

Executive Summary: Guess the Elo – Investigating Correlation Between Game Performance and Player Rating in Chess

Team: Dorian Soergel, Foivos Chnaras, Lang Song

GitHub: <https://github.com/foivoshn/Guess-The-Elo>

Background and Project Overview:

Our project, inspired by the popular chess YouTube series "Guess the Elo," explores the relationship between game performance and player ratings in chess. Elo is a widely used rating system that measures a player's skill based on their game results. With increasing allegations of cheating in chess, often justified using game performance metrics, our work seeks to investigate the validity of such claims and explore the potential for developing predictive and anti-cheating tools.

Our Goal:

- Analyze the correlation between chess performance metrics and Elo ratings.
- Develop insights into whether single-game performance metrics can indicate cheating.
- Predict Elo ratings using performance metrics aggregated across multiple games.

Stakeholders:

- Chess platforms and organizations for anti-cheating insights.
- Developers and researchers interested in Elo prediction algorithms.
- The general chess community and enthusiasts for engagement and learning.

Methodology:

- **Data Collection:** Analyzed 1 million games from *The Week in Chess*, with a focus on games around 2200 Elo.
- **Game Processing:** Used Stockfish, a state-of-the-art chess engine, to compute centipawn loss and evaluate positions. Approximately 30k games were processed at depth 20 for higher accuracy.
- **Metrics Development:**
 - Replicated Lichess.org's Accuracy score and analyzed its correlation with Elo.
 - Developed custom metrics, including "Winning Chance Loss," which quantifies the severity of mistakes by evaluating changes in winning probability between moves.
 - Classified errors into bins of 5% for nuanced analysis.

Results:

- **Single Game Analysis:** Showed weak correlation between performance metrics and Elo, questioning the reliability of using single-game metrics for cheating accusations.

- **Multiple Game Analysis:** Aggregated performance metrics across 10 games per player. Regression analysis revealed significantly stronger correlations, indicating that long-term patterns are better predictors of Elo.
- **Sample Size Concerns:** Despite starting with a large database, grouping games by player reduced the effective sample size to 20k, which limits the robustness of findings.

Future Work:

- Expand the database and analyze games at greater depths.
- Incorporate additional features, such as opening and endgame mistakes.
- Leverage deep learning to explore position complexity and refine predictive models.
- Develop a web application for Elo prediction, accessible to the general public.

Conclusion:

While single-game metrics lack reliability in predicting Elo or detecting cheating, aggregated data across multiple games holds significant potential. With further research and advanced computational techniques, this work could contribute to fair play in chess and enhance the experience of chess enthusiasts worldwide.