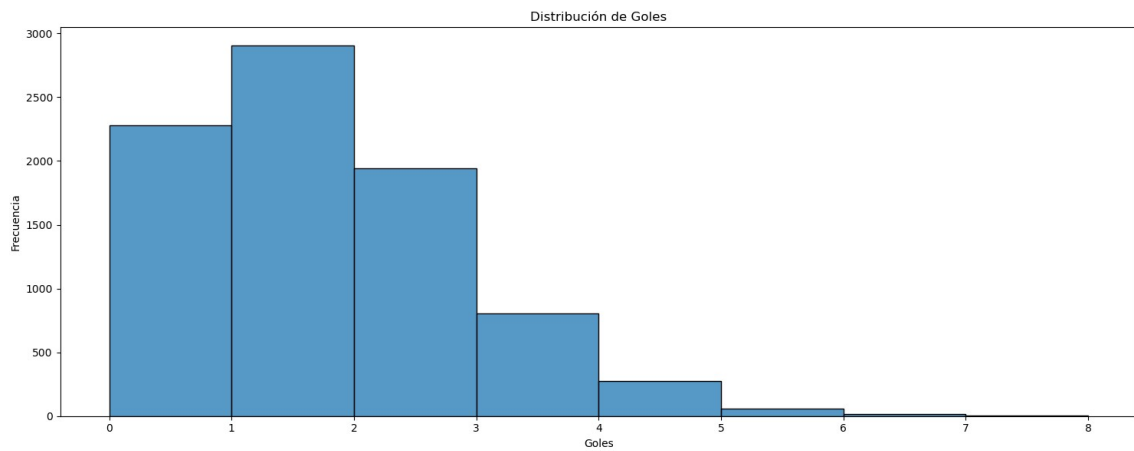
- los títulos de columna están hechos de manera poco intuitiva para hacer un parser, donde hay un título y hay que separar los nombre del título para poder nombrar correctamente a la columna correspondiente
- opté por evitar los partidos de selección, porque tienen cantidad de columnas distintas



1. **Número de Partidos por Mes** Este gráfico muestra la distribución temporal de los partidos desde 2017 hasta 2024:

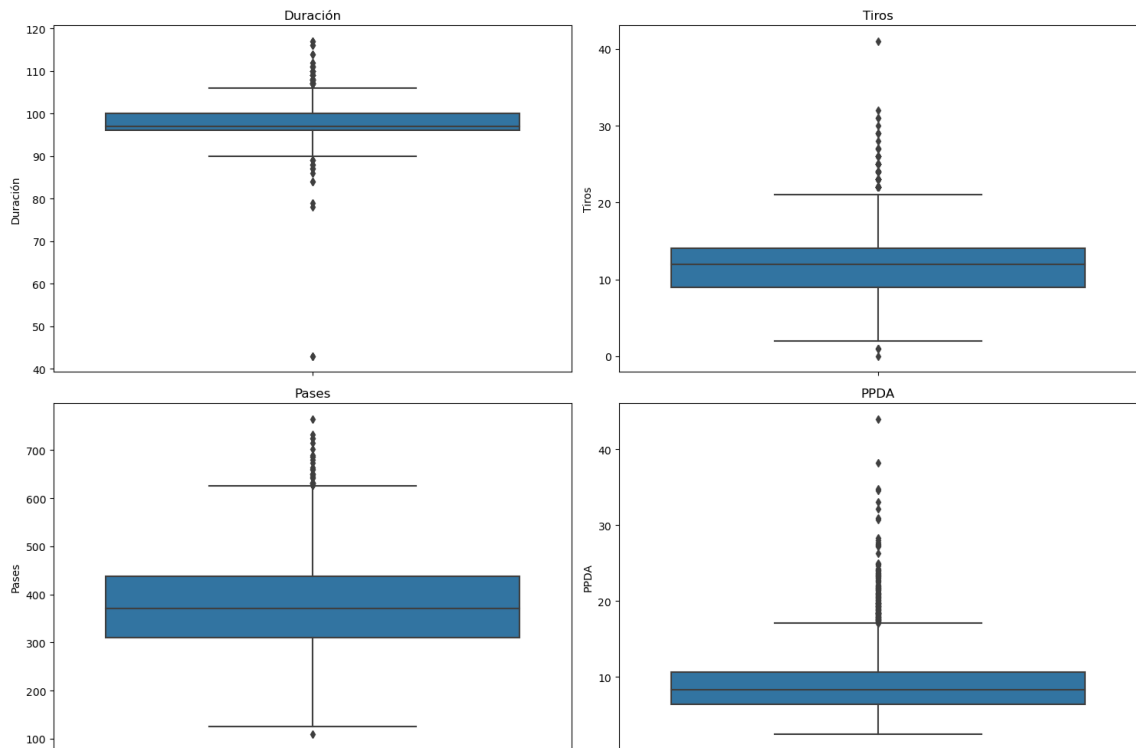
- Se observa una consistencia en el número de partidos por mes, típicamente entre 30-40 partidos mensuales
- Hay algunos valles que probablemente corresponden a:

- Pausas por fechas FIFA
- Períodos entre temporadas
- Posible pausa por COVID-19 (habría que verificar alrededor de 2020)
- La regularidad en el número de partidos sugiere un calendario bien estructurado de la liga



2. **Distribución de Goles** Este histograma muestra la frecuencia de goles por partido:

- La distribución parece seguir una distribución de Poisson (típica en goles de fútbol)
- La mayoría de los partidos tienen entre 0-3 goles
- El pico está en 1-2 goles por equipo
- Hay pocos partidos con más de 4 goles, siendo estos casos atípicos
- Esta distribución es consistente con el promedio de goles en ligas profesionales



3. Métricas de Rendimiento (Boxplots)

Duración:

- La mediana está alrededor de los 95-97 minutos
- Hay varios outliers por encima de los 100 minutos, indicando partidos con tiempo adicional significativo
- Algunos valores atípicos bajos (menos de 90 minutos) podrían indicar partidos suspendidos

PPDA (Passes Per Defensive Action):

- Este es un indicador de la intensidad del pressing

- La mayoría de los valores están entre 5-15
- Hay varios outliers por encima de 20, indicando partidos con pressing menos intenso
- Los valores más bajos (menos de 5) indican partidos con pressing muy agresivo

Análisis:

1. Análisis Local vs Visitante:

- Diferencia en Promedio de Goles:

- Locales: 1.48 goles por partido
- Visitantes: 1.14 goles por partido
- Existe una clara ventaja local con aproximadamente 0.34 goles más por partido (29.3% más goles)

- Dispersión de Datos:

- Locales tienen mayor desviación estándar (1.19 vs 1.07), indicando resultados más variables
- El máximo de goles es similar (6 local vs 7 visitante)

- Distribución:

- Locales: Mediana = 1.0, con 75% de partidos teniendo 2 o menos goles
- Visitantes: Mediana = 1.0, pero con 25% de partidos sin goles (vs local donde el Q1 es 1.0)

- Esto sugiere que los equipos visitantes tienen más probabilidad de irse en blanco

2. Análisis de Top 5 Equipos:

- Universidad Católica:

- Mejor promedio goleador (1.56 goles por partido)
- xG muy similar a goles reales (1.57), indicando consistencia
- Mayor posesión promedio (55.22%), significativamente superior al resto

- Ñublense:

- Segundo mejor promedio goleador (1.49)
- Ligera sobre performance vs xG (goles 1.49 vs xG 1.51)
- Posesión más baja del top 5 (51.18%)

- Colo Colo:

- Tercer mejor promedio (1.48 goles)
- Rendimiento casi exacto a su xG (1.47)
- Buena posesión (52.40%)

- Palestino:

- Cuarto en goles (1.42)
- Significativa subperformance vs xG (goles 1.42 vs xG 1.58)
- Sugiere problemas de definición o mala suerte

- Universidad de Chile:

- Cierra el top 5 (1.39 goles)
- Rendimiento acorde a su xG (1.38)

- Posesión similar a Colo Colo (52.39%)

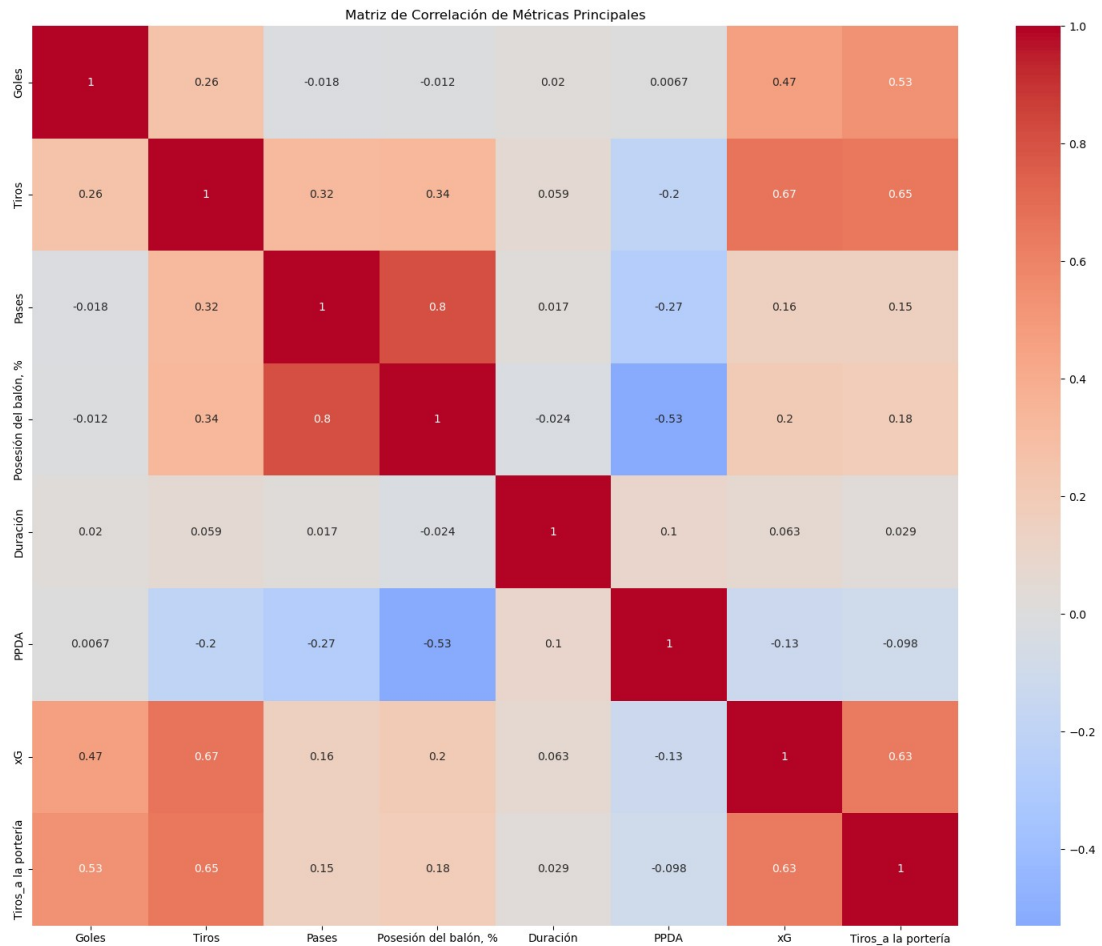
Observaciones Adicionales:

1. La diferencia entre el mejor y peor del top 5 es relativamente pequeña (0.17 goles por partido)
2. Todos los equipos del top 5 mantienen posesiones superiores al 51%
3. Palestino destaca por tener el mejor xG pero no traducirlo en goles
4. Universidad Católica muestra dominancia clara en todas las métricas

Implicaciones Tácticas

1. La ventaja local es significativa y debería considerarse en análisis tácticos
2. Los equipos top tienden a mantener posesión alta
3. Hay una correlación clara entre posesión y efectividad goleadora
4. La mayoría de equipos top tienen un rendimiento cercano a su xG, excepto Palestino

Matriz de correlación:



Correlaciones Fuertes Positivas (>0.6):

1. Pases y Posesión del balón (0.8)

- Indica una relación muy fuerte y lógica: más pases implican mayor posesión

2. Tiros y Tiros a la portería (0.66)

- Correlación esperada: más tiros totales resultan en más tiros al arco

3. xG y Tiros a la portería (0.64)

- Muestra que los tiros al arco son buenos predictores del xG

4. Tiros y xG (0.68)

- Similar al anterior, confirma que la cantidad de tiros influye en el xG

Correlaciones Moderadas Positivas (0.3-0.6):

1. Goles y xG (0.47)

- Correlación moderada que sugiere que el xG es un predictor razonable de goles

2. Goles y Tiros a la portería (0.54)

- Mayor correlación que con xG, sugiriendo que los tiros al arco son mejores predictores

3. Tiros y Posesión (0.34)

- Mayor posesión lleva a más oportunidades de tiro

Correlaciones Negativas Significativas:

1. PPDA y Posesión (-0.53)

- La correlación negativa más fuerte
- Sugiere que equipos con menor PPDA (pressing más intenso) tienden a tener más posesión

2. PPDA y Pases (-0.25)

- Consistente con lo anterior

3. PPDA y Tiros (-0.20)

- Equipos con pressing más intenso generan más tiros

Correlaciones Débiles o Nulas:

1. Duración con casi todas las variables

- Sugiere que la duración del partido no influye significativamente en otras métricas

2. Goles y Posesión (-0.012)

- Interesantemente, la posesión no correlaciona con goles

3. Goles y Pases (-0.014)

- Similar al anterior, sugiriendo que el estilo de posesión no garantiza efectividad

Implicaciones Tácticas:

1. El pressing intenso (PPDA bajo) parece ser una táctica efectiva para:

- Mantener la posesión
- Generar más tiros
- Controlar el juego

2. La eficiencia en ataque es clave:

- La posesión no correlaciona con goles
- Los tiros a puerta son mejores predictores que el volumen de juego

3. El xG parece ser un buen predictor:

- Correlaciona bien con goles
- Tiene fuerte relación con tiros y tiros a puerta

Recomendaciones basadas en los datos:

1. Enfocarse en la calidad de los tiros más que en la cantidad
2. Implementar pressing intenso para controlar el juego
3. No sobrevalorar la posesión como táctica para marcar goles
4. Priorizar la creación de ocasiones claras (xG alto) sobre el volumen de juego

4.2. Metodología

1. Recopilación y Selección de Datos

- **Fuente de Datos:** La base de datos para este proyecto se construyó a partir de registros históricos de partidos de la liga profesional de fútbol chileno. Esta recopilación abarcó un período significativo, desde el 31 de mayo de 2017 hasta el 10 de noviembre de 2024, permitiendo capturar una amplia gama de dinámicas y rendimientos de los equipos a lo largo del tiempo. El conjunto de datos resultante comprendió un total de 2,548 partidos, cada uno con 109 características rastreadas, lo que proporcionó una rica fuente de información para el desarrollo de los modelos predictivos.
- **Preprocesamiento:** Previo al entrenamiento de los modelos, se llevó a cabo un exhaustivo proceso de preprocesamiento de los datos. Esto incluyó la identificación y tratamiento de posibles errores o inconsistencias en el conjunto de datos. Dada la observación de títulos de columna poco intuitivos que dificultaban el parsing de la información, se realizó una etapa de reestructuración y normalización de los nombres de las columnas para asegurar una manipulación y análisis eficientes. Adicionalmente, se optó por excluir los partidos de selecciones nacionales debido a la heterogeneidad en la cantidad y naturaleza de las columnas registradas para estos encuentros, lo que podría introducir ruido y dificultar la generalización de los modelos entrenados con datos de la liga profesional.
- **Selección de Características:** La selección de características relevantes para la predicción del resultado de los partidos fue una etapa crucial en el desarrollo de este proyecto. En los experimentos realizados, se exploraron diferentes estrategias para identificar los atributos con mayor poder predictivo. En el Experimento 1, se llevó a cabo un análisis de correlación entre las variables predictoras y la variable objetivo (resultado del

partido), seleccionando las 15 características que presentaron las correlaciones más significativas . En el Experimento 2, se adoptó un enfoque basado en la importancia de las características evaluada por el algoritmo Random Forest, identificando las 15 variables que este modelo consideró más influyentes en la predicción . Estas selecciones se basaron en la premisa de que las características con mayor correlación o importancia para un modelo específico tendrían un mayor impacto en la precisión de las predicciones futuras

- Ventana Movil: Para incorporar la dinámica temporal del rendimiento de los equipos, se implementó el concepto de ventanas móviles en el cálculo de las características. En los experimentos, se utilizaron ventanas de tamaño 10 y 5 partidos hacia atrás para calcular estadísticas acumuladas o rolling features de los equipos . La elección de estas ventanas se basó en la idea de que el rendimiento reciente de un equipo es un indicador más relevante de su desempeño futuro que datos históricos muy lejanos . Experimentar con diferentes tamaños de ventana permitió evaluar la sensibilidad de los modelos a la cantidad de información histórica reciente considerada
- El objetivo principal de este trabajo es desarrollar un modelo predictivo que permita anticipar los resultados de partidos de fútbol basándose en datos históricos de desempeño de los equipos. Para lograr esto, se construye una variable objetivo, Resultado_Num, que representa el resultado del partido (victoria, empate o derrota desde la perspectiva del equipo local). A partir de esta variable objetivo, se aplican técnicas de análisis, limpieza y modelado de datos para construir un sistema que evalúe el impacto de diversas características en el resultado de los encuentros.

2. División de Datos

- **Entrenamiento y Prueba:** Para evaluar la capacidad de generalización de los modelos desarrollados, el conjunto de datos históricos se dividió en conjuntos de entrenamiento y prueba. Se aseguró que esta división se realizara de manera cronológica, reservando los partidos más recientes para el conjunto de prueba. Esta estrategia es fundamental para mantener la integridad temporal de las predicciones, evitando el uso de información futura para predecir resultados pasados, una práctica que podría llevar a una sobreestimación del rendimiento real de los modelos
- **Validación Cruzada:** Con el fin de obtener una evaluación más robusta del rendimiento de los modelos y minimizar el riesgo de sobreajuste, se emplearon técnicas de validación cruzada. Específicamente, se utilizó la validación cruzada estratificada de k-pliegues (k-fold cross-validation), donde el conjunto de entrenamiento se dividió en k subconjuntos o pliegues. Cada pliegue se utilizó como conjunto de validación de forma iterativa, mientras que los k-1 pliegues restantes se utilizaron para el entrenamiento. La estratificación aseguró que cada pliegue mantuviera aproximadamente la misma proporción de las diferentes clases del resultado del partido que el conjunto de datos original, lo cual es especialmente importante en escenarios con desbalance de clases

3. Implementación de Modelos

- **Modelos a Evaluar:** En este proyecto, se implementaron y compararon diversos algoritmos de aprendizaje automático ampliamente utilizados en tareas de clasificación y predicción. Los modelos seleccionados fueron: Random Forest, conocido por su robustez y capacidad para manejar grandes conjuntos de datos y relaciones no lineales; XGBoost (Extreme Gradient Boosting), un algoritmo de boosting que ha demostrado un excelente rendimiento en diversas competiciones y problemas de predicción; CatBoost, otro algoritmo de gradient boosting que se destaca

por su manejo de variables categóricas y su robustez a la hora de ajustar hiperparámetros; y Regresión Logística, un modelo lineal que sirvió como punto de referencia (baseline) para comparar el rendimiento de los modelos más complejos

4. Entrenamiento y Optimización de Modelos

- **Entrenamiento:** El entrenamiento de cada uno de los modelos seleccionados se realizó utilizando el conjunto de datos de entrenamiento preprocesado. Durante esta etapa, se ajustaron los parámetros internos de cada algoritmo para aprender los patrones subyacentes en los datos que permitieran predecir el resultado de los partidos. Dada la importancia de la configuración de los hiperparámetros en el rendimiento de los modelos, se implementaron estrategias de optimización para encontrar la combinación de hiperparámetros que mejor maximizara una métrica de rendimiento predefinida, como la precisión (accuracy) o el F1-score.
- **Evaluación de Modelos:** La evaluación del rendimiento de los modelos entrenados se llevó a cabo utilizando diversas métricas de clasificación, incluyendo la precisión (accuracy), el AUC-ROC (área bajo la curva ROC), el F1-score, la precisión (precision) y el recuerdo (recall) para cada una de las clases (Victoria Local, Empate, Victoria Visitante). Estas métricas permitieron cuantificar la capacidad de los modelos para predecir correctamente los resultados de los partidos y evaluar su desempeño en la detección de cada una de las clases, especialmente las minoritarias como el empate y la victoria visitante

5. Análisis de Resultados

- **Comparación de Modelos:** Una vez entrenados y evaluados los diferentes modelos, se realizó una comparación exhaustiva de sus resultados en términos de las métricas de rendimiento obtenidas en el conjunto de

prueba y mediante la validación cruzada. Esta comparación permitió identificar el modelo que ofrecía el mejor equilibrio entre precisión y capacidad de generalización, así como analizar las fortalezas y debilidades de cada algoritmo en la predicción de las diferentes clases del resultado del partido.

- **Limitaciones Identificadas:** Durante el proceso de modelado y análisis de resultados, se identificaron diversas limitaciones inherentes al problema y a las metodologías aplicadas. Estas limitaciones incluyeron la calidad y disponibilidad de los datos, la naturaleza estocástica e impredecible de los eventos deportivos, y la dificultad de incorporar factores cualitativos relevantes para el resultado de un partido, como la moral del equipo o la dinámica de los jugadores. Estas limitaciones son cruciales para contextualizar el rendimiento de los modelos y comprender los desafíos existentes en la predicción de resultados de fútbol.

6. Conclusiones y Recomendaciones

- **Conclusiones:** Las conclusiones principales del estudio se basaron en el análisis comparativo del rendimiento de los modelos evaluados. Se resumieron los hallazgos clave en términos de la precisión predictiva alcanzada, la capacidad de los modelos para identificar las diferentes clases de resultados, y las limitaciones encontradas durante el proceso de investigación. Se identificó el modelo con mejor desempeño general y se destacaron las clases de resultados que resultaron más difíciles de predecir.
- **Recomendaciones:** Finalmente, se formularon recomendaciones para el uso práctico del modelo desarrollado, considerando sus fortalezas y limitaciones. Asimismo, se sugirieron áreas de mejora y futuras líneas de investigación, incluyendo la posible incorporación de nuevas fuentes de datos en tiempo real, la exploración de modelos híbridos, y la integración

de factores cualitativos que podrían enriquecer la capacidad predictiva de los modelos

7. Documentación y Presentación de Resultados

- **Documentación Completa:** Se generó una documentación de todo el proceso de investigación, incluyendo la descripción detallada de la recopilación y preprocesamiento de los datos, la implementación y optimización de los modelos, el código utilizado para el análisis, y los resultados obtenidos. Esta documentación tiene como objetivo asegurar la transparencia y la reproducibilidad del estudio
- **Visualización de Resultados:** Para facilitar la comprensión de los hallazgos y la comparación entre los modelos, los resultados se presentaron utilizando gráficos, tablas y matrices de confusión. Estas visualizaciones permitieron ilustrar de manera clara y concisa el rendimiento de cada modelo en las diferentes métricas y clases, así como identificar patrones de error y áreas de mejora

5. Resultados

Experimento 1

- Se utilizan para este experimento una ventana móvil de 10 partidos para calcular rolling features
- se utilizan como características para la predicción las 15 primeras características resultantes de un análisis de correlación:

Feature	F Score	p-value
Visitante_porcentaje_Duelos_ganados_rolling	9.1784	0.0001
	38	25
Local_porcentaje_Pases hacia adelante_logrados_rolling	8.0637	0.0003
	39	65
Visitante_porcentaje_Duelos ofensivos_ganados_rolling	7.5558	0.0005
	45	96
Local_porcentaje_Pases_logrados_rolling	6.8837	0.0011
	71	41
Visitante_Saques laterales_rolling	6.8669	0.0011
	77	6
Local_porcentaje_Pases en el último tercio_logrados_rolling	6.5688	0.0015
	27	48
Local_Pases hacia adelante_logrados_rolling	5.9013	0.0029
	78	62
Local_Longitud media pases_rolling	5.8597	0.0030
	12	85
Local_porcentaje_Pases progresivos_precisos_rolling	5.8583	0.0030
	98	89
Local_Pases_logrados_rolling	5.6796	0.0036

	91	76
Visitante_porcentaje_Duelos aéreos_ganados_rolling	5.6404	0.0038
	38	19
Visitante_Saques laterales_logrados_rolling	5.4508	0.0045
	49	94
Visitante_porcentaje_Centros_precisos_rolling	5.3923	0.0048
	5	64
Local_Pases_rolling	5.3257	0.0051
	27	91
Visitante_Desmarques_rolling	5.2486	0.0055
	78	96

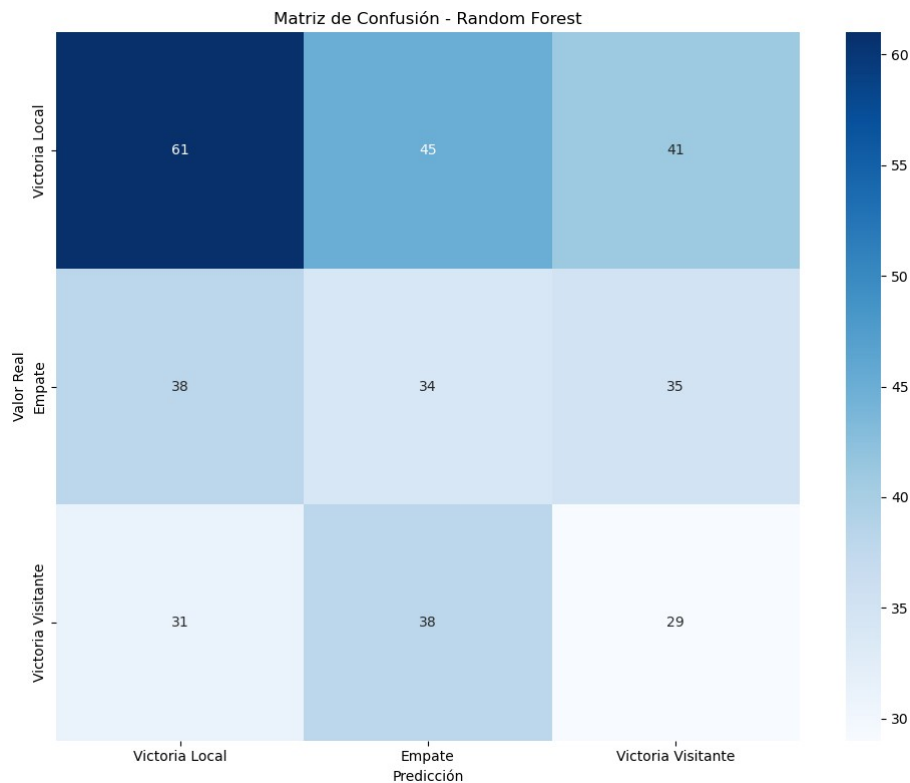
Evaluando modelo: Random Forest

El modelo de Random Forest demostró un accuracy promedio de 0.4250 y un F1-score promedio de 0.4167 durante la validación cruzada. Sin embargo, al ser evaluado en el conjunto de prueba, estas métricas disminuyeron a un accuracy de 0.3523 y un F1-score de 0.35582 . Este descenso sugiere un posible sobreajuste del modelo a los datos de entrenamiento o una menor capacidad de generalización a datos no vistos. Al observar el reporte de clasificación, se aprecia que la clase "Victoria Local" obtuvo la precisión y el recall más altos (0.47 y 0.41 respectivamente), mientras que las clases "Empate" y "Victoria Visitante" presentaron métricas considerablemente más bajas . Esto ya anticipa una de las dificultades centrales del proyecto: la predicción precisa de resultados distintos a la victoria del equipo local.

Class	Precision	Recall	F1-Score	Support
Victoria Local	0.47	0.41	0.44	147
Empate	0.29	0.32	0.3	107
Victoria	0.28	0.3	0.29	98
Visitante				
accuracy			0.35	352
macro avg	0.35	0.34	0.34	352

weighted avg 0.36 0.35 0.36 352

Matriz de Confusión Random Forest:



Evaluando modelo: XGBoost

El modelo XGBoost, en su evaluación sobre el conjunto de prueba, alcanzó un accuracy de 0.3693 y un F1-score de 0.36953 . Comparativamente con el Random Forest en el conjunto de prueba, XGBoost muestra un ligero aumento en ambas métricas. El reporte de clasificación revela una precisión de 0.44 y un recall de 0.41 para la clase "Victoria Local", valores similares a los del Random Forest . No obstante, para la clase "Empate", XGBoost exhibe una precisión y un recall ligeramente superiores (0.31 y 0.38 respectivamente), sugiriendo una mejor capacidad para identificar este resultado en comparación con el modelo anterior³ . La clase "Victoria Visitante" aún presenta métricas bajas (precisión de 0.35 y recall de 0.29), aunque marginalmente mejores que las del Random Forest

Accuracy: 0.3693

F1-score: 0.3695

Class	Precisi	Reca	F1-	Suppo
	on	ll	Score	rt
Victoria Local	0.44	0.41	0.43	147
Empate	0.31	0.38	0.34	107
Victoria	0.35	0.29	0.31	98
Visitante				
accuracy			0.37	352
macro avg	0.37	0.36	0.36	352
weighted avg	0.37	0.37	0.37	352

Evaluando modelo: Regresión Logística

Accuracy promedio: 0.3833

F1-score promedio: 0.3801

Evaluación en el conjunto de prueba:

Accuracy: 0.3352

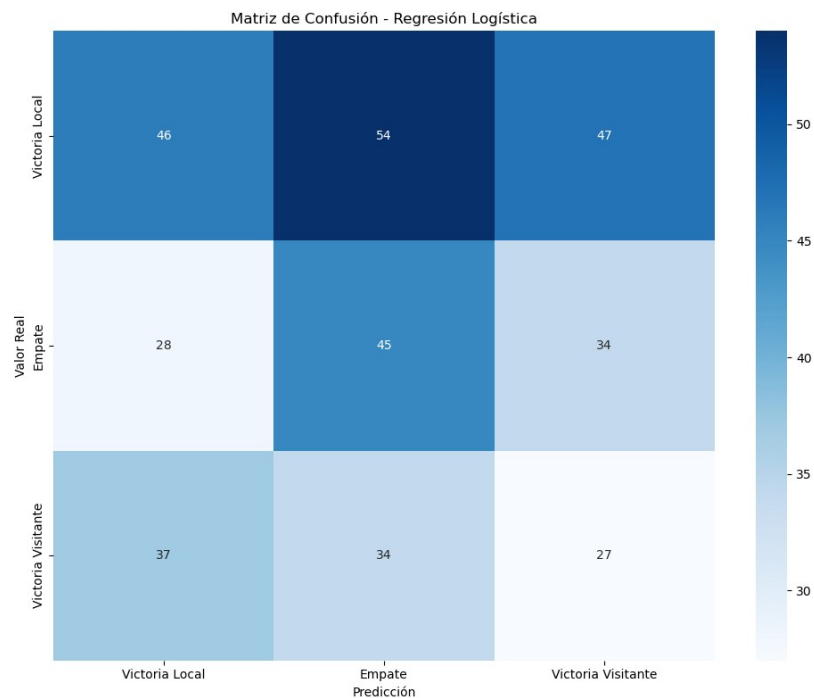
F1-score: 0.3359

El modelo de Regresión Logística, durante la validación cruzada, obtuvo un accuracy promedio de 0.3833 y un F1-score promedio de 0.3801. Al ser evaluado en el conjunto de prueba, su rendimiento decreció, alcanzando un accuracy de

0.3352 y un F1-score de 0.3359. El reporte de clasificación indica que la clase "Empate" logró una precisión relativamente buena de 0.34 y un recall de 0.42, siendo el recall más alto entre los tres modelos para esta clase. Sin embargo, la clase "Victoria Local" presenta una precisión y un recall más bajos en comparación con los modelos de ensamble (0.41 y 0.31 respectivamente), y la clase "Victoria Visitante" muestra las métricas más deficientes (precisión de 0.25 y recall de 0.28)

Reporte de Clasificación:

Class	Precisi on	Reca ll	F1- Score	Suppo rt
Victoria Local	0.41	0.31	0.36	147
Empate	0.34	0.42	0.37	107
Victoria Visitante	0.25	0.28	0.26	98
accuracy			0.34	352
macro avg	0.33	0.34	0.33	352
weighted avg	0.35	0.34	0.34	352



Análisis de Resultados Experimento 1

1. Resultados Generales de los Modelos

- Los resultados del Experimento 1 revelan un **rendimiento predictivo limitado en general para los tres modelos evaluados**, con accuracys en el conjunto de prueba que oscilan entre 0.3352 y 0.36931 Si consideramos que un clasificador aleatorio tendría una probabilidad de acierto de aproximadamente 0.33 en un problema de tres clases, estos resultados sugieren que **los modelos logran una mejora modesta sobre la predicción aleatoria**. El Random Forest demostró el accuracy promedio más alto durante la validación cruzada (0.4250), pero esta ventaja no se mantuvo en el conjunto de prueba XGBoost, por otro lado, mostró un rendimiento ligeramente superior en el conjunto de prueba en términos de accuracy y F1-score3 . La Regresión Logística fue el modelo con el rendimiento más bajo en el conjunto de prueba, a pesar de haber mostrado un accuracy promedio razonable durante la validación cruzada .

2. Comparación entre Matrices de Confusión

- El análisis de las matrices de confusión proporciona una visión más detallada del rendimiento de cada modelo por clase En el caso del **Random Forest**, se observa una tendencia a predecir con mayor precisión la clase "Victoria Local" (61 aciertos), aunque también incurre en un número considerable de errores al clasificar empates y victorias visitantes como victorias locales. El rendimiento en la clase "Victoria Visitante" es particularmente bajo, con pocas predicciones correctas . **XGBoost** presenta un patrón similar para la "Victoria Local" (61 aciertos), pero muestra una ligera mejora en la clasificación de la "Victoria Visitante" (28 aciertos), sugiriendo un **mejor equilibrio entre las clases en comparación con el Random Forest** . La **Regresión Logística** exhibe el **peor desempeño general**, con una distribución de errores más amplia y una tendencia significativa a confundir "Empate" con "Victoria Local" (54 errores de este tipo) .

3. Conclusión del Análisis

- En base a los resultados del Experimento 1, **XGBoost emerge como el modelo con un rendimiento ligeramente superior en el conjunto de prueba**, mostrando un mejor equilibrio entre las métricas de precisión, recall y F1-score para las diferentes clases,

especialmente para la clase "Empate" . Si bien el Random Forest obtuvo el mejor rendimiento promedio durante la validación cruzada para la "Victoria Local", su desempeño en el conjunto de prueba fue inferior al de XGBoost. La Regresión Logística, a pesar de su capacidad para identificar algunos empates, demostró ser el modelo menos efectivo en general, Estos resultados iniciales **subrayan la dificultad inherente en la predicción de resultados distintos a la victoria local**, un desafío que se evidenciará consistentemente a lo largo de este proyecto, La **calidad y disponibilidad de datos relevantes**, como se menciona en la literatura,, podría ser una limitación importante, así como la **naturaleza impredecible de los eventos deportivos** y la dificultad de incorporar **factores cualitativos** en los modelos

Experimento 2

Cambios en los parametros del experimento:

- 5 partidos para el calculo de features en ventana móviles
- se utilizan 15 características seleccionadas por random forest

Top 15 características según Importancia de Random Forest:

Feature	Importancia
Visitante_porcentaje_Duelos_ganados_rolling	0.013384
Visitante_porcentaje_Duelos aéreos_ganados_rolling	0.010382
Local_Pases largos_rolling	0.008998
Local_Balones perdidos_rolling	0.008957
Local_porcentaje_Pases largos_logrados_rolling	0.00882
Local_porcentaje_Córneres_con remate_rolling	0.008583
Local_Intensidad de paso_rolling	0.008524
Local_Pases_rolling	0.008227
Visitante_Entradas a ras de suelo_rolling	0.008121
Visitante_porcentaje_Entradas a ras de suelo_logradas_rolling	0.007998

Local_Promedio pases por posesión del balón_rolling	0.007992
Visitante_porcentaje_Centros_precisos_rolling	0.007618
Visitante_porcentaje_Jugadas a balón parado_con remate_rolling	0.007334
Visitante_Balones perdidos_rolling	0.007325
Local_Duelos defensivos_ganados_rolling	0.007237

Evaluando modelo: Random Forest

Accuracy promedio: 0.4364

F1-score promedio: 0.4278

Evaluación en el conjunto de prueba:

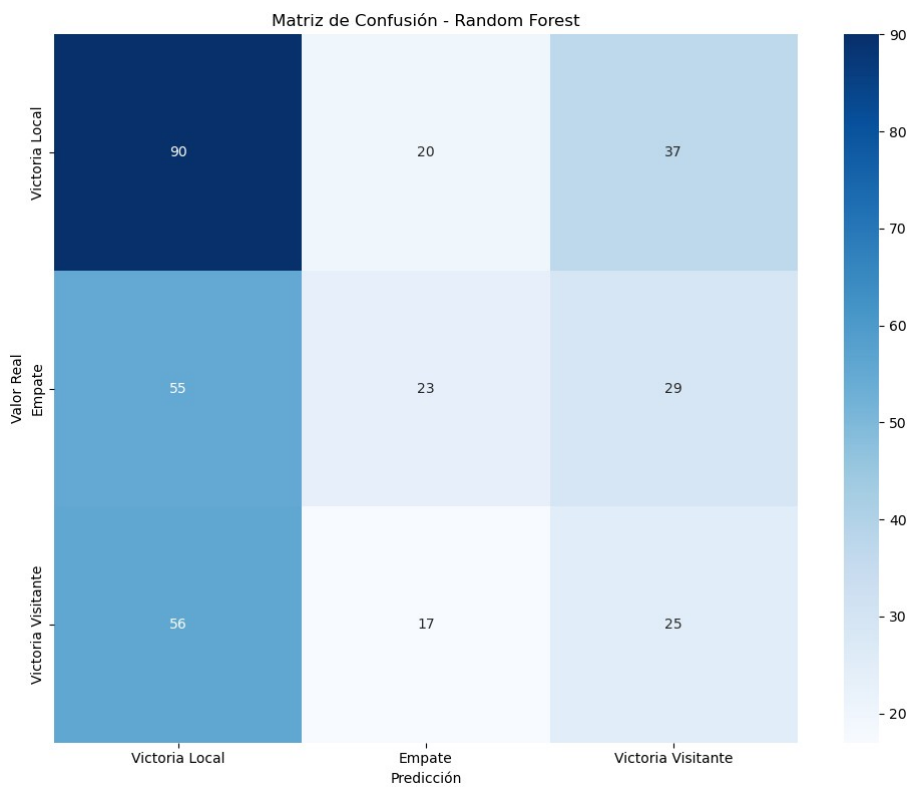
Accuracy: 0.3920

F1-score: 0.3734

Random Forest:

El modelo de Random Forest en el Experimento 2, utilizando 5 partidos para la ventana móvil y las 15 características seleccionadas por su importancia en este mismo modelo, logró un desempeño en el conjunto de prueba con un accuracy de 0.3920 y un F1-score de 0.3734. Se observa un buen desempeño en la predicción de "Victoria Local" con 90 casos correctos. Sin embargo, las clases "Empate" y "Victoria Visitante" continúan presentando desafíos significativos, con un número considerable de errores en la clasificación, especialmente confundiendo estas clases con "Victoria Local" (55 empates y 56 victorias visitantes clasificados incorrectamente como victoria local)

Class	Precisi on	Reca ll	F1- Score	Suppo rt
Victoria Local	0.45	0.61	0.52	147
Empate	0.38	0.21	0.28	107
Victoria Visitante	0.27	0.26	0.26	98
accuracy			0.39	352
macro avg	0.37	0.36	0.35	352
weighted avg	0.38	0.39	0.37	352



Evaluando modelo: XGBoost

Accuracy promedio: 0.4156

F1-score promedio: 0.4076

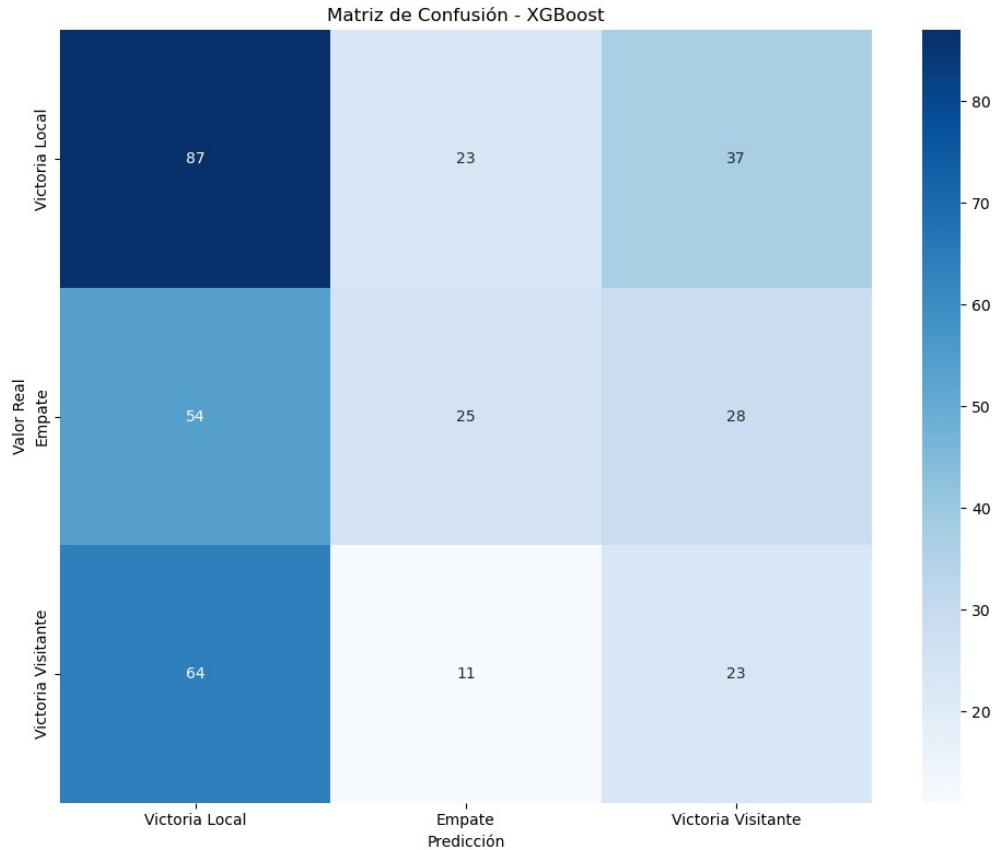
Evaluación en el conjunto de prueba:

Accuracy: 0.3835

F1-score: 0.3668

El modelo XGBoost, bajo las mismas condiciones del Experimento 2, alcanzó un accuracy de 0.3835 y un F1-score de 0.3668 en el conjunto de prueba. Similar al Random Forest, obtuvo un buen número de predicciones correctas para "Victoria Local" (87 casos). No obstante, mostró una mayor tendencia a fallar en la predicción de "Victoria Visitante", con 64 errores clasificados como "Victoria Local". La clase "Empate" también presenta un rendimiento moderado, reflejando la dificultad persistente en la identificación de este resultado.

Class	Precisi on	Reca ll	F1- Score	Suppo rt
Victoria Local	0.42	0.59	0.49	147
Empate	0.42	0.23	0.3	107
Victoria Visitante	0.26	0.23	0.25	98
accuracy			0.38	352
macro avg	0.37	0.35	0.35	352
weighted avg	0.38	0.38	0.37	352



Evaluando modelo: Regresión Logística

Accuracy promedio: 0.3740

F1-score promedio: 0.3760

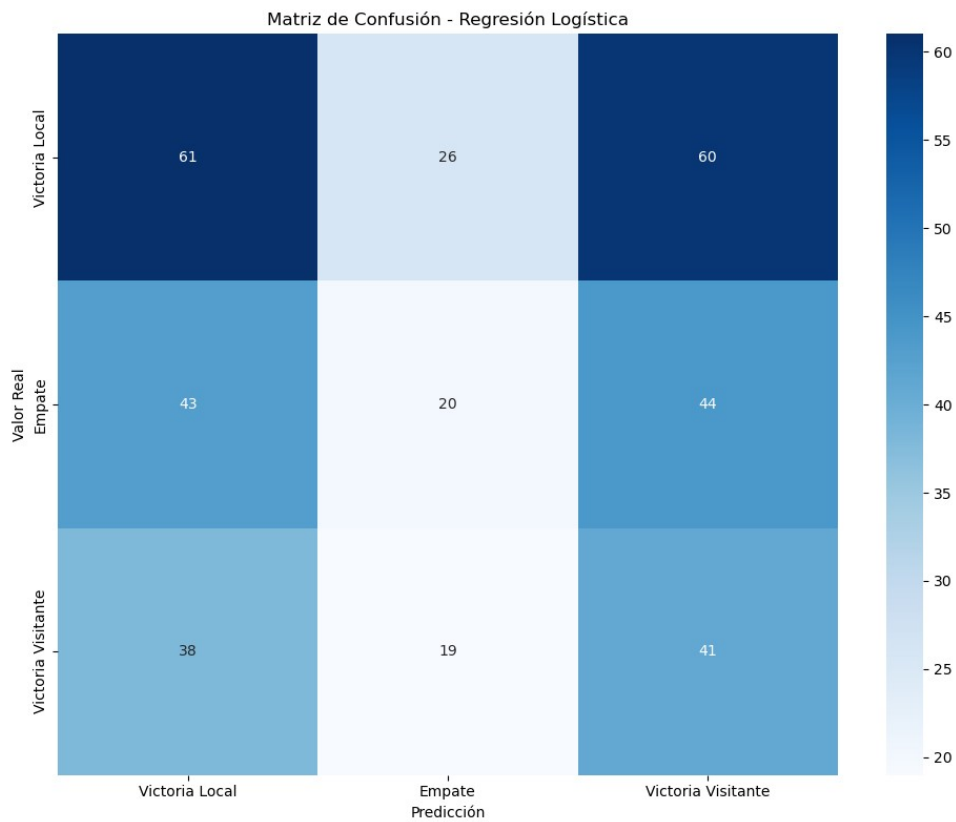
Evaluación en el conjunto de prueba:

Accuracy: 0.3466

F1-score: 0.3409

La Regresión Logística en el Experimento 2 demostró un accuracy de 0.3466 y un F1-score de 0.3409 en el conjunto de prueba. A diferencia de los modelos de ensamble, este modelo muestra un mayor equilibrio en la clasificación entre las tres clases, aunque a costa de una menor precisión general. Se observan errores significativos en la clasificación tanto de "Victoria Local" (60 casos clasificados como "Victoria Visitante") como de "Victoria Visitante"

Class	Precisi on	Reca ll	F1- Score	Suppo rt
Victoria Local	0.43	0.41	0.42	147
Empate	0.31	0.19	0.23	107
Victoria Visitante	0.28	0.42	0.34	98
accuracy			0.35	352
macro avg	0.34	0.34	0.33	352
weighted avg	0.35	0.35	0.34	352



Análisis:

Comparación de Modelos

Métrica	Random Forest	XGBoost	Regresión Logística
Accuracy Promedio	0.4364	0.4156	0.3740
F1-score Promedio	0.4278	0.4076	0.3760
Accuracy Prueba	0.3920	0.3835	0.3466
F1-score Prueba	0.3734	0.3668	0.3409

Detalles por Clase

Modelo	Clase	Precisión	Recall	F1-Score
Random Forest	Victoria Local	0.45	0.61	0.52
	Empate	0.38	0.21	0.28
	Victoria Visitante	0.27	0.26	0.26
	Victoria Local	0.42	0.59	0.49
XGBoost	Empate	0.42	0.23	0.30
	Victoria Visitante	0.26	0.23	0.25
	Victoria Local	0.43	0.41	0.42
	Empate	0.31	0.19	0.23
Regresión Logística	Victoria Visitante	0.28	0.42	0.34

Análisis de los resultados Experimento 2

Al comparar los modelos en el Experimento 2, Random Forest presenta el mejor rendimiento general en términos de accuracy y F1-score en el conjunto de prueba, particularmente en la predicción de "Victoria Local". XGBoost muestra un rendimiento muy similar, aunque con una mayor propensión a errores en la clase "Victoria Visitante". La Regresión Logística, si bien exhibe un equilibrio en la clasificación, se sitúa como el modelo con el rendimiento más bajo. De manera consistente con el Experimento 1, las clases "Empate" y "Victoria Visitante" siguen siendo difíciles de predecir, evidenciado por el bajo recall y F1-score para estas categorías en los tres modelos. Esto refuerza la hipótesis de que las características actualmente utilizadas podrían no ser suficientemente representativas de los factores que determinan estos resultados. La falta de información diferenciadora y la dificultad intrínseca en la predicción del empate, posiblemente debido a su menor frecuencia y a la influencia de factores cualitativos difíciles de cuantificar, continúan siendo limitaciones importantes.

Experimento 3

Parámetros del experimento:

Ventana Movil de 5 partidos para calcular rolling features

Métricas del modelo Random Forest

Accuracy: 0.3982

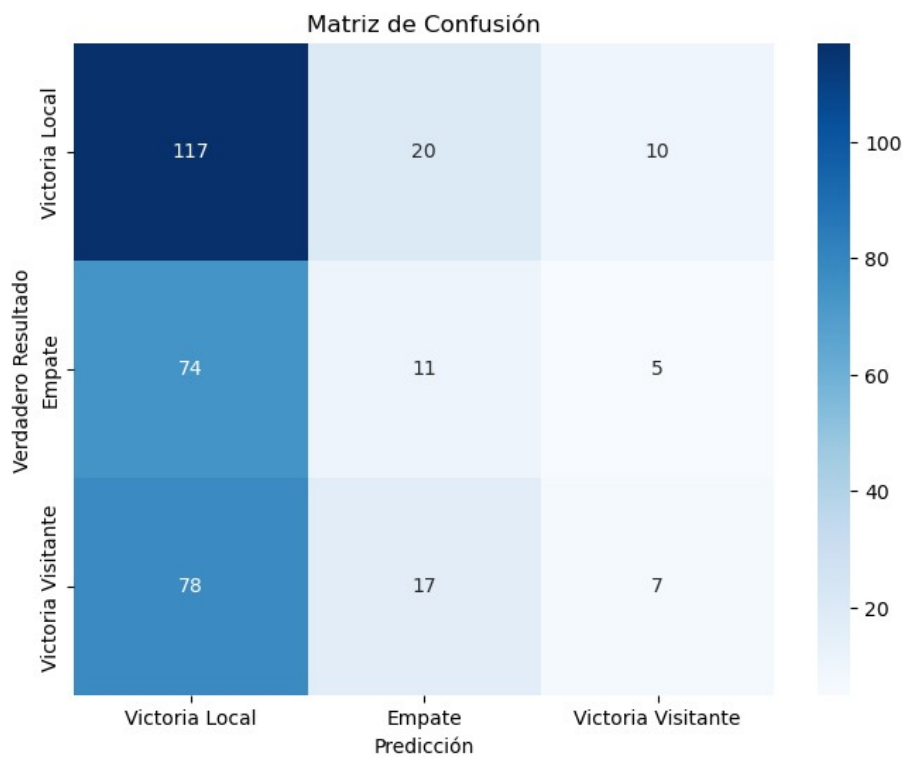
Precision: 0.3452

Recall: 0.3982

F1 Score: 0.3202

Class	Precisi on	Reca ll	F1- Score	Suppo rt
Victoria Local	0.43	0.8	0.56	147

Empate	0.23	0.12	0.16	90
Victoria	0.32	0.07	0.11	102
Visitante				
accuracy			0.4	339
macro avg	0.33	0.33	0.28	339
weighted avg	0.35	0.4	0.32	339



Optimización: Usando SMOTE para balancear la data:

Métricas:

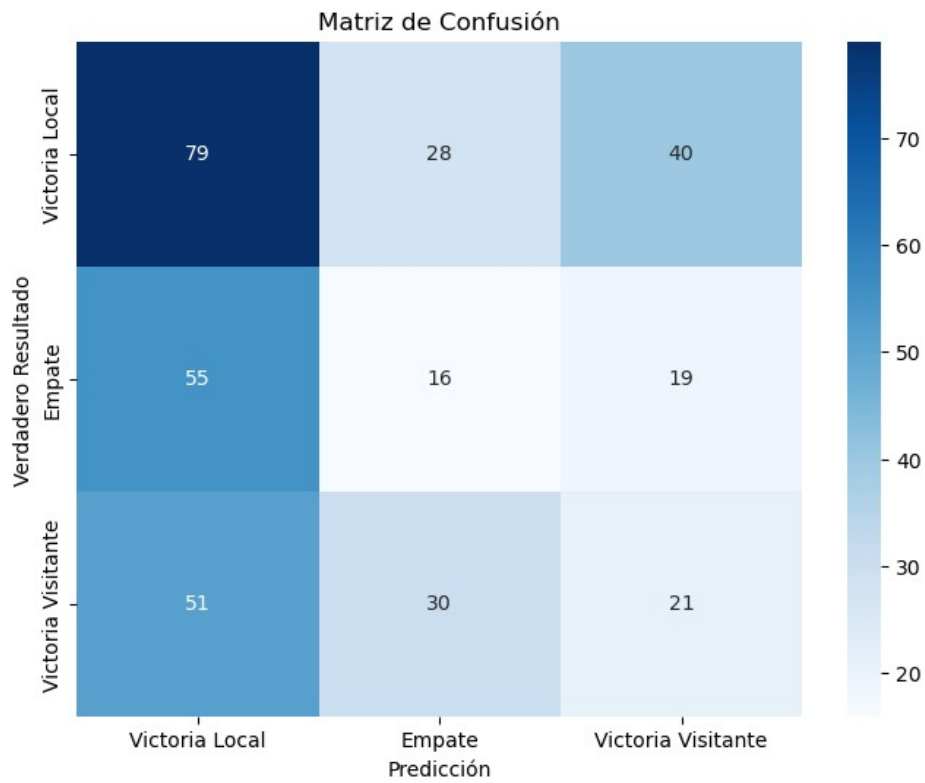
Accuracy: 0.3422

Precision: 0.3216

Recall: 0.3422

F1 Score: 0.3276

Class	Precisi on	Reca ll	F1- Score	Suppo rt
Victoria Local	0.43	0.54	0.48	147
Empate	0.22	0.18	0.2	90
Victoria Visitante	0.26	0.21	0.23	102
accuracy			0.34	339
macro avg	0.3	0.31	0.3	339
weighted avg	0.32	0.34	0.33	339



Optimización:

se realiza búsqueda de hiperparámetros para encontrar la mejor configuración del modelo RandomForestClassifier utilizando RandomizedSearchCV.

Parámetros del Modelo:

Se definen diferentes valores para probar en el modelo, como:

- Número de árboles (**n_estimators**): 50, 100, 200, etc.
- Profundidad máxima de los árboles (**max_depth**): Ningún límite, 10, 20, etc.
- Otros ajustes como tamaño mínimo de hojas (**min_samples_leaf**) y uso de muestreo (**bootstrap**).

2. RandomizedSearchCV:

En lugar de probar todas las combinaciones posibles de parámetros (lo que es muy lento), selecciona **50 combinaciones aleatorias** (**n_iter=50**).

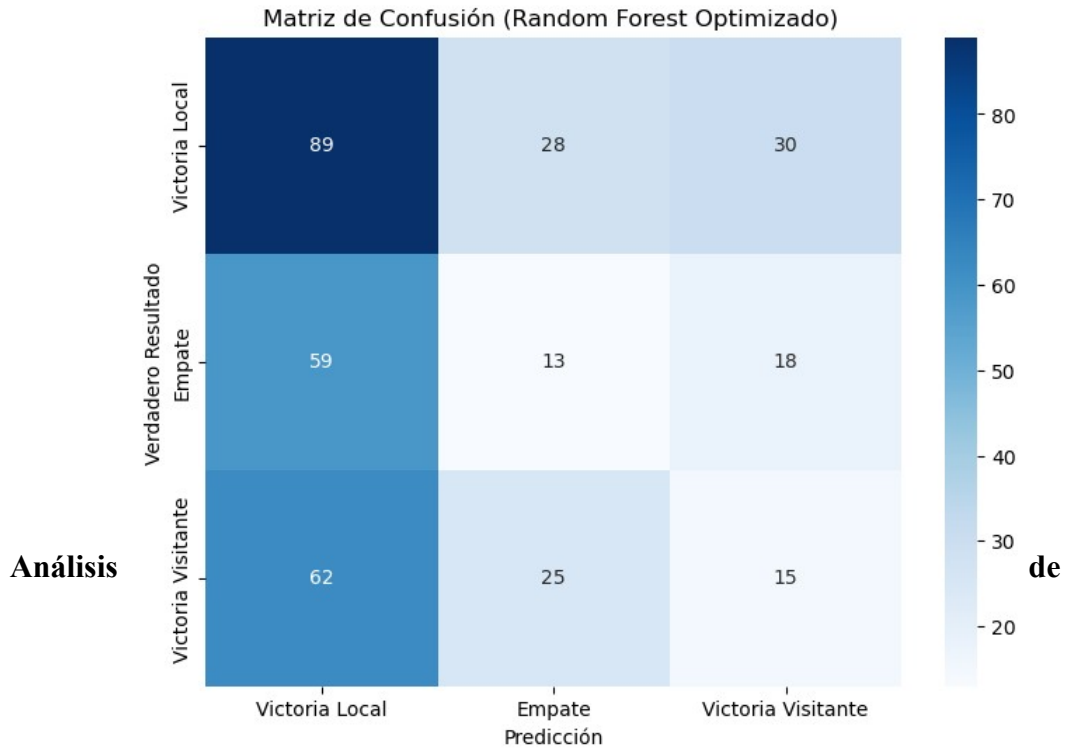
3. Entrenamiento y Validación:

- Usa **validación cruzada de 5 partes** (**cv=5**) para evaluar cada combinación.
- Optimiza el modelo buscando la **mayor precisión**.

Precisión del modelo optimizado: 0.3451

Class	Precisi on	Reca ll	F1- Score	Suppo rt
Victoria Local	0.42	0.61	0.5	147
Empate	0.2	0.14	0.17	90
Victoria Visitante	0.24	0.15	0.18	102

accuracy			0.35	339
macro avg	0.29	0.3	0.28	339
weighted avg	0.31	0.35	0.32	339



Resultados Experimento 3

El Experimento 3 se centró en la optimización del modelo Random Forest, inicialmente con una ventana móvil de 5 partidos. En la evaluación del modelo sin balanceo, se obtuvo un accuracy cercano al 40% (0.3982), con un F1 Score global de 0.320221 . Un análisis por clase revela un fuerte sesgo hacia la predicción de "Victoria Local", con un recall del 0.80, lo que indica que el modelo identifica la mayoría de estos casos, aunque también comete errores al clasificar otros resultados como victorias locales . Las clases "Empate" y "Victoria Visitante" mostraron un rendimiento muy deficiente, con recalls de 0.12 y 0.07 respectivamente, sugiriendo que el modelo apenas logra identificar correctamente estos resultados . Esto confirma la dificultad en la predicción de las clases minoritarias, probablemente debido al desequilibrio en la distribución de los datos .

La primera optimización consistió en el uso de SMOTE para balancear los datos de entrenamiento. Tras esta técnica, el accuracy global disminuyó ligeramente a 0.3422, aunque se observó una mejora en el recall de las clases minoritarias . El recall para "Empate" aumentó a 0.18 y para "Victoria Visitante" a 0.21, lo que indica una mayor capacidad del modelo para detectar estos resultados después del balanceo . Sin embargo, esta mejora en la detección de las clases minoritarias vino acompañada de una disminución en el rendimiento de la clase mayoritaria ("Victoria Local") y en el accuracy global . Esto es un patrón común al balancear clases: el modelo se vuelve menos sesgado hacia la clase dominante, pero puede perder precisión general .

Finalmente, se realizó una búsqueda de hiperparámetros utilizando RandomizedSearchCV sobre el modelo Random Forest con los datos balanceados por SMOTE . La precisión del modelo optimizado fue de 0.3451, ligeramente superior a la versión con SMOTE puro, pero aún menor que el modelo inicial sin balanceo . Al analizar las métricas por clase, se observó una ligera mejora en el recall de "Victoria Local" (0.61) en comparación con el modelo con SMOTE . Sin embargo, las métricas para "Empate" (recall de 0.14) y "Victoria Visitante" (recall de 0.15) no mostraron mejoras significativas, e incluso se situaron ligeramente por debajo de los resultados obtenidos con SMOTE sin optimización de hiperparámetros . Estos resultados sugieren que, si bien la optimización de hiperparámetros puede ofrecer mejoras marginales, la limitación principal reside en la capacidad de las variables utilizadas para discriminar entre las clases, especialmente el "Empate" . La calidad de las variables y la complejidad inherente del problema parecen ser factores más determinantes que la simple manipulación del balance de clases o los hiperparámetros del modelo

Conclusiones del Análisis:

1. Desequilibrio entre Clases:

El problema presenta un fuerte desequilibrio, sobre todo en la identificación de empates y victorias visitantes. La aplicación de SMOTE mejoró la capacidad del modelo para detectar las clases menos frecuentes (aumentando su Recall), pero redujo la exactitud global. Esto es esperable, pues el modelo deja de estar tan sesgado hacia la clase mayoritaria.

2. Calidad de las Variables:

A pesar de la ingeniería de características mediante rolling features, los

resultados sugieren que el conjunto de variables puede no ser suficiente para capturar la complejidad real del problema. El fútbol es altamente impredecible, y las condiciones contextuales (lesiones, estado anímico, clima, importancia del partido, alineaciones, etc.) no se capturan fácilmente en las métricas utilizadas. Esto podría explicar la dificultad para mejorar el rendimiento del modelo incluso con optimización y balanceo.

3. Más Datos o Más Contexto:

La mejora podría depender de añadir más datos históricos, características adicionales (por ejemplo, clasificación en la liga, estado del local frente al visitante en los últimos partidos, información sobre jugadores clave), o incluso datos que capturen aspectos cualitativos.

4. Expectativas Realistas:

La predicción de resultados deportivos es notoriamente complicada. Un 40% de exactitud en un problema con 3 clases (donde el azar sería alrededor del 33%) no es malo, pero indica que el modelo solo ofrece una ligera mejora sobre la predicción aleatoria. Mejorar en la clase "Empate" suele ser especialmente complejo, dado que es el resultado menos predecible.

En resumen, el modelo enfrenta un problema difícil con información limitada. Aunque se han aplicado estrategias para mejorar la clasificación, los resultados indican la necesidad de mayor variedad y calidad de datos, o el uso de otras metodologías, para lograr una mejora significativa.

Experimento 4

- Ventana de tiempo para los rolling features de 10
- se utilizan 29 features preseleccionadas del análisis anterior

- Aumento de Muestra de Datos usando ADASYN, Validación Cruzada Estratificada K-Fold, Inyección de Ruido y Escalador Robusto

1.- Escalado de Características con RobustScaler

Se aplicó el escalador robusto (RobustScaler) para escalar las características, eliminando la influencia de valores atípicos. El conjunto de entrenamiento se ajustó y transformó, y el conjunto de prueba se transformó usando el mismo escalador.

2.- Aplicación de ADASYN para Balancear el Conjunto de Datos

Se utilizó ADASYN para crear muestras sintéticas y balancear las clases en el conjunto de entrenamiento, reduciendo el sesgo hacia la clase mayoritaria.

3.- Inyección de Ruido para Crear Muestras Sintéticas Adicionales

Se implementó una función para generar copias ruidosas de los datos balanceados, creando muestras adicionales mediante la inyección de ruido gaussiano en diferentes niveles.

4.- Validación Cruzada Estratificada K-Fold y Optimización de Hiperparámetros

Se definió una función de objetivo para la validación cruzada estratificada utilizando StratifiedKFold con 5 divisiones. Los hiperparámetros del modelo RandomForestClassifier se ajustaron usando optimización bayesiana para maximizar la precisión promedio obtenida en las validaciones cruzadas.

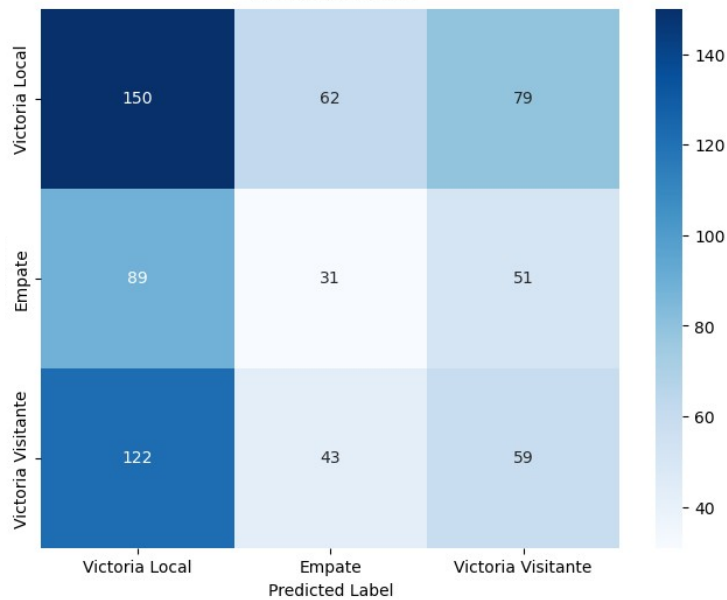
5.- Optimización Bayesiana con Optuna

Se utilizó la biblioteca Optuna para realizar la optimización de hiperparámetros mediante la búsqueda bayesiana. El objetivo era maximizar la precisión promedio obtenida durante la validación cruzada en cinco pruebas de optimización.

Clase	Precisión	Recall	F1-Score	Soprote
Victoria Local	0.42	0.52	0.46	291
Empate	0.23	0.18	0.2	171

Victoria Visitante	0.31	0.26	0.29	224
Exactitud (Accuracy)				686
Prom. Macro (avg)	0.32	0.32	0.32	686
Prom. Ponderado (avg)	0.34	0.35	0.34	686

Matriz de confusión Random Forest



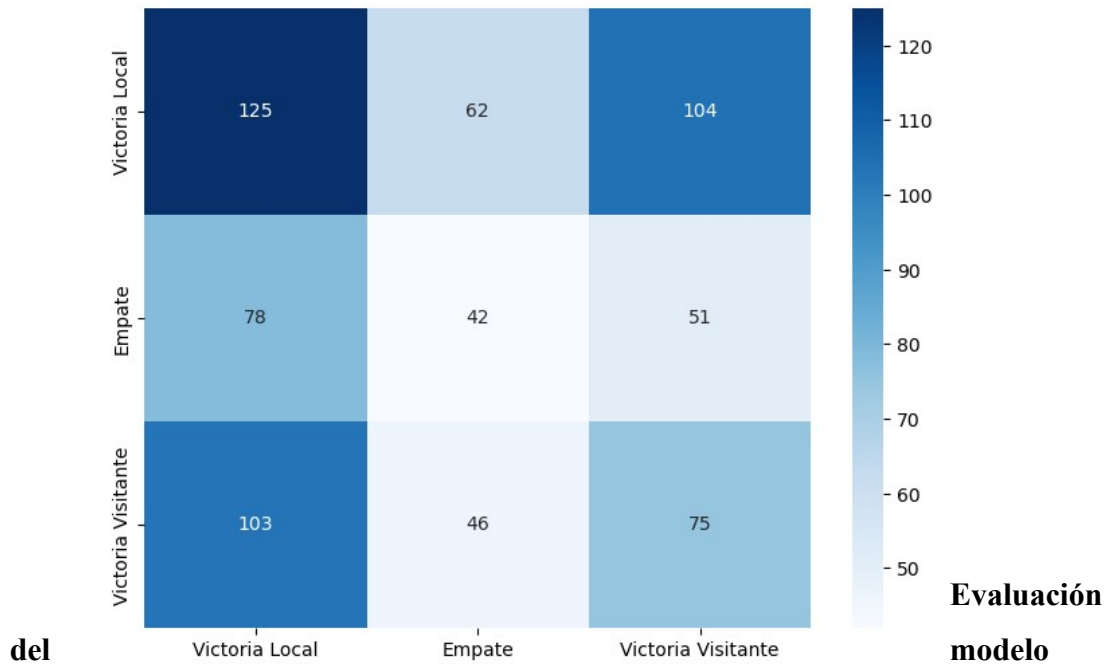
Evaluando el

modelo XGBoost Classifier

Clase	Precisión	Recall	F1-Score	Support
Victoria Local	0.41	0.43	0.42	291
Empate	0.28	0.25	0.26	171
Victoria Visitante	0.33	0.33	0.33	224
Exactitud (Accuracy)				686

Prom. Macro (avg)	0.34	0.34	0.34	686
Prom. Ponderado (avg)	0.35	0.35	0.35	686

Matriz de confusión XGBoost

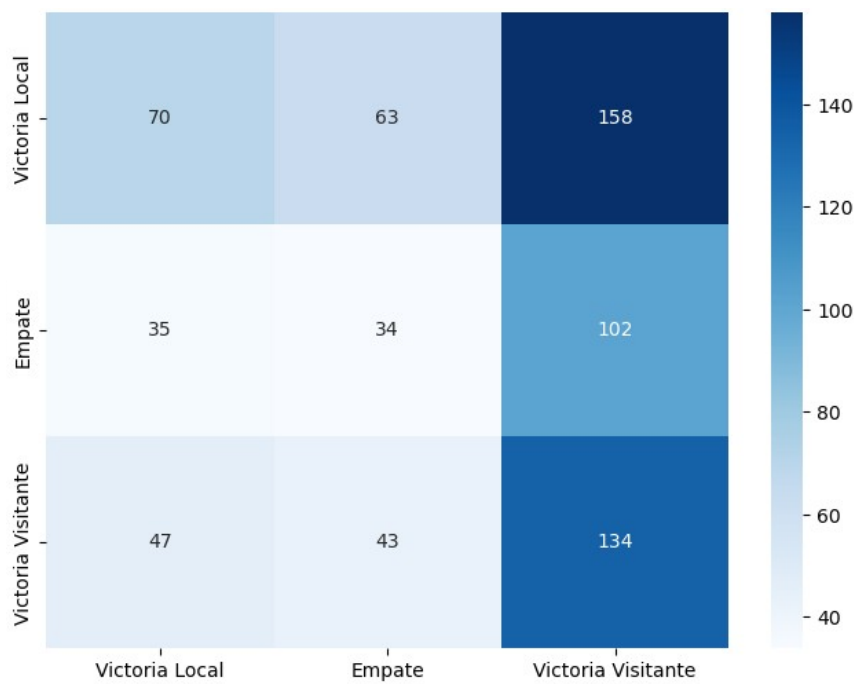


CatBoost

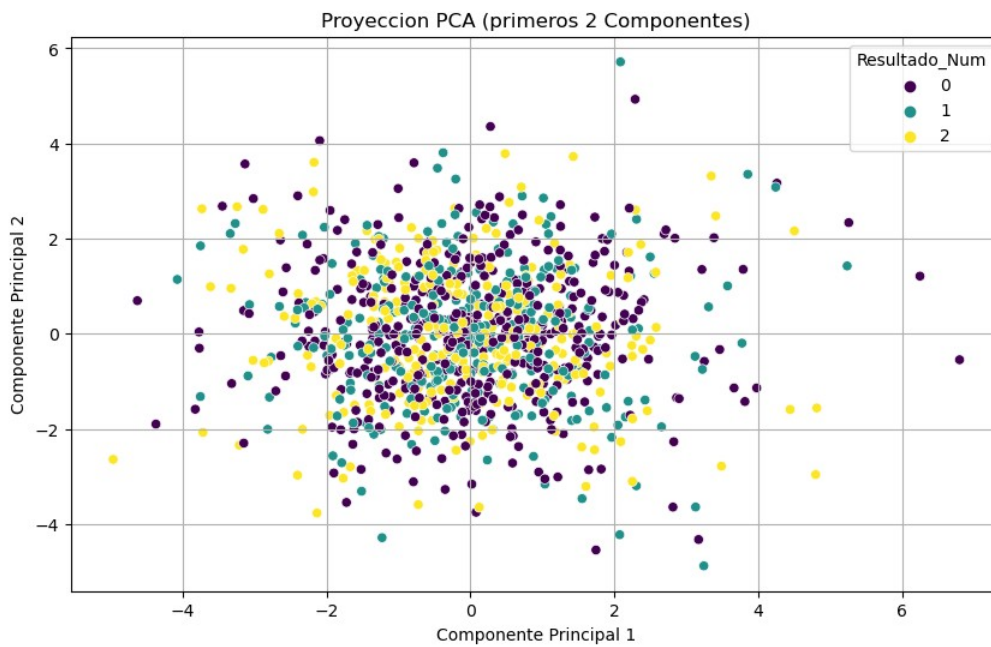
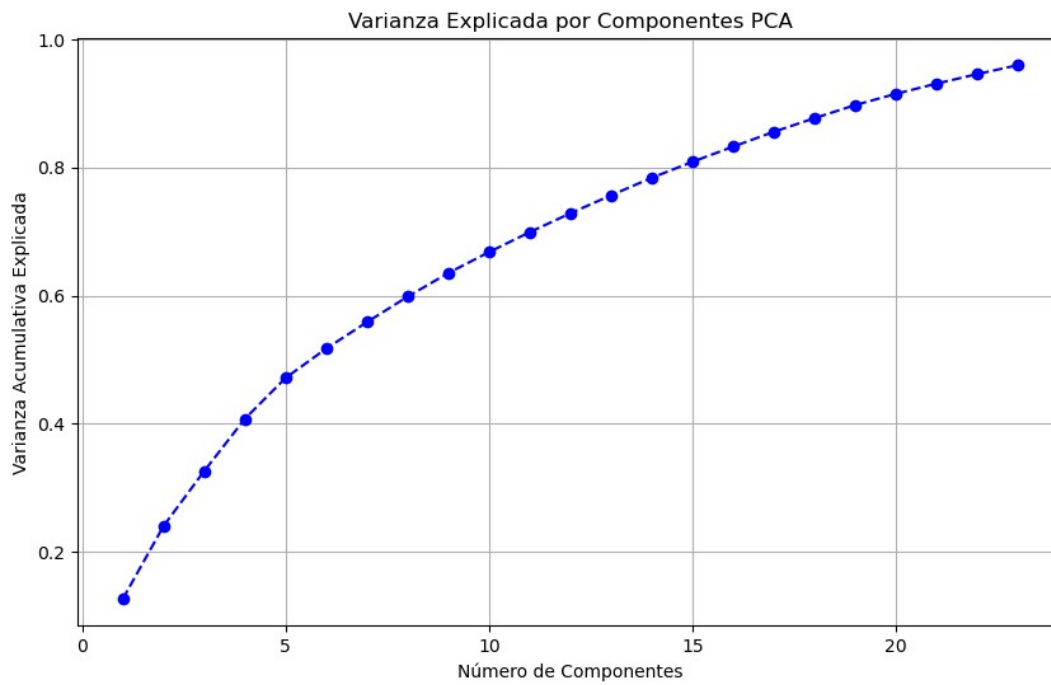
Clase	Precisión	Recall	F1-Score	Sporte
Victoria Local	0.46	0.24	0.32	291
Empate	0.24	0.2	0.22	171
Victoria Visitante	0.34	0.6	0.43	224

Exactitud				686
(Accuracy)				
Prom. Macro (avg)	0.35	0.35	0.32	686
Prom. Ponderado (avg)	0.37	0.35	0.33	686

Matriz de Confusión Modelo Catboost



Aplicar PCA y Reducir Dimensionalidad

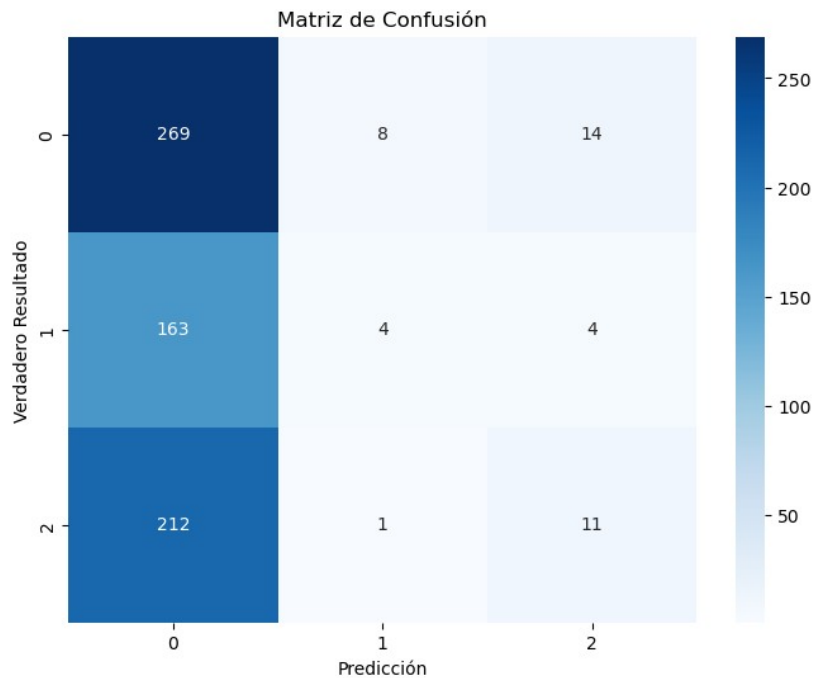


Evaluando Modelo Random Forest con dimensionalidad Reducida:

Precisión del set Test : 0.4140

Clase	Precisión	Recall	F1-Score	Soporte
0	0.42	0.92	0.58	291
1	0.31	0.02	0.04	171
2	0.38	0.05	0.09	224
Exactitud (Accuracy)				686
Prom. Macro (avg)	0.37	0.33	0.24	686
Prom. Ponderado (avg)	0.38	0.41	0.28	686

Matriz de Confusión de Random Forest con dimensionalidad Reducida



Contexto y Objetivo del Experimento:

En este experimento (Experimento 4) se buscó predecir el resultado de partidos (Victoria

Local, Empate, Victoria Visitante) a partir de un conjunto de datos al que se aplicaron diversos procesos: selección de 29 características relevantes, escalado robusto (RobustScaler), balanceo de clases mediante ADASYN, inyección de ruido para aumentar la muestra y validación cruzada estratificada con optimización bayesiana. Además, se evaluaron diferentes clasificadores (Random Forest, XGBoost, CatBoost) y posteriormente se exploró la reducción de dimensionalidad por PCA.

1. Resultados Generales de las Métricas

Los resultados del Experimento 4, que aplicó un conjunto de técnicas de preprocesamiento y aumento de datos a un conjunto de 29 características, mostraron una exactitud general de los modelos (Random Forest, XGBoost, CatBoost) que se sitúa alrededor del 35% al 40% en el conjunto de datos aumentado Este nivel de precisión no representa una mejora sustancial en comparación con los experimentos anteriores, y se mantiene cercano al rendimiento que se podría esperar de un clasificador ingenuo dada la distribución de las tres clases .

2. Precisión, Recall y F1 por Clase

Al analizar las métricas por clase, se mantiene la tendencia observada previamente . La clase "Victoria Local" consistentemente presenta las mejores métricas de precisión y, en algunos casos, de recall, lo que sugiere que los modelos siguen teniendo una mayor facilidad para identificar este resultado . La clase "Empate" continúa mostrando las métricas más bajas en precisión y recall, lo que indica que, a pesar de las técnicas de balanceo y aumento de datos aplicadas, los modelos aún luchan por capturar los patrones característicos de los partidos que terminan en empate . La clase "Victoria Visitante" exhibe métricas intermedias, superiores a las del empate pero aún lejos de un rendimiento satisfactorio .

3. Métricas Promedio y Matrices de Confusión

Las métricas promedio macro F1 se mantienen en un rango de 0.32 a 0.34, lo que evidencia una clasificación global que sigue siendo deficiente y desbalanceada . Las matrices de confusión de los diferentes modelos (Random Forest, XGBoost, CatBoost) revelan patrones similares de confusión. En general, la clase "Victoria Local" es la que acumula la mayor cantidad de aciertos, pero también se observa una frecuente confusión con las clases "Empate" y "Victoria Visitante". CatBoost, en particular, mostró una ligera tendencia a identificar más "Victoria Visitante" en

algunos casos, pero a expensas de un menor rendimiento en la predicción de "Victoria Local", sin que esto se tradujera en una mejora general significativa . Estos resultados sugieren que el problema subyacente no radica únicamente en el desbalance de clases, sino en la posible falta de características verdaderamente discriminativas en el conjunto de datos utilizado .La dificultad intrínseca de predecir el empate, influenciada por factores cualitativos no modelados , podría ser una limitación fundamental.

Conclusiones

Los resultados del Experimento 4 refuerzan la dificultad intrínseca del problema de predicción con el conjunto de datos y características disponibles. A pesar de la aplicación de técnicas avanzadas de balanceo (ADASYN) y generación de datos sintéticos con ruido, el problema del desequilibrio y la falta de separabilidad clara entre las clases persiste. Los modelos ensambladores evaluados, aunque potentes, no lograron un avance significativo en las métricas predictivas. La reducción de dimensionalidad mediante PCA, lejos de mejorar la situación, pareció exacerbar el sesgo hacia la clase mayoritaria, lo que sugiere que la información relevante para la discriminación de las clases minoritarias se distribuye en múltiples dimensiones y se pierde al reducir drásticamente la dimensionalidad. En conjunto, los resultados indican que se requiere una exploración más profunda de las características utilizadas y posiblemente la incorporación de nuevas fuentes de información para superar las limitaciones actuales

Experimento 5

el objetivo de este experimento es ver si aplicando el mismo proceso a otra liga , en este caso la liga premier de inglaterra se sigue dando estos mismos resultados usando el mismo proceso

Evaluación Modelo Random Forest

Métricas:

Accuracy: 0.5470

Precision: 0.5350

Recall: 0.5470

F1 Score: 0.4864

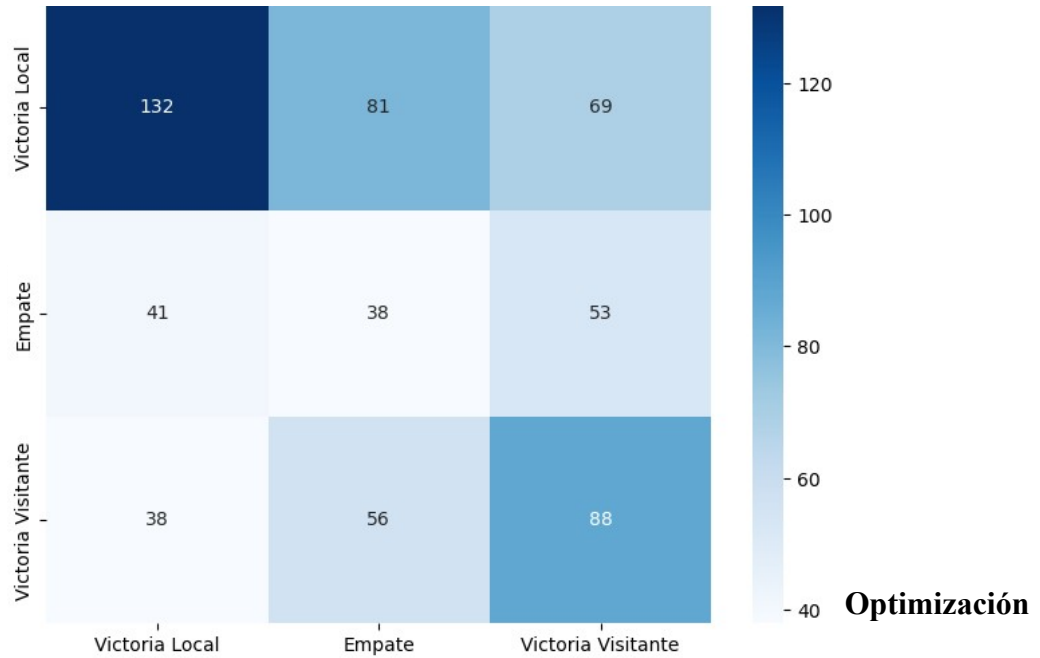
Resultados

Category	Precisi on	Reca ll	F1- Score	Suppo rt
Victoria Local	0.56	0.8	0.66	282
Empate	0.5	0.04	0.07	132
Victoria Visitante	0.53	0.52	0.52	182
accuracy		0.55		596
macro avg	0.53	0.45	0.42	596
weighted avg	0.53	0.55	0.49	596

Matriz de Confusión Random Forest:



Después de balancear la data con SMOTE

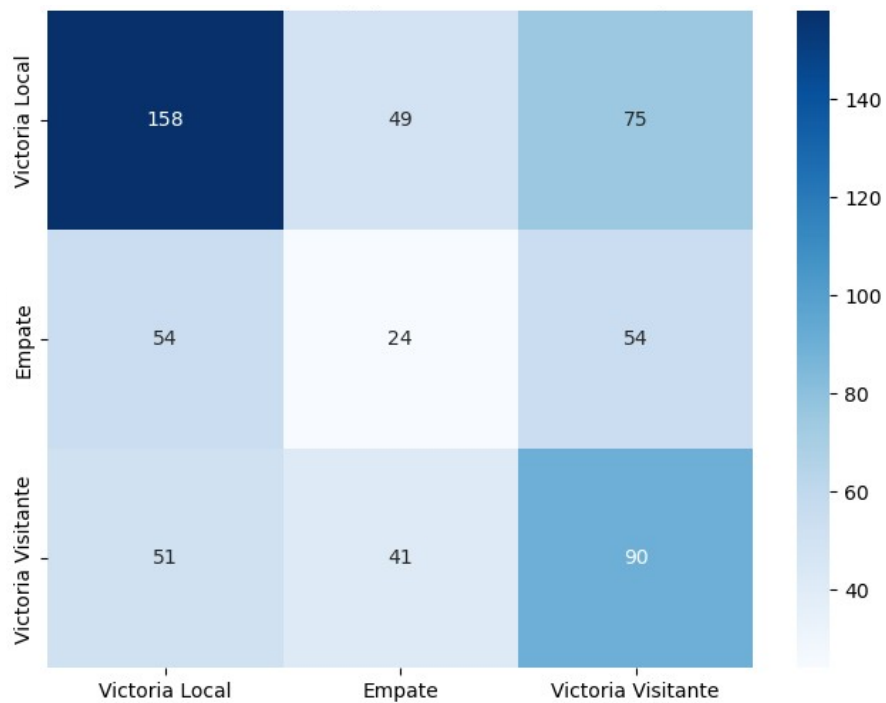


de

HiperParametros con RandomizedSearchCV

Category	Precisi on	Reca ll	F1- Score	Suppo rt
Victoria Local	0.6	0.56	0.58	282
Empate	0.21	0.18	0.2	132
Victoria Visitante	0.41	0.49	0.45	182
accuracy		0.46		596
macro avg	0.41	0.41	0.41	596
weighted avg	0.46	0.46	0.45	596

Matriz de confusión Random Forest Optimizado:

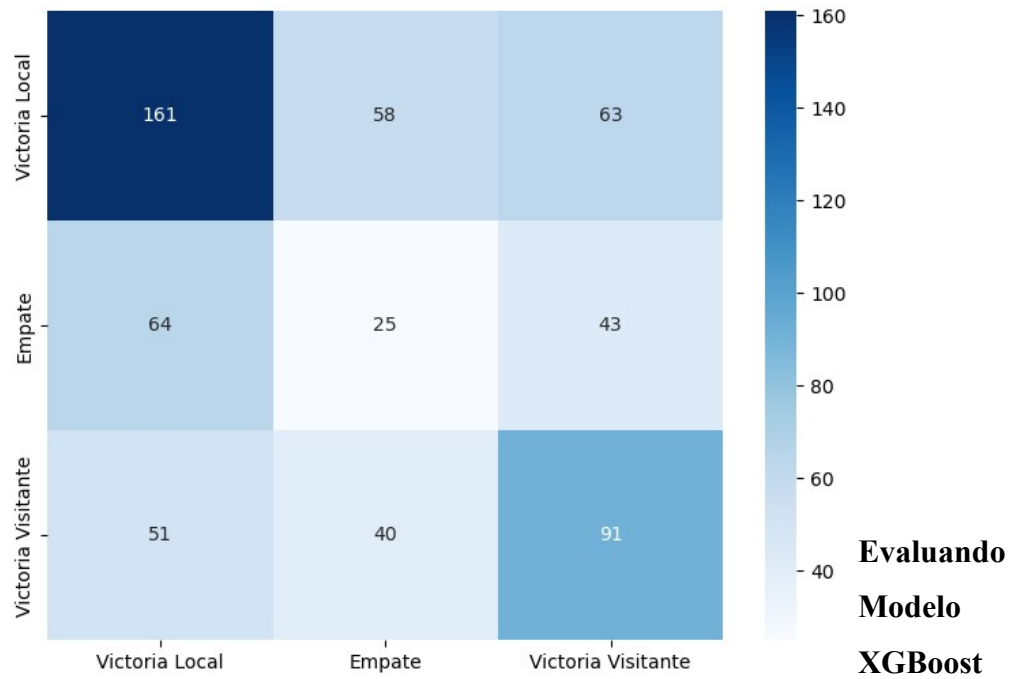


Optimización: Aumento de la Muestra de Datos utilizando ADASYN, Validación Cruzada Estratificada K-Fold, Inyección de Ruido y Scaler Robusto

Evaluación modelo RandomForest

Victoria Local	0.58	0.57	0.58	282
Empate	0.2	0.19	0.2	132
Victoria	0.46	0.5	0.48	182
Visitante				
Accuracy			0.46	596
Macro Avg	0.42	0.42	0.42	596
Weighted Avg	0.46	0.46	0.46	596

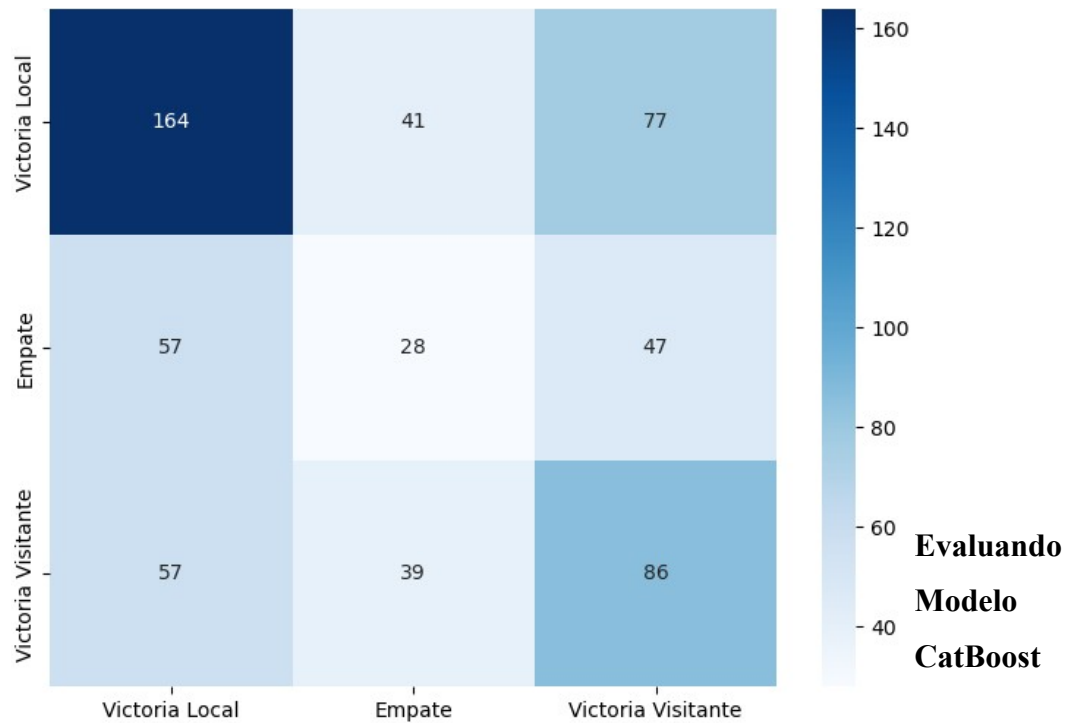
Matriz de Confusión Random Forest con más Data Generada



Classifier

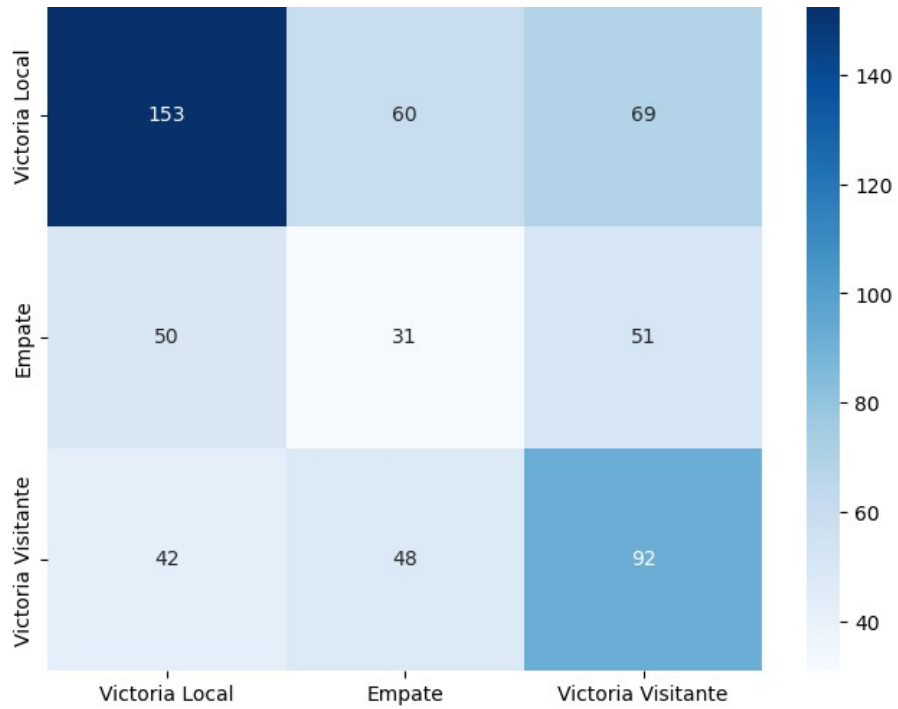
Clase	Precisi on	Reca ll	F1- Score	Suppo rt
Victoria Local	0.59	0.58	0.59	282
Empate	0.26	0.21	0.23	132
Victoria Visitante	0.41	0.47	0.44	182
Accuracy			0.47	596
Macro Avg	0.42	0.42	0.42	596
Weighted Avg	0.46	0.47	0.46	596

Matriz de Confusión XGBoost

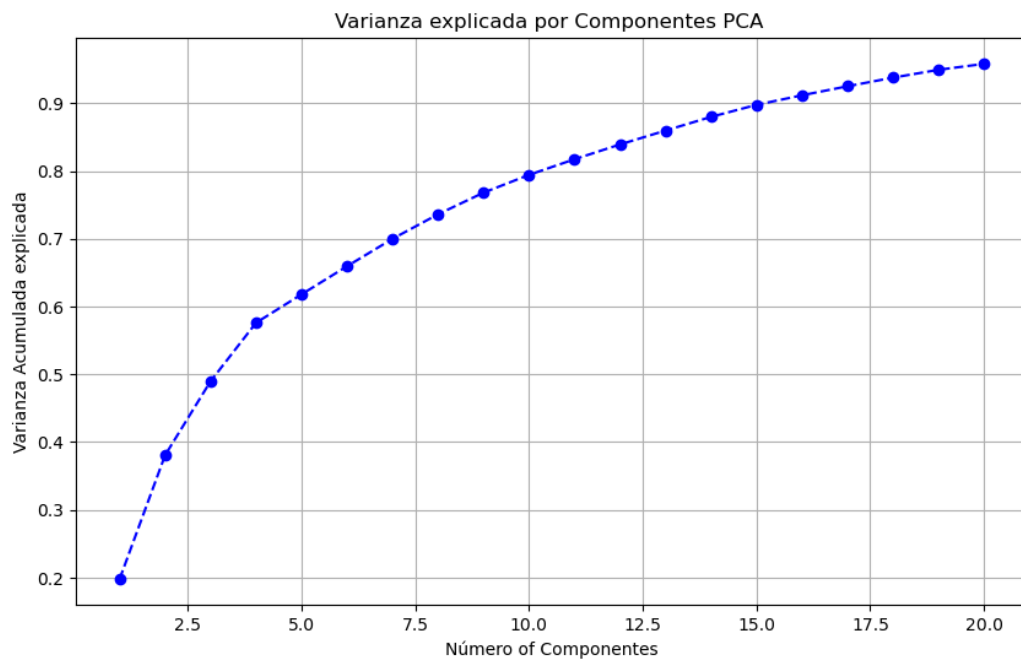


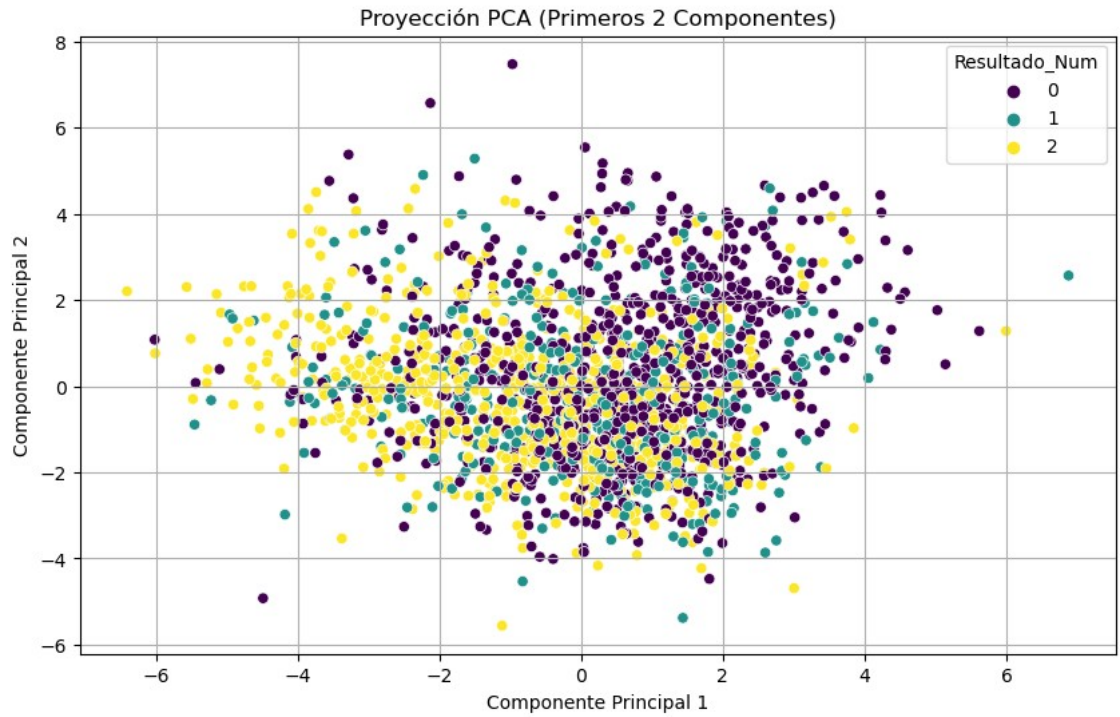
Clase	Precisi on	Reca ll	F1- Score	Suppo rt
Victoria Local	0.62	0.54	0.58	282
Empate	0.22	0.23	0.23	132
Victoria Visitante	0.43	0.51	0.47	182
Accuracy			0.46	596
Macro Avg	0.43	0.43	0.43	596
Weighted Avg	0.48	0.46	0.47	596

Matriz de Confusión CatBoost



Análisis PCA y Reducir Dimensionalidad

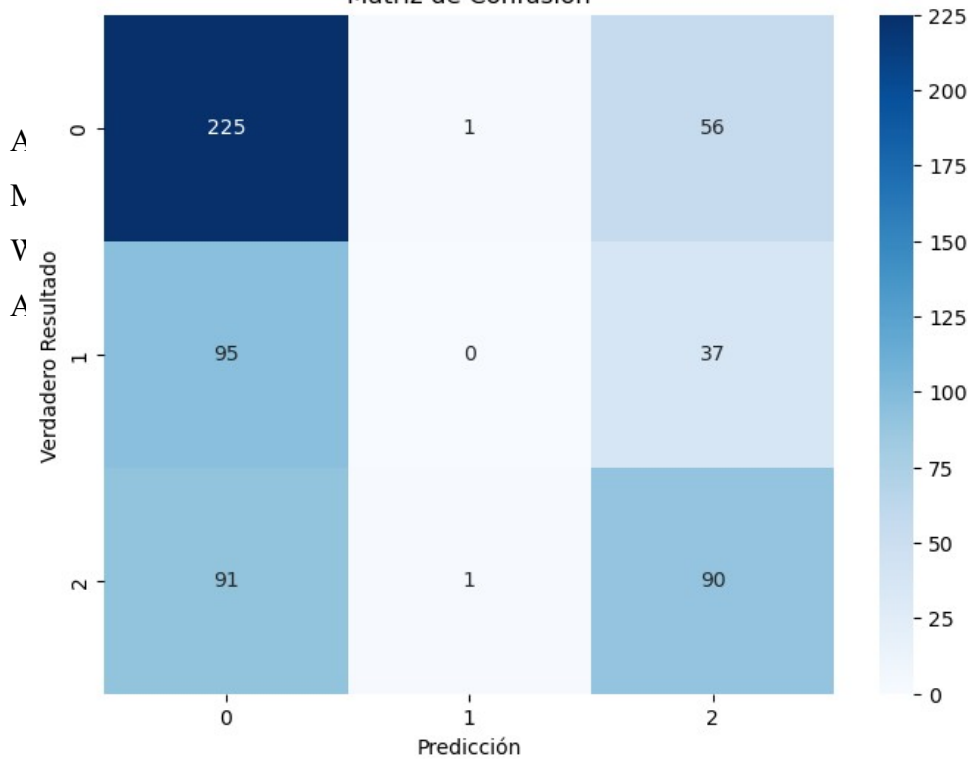




Evaluando Modelo Random Forest con dimensionalidad reducida

Clase	Precisi	Reca	F1-	Suppo
	on	ll	Score	rt
0	0.55	0.8	0.65	0.82

Matriz de Confusión



Análisis resultados experimento 5:

El Experimento 5 buscó verificar la consistencia de los resultados obtenidos en la liga chilena al aplicar el mismo proceso metodológico a datos de la Premier League de Inglaterra . Los resultados iniciales con el modelo Random Forest arrojaron un accuracy de 0.5470, mostrando un patrón similar al observado en la liga chilena: la clase "Victoria Local" fue la mejor clasificada, mientras que la clase "Empate" presentó las métricas más bajas y fue la que más se confundió con otras clases . Las métricas iniciales, aunque ligeramente superiores a las de la liga chilena, confirmaron la dificultad persistente en la predicción del empate.

Tras aplicar técnicas de balanceo con SMOTE, la optimización de hiperparámetros con RandomizedSearchCV y el aumento de datos con ADASYN, no se observaron mejoras sustanciales en la capacidad predictiva del modelo, especialmente para la clase "Empate" Las métricas fluctuaron ligeramente, pero el empate continuó siendo la clase más problemática, con una precisión y un recall consistentemente bajos . La evaluación con otros modelos como XGBoost y CatBoost reveló un patrón similar al de Random Forest . A pesar de las diferencias en los algoritmos, la dificultad para predecir el empate se mantuvo, sugiriendo que el problema no reside principalmente en el modelo utilizado, sino en la separabilidad de las clases en el espacio de características y en la complejidad inherente del dominio

El análisis de PCA en los datos de la Premier League también mostró una superposición significativa entre las clases, sin una separación clara en las dimensiones principales . La reducción de dimensionalidad antes del entrenamiento del modelo no condujo a mejoras, y en algunos casos, empeoró la clasificación de la clase "Empate" . Los resultados obtenidos en la Premier League, al replicar el proceso utilizado en la liga chilena, confirman la dificultad estructural de predecir empates y, en menor medida, victorias visitantes, lo que sugiere que este es un desafío inherente a la dinámica del fútbol y a la naturaleza de las variables empleadas, y no un problema específico de un conjunto de datos o liga. Esto refuerza la necesidad de explorar características adicionales, más específicas y de mayor calidad para capturar patrones predictivos más robustos, especialmente para el resultado del empate

Conclusiones Generales del Análisis:

- 1. Clase Empate Difícil de Predecir:** En todos los modelos y experimentos, la clase de Empate es la que presenta peores métricas. Ni el balanceo de datos ni la optimización de hiperparámetros han logrado aumentar significativamente su F1-Score.
- 2. Victoria Local Es la Clase Más Fácilmente Identificable:** Los modelos consistentemente muestran una mayor precisión y recall al predecir victorias locales, posiblemente debido a que son más comunes o presentan patrones más marcados.
- 3. PCA y Reducción de Dimensionalidad No Ayudan a Separar Mejor las Clases:** La visualización en 2D de la proyección PCA demuestra que las clases no se separan fácilmente, lo que sugiere que el espacio de atributos no presenta estructuras lineales claras que faciliten la clasificación.
- 4. Cambio de Modelo No Soluciona el Problema:** Ya sea Random Forest, XGBoost o CatBoost, el patrón se mantiene. Esto indica que el principal problema es la naturaleza del dataset y las variables empleadas, más que el algoritmo elegido.
- 5. Estrategias de Balanceo y Aumento de Datos (SMOTE, ADASYN), Inyección de Ruido y Escalados:** A pesar de aplicar múltiples estrategias, no se observan mejoras significativas. Esto refuerza la idea de que las características actuales no capturan adecuadamente la complejidad del fenómeno "empate".

6. Conclusiones

A continuación se presentan las conclusiones generales integrando los hallazgos de experimentos realizados:

Conclusiones Generales y Acciones Inmediatas:

Si bien los experimentos realizados demuestran la complejidad inherente a la predicción de resultados de fútbol con el conjunto de datos actual, varias conclusiones emergen con implicaciones para acciones inmediatas. En primer lugar, el rendimiento global de los modelos se mantiene limitado, sugiriendo que el conjunto de características actual no captura la variabilidad suficiente para superar un umbral cercano a la predicción aleatoria. Una acción inmediata podría ser realizar una revisión exhaustiva de las variables disponibles, evaluando su relevancia estadística y su potencial predictivo individualmente. Adicionalmente, se podría experimentar con la creación de características compuestas que combinen la información de múltiples variables para intentar capturar interacciones más complejas.

En segundo lugar, la consistente dificultad para predecir la clase "Empate" señala una necesidad urgente de enfocar los esfuerzos en identificar factores distintivos de este resultado. Una acción inmediata sería analizar en detalle los partidos que resultaron en empate, buscando patrones en las estadísticas previas que no se hayan destacado en el análisis general de correlaciones. Podría ser útil segmentar los datos de los partidos que terminaron en empate y compararlos directamente con los de victoria local y visitante para identificar qué variables presentan diferencias significativas.

Finalmente, aunque la "Victoria Local" es la clase mejor predicha, el rendimiento general aún es susceptible de mejora. Una acción inmediata podría ser evaluar si existe un desbalance excesivo en los datos originales antes de aplicar técnicas de balanceo sintético como SMOTE o ADASYN, ya que estos métodos, como se observó, no siempre generaron mejoras sustanciales y en algunos casos incluso disminuyeron la

precisión global. Se podría experimentar con diferentes estrategias de ponderación de clases durante el entrenamiento de los modelos para mitigar el sesgo hacia la clase mayoritaria sin recurrir necesariamente a la generación de datos sintéticos en primera instancia.

Limitaciones y posibles líneas futuras

Limitaciones Identificadas y Recomendaciones Concretas:

Las limitaciones del modelo para predecir con precisión, especialmente la dificultad con la clase "Empate", resaltan la carencia de información verdaderamente diferenciadora en el conjunto de características actual. Para abordar esta limitación de forma inmediata, se recomienda iniciar un proceso de recopilación de nuevas fuentes de datos. Esto podría incluir incorporar variables contextuales fácilmente accesibles, como el día de la semana del partido o información básica sobre las condiciones climáticas reportadas el día del encuentro, ya que estos factores podrían influir sutilmente en el desarrollo del juego. Asimismo, se sugiere explorar la disponibilidad de estadísticas de rendimiento más avanzadas que puedan ofrecer una visión más profunda del juego, como métricas sobre la efectividad de los equipos en diferentes fases del juego (ataque posicional, contraataque, jugadas a balón parado).

La dificultad intrínseca en la predicción del empate sugiere que este resultado podría depender de una combinación de factores que las estadísticas agregadas a nivel de equipo quizás no capturen adecuadamente. Una recomendación concreta en este sentido es investigar la viabilidad de incorporar información a nivel de jugador. Esto podría comenzar con la inclusión de datos sobre jugadores clave lesionados o suspendidos para cada partido, ya que su ausencia podría tener un impacto significativo en el rendimiento del equipo y potencialmente aumentar la probabilidad de un resultado más paado. A

futuro, se podría considerar la incorporación de métricas de rendimiento individual de los jugadores en los partidos recientes.

Finalmente, si bien se exploraron diversas técnicas de preprocesamiento y modelado, la falta de una mejora sustancial indica que el problema podría requerir un enfoque de modelado más sofisticado. Como una acción a corto plazo, se podría experimentar con la implementación de modelos de clasificación sensible a los costos, donde se asigna un mayor "costo" a los errores de clasificación en la clase "Empate" durante el entrenamiento del modelo. Esto podría incentivar al modelo a prestar más atención a esta clase minoritaria. A más largo plazo, se debería considerar la exploración de arquitecturas de aprendizaje automático más complejas, como redes neuronales sencillas, que podrían ser capaces de aprender relaciones no lineales entre las características y el resultado del partido, aunque esto requeriría una mayor inversión en tiempo y recursos computacionales.

Bibliografia

- [1] O. Hubacek, G. Sourek, and F. Zelezny. Learning to predict soccer results from relational data with gradient boosted trees. *Mach. Learn.*, 108:29–47, 2019. (Citado como Hubáček et al.)
- [2] Hubacek, O., Sourek, G., & Zelezny, F. (2022). Forty years of score-based soccer match outcome prediction: An experimental review. *IMA Journal of Management Mathematics*, 33, 1-18. <https://doi.org/10.1093/imaman/dpab029>
- [3] Y. F. Alfredo and S. M. Isa. Football Match Prediction with Tree Based Model Classification. *I. J. Intelligent Systems and Applications*, 11(7):20-28, 2019.
- [4] Milad Keshtkar Langaroudi et al./Sports Result Prediction Based on Machine Learning and Computational Intelligence Approaches: A Survey. *J. ADV COMP ENG TECHNOL*, 5(1) Winter 2019
- [5] Igiri, C. P., & Nwachukwu, E. O. (2014). An improved prediction system for football a match result. *IOSR Journal of Engineering (IOSRJEN)*, 4(12), 12-20. <https://doi.org/10.9790/3021-041212220>
- [6] Choi, B. S., Foo, L. K., & Chua, S. L. (2023). Predicting football match outcomes with machine learning approaches. *Mendel*, 29(4), 229-243. <https://doi.org/10.13164/mendel.2023.k.229>
- [7] Horvat, T., & Job, J. (2020). The use of machine learning in sport outcome prediction: A review. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 10(e1380). <https://doi.org/10.1002/widm.1380>
- [8] Fialho, G., Manhães, A., & Teixeira, J. (2019). Predicting sports results with artificial intelligence: A proposal framework for soccer games. *Procedia Computer Science*, 164, 131-136. <https://doi.org/10.1016/j.procs.2019.12.164>
- [9] Razali, N., Mustapha, A., Utama, S., & Din, R. (2018). A review on football match outcome prediction using Bayesian networks. *Journal of Physics: Conference Series*, 1020(1), 012004. <https://doi.org/10.1088/1742-6596/1020/1/012004>
- [10] Raju, M., Mia, M., Sayed, M. A., & Uddin, M. (2020). Predicting the outcome of English Premier League matches using machine learning. In *Proceedings of the 2nd*

International Conference on Sustainable Technologies for Industry 4.0 (STI). IEEE.

<https://doi.org/10.1109/STI50764.2020.9350327>

[11] Bunker, R., Yeung, C., & Fujii, K. (2024). Machine learning for soccer match result prediction. arXiv. <https://doi.org/10.48550/arXiv.2403.07669>

[12] Sarmiento, H., Marcelino, R., Anguera, M. T., Campaniço, J., Matos, N., & Leitão, J. C. (2014). Match analysis in football: a systematic review. *Journal of Sports Sciences*, 32(20), 1831–1843. <https://doi.org/10.1080/02640414.2014.898852>

[13] Ati, A., Bouchet, P., & Ben Jeddou, R. (2024). Using multi-criteria decision-making and machine learning for football player selection and performance prediction: A systematic review. *Data Science and Management*, 7, 79-88.

<https://www.keaipublishing.com/en/journals/data-science-and-management>

[14] Bunker, R., & Susnjak, T. (2022). The application of machine learning techniques for predicting match results in team sport: A review. *Journal of Artificial Intelligence Research*, 73, 1285-1322. <https://doi.org/10.48550/arXiv.1912.11762>

[15] Yeung, C. K. C., Bunker, R., & Fujii, K. (2023). A framework of interpretable match results prediction in football with FIFA ratings and team formation. *PLoS ONE*, 18(4), e0284318. <https://doi.org/10.1371/journal.pone.0284318>