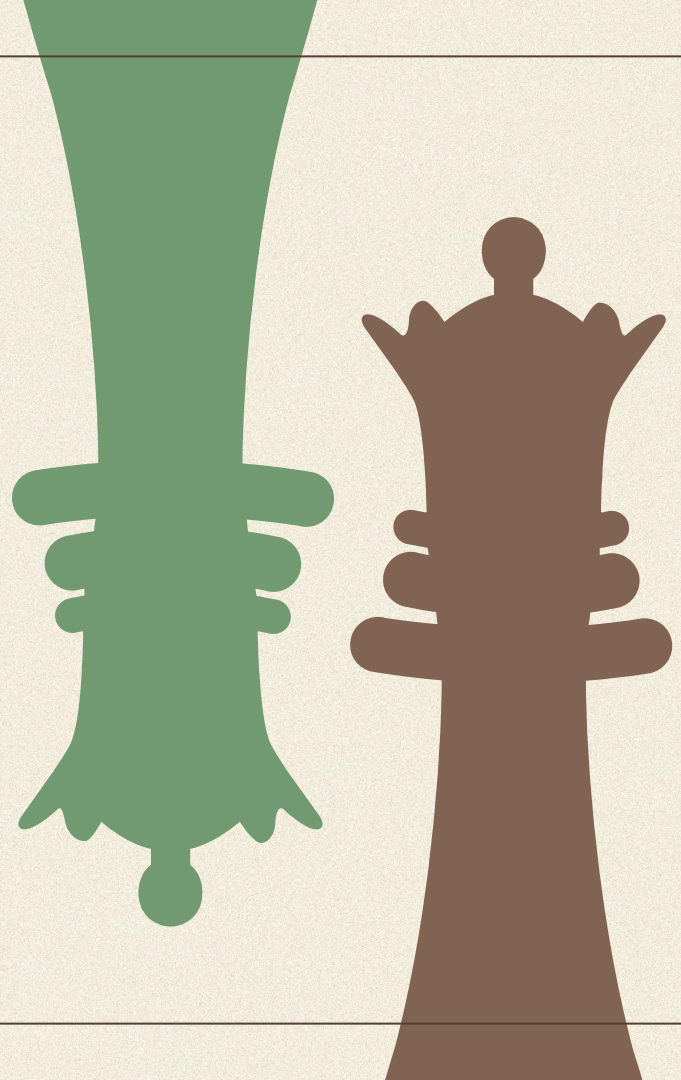
Chess.com Classification

Braden Taack

TABLE OF CONTENTS

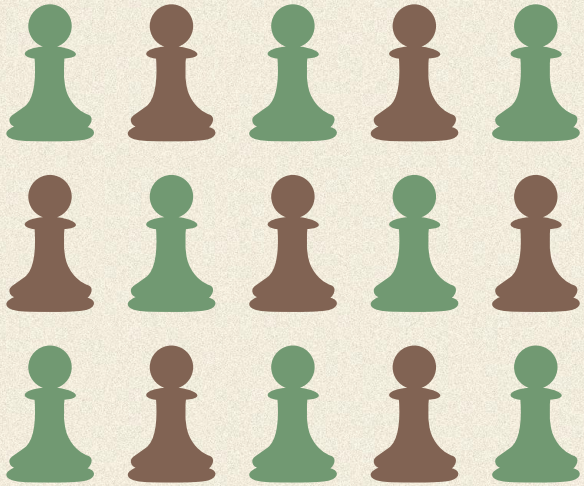


01

Problem Overview

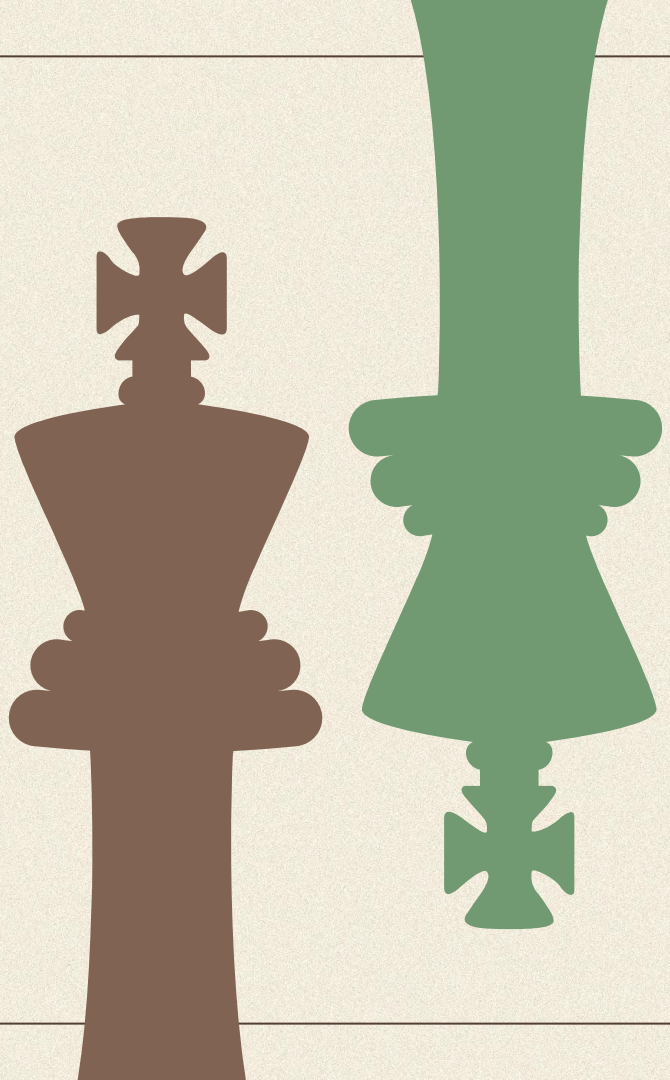
Problem Overview

Chess is a classic game and lots of fun. The goal of this project is to classify a game as a win or a loss based on player statistics.



02

Data



Data

Well, actually...

Games

All historical games from streamer archives

Streamers

Normally dist ratings
Lots of playtime

Features/Player

Game Type Stats
Lessons
Puzzles

Data Subset for Modeling

30K

Games

Randomly sampled with binned streamer subset

13K

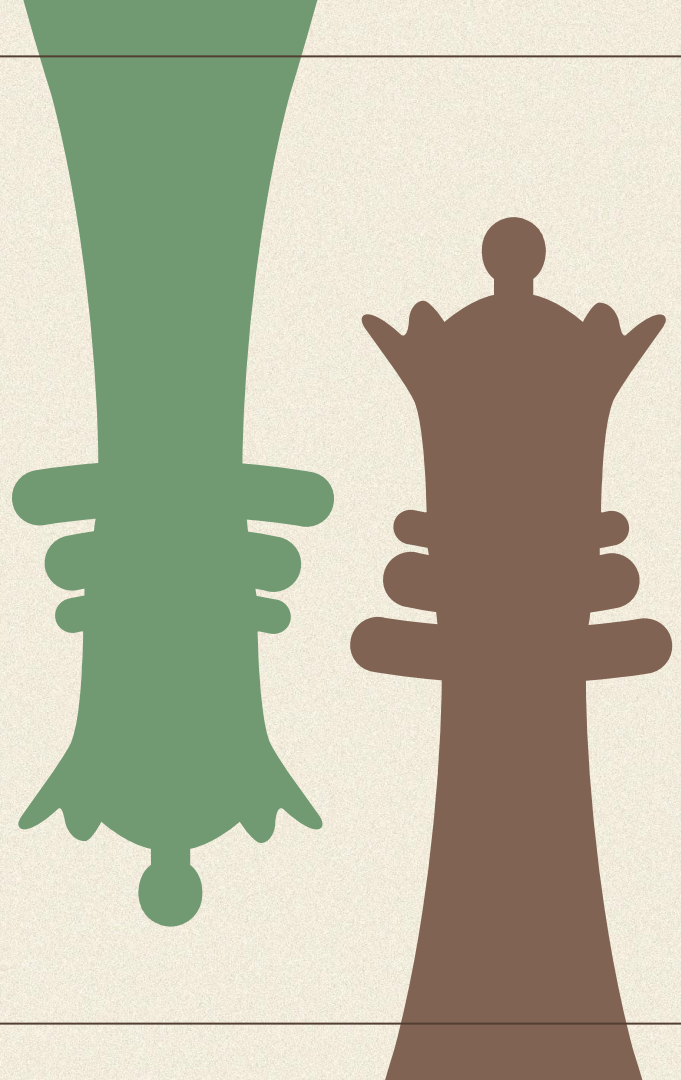
Players

Created pipeline to gather all opponent statistics (not all opponents were streamers)

20

Features/Player

Game Type Stats
Lessons
Puzzles

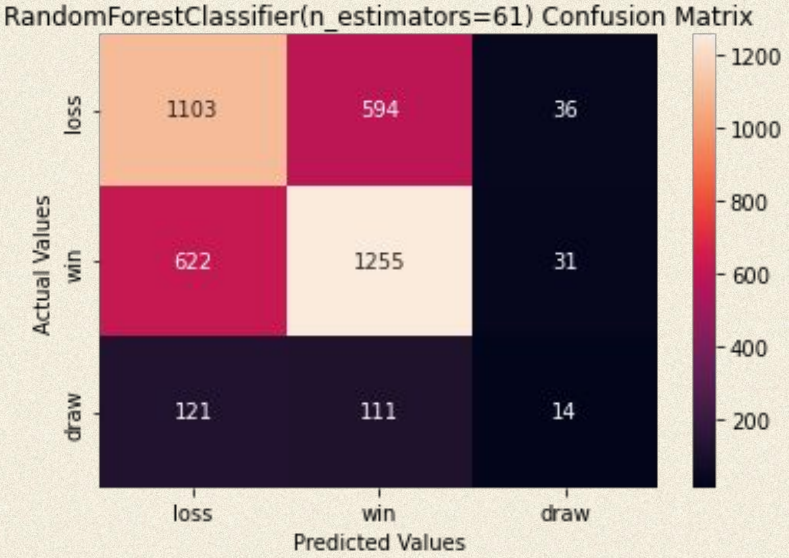


03

Modeling

Modeling – Class Imbalance Issue

- Draw class is rare
- ~0.03% Draw Pred Accuracy
- Solutions Tested on 5 Models:
  - Overbalanced Sampling
  - SMOTE Sampling
  - Class Weight Changes
- Results:
  - More Draw predictions
  - Drop in accuracy and F1

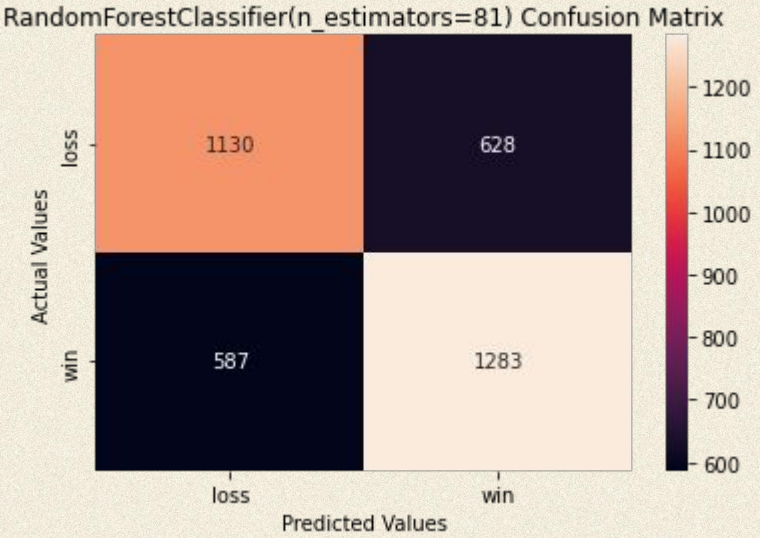


“Draws are boring.”

— **Me**

Modeling - Updated Classes

- Removed ~4% of data when removing draws
- New classes: win or loss



Rating Only - Base Model Choices

KNN

Accuracy: 0.69

F1 Value: 0.69

Random Forest

Accuracy: 0.66

F1 Value: 0.66

Logistic Reg

Accuracy: 0.70

F1 Value: 0.70

Decision Trees

Accuracy: 0.65

F1 Value: 0.65

XGBoost

Accuracy: 0.58

F1 Value: 0.65

Model Engineering

Feature Engineering

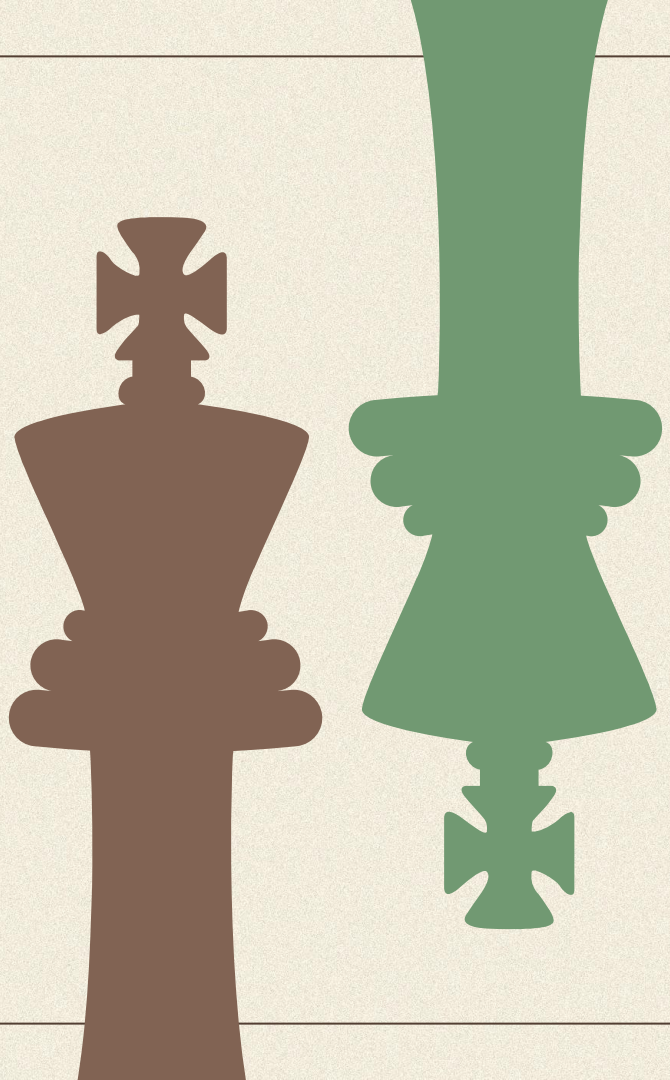
- Categorical Features
  - Time Class
  - Rules
  - Game Type
- Date since best rating
- Win %
- Current rating/best rating

Model Tuning

- GridSearchCV
- Hyperparameters
  - C
  - Solver
  - Cost Function

04

Results



Final Model

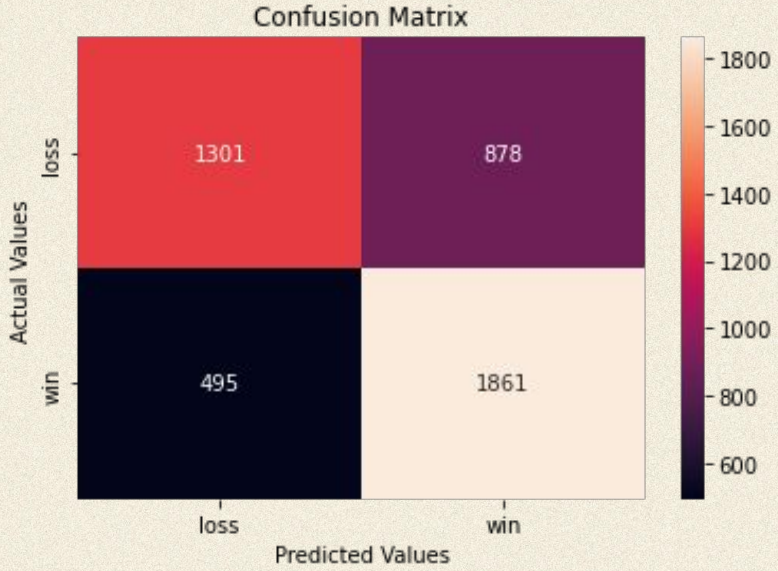
Metrics

Accuracy

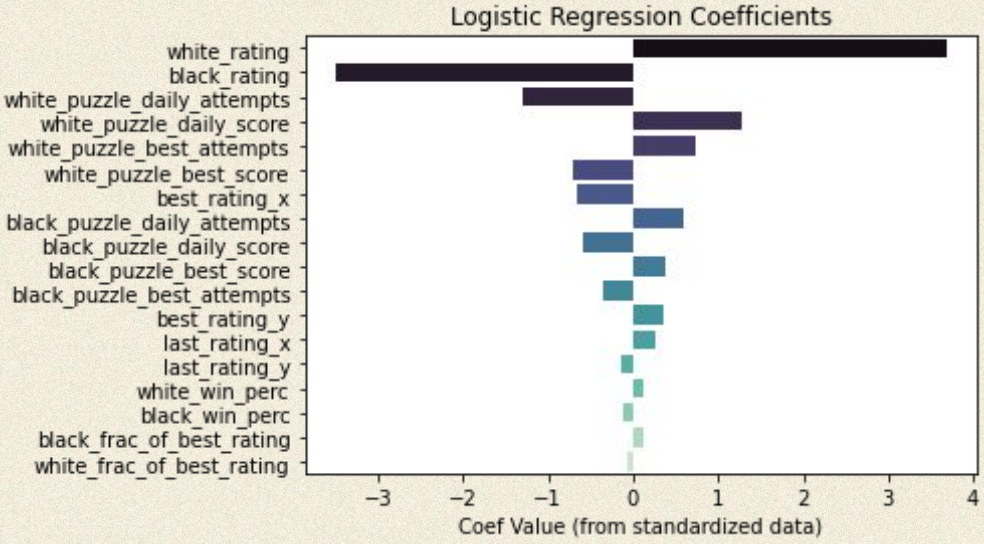
0.7

F1 Score

0.73



Top Features



Glicko Rating

Puzzle Stats

Best Rating

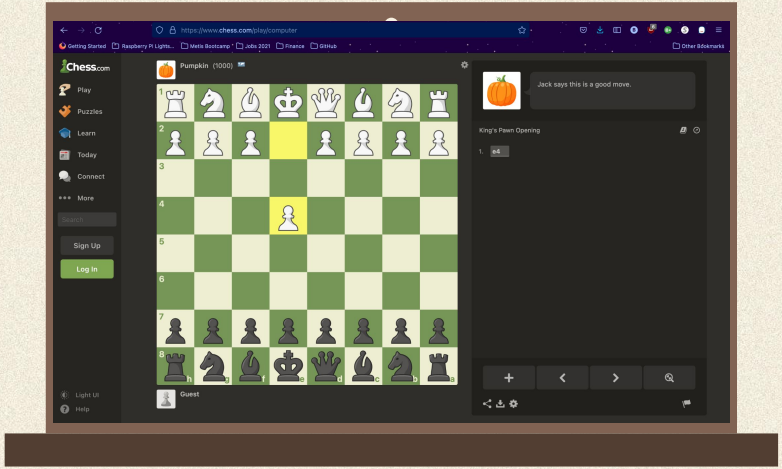
Example Game

	GAME:	Chess Rules
	TYPE:	Chess Blitz
	TIME CONTROL:	3 Minutes

	PLAYER	Rating	Puzzle Attempts	Pred / Prob	RESULT
WHITE:	Josiwales (streamer)	2621	57	Win (61.2%)	WIN
BLACK:	Agent-JL	2560	63	Loss (38.8%)	LOSS

Conclusion

- Logistic Regression Model worked best
- Glicko Ratings difficult features to outperform
- Future Work
  - Design more advanced feature
  - NLP to analyze gameplay
  - More Data
  - Spark



THANKS

QUESTIONS?