



FootBase

Grado en Ingeniería Informática, Universidad de León

Sistemas de Información de Gestión y Business Intelligence

Trabajo realizado y estudiado por Airam Diez Flecha

León, 23 de diciembre de 2025

1. Contenido

2.	Descripción del proyecto	4
3.	Herramientas y tecnologías utilizadas	6
3.1	Base de datos: Neo4j	6
3.2	Inteligencia Artificial: Ollama y Groq	7
3.3	Interfaz de usuario: Streamlit	7
3.4	Lenguaje de programación e integración	7
4.	Origen de los datos y fuentes de información	8
5.	Arquitectura y diseño de la solución	9
5.1	Visión general de la arquitectura	9
5.2	Organización del proyecto	9
5.4	Integración de los componentes	10
6.	Funcionamiento de la aplicación	12
6.1	Obtención y preprocesado de los datos	12
6.2	Cálculo del rendimiento individual	12
6.3	Modelo de química entre jugadores	13
6.4	Construcción inicial de la convocatoria	13
6.5	Evaluación del equipo y función objetivo	13
6.6	Mejora iterativa de la solución	14
6.7	Resultado final	14
7.	Guía de usuario	15
7.1	Inicio de la aplicación	15
7.2	Tipos de consultas soportadas	15
7.3	Respuesta del sistema	17
8.	Casos de uso y escenarios de prueba	18
8.1	Caso de uso 1 – Escenario verosímil	18
8.2	Caso de uso 2 – Escenario optimista	19
8.3	Caso de uso 3 – Escenario pesimista	20
9.	Seguridad y ciberseguridad	22
9.1	Seguridad de los datos	22
9.2	Seguridad en el uso de modelos de lenguaje	22
9.3	Consideración del uso de API Key	23
9.4	Seguridad en un posible despliegue en servidor	23
9.5	Alcance de la seguridad del sistema	24
10.	Análisis crítico del sistema	24
10.1	Análisis DAFO	24
10.2	Limitaciones del proyecto	26

10.3 Valoración crítica global.....	26
11. Lecciones aprendidas y líneas de futuro	28
11.1 Lecciones aprendidas.....	28
11.2 Líneas de futuro.....	29
12. Enlaces y bibliografía	30

2. Descripción del proyecto

El presente proyecto consiste en el desarrollo de un sistema inteligente de recomendación de convocatorias de fútbol, planteado desde la perspectiva de un seleccionador nacional real. El objetivo principal es asistir en la toma de decisiones a la hora de confeccionar una convocatoria óptima de jugadores, basándose en datos objetivos y reduciendo la subjetividad inherente a este proceso.

En el fútbol profesional actual, los seleccionadores disponen de una gran cantidad de información estadística sobre los jugadores, sin embargo, el análisis manual de estos datos resulta complejo, lento y difícil de optimizar cuando se deben considerar múltiples criterios simultáneamente. El problema no se limita a seleccionar a los jugadores con mejores estadísticas individuales, sino a construir un equipo equilibrado, donde el rendimiento individual y la química entre jugadores maximicen el rendimiento colectivo.

Para abordar este problema, el proyecto utiliza una base de datos de grafos que contiene información estadística y relacional de aproximadamente 1000 futbolistas ficticios. A partir de estos datos, el sistema evalúa distintas combinaciones de jugadores con el fin de maximizar un score total, calculado como una combinación ponderada entre:

- el rendimiento individual de los futbolistas
- la química existente entre ellos.

El sistema permite además que el usuario especifique distintos parámetros antes de realizar la consulta, como el número de jugadores, la nacionalidad y el estilo de juego deseado (por ejemplo, ofensivo o defensivo). En función de estas preferencias, el algoritmo ajusta dinámicamente los pesos asociados a las estadísticas relevantes, priorizando distintos aspectos del juego según el contexto táctico.

Desde el punto de vista técnico, el proyecto combina varias tecnologías actuales: una base de datos orientada a grafos para modelar relaciones complejas, un algoritmo de recomendación basado en ponderaciones y exploración de combinaciones, y un sistema de inteligencia artificial conversacional que permite interactuar con la aplicación mediante un chatbot. Todo ello se integra en una interfaz sencilla y accesible, orientada tanto a la demostración académica como a una posible aplicación real en el ámbito deportivo.

Este proyecto surge de la motivación personal por el fútbol y la tecnología, y pretende demostrar cómo técnicas de ingeniería informática, inteligencia artificial y análisis de datos pueden aplicarse a un problema realista, ofreciendo una solución objetiva, flexible y extensible.

3. Herramientas y tecnologías utilizadas

Para el desarrollo del sistema de recomendación de convocatorias se ha optado por una arquitectura basada en tecnologías actuales, seleccionadas en función de la naturaleza del problema y de los requisitos del proyecto. A continuación, se describen las principales herramientas utilizadas y la justificación de su elección.

3.1 Base de datos: Neo4j

La base de datos utilizada es Neo4j, una base de datos orientada a grafos. Esta elección resulta especialmente adecuada para el problema planteado, ya que el dominio del fútbol presenta relaciones complejas y altamente interconectadas entre entidades como jugadores, equipos, ligas y países.

En el grafo, los jugadores se modelan como nodos principales, con un gran número de atributos estadísticos asociados (ofensivos, defensivos, físicos y de rendimiento general). Además, se crean nodos adicionales para equipos, ligas y países, junto con relaciones que permiten representar:

- el equipo en el que juega un futbolista
- la liga a la que pertenece dicho equipo
- la selección nacional que representa el jugador
- las relaciones de compañerismo entre jugadores que comparten equipo

Este modelo facilita la explotación de la química entre jugadores, ya que las relaciones del grafo permiten analizar conexiones directas e indirectas de forma eficiente, algo que resultaría mucho más complejo de implementar en una base de datos relacional tradicional.

3.2 Inteligencia Artificial: Ollama y Groq

Para la integración de inteligencia artificial conversacional se ha utilizado Ollama ejecutándose en local. Esta herramienta permite disponer de un modelo de lenguaje sin depender exclusivamente de servicios externos, lo que aporta mayor control sobre la aplicación y sus datos.

No obstante, debido a las limitaciones de almacenamiento y recursos locales, se ha integrado Groq como soporte para poder utilizar un modelo de lenguaje de mayor tamaño (LLaMA 3B). De este modo, el sistema puede beneficiarse de un modelo más potente sin requerir grandes recursos hardware en local.

El modelo de lenguaje cumple una doble función dentro de la aplicación:

- interpretar las peticiones del usuario expresadas en lenguaje natural,
- generar consultas estructuradas que permiten interactuar con la base de datos de grafos.

3.3 Interfaz de usuario: Streamlit

La interfaz de la aplicación se ha desarrollado utilizando Streamlit, una herramienta que permite crear aplicaciones web interactivas de forma rápida y sencilla a partir de código Python.

Streamlit se ha elegido por su facilidad de uso, su rapidez de desarrollo y su idoneidad para la demostración de prototipos funcionales, permitiendo implementar un chatbot accesible e intuitivo sin necesidad de desarrollar una interfaz web compleja desde cero.

3.4 Lenguaje de programación e integración

Todo el proyecto ha sido desarrollado en Python, integrando en una única aplicación la base de datos, la lógica de recomendación, la inteligencia artificial y la interfaz de usuario. Esta integración facilita el mantenimiento del sistema y permite una comunicación fluida entre los distintos componentes de la aplicación.

4. Origen de los datos y fuentes de información

Los datos utilizados en el proyecto no proceden de fuentes reales externas, sino que han sido generados de forma sintética mediante un script en Python. Esta decisión se tomó de manera consciente tanto por motivos técnicos como académicos y de seguridad.

La base de datos está compuesta por aproximadamente 1000 futbolistas ficticios, cada uno de ellos con un conjunto amplio de estadísticas que cubren distintos aspectos del juego. Entre los datos generados se incluyen estadísticas ofensivas, defensivas, físicas, de pases y específicas de portero, lo que permite simular un entorno realista y completo para el análisis y la recomendación de convocatorias.

El uso de datos generados presenta varias ventajas relevantes para el proyecto:

- Permite evitar problemas legales y de privacidad asociados al uso de datos reales de jugadores profesionales.
- Garantiza un entorno de pruebas seguro, sin exponer información sensible.
- Facilita la generación de conjuntos de datos amplios y equilibrados, adecuados para evaluar el comportamiento del algoritmo de recomendación.
- Proporciona libertad para adaptar la estructura de los datos a las necesidades del modelo, sin depender de formatos o limitaciones externas.

Además, la generación sintética de los datos asegura la coherencia estadística necesaria para simular escenarios realistas, permitiendo que el sistema evalúe combinaciones de jugadores de forma verosímil. De este modo, el proyecto se centra en el diseño y funcionamiento del sistema de recomendación, más que en la obtención o limpieza de datos reales.

Este enfoque resulta especialmente adecuado para un contexto académico y de demostración, donde el objetivo principal es validar la arquitectura, la lógica de selección y la integración de las distintas tecnologías empleadas.

5.Arquitectura y diseño de la solución

La arquitectura del sistema ha sido diseñada siguiendo un enfoque modular y claramente estructurado, integrando en una única aplicación todos los componentes necesarios para la recomendación de convocatorias, pero manteniendo una separación lógica entre responsabilidades.

5.1 Visión general de la arquitectura

La aplicación sigue un flujo de información bien definido que comienza en la interacción del usuario y finaliza en la generación de una recomendación óptima. A alto nivel, el funcionamiento puede resumirse en los siguientes pasos:

- El usuario interactúa con la aplicación a través de una interfaz web en forma de chatbot.
- La petición en lenguaje natural es procesada por el sistema de inteligencia artificial.
- A partir de la interpretación de la consulta, se generan consultas estructuradas contra la base de datos.
- Los datos obtenidos se procesan mediante la lógica de recomendación.
- El sistema devuelve al usuario una convocatoria óptima basada en los criterios solicitados.

Este flujo permite desacoplar la interacción con el usuario de la lógica interna del sistema, facilitando la extensibilidad y el mantenimiento de la aplicación.

5.2 Organización del proyecto

El proyecto se organiza en tres bloques principales, claramente diferenciados en la estructura de directorios:

- backend/

Contiene la lógica principal de la aplicación. En este módulo se integra:

- la conexión con la base de datos Neo4j
- la lógica de recomendación
- la comunicación con el modelo de lenguaje
- y los endpoints necesarios para el funcionamiento del sistema.

- datasets/

Alberga los conjuntos de datos generados mediante scripts en Python. Estos datasets se utilizan como base para la carga inicial de jugadores y estadísticas en la base de datos de grafos.

- neo4j/

Contiene los datos persistentes de la base de datos de grafos, incluyendo:

- bases de datos
- índices
- esquemas
- transacciones y metadatos del sistema

Esta separación permite aislar claramente los datos, la lógica de negocio y la persistencia, favoreciendo un diseño limpio y escalable.

5.4 Integración de los componentes

Todos los componentes del sistema se encuentran integrados en una única aplicación Python, lo que simplifica la comunicación entre módulos y reduce la complejidad de despliegue. No obstante, a nivel lógico, cada parte cumple una función bien delimitada:

- la interfaz gestiona la interacción con el usuario
- la inteligencia artificial interpreta las consultas
- la base de datos proporciona los datos necesarios
- y la lógica interna calcula la mejor convocatoria posible.

Este enfoque híbrido permite beneficiarse de una integración sencilla sin renunciar a una estructura modular clara.

6. Funcionamiento de la aplicación

El funcionamiento de la aplicación se basa en la generación automática de una convocatoria óptima de futbolistas a partir de un conjunto de criterios definidos por el usuario. El sistema combina información estadística individual con un modelo de química entre jugadores, resolviendo el problema como un proceso de optimización iterativa.

6.1 Obtención y preprocesado de los datos

En primer lugar, el sistema recupera desde la base de datos Neo4j todos los jugadores que cumplen los criterios básicos de la consulta, como la nacionalidad y, opcionalmente, el estado de lesión. Estos datos se cargan en memoria y se transforman a una estructura tabular para facilitar su procesamiento.

Las variables numéricas asociadas a cada jugador se normalizan, de forma que todas las estadísticas se encuentran en un rango comparable. Este paso es fundamental para evitar que métricas con escalas diferentes influyan de forma desproporcionada en el cálculo del rendimiento.

6.2 Cálculo del rendimiento individual

Cada jugador recibe un score individual que representa su rendimiento global. Este score se calcula como una suma ponderada de estadísticas, donde los pesos dependen de dos factores principales:

- la posición general del jugador (portero, defensa, centrocampista o delantero)
- y el estilo de juego seleccionado (ofensivo, defensivo o balanceado)

De este modo, el sistema puede priorizar distintos atributos según el contexto táctico. Por ejemplo, en un estilo ofensivo se da mayor importancia a goles, asistencias o pases clave, mientras que en un estilo defensivo se priorizan entradas, intercepciones o porterías a cero. Este diseño permite adaptar la convocatoria a diferentes planteamientos estratégicos.

6.3 Modelo de química entre jugadores

Además del rendimiento individual, el sistema introduce un modelo de química entre jugadores, que mide la compatibilidad entre pares de futbolistas. Esta química se define en función de relaciones contextuales simples pero efectivas, como:

- pertenecer al mismo equipo,
- competir en la misma liga.

Cada pareja de jugadores recibe un valor de química que posteriormente se agrega a nivel de equipo. Este enfoque permite favorecer convocatorias donde los jugadores presentan mayor familiaridad y cohesión, simulando sinergias reales del fútbol profesional.

6.4 Construcción inicial de la convocatoria

A partir de los scores individuales, el sistema construye una convocatoria inicial respetando una estructura predefinida de posiciones específicas (por ejemplo, número de porteros, defensas centrales, laterales, mediocentros o delanteros). Esta estructura puede ajustarse dinámicamente y también permite fijar jugadores concretos que deban estar incluidos en la convocatoria.

La selección inicial se realiza priorizando a los jugadores con mayor rendimiento individual dentro de cada posición, garantizando así una base competitiva.

6.5 Evaluación del equipo y función objetivo

El rendimiento global de una convocatoria se evalúa mediante una función objetivo que combina dos componentes:

- el rendimiento medio individual de los jugadores seleccionados,
- y la química media entre todos los pares de jugadores del equipo.

Ambos componentes se combinan mediante pesos globales, permitiendo controlar la importancia relativa del rendimiento frente a la cohesión del grupo. El resultado es un score total del equipo, que sirve como métrica de comparación entre distintas convocatorias.

6.6 Mejora iterativa de la solución

Una vez construida la convocatoria inicial, el sistema aplica un proceso de mejora iterativa. En cada iteración, se evalúan posibles sustituciones de jugadores dentro de la misma posición, comprobando si el cambio produce una mejora en el score total del equipo.

Este proceso continúa hasta que no se encuentran mejoras adicionales o se alcanza un número máximo de iteraciones. De este modo, el sistema converge hacia una solución localmente óptima sin necesidad de explorar exhaustivamente todas las combinaciones posibles, lo que mantiene un equilibrio entre calidad de la solución y eficiencia computacional.

6.7 Resultado final

Finalmente, la aplicación devuelve al usuario:

- la lista de jugadores seleccionados,
- el estilo de juego utilizado,
- el score total del equipo,
- el rendimiento medio,
- y la química media de la convocatoria.

Este enfoque permite obtener recomendaciones coherentes, adaptables y explicables, alineadas con la forma de trabajo de un seleccionador nacional real.

7. Guía de usuario

La aplicación está diseñada para ser utilizada de forma intuitiva a través de una interfaz conversacional, sin necesidad de conocimientos técnicos por parte del usuario. El acceso a todas las funcionalidades se realiza mediante un chatbot que interpreta las peticiones escritas en lenguaje natural.

7.1 Inicio de la aplicación

Al iniciar la aplicación, el usuario accede directamente a una pantalla principal que presenta el asistente SIBI (Seleccionador Inteligente Basado en IA). Desde este momento, el sistema está preparado para recibir consultas sin necesidad de configuración previa.

El historial de conversación se mantiene visible durante toda la sesión, permitiendo al usuario revisar consultas anteriores y entender la evolución de las recomendaciones.

7.2 Tipos de consultas soportadas

El sistema distingue automáticamente entre tres tipos principales de consultas:

a) Generación de convocatorias

El usuario puede solicitar la generación de una convocatoria indicando parámetros como:

- nacionalidad de los jugadores
- número total de convocados
- estilo de juego (ofensivo, defensivo o balanceado),
- exclusión de jugadores lesionados

Ejemplos de consultas válidas:

“Hazme la mejor convocatoria de 23 jugadores españoles con estilo ofensivo”

“Dame una convocatoria defensiva de 20 jugadores argentinos sin lesionados”

El sistema interpreta la consulta, extrae automáticamente los parámetros relevantes y genera una convocatoria óptima basada en rendimiento y química.

b) Ajustes sobre una convocatoria existente

Una vez generada una convocatoria, el usuario puede realizar ajustes incrementales utilizando expresiones como:

“Hazla más ofensiva”

“Ajusta la convocatoria”

Gracias al uso de memoria de contexto, el sistema aplica los cambios solicitados sobre la última convocatoria generada, manteniendo coherencia entre las interacciones.

c) Consultas informativas libres

El usuario también puede realizar consultas generales sobre los datos almacenados, como:

- nombres de jugadores,
- equipos,
- ligas,
- nacionalidades.

En este caso, el sistema utiliza inteligencia artificial para generar consultas seguras contra la base de datos y devuelve únicamente información permitida, sin exponer datos sensibles ni estructuras internas.

7.3 Respuesta del sistema

Las respuestas del sistema se presentan de forma estructurada y legible, incluyendo:

- resumen de la convocatoria,
- score total del equipo,
- rendimiento medio,
- química media,
- y listado de jugadores agrupados por posición.

Adicionalmente, el sistema puede mostrar un análisis textual de la convocatoria, facilitando la interpretación de los resultados por parte del usuario.

8. Casos de uso y escenarios de prueba

En este apartado se presentan tres casos de uso reales ejecutados sobre la aplicación, con el objetivo de evaluar su comportamiento en distintos escenarios: uno verosímil, uno optimista (alta dificultad de optimización) y uno pesimista (escenario restrictivo). Todos los casos han sido realizados mediante interacción directa con el chatbot, utilizando lenguaje natural y aprovechando la memoria de contexto del sistema.

8.1 Caso de uso 1 – Escenario verosímil

Generación de una convocatoria estándar

Consulta realizada

“Hazme la mejor convocatoria de 23 jugadores españoles”

Descripción del escenario

Este caso representa la situación más habitual para un seleccionador nacional: la creación de una convocatoria estándar sin restricciones adicionales. El sistema trabaja con un conjunto amplio de jugadores y aplica el estilo balanceado por defecto.

Resultados obtenidos

La aplicación genera una convocatoria de 23 jugadores con las siguientes métricas globales:

- Score total: 0.2855
- Rendimiento medio: 0.2828
- Química media: 0.2964

La convocatoria presenta una distribución equilibrada de posiciones y una representación diversa de clubes y ligas, destacando especialmente la presencia de varios jugadores del mismo club, lo que incrementa la química del equipo. Este

comportamiento es coherente con el modelo implementado, ya que favorece sinergias reales entre futbolistas.

Análisis

El sistema demuestra su capacidad para generar una convocatoria realista, equilibrando rendimiento individual y cohesión grupal. Este caso sirve como referencia base para los escenarios posteriores.

8.2 Caso de uso 2 – Escenario optimista

Ajuste táctico mediante memoria de contexto

Consulta realizada

“Cámbiala a estilo ofensivo”

Descripción del escenario

En este caso, el usuario solicita un ajuste táctico sobre la convocatoria previamente generada, sin redefinir parámetros como nacionalidad o número de jugadores. Este escenario se considera optimista porque el sistema dispone de una solución inicial válida y puede refinarla mediante un reajuste de pesos, enfrentándose a un problema de optimización más complejo.

Resultados obtenidos

El sistema reutiliza la convocatoria anterior y recalcula los scores aplicando pesos ofensivos:

- Score total: 0.2787
- Rendimiento medio: 0.2743
- Química media: 0.2964

Se observa una ligera reducción del rendimiento medio, mientras que la química se mantiene constante, ya que las relaciones entre jugadores no varían. Los jugadores

ofensivos incrementan su relevancia dentro del equipo, reflejando correctamente el cambio de estilo solicitado.

Análisis

Este caso demuestra de forma clara el uso de memoria de contexto, permitiendo modificar una convocatoria existente mediante instrucciones naturales. El sistema actúa como un asistente inteligente real, ajustando la solución sin necesidad de repetir la consulta inicial, lo que mejora notablemente la experiencia de usuario.

8.3 Caso de uso 3 – Escenario pesimista

Convocatoria con restricciones fuertes

Consulta realizada

“Ahora haz una convocatoria defensiva de 18 jugadores italianos sin lesionados”

Descripción del escenario

Este escenario impone restricciones severas: reducción del número de jugadores, exclusión de lesionados y priorización de un estilo defensivo. Estas condiciones reducen significativamente el espacio de búsqueda y dificultan la obtención de una solución óptima.

Resultados obtenidos

La aplicación genera una convocatoria válida con las siguientes métricas:

- Score total: 0.2787
- Rendimiento medio: 0.2315
- Química media: 0.4673

Aunque el rendimiento medio disminuye debido a las restricciones, se observa un aumento notable de la química, consecuencia de la concentración de jugadores en un número reducido de clubes y ligas.

Análisis

Este caso pone de manifiesto la robustez del sistema, que es capaz de generar soluciones coherentes incluso en condiciones desfavorables. El aumento de la química compensa parcialmente la pérdida de rendimiento individual, reflejando un comportamiento lógico y alineado con el modelo implementado.

Conclusión de los casos de uso

Los tres escenarios permiten comprobar que la aplicación:

- genera convocatorias realistas en condiciones normales
- se adapta dinámicamente mediante memoria de contexto
- mantiene estabilidad ante restricciones severas.

Estos resultados validan el enfoque adoptado y demuestran que el sistema se comporta de manera coherente y consistente en distintos niveles de complejidad.

9. Seguridad y ciberseguridad

La seguridad y la protección de la información son aspectos relevantes en cualquier sistema que integre bases de datos, inteligencia artificial y acceso mediante interfaces web. Aunque el presente proyecto se ha desarrollado como un prototipo académico, se han tenido en cuenta diversas consideraciones de seguridad durante su diseño e implementación.

9.1 Seguridad de los datos

Uno de los primeros aspectos considerados fue el uso de datos sintéticos en lugar de información real. Esta decisión elimina problemas relacionados con la privacidad, la protección de datos personales y el cumplimiento normativo, permitiendo trabajar en un entorno seguro y controlado.

El acceso a la base de datos Neo4j se realiza exclusivamente desde el backend de la aplicación, evitando exposiciones directas al usuario final. De este modo, todas las consultas están mediadas por la lógica del sistema, reduciendo el riesgo de accesos no autorizados.

9.2 Seguridad en el uso de modelos de lenguaje

El sistema integra un modelo de lenguaje para interpretar consultas en lenguaje natural y generar consultas estructuradas contra la base de datos. Este componente introduce posibles riesgos de seguridad, por lo que se han aplicado medidas de control como:

- validación previa de las consultas generadas,
- restricción explícita de operaciones peligrosas (creación, modificación o eliminación de datos),
- y limitación del tipo de consultas permitidas.

Estas medidas evitan que el modelo de lenguaje pueda comprometer la integridad de la base de datos o acceder a información no deseada.

9.3 Consideración del uso de API Key

Durante el diseño del sistema se valoró la incorporación de un mecanismo de autenticación mediante API Key para controlar el acceso a la aplicación. Este mecanismo habría permitido:

- restringir el acceso únicamente a usuarios autorizados,
- identificar el origen de cada petición,
- aplicar límites de uso y prevenir abusos,
- y mejorar la trazabilidad de las consultas realizadas.

Finalmente, esta opción fue descartada debido al carácter académico y demostrativo del proyecto, priorizando la facilidad de acceso y uso. No obstante, la arquitectura del sistema permite integrar este mecanismo de forma sencilla en un entorno de producción.

9.4 Seguridad en un posible despliegue en servidor

En un escenario de despliegue real, sería necesario aplicar medidas adicionales de seguridad, entre las que se incluyen:

- uso de variables de entorno para credenciales y claves sensibles,
- aislamiento de servicios mediante contenedores,
- configuración de firewalls y restricciones de red,
- autenticación y autorización de usuarios,
- y monitorización de accesos, errores y actividad sospechosa.

Estas medidas permitirían proteger tanto la base de datos como los servicios de inteligencia artificial frente a accesos no autorizados.

9.5 Alcance de la seguridad del sistema

Dado el contexto académico del proyecto, no se puede afirmar que el sistema sea completamente seguro frente a ataques avanzados. Sin embargo, se han aplicado buenas prácticas básicas de seguridad, adecuadas al alcance del proyecto y coherentes con su finalidad.

El sistema se ha diseñado con una clara separación de responsabilidades y con controles suficientes para minimizar riesgos evidentes, sentando una base sólida para futuras mejoras en materia de ciberseguridad.

10. Análisis crítico del sistema

Este apartado presenta un análisis crítico del proyecto desarrollado, evaluando tanto sus fortalezas como sus limitaciones. El objetivo no es únicamente destacar los logros alcanzados, sino también identificar debilidades, riesgos y aspectos susceptibles de mejora, adoptando una visión realista y reflexiva del sistema.

10.1 Análisis DAFO

- **Fortalezas**

- Uso de una base de datos orientada a grafos, especialmente adecuada para modelar relaciones complejas entre jugadores, equipos y ligas.
- Integración efectiva de rendimiento individual y química colectiva, proporcionando recomendaciones más realistas que sistemas basados únicamente en estadísticas.
- Sistema flexible y adaptable, capaz de ajustar convocatorias según el estilo de juego solicitado.

- Incorporación de memoria de contexto, que permite una interacción conversacional natural y progresiva.
- Arquitectura modular e integrada en una única aplicación, lo que facilita el mantenimiento y la extensibilidad.
- Uso de datos sintéticos que garantiza un entorno de pruebas seguro y controlado.

- **Debilidades**

- Dependencia de pesos definidos manualmente, lo que introduce cierto grado de subjetividad en la valoración de estadísticas.
- Modelo de química simplificado, basado únicamente en coincidencias de club o liga.
- Uso de datos ficticios, que limita la validez directa de los resultados en un entorno profesional real.
- Optimización basada en heurísticas locales, sin garantía de alcanzar el óptimo global.
- Escalabilidad limitada si el número de jugadores o combinaciones crece de forma significativa.

- **Oportunidades**

- Aplicación directa en ámbitos como análisis deportivo, scouting o simulación táctica.
- Posibilidad de integrar datos reales y actualizados procedentes de fuentes oficiales.
- Evolución hacia algoritmos más avanzados (optimización multiobjetivo, aprendizaje automático).

- Mejora del sistema explicativo para ofrecer recomendaciones más transparentes y justificadas.
- Despliegue como servicio web multiusuario o herramienta de apoyo a cuerpos técnicos.

- **Amenazas**

- Riesgos asociados al uso de modelos de lenguaje, como interpretaciones incorrectas de las consultas del usuario.
- Posibles problemas de rendimiento en escenarios con grandes volúmenes de datos.
- Dependencia de servicios externos para la inferencia del modelo de lenguaje.
- Necesidad de cumplir normativas de protección de datos en caso de utilizar información real.

10.2 Limitaciones del proyecto

El proyecto presenta varias limitaciones que deben ser tenidas en cuenta. En primer lugar, el uso de datos sintéticos, aunque adecuado para un contexto académico y de seguridad, impide validar el sistema con resultados reales de competición. Además, la química entre jugadores se modela de forma simplificada, sin considerar factores tácticos complejos o históricos de juego conjunto más detallados.

Desde el punto de vista algorítmico, la estrategia de mejora iterativa utilizada prioriza la eficiencia computacional frente a la exhaustividad, lo que puede dar lugar a soluciones localmente óptimas. Asimismo, la asignación manual de pesos requiere experiencia previa y puede no reflejar todos los contextos posibles.

10.3 Valoración crítica global

A pesar de las limitaciones identificadas, el sistema cumple de forma sólida los objetivos planteados. El proyecto demuestra que es posible integrar bases de datos de grafos, algoritmos de recomendación e inteligencia artificial conversacional en una solución coherente y funcional.

El enfoque adoptado es especialmente adecuado para un entorno académico, donde se prioriza la comprensión del problema, la correcta elección de herramientas y la capacidad de análisis crítico. El sistema no pretende sustituir la toma de decisiones humana, sino actuar como una herramienta de apoyo, ofreciendo recomendaciones objetivas que pueden complementar el criterio de un seleccionador real.

En conclusión, el proyecto presenta un equilibrio adecuado entre ambición técnica y realismo, sentando una base sólida para futuras mejoras y demostrando una correcta aplicación de conceptos avanzados de ingeniería informática.

11. Lecciones aprendidas y líneas de futuro

A lo largo del desarrollo de este proyecto se han extraído diversas lecciones tanto a nivel técnico como conceptual, que han contribuido a consolidar los conocimientos adquiridos y a identificar posibles mejoras y ampliaciones del sistema.

11.1 Lecciones aprendidas

Una de las principales lecciones aprendidas es la importancia de modelar adecuadamente los datos en función del problema a resolver. El uso de una base de datos orientada a grafos ha demostrado ser especialmente eficaz para representar relaciones complejas entre entidades, permitiendo explotar conceptos como la química entre jugadores de una forma natural y eficiente.

Asimismo, se ha comprobado que la combinación de rendimiento individual y cohesión grupal es clave para obtener recomendaciones realistas. El proyecto pone de manifiesto que maximizar únicamente estadísticas individuales no conduce necesariamente a la mejor solución colectiva, y que introducir factores relacionales mejora significativamente la calidad de los resultados.

Desde el punto de vista algorítmico, el desarrollo del sistema ha permitido comprender las ventajas de los enfoques heurísticos e iterativos frente a soluciones exhaustivas. Este tipo de estrategias ofrece un equilibrio adecuado entre calidad de la solución y coste computacional, especialmente en problemas con grandes espacios de búsqueda.

Otra lección relevante ha sido la integración de inteligencia artificial conversacional con sistemas clásicos de información. El uso de memoria de contexto ha demostrado ser fundamental para ofrecer una interacción natural, facilitando ajustes progresivos y reduciendo la carga cognitiva del usuario.

Finalmente, el proyecto ha puesto de relieve la necesidad de prestar atención a aspectos como la robustez del sistema, la validación de entradas y el control de errores, especialmente cuando se combinan modelos de lenguaje con bases de datos reales.

11.2 Líneas de futuro

El sistema desarrollado constituye una base sólida sobre la que se pueden implementar numerosas mejoras y ampliaciones. Entre las principales líneas de futuro destacan:

- Enriquecimiento del modelo de química, incorporando más factores como historial de partidos compartidos, compatibilidad de estilos o relaciones temporales.
- Optimización del algoritmo de selección, explorando técnicas más avanzadas como algoritmos genéticos o métodos de optimización multiobjetivo.
- Uso de datos reales, integrando fuentes oficiales y actualizadas que permitan aplicar el sistema en entornos profesionales, siempre teniendo en cuenta aspectos legales y de privacidad.
- Mejora de la explicación de resultados, proporcionando justificaciones más detalladas de por qué se selecciona a cada jugador.
- Ampliación del sistema conversacional, permitiendo consultas más complejas y mayor control sobre los parámetros de la convocatoria.
- Despliegue en entorno productivo, adaptando la arquitectura para su uso en servidores y escenarios multiusuario.
- En conjunto, estas líneas de futuro permitirían evolucionar el proyecto desde un prototipo académico hacia una herramienta más completa y cercana a una aplicación real en el ámbito del análisis deportivo.

12. Enlaces y bibliografía

En este apartado se recogen los principales recursos utilizados como referencia durante el desarrollo del proyecto, así como materiales recomendados para profundizar en las tecnologías empleadas.

- Neo4j.
Neo4j Documentation.
Documentación oficial de Neo4j, utilizada como referencia para el modelado de grafos, definición de nodos, relaciones y consultas en Cypher.
<https://neo4j.com/docs/>
- Robinson, I., Webber, J., & Eifrem, E.
Graph Databases. O'Reilly Media.
Libro de referencia sobre bases de datos de grafos, sus ventajas frente a modelos relacionales y casos de uso reales.
- Ollama.
Ollama Documentation.
Guía oficial para la ejecución local de modelos de lenguaje y su integración en aplicaciones.
<https://ollama.com/docs>
- Groq.
Groq API Documentation.
Documentación utilizada para la integración de inferencia de modelos de lenguaje de gran tamaño.
<https://console.groq.com/docs>
- Russell, S., & Norvig, P.
Artificial Intelligence: A Modern Approach. Pearson.
Obra clásica de referencia sobre inteligencia artificial, razonamiento, búsqueda y toma de decisiones.
- Goodfellow, I., Bengio, Y., & Courville, A.
Deep Learning. MIT Press.
Libro de referencia sobre aprendizaje profundo y fundamentos teóricos de los modelos actuales.
- Python.
Python Official Documentation.
<https://docs.python.org/3/>
- Streamlit.
Streamlit Documentation.
Documentación oficial para el desarrollo de interfaces web interactivas en Python.
<https://docs.streamlit.io/>
- Neo4j Python Driver.
Neo4j Python Driver Documentation.
<https://neo4j.com/docs/python-manual/current/>
- OWASP Foundation.
OWASP Top 10.

Referencia sobre las principales vulnerabilidades en aplicaciones web y buenas prácticas de seguridad.

<https://owasp.org/www-project-top-ten/>

- Stallings, W.
Network Security Essentials. Pearson.
Libro de referencia sobre fundamentos de seguridad informática y protección de sistemas.