EloMetrics: Advanced Outcome Prediction for Chess Matches with Elo Ratings and Logistic Regression
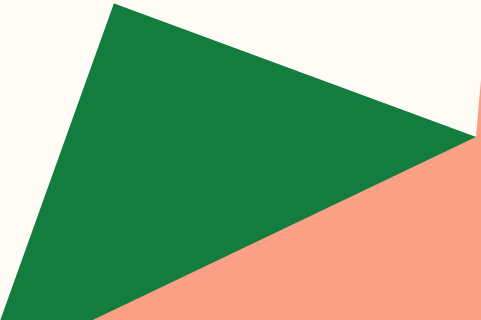
Adam Stein, CJ Jones
3/17/2025

#	↑↓	Name	🚩	Rating	+/−	🔍	12m	Delta		Age	
1		Carlsen, Magnus	🇳🇴	2837.1	+4.1	🔍		+7.1	📊	34	
2		Nakamura, Hikaru	🇺🇸	2796.8	−5.2	🔍	▲ 1	+7.8	📊	37	
3	👑	Gukesh D	🇮🇳	2787.0	0.0		▲ 12	+40.0	📊	18	
4		Caruana, Fabiano	🇺🇸	2780.4	−2.6	🔍	▼ 2	−23.6	📊	32	
5		Erigaisi Arjun	🇮🇳	2777.0	0.0		▲ 9	+29.0	📊	21	
6		Abdusattorov, Nodi	🇺🇿	2773.0	0.0		▲ 5	+23.0	📊	20	
7		Wei, Yi	🇨🇳	2758.2	−1.8	🔍	▲ 2	+3.2	📊	25	
8		Praggnanandhaa R	🇮🇳	2758.2	+0.2	🔍	▲ 8	+11.2	📊	19	
9		Firouzja, Alireza	🇫🇷	2757.0	0.0		▼ 3	−3.0	📊	21	
10		Nepomniachtchi, Ia	🏳️	2756.5	+3.5	🔍	▼ 3	−1.5	📊	34	
11		Mamedyarov, Shakl	🇦🇿	2748.0	0.0		▲ 8	+14.0	📊	39	
12		So, Wesley	🇺🇸	2747.8	−0.2	🔍	▼ 4	−9.2	📊	31	
13		Aronian, Levon	🇺🇸	2746.4	−1.6	🔍	▲ 13	+24.4	📊	42	
14	↑8	Aravindh, Chithamb	🇮🇳	2743.0	+12.0	🔍	▲ 58	+81.0	📊	25	
15	↓1	Anand, Viswanatha	🇮🇳	2743.0	0.0		▼ 5	−8.0	📊	55	
16	↓1	Dominguez Perez, l	🇺🇸	2741.2	+0.2	🔍	▼ 3	−7.8	📊	41	

Source: https://2700chess.com/

What is the Elo Rating?

Measures the relative skill levels in zero-sum games, like chess, that changes based on a player's wins and losses

Goal: How can we accurately predict players' chances of winning a game of chess?

Data Collection and Processing

Data Acquisition

- Chess Games Dataset (Lichess), was obtained from Kaggle and contains over 20,000 games from the website lichess.org
- Data has features such as start time, end time, elo rating of each player, first move, players' moves for the entire game in standard chess notation, and the game's winner

Data Cleaning

- Data was de-duped and excludes "draws"
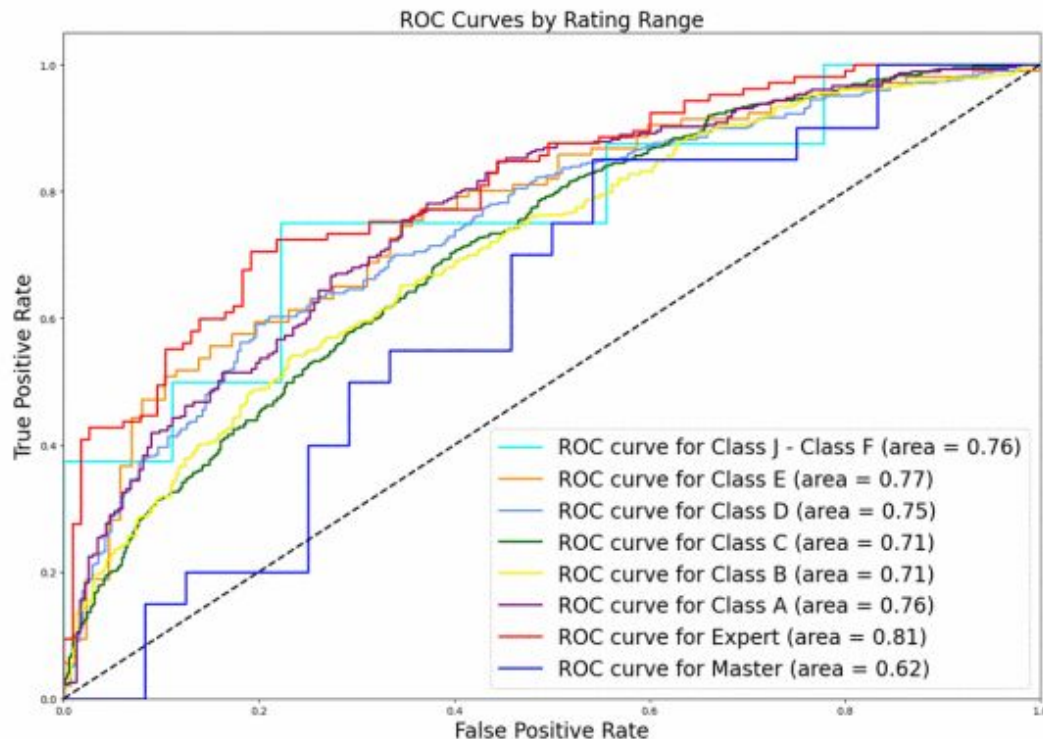- Games with incomplete and missing data were removed

Data Processing

- Created the variable rating_diff, which indicates the difference in ratings between the white and black players
- average rating for each game was computed, and games were categorized into Elo rating ranges through binning
- Games in the "2400+" elo rating range were removed

A binary logistic regression model was used to classify each game as win/loss with respect to white player

Model Results

Model has an overall accuracy of 68.18%

Rating Classification	Rating Range
Senior Master	2400+
Master	2200 to 2399
Expert	2000 to 2199
Class A	1800 to 1999
Class B	1600 to 1799
Class C	1400 to 1599
Class D	1200 to 1399
Class E	1000 to 1199
Class F	800 to 999
Class G	600 to 799
Class H	400 to 599
Class I	200 to 399
Class J	0 to 199

ROC Curves by Rating Range

- ROC curve for Class J - Class F (area = 0.76)
- ROC curve for Class E (area = 0.77)
- ROC curve for Class D (area = 0.75)
- ROC curve for Class C (area = 0.71)
- ROC curve for Class B (area = 0.71)
- ROC curve for Class A (area = 0.76)
- ROC curve for Expert (area = 0.81)
- ROC curve for Master (area = 0.62)

Conclusion

- Research developed a predictive model for chess game outcomes using logistic regression, focusing on the **total number of moves** and **Elo rating differences** as significant variables

- Predictive accuracy decreased for players in the **lower (0–1199) and higher (2000–2399) Elo ranges**, highlighting potential limitations in predictive value

- **Moderate Elo ranges produced accurate predictions**, but lower and higher ranges presented issues such as inconsistent results

- Future models should consider more variables such as **player openings, time limits, and psychological factors** may give a more comprehensive view of game results