

Que no se coman a tu Rey

Integrantes

- García Alanís A.U.
- Estrada Dávila J.E.
- Paz Reyes N.M.
- González García E.
- Sobrevilla Ruiz R

Introducción

El ajedrez es un juego con muchos años de antigüedad, sencillo de entender pero con una dificultad estratégica muy interesante. Se cree que los grandes reyes y estrategas militares veían al ajedrez como una forma de estudiar a sus enemigos y mejorar como líderes.

El ajedrez se ha popularizado últimamente gracias a series y películas que engrandecen este deporte, por lo que ha aumentado la cantidad de nuevos jugadores, los cuales no saben que estrategias seguir, ademas que los mas veteranos del juego quieren mejorar sus estrategias día con día, por eso, el estudio de las jugadas y ver tu impacto en el resultado del juego ayudaría a estas personas a mejorar y entender más el ajedrez.

Objetivos

OBJETIVO PRINCIPAL

Clasificar las mejores aperturas del ajedrez, dependiendo del "elo" que tengan los jugadores y la eficacia de las mismas para la creación de estrategias ganadoras.

OBJETIVOS SECUNDARIOS

- Determinar si existe algún tipo de relación de los distintos errores entre el jugador de negras y el de blancas.
- Categorizar a los jugadores dependiendo de los errores que estos cometan.

Apertura	Ganadas	Frecuencia	Puntaje
4 Queen's Pawn Game: Mason Attack	0.536612	0.018130	0.009729
2 Philidor Defense	0.587308	0.016230	0.009532
69 Scandinavian Defense: Mieses-Kotrov Variation	0.575330	0.013052	0.007509
57 Sicilian Defense	0.413096	0.013966	0.005769
117 Van't Kruisj Opening	0.407439	0.013629	0.005553

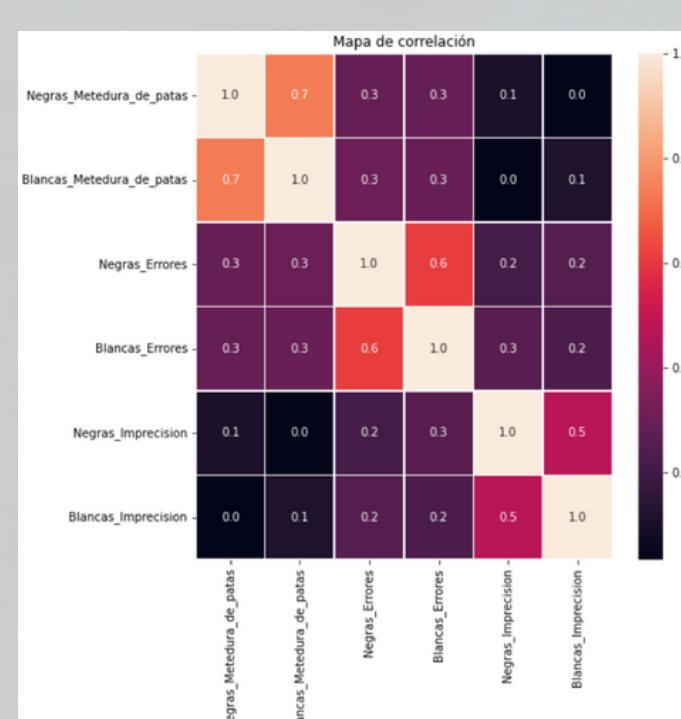
Tabla 1. Mejores aperturas Low rating

Apertura	Ganadas	Frecuencia	Puntaje
118 Indian Game	0.488273	0.003454	0.001687
5 Pirc Defense	0.525241	0.002331	0.001224
128 Queen's Pawn Game: Mason Attack	0.546816	0.002179	0.001191
35 Caro-Kann Defense	0.485491	0.002359	0.001145
150 Modern Defense	0.539073	0.002019	0.001088

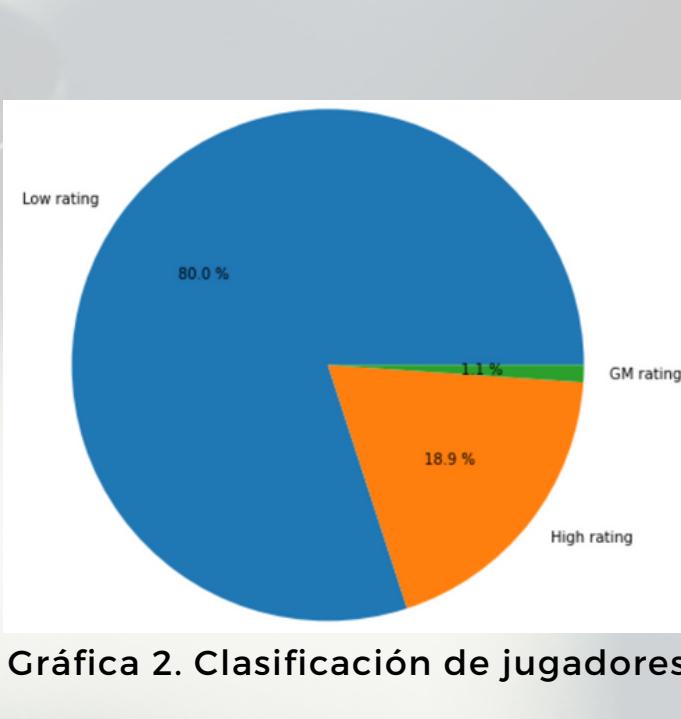
Tabla 2. Mejores aperturas High rating

Apertura	Ganadas	Frecuencia	Puntaje
8 Indian Game	0.558052	0.000143	0.000080
60 Modern Defense	0.573059	0.000117	0.000067
142 Caro-Kann Defense	0.593407	0.000097	0.000058
20 Sicilian Defense: Najdorf Variation	0.604294	0.000087	0.000053
260 Queen's Pawn	0.595679	0.000087	0.000052

Tabla 3. Mejores aperturas GM rating



Gráfica 1. Correlación de errores



Gráfica 2. Clasificación de jugadores



Árbol 1. 1er árbol de decisión



Árbol 2. 2do árbol de decisión

Recursos:



Base de datos



Notebook



jupyter
seaborn



matplotlib



Python



colab



scikit
learn



pandas

Metodología y resultados

En la base de datos hay más de 3 millones de filas, y 40 columnas. Cada fila representa un enfrentamiento entre 2 personas y las columnas nos revelan varios datos, como el tipo de juego, nombre de aperturas, número de movimientos, entre otros. Al haber tantos datos, se descartaron las siguientes columnas:

- GAME, Date, Site, BlackRatingDiff, WhiteRatingDiff, starting_time, increment, TimeControl, UTCTime, ECO, Black_inferior_moves, White_inferior_moves, Black_ts_moves, White_ts_moves, Black_ts_blunders, White_ts_blunders, Black_ts_mistakes, White_ts_mistake, Black_long_moves, White_long_moves, Black_bad_long_moves, White_bad_long_moves, Game_flips, Game_flips_ts, Event.

Estas columnas se descartaron debido a que su información no nos era de utilidad a la hora querer dar respuesta a nuestros objetivos.

Para resolver nuestro objetivo principal, dividimos la base de datos, ya modificada, por categoría de jugador (Low, High y GM rating) con respecto al jugador de las blancas, quien es nuestro objetivo principal. Una vez realizada la división, evaluamos cada apertura con la siguiente fórmula:

$$\text{Puntaje} = (\text{Eficacia}) * (\text{Frecuencia})$$

Donde la "Eficacia" representa el porcentaje de victoria de la apertura con respecto a los juegos que se jugaron con la misma, y la "Frecuencia" es la tasa de veces que se jugó esta apertura con respecto a todos los juegos a analizar, así castigamos las aperturas que se jugaron muy raramente y premiamos a las aperturas más populares con respecto a la eficacia que estas tengan. Con lo anterior, se creó una tabla con el puntaje de todas las aperturas por categoría de jugador, y ordenando estas mismas aperturas de mejor a peor puntuadas, sacamos el Top 5 de cada categoría, quedando de la siguiente manera:

Conclusiones

Con respecto a nuestro objetivo principal, vemos que realmente pudimos clasificar las mejores aperturas del ajedrez, dependiendo de la categoría de los jugadores, los cuales se pueden agrupar de la siguiente manera:

Top 5 mejores aperturas del ajedrez		
Low rating	High rating	GM rating
Queen's Pawn Game: Mason Attack	Indian Game	Indian Game
Philidor Defense	Pirc Defense	Modern Defense
Scandinavian Defense: Mieses-Kotrov Variation	Queen's Pawn Game: Mason Attack	Caro-Kann Defense
Sicilian Defense	Caro-Kann Defense	Sicilian Defense: Najdorf Variation
Van't Kruisj Opening	Modern Defense	Queen's Pawn

Tabla 4. Agrupación de mejores aperturas

Vemos que algunas aperturas no tienen mas del 50% de tasa de victoria, pero estas aparecen en esta lista debido a su gran frecuencia en su categoría, haciendo que escalen posiciones para que los jugadores no se sorprendan a la hora de que se enfrenten a estas aperturas o si las quieren utilizar. Además, se ve una tendencia de jugar mas defensas en "elo's" mas altos, mientras que los novatos tienen tendencias a jugar de forma agresiva, es decir, al ataque.

Ahora, con nuestro primer objetivo secundario y viendo la Gráfica 1, podemos ver que el error de "Metedura de pata" aparenta tener una relación muy fuerte con los jugadores. Se puede decir entonces que, la cantidad de "Meteduras de pata" de tu rival puede depender de las "Meteduras de pata" que hayas hecho, o que uno puede cometer ciertas "Meteduras de pata" dependiendo de las cometidas por el rival. Lo anterior igualmente aplica con las "imprecisiones" y los "errores", pero en menor medida.

Por último, al querer responder nuestro 2do objetivo secundario, nos encontramos con lo siguiente. Vemos que el Árbol 1 tiene una efectividad del 80% según nuestras métricas, 24 niveles de profundidad y la característica que más relevancia tiene son las meteduras de pata. El problema es que el tamaño del árbol es realmente grande, y puede que esté sesgado por la gran cantidad de jugadores en "Low rating" o por la poca gente en "GM rating", como se puede ver en la Gráfica 1, alargando más el árbol y haciéndolo menos específico. Por el contrario, al tratar de distribuir nuestra población de jugadores y ser más específicos, este Árbol 2 sufre, ya que nuestras métricas arrojan una efectividad del 22.3%. Por lo que, mientras más específico seamos, peor será nuestro árbol, pero si generalizamos, este se sesga por los jugadores, por lo que tus errores no definen al 100% la calidad de jugador que eres, es decir, hay más factores que definen tu nivel de juego, como las aperturas que juegas, tus movimientos totales, el tipo de juego, entre otros factores.

Trabajo a futuro

- Automatizar nuestros cálculos "arcaicos" que puntuán para aumentar la eficiencia de la obtención de las mejores aperturas, además de que puede personalizarse tanto para torneos como para jugadores individuales y utilizarse como análisis personal o como institución que organiza un torneo para el análisis de estos.
- Usar más variables para poder clasificar a los jugadores de mejor manera, es decir, sin sesgos y con efectividad alta. Se puede hacer uso de la técnica de "bosques aleatorios".
- Usar métodos predictivos que relacionen los errores (meteduras de pata, errores e imprecisiones) de un jugador para predecir los errores del otro.