Predicting Chess Endings

Devin Belden
2/3/2020

How can we predict the winner of a chess game?

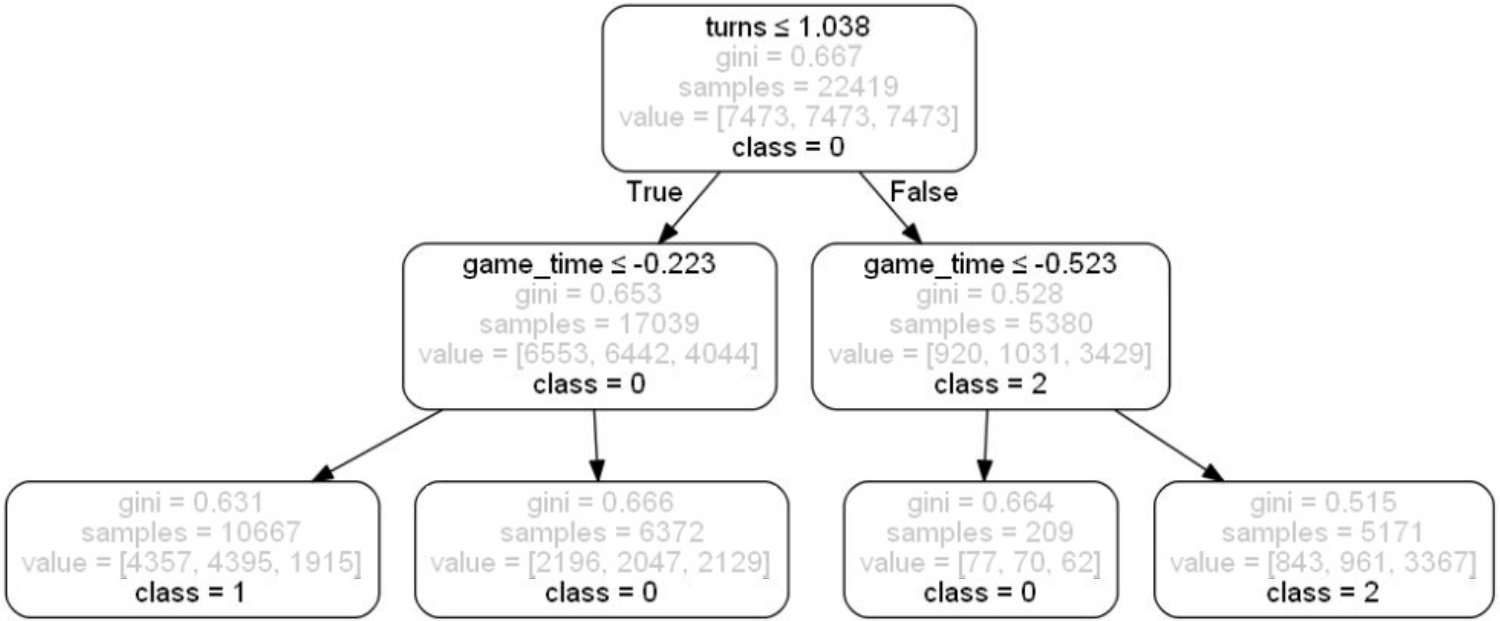
Overview

- Analyzed 20,000 games from LiChess.org, from low FIDE-ranked (~800) to highly ranked (2700+)
- Used classification modeling techniques to predict whether games were won by White, Black, or neither
- Used the following features:
  - Game opening
  - Turn count
  - Game time
  - Player ratings
  - Ranked match vs. Unranked match

Methodology

- Created a Decision Tree algorithm that determined classification thresholds for each factor
- Used Gradient Boosting to increase performance of each successive tree in the “forest”

Decision Tree Example



Predicted winner: Black

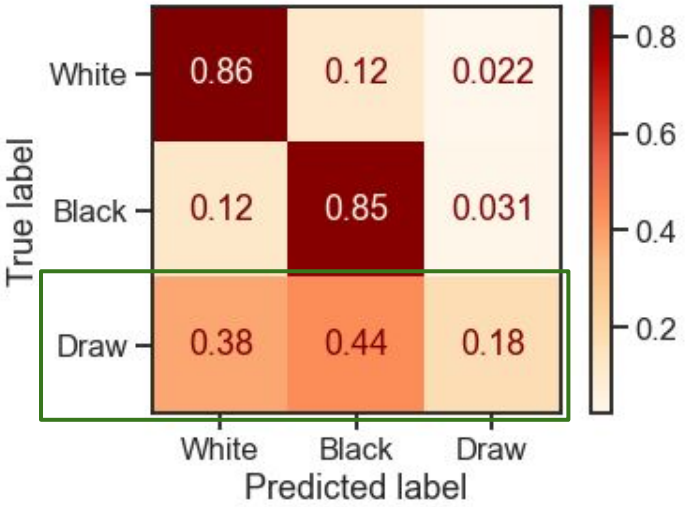
Predicted winner: White

Predicted winner: White

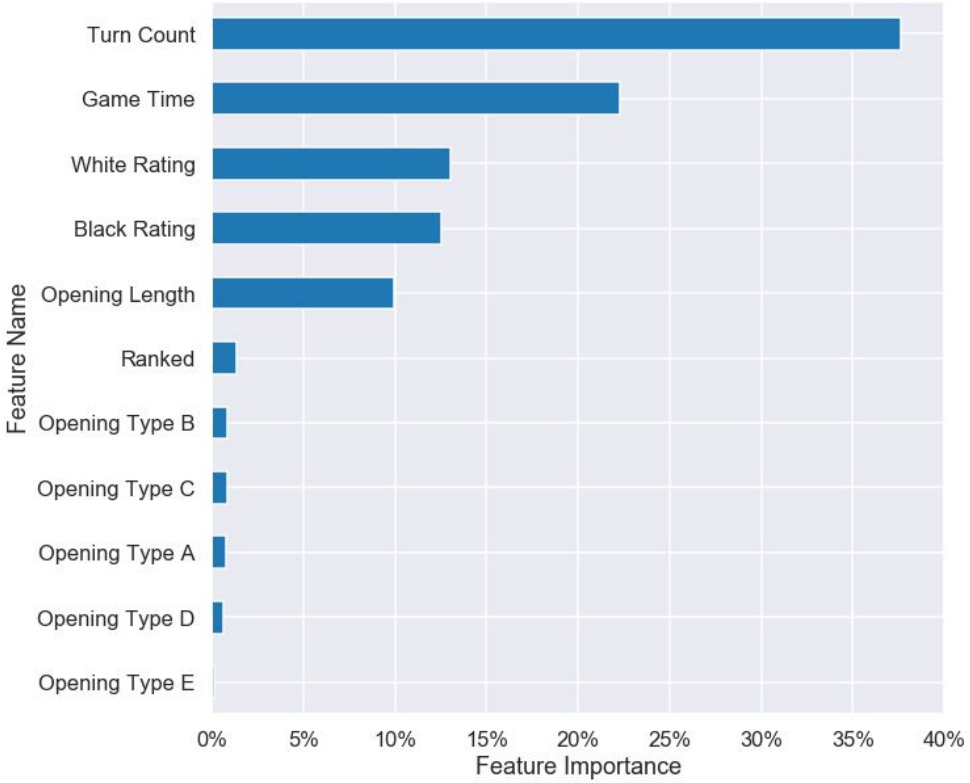
Predicted winner: Draw

Overall Model Performance

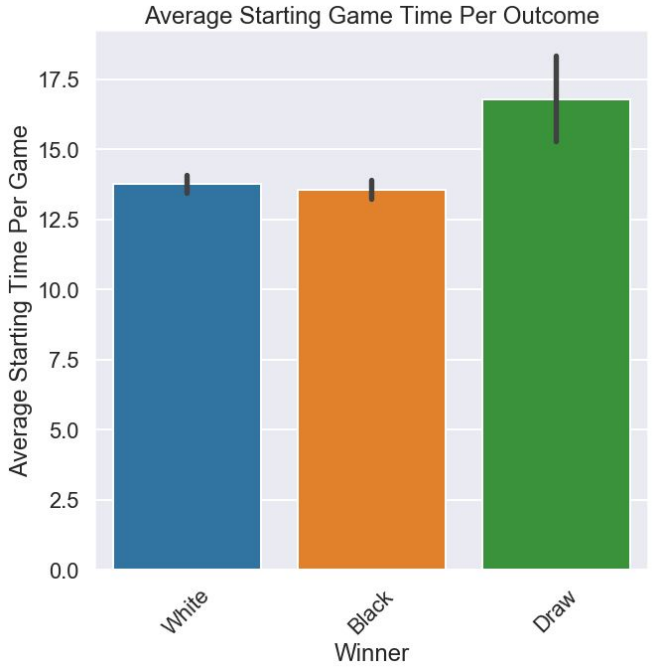
- Model is 81% accurate overall
  - Random chance is 33% accurate
- Model accurately predicts:
  - White: 86%
  - Black: 85%
  - Draw: 18%
- Considerable difficulty predicting draws compared to predicting winner colors

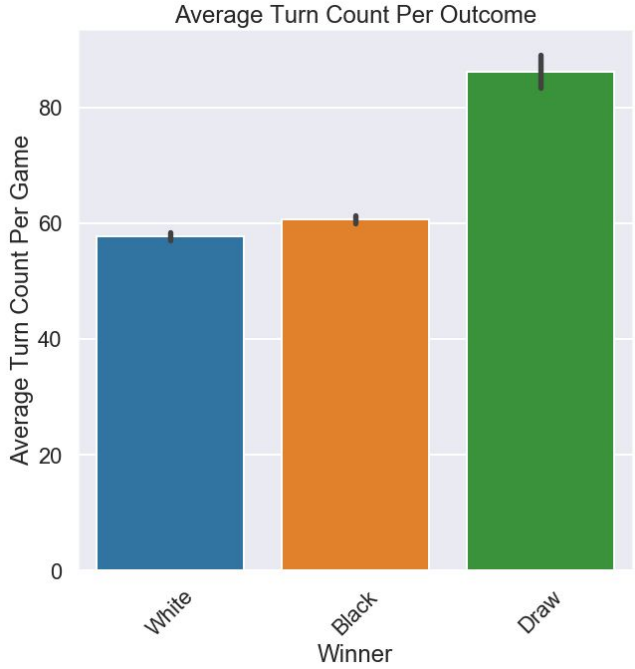


Results



Feature Importance: The most important factors in determining how the game ended were the turn count and the starting game time





A lower starting game time might prevent draws, but higher turn counts slightly favor Black

Conclusions

- Prediction is 81% accurate, compared to 33% accuracy for random guessing

- Turn count and game time accounted for ~60% of total prediction weight
  - High turn count favors Black over White, high game time encourages Draw
  - For example, when playing as Black, aiming to extend the game's turn count will give a statistical advantage

Further Research

- Analyze impact of further granulation in such factors as openings, etc.
- Narrow scope to include only high-level games to investigate the existence/structure of a top-down metagame
- Widen scope to include mid- and end-game conditions to further predict game winners

Thank you!

Source: https://www.kaggle.com/datasnaek/chess/data