

Análisis del contexto y la normatividad

Normativa asociada al tipo de datos utilizados

El dataset utilizado en este proyecto proviene de Kaggle, una plataforma que ofrece datasets públicos para análisis y desarrollo de modelos. Este conjunto de datos está relacionado con partidos de fútbol de la Premier League, y aunque no contiene información personal identificable (como nombres, direcciones o datos financieros), es importante cumplir con las normativas relacionadas con la recolección y uso de datos deportivos y analíticos.

En Europa, por dar un ejemplo entre los tantos que pueden haber, el Reglamento General de Protección de Datos (GDPR) establece regulaciones estrictas sobre el uso de datos personales. Aunque los datos deportivos utilizados no incluyen información personal, es relevante asegurar que, de incluirse datos sensibles, estos sean manejados de acuerdo con las normativas (Álvarez, 2018).

Uso de los datos y cumplimiento normativo

En este proyecto, los datos fueron obtenidos a través de Kaggle, que asegura la legalidad de los datasets subidos a su plataforma bajo los términos y condiciones establecidos por ellos mismos. Kaggle menciona que los datasets subidos no deben violar derechos de terceros ni la normativa vigente de protección de datos.

Al utilizar estos datos, el proyecto respeta las políticas de Kaggle y no infringe las normativas de protección de datos, ya que no se procesan datos personales de usuarios o participantes (Kaggle, 2024).

Documentos respaldatorios:

- Términos de Uso: Kaggle
- Reglamento General de Protección de Datos (GDPR)

Cumplimiento normativo de la herramienta y manejo ético

En cuanto a la herramienta, XGBoost es un algoritmo ampliamente utilizado en la industria para resolver problemas de clasificación y regresión. No genera sesgos inherentes, pero como cualquier algoritmo, su desempeño y ética dependen del diseño del proyecto y el manejo de los datos.

En este caso, el dataset utilizado no contiene información que pueda generar sesgos éticos hacia grupos vulnerables. El enfoque en la clasificación de resultados deportivos reduce la posibilidad de afectar a personas o comunidades. Sin embargo, se tiene en cuenta que:

- Las decisiones del modelo no impactan directamente en las personas.
- Los datos se procesan de forma anónima y no identifican a individuos.
- No se generan predicciones que puedan influir en eventos futuros de manera desproporcionada o inadecuada.

Escenarios de mal uso ético

A pesar de que este modelo fue desarrollado con un enfoque ético, su mal uso o la negligencia pueden conllevar faltas éticas. Ya se mencionó que en general los datos que se contienen son obtenidos de manera legal y, por la naturaleza de los mismos, no deberían de ser una gran implicación éticamente hablando. Aún así, si podría imaginar algunos escenarios que pudieran significar faltas a la ética. Por ejemplo:

- Si el modelo es alterado para manipular apuestas deportivas o influir en el rendimiento de equipos de manera fraudulenta, podría resultar en implicaciones éticas graves.
- El uso del modelo sin informar de sus limitaciones o con fines opacos podría generar decisiones injustas, como influenciar mercados sin considerar el contexto deportivo completo.
- La falta de validación continua del modelo podría generar resultados sesgados si las características del deporte cambian, por ejemplo, si un equipo evoluciona en su estrategia o el dataset se vuelve obsoleto.

Referencias

Álvares, C. (2018). *GDPR: La nueva ley europea de protección de datos*. BBVA. Obtenido el 9 de Septiembre de 2024, de

<https://www.bbva.com/es/gdpr-nueva-ley-europea-proteccion-datos/>

Kaggle. (2014). *Terms of service*. Kaggle. Obtenido el 9 de Septiembre de 2024, de

<https://www.kaggle.com/terms>