

Título del Proyecto: Análisis de aperturas de ajedrez para mejorar la tasa de victoria:

Estudiante: Sebastián Colorado Barrientos

e-mail: s.colorado@udea.edu.co

GitHub: <https://github.com/TensaiSup>

1. Descripción del Problema

El problema de predicción que se abordará en este proyecto es el de predecir la tasa de victoria de un jugador en función de la apertura que elige al inicio del juego. La elección de la apertura es una de las decisiones más importantes que un jugador de ajedrez debe tomar al comienzo de la partida, ya que puede influir significativamente en el resultado final del juego. Por lo tanto, es importante para los jugadores de ajedrez y los entrenadores conocer qué aperturas tienen una mayor tasa de victoria para poder incluirlas en su estrategia de juego. El problema de negocio que se aborda es el de ofrecer recomendaciones de apertura a los usuarios de una plataforma de ajedrez en línea, con el objetivo de mejorar su tasa de victoria y, por lo tanto, su experiencia de juego en la plataforma.

2. Descripción del dataset

El dataset utilizado en este proyecto proviene de la plataforma de ajedrez en línea Lichess.org y se recopiló utilizando la API de Lichess. La URL para acceder a la página de descarga del dataset es la siguiente: <https://www.kaggle.com/datasnaek/chess>.

El dataset consta de un solo archivo CSV que contiene más de 20,000 partidas de ajedrez. Cada fila representa una partida de ajedrez y contiene información como el ID de la partida, si fue clasificada o no, la hora de inicio y finalización de la partida, el número de turnos, el estado de la partida, el ganador, el incremento de tiempo, la ID y la clasificación de los jugadores blancos y negros, todos los movimientos en la notación estándar de ajedrez, la apertura (identificada por su código ECO y su nombre) y el número de jugadas en la apertura.

Además, hay columnas para el rating de los jugadores, la fecha de la partida y la velocidad del juego. El archivo CSV tiene un tamaño de aproximadamente 6,5 MB y consta de 15 columnas y más de 20,000 filas. Este dataset proporciona una gran cantidad de información valiosa para analizar las aperturas de ajedrez y su relación con la tasa de victoria de los jugadores.

3. Métricas de desempeño

Para el problema de predicción de la tasa de victoria en función de la apertura de ajedrez elegida por un jugador, podemos considerar tanto las métricas de desempeño del aprendizaje automático como las métricas de negocio.

Métricas de desempeño de aprendizaje automático:

- **Precisión (Accuracy):** la proporción de predicciones correctas realizadas por el modelo en comparación con el número total de predicciones realizadas. En el contexto de este proyecto, una alta precisión indicaría que el modelo puede predecir con precisión la tasa de victoria de un jugador en función de la apertura de ajedrez elegida.

- Sensibilidad (Recall): la proporción de instancias positivas (en este caso, las partidas que realmente ganó el jugador) que se identifican correctamente como positivas. En este contexto, una alta sensibilidad indicaría que el modelo puede identificar correctamente cuando un jugador ganó una partida en función de la apertura de ajedrez que eligió.

Métricas de negocio:

- Tasa de conversión: la proporción de jugadores que, después de recibir recomendaciones basadas en el modelo, eligen utilizar las aperturas recomendadas y logran una tasa de victoria más alta.
- ROI (Retorno de la inversión): un indicador de la rentabilidad del proyecto. Este indicador puede calcularse restando los costos del proyecto de los ingresos obtenidos por el aumento de la tasa de victoria de los jugadores que adoptan las recomendaciones basadas en el modelo.

4. Criterios de desempeño

Criterios de desempeño para el aprendizaje automático:

- Precisión (Accuracy): el modelo debe tener una alta precisión en sus predicciones para que sea útil en la toma de decisiones.
- Estabilidad: el modelo debe ser estable y consistente en sus predicciones, incluso cuando se le presentan nuevos datos de entrada.
- Velocidad de entrenamiento: el modelo debe entrenarse en un tiempo razonable, para que sea práctico su uso en un entorno empresarial.

Criterios de desempeño para el problema de negocio:

- Incremento en la tasa de victoria: el objetivo principal de este proyecto es incrementar la tasa de victoria de los jugadores que utilizan las recomendaciones basadas en el modelo. Por lo tanto, este es un criterio clave para evaluar el impacto del proyecto.
- Costo del proyecto: el proyecto debe tener un costo razonable en comparación con los beneficios que proporciona.
- Fácil implementación: las recomendaciones basadas en el modelo deben ser fáciles de implementar en el entorno de los jugadores y del negocio, para que se puedan adoptar y aprovechar los beneficios.