CS210 Data Science Project

my_chess_probe

Ata Çavuşoğlu

30997

Introduction

I've started playing chess about August of 2023 with the chess.com site, and I've encountered the plateau rating of mine and got stuck there for a while. That's why this project is aiming to find the hamartia of my chess.com games. First, I've gathered the data with Selenium and Bs4 by scraping my profile and the all the Blitz and Rapid type games' reviews which are also evaluated by Stockfish except a few minorities. The data included several variables like the ratings, accuracy of both ends, evaluations, moves, blunders, misses, mistakes, date, time, opening and dependent variable result. In EDA testing part, missing values are handled, and new variables such as detecting the characteristics of the games based on the evaluations and moves is tested. Time of the day, rating differences are measured for their impact on the result of the games. In the hypothesis testing, for my initial hypothesis, T-Test and Whittney U Test is done on Italian like openings for both white and black (mirrored) with the accuracy associated with my games since the most frequently moves I play is that. Chi-Square test on time of the day and the result, and openings with result is tested. For ML part, first blundered move and win binary variable is checked with logistic regression. Lastly, I wanted to check whether my predictors of the result are affective on the data; thereupon, by tuning hyperparameters, decision tree classifier is established with 86% accuracy with max depth 10 and min sample split 2 parameters.
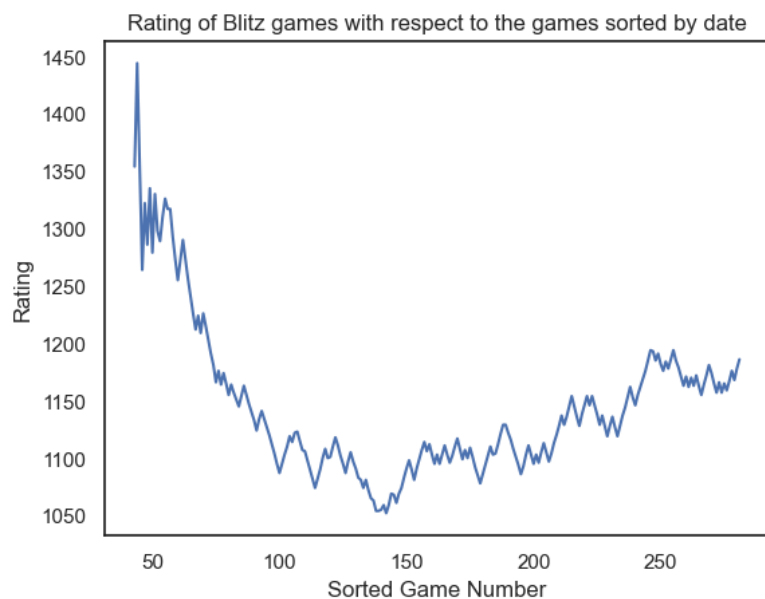
Data:

Firstly, the data of my chess games is accumulated from 43 Rapid games and 239 Blitz games. Thereupon, RatingM showing the density of games corresponding to the rating of mine is clearly divided as a 2 sample (appendix 1). However, since the average of people having 1500 points in Rapid is around 1100-1200 in Blitz, there is no need for an immediate division.
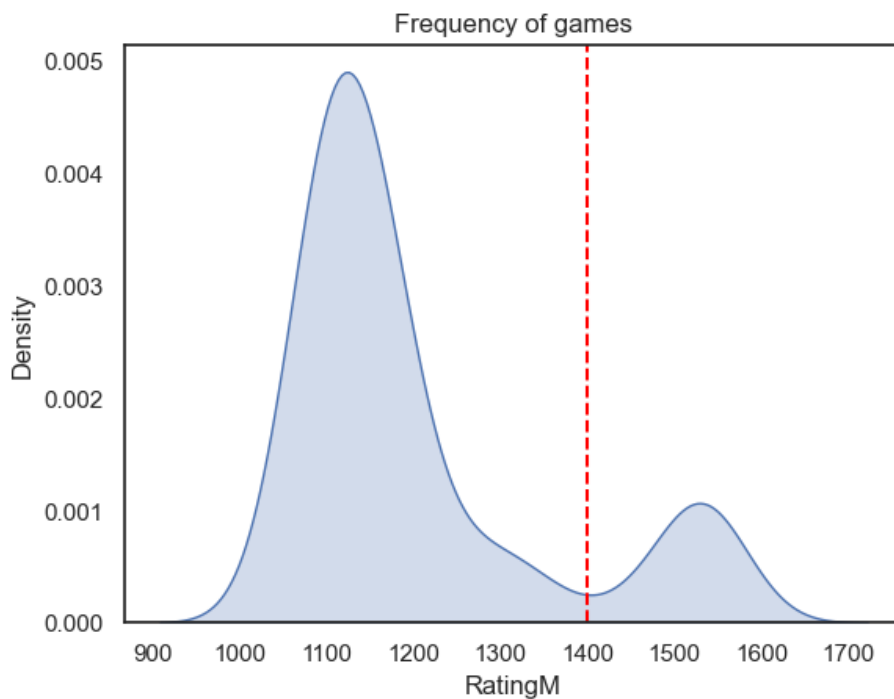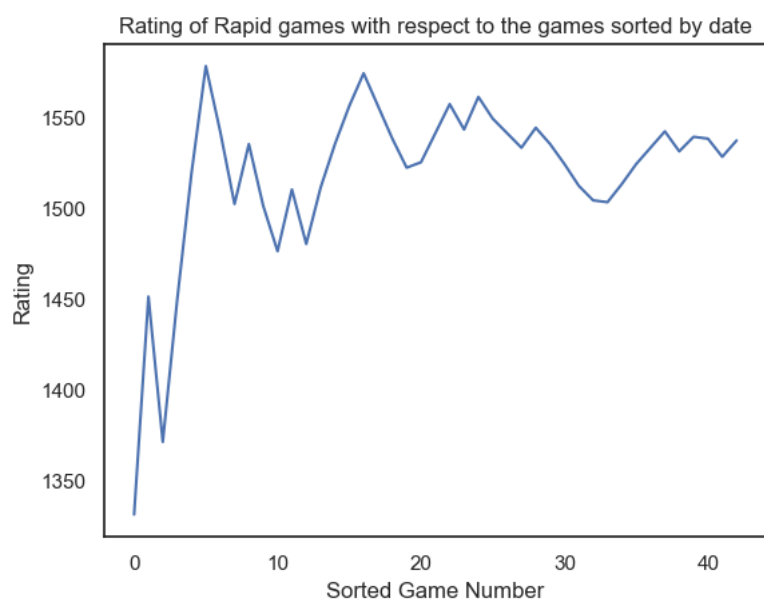

(appendix 1)

There has been a drop in the first 150 games of Blitz type, since the account is fairly new, it was coming to a balanced state.
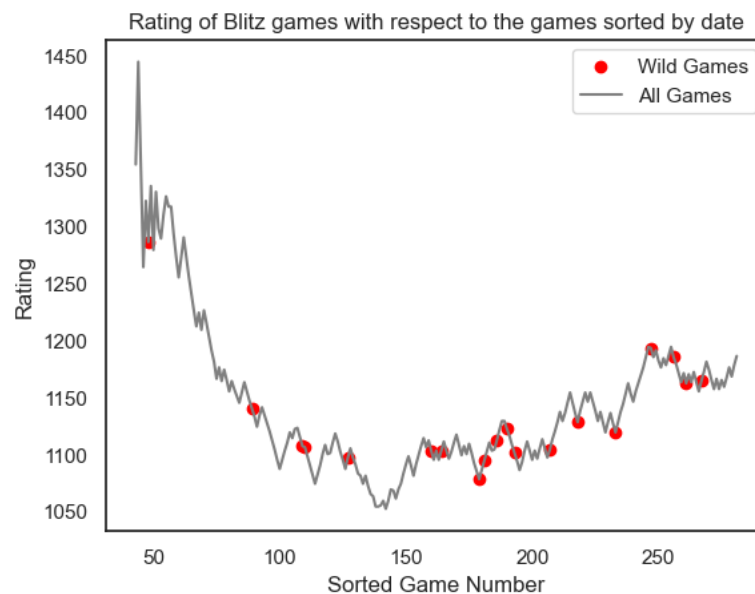
Rating of Blitz games with respect to the games sorted by date

(appendix 2)

Rating of Rapid games with respect to the games sorted by date

(appendix 3)

Analysis:

By taking the mean of evaluations, identifying sudden changes and outliers, the characteristics of the games are categorized as "Easy",

Rating of Blitz games with respect to the games sorted by date

"Mild" and "Wild".
(appendix 4)

This also showed that even though "Wild" is mostly associated with the games after the 150th one, 12/19 of the games are lost.

Moreover, the time played over the day resulted in a significant way.

Count of Results with respect to the hour
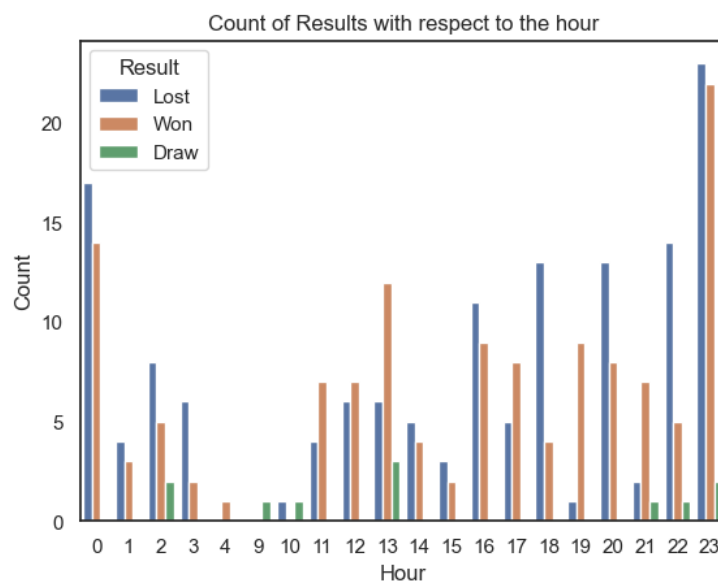
(appendix 5)

It seems that there is a peak around 1 pm that results positive for winning with prominent counts. By categorizing the hours 3-13, 14-22,23-03, there happened to be a great relation between the result and the time of the day.
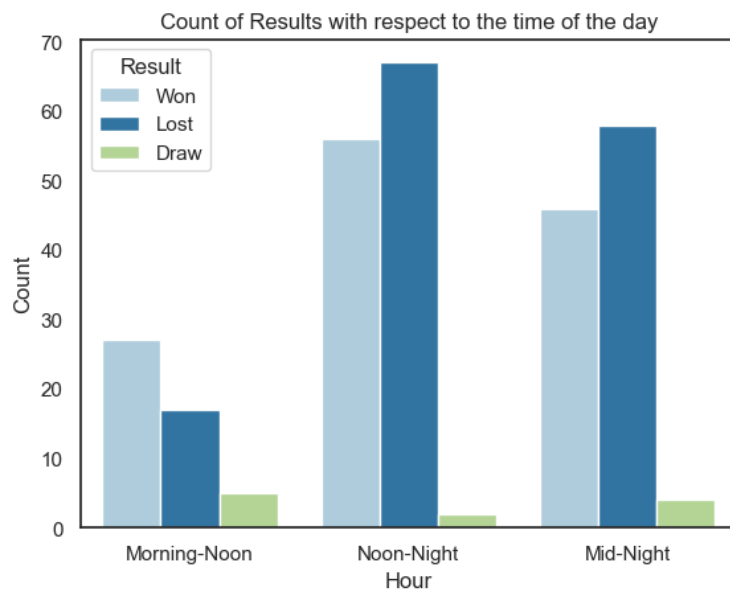
(appendix 6)

Evaluating Chi square test, null hypothesis being that "There is no significant relation between time of day and hour." We reject it.

```
Chi-square value: 10.783628571208
P-value: 0.029106431026284182

Significance level: 0.05
We reject the null hypothesis. There is a significant association between Result and TimeOfDay.
```
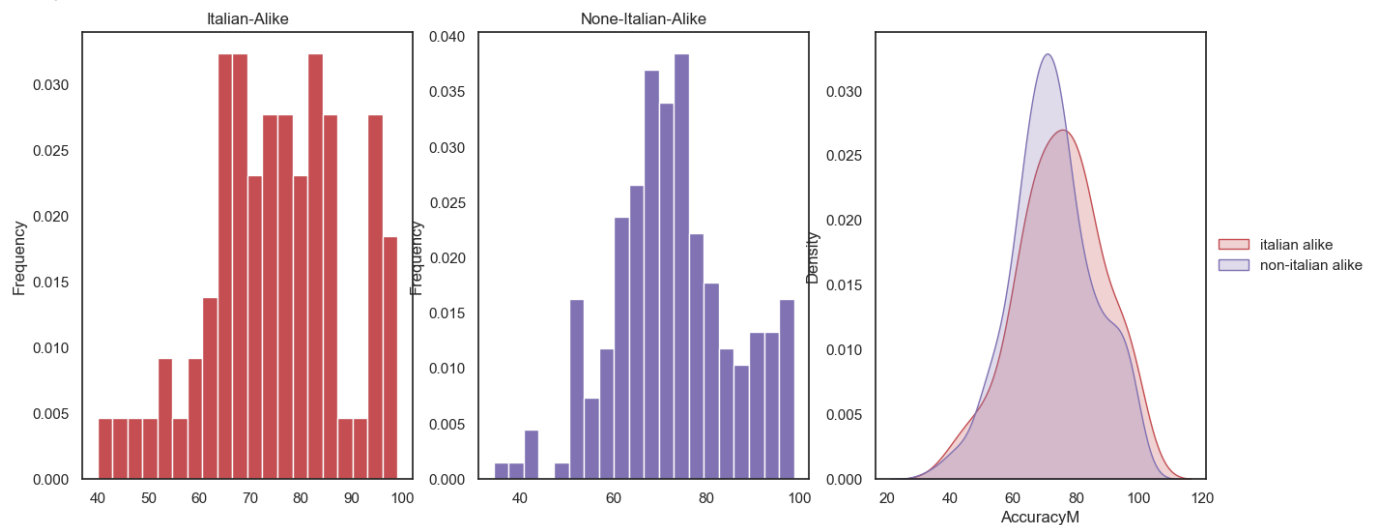
(appendix 7)

Furthermore, since the mostly played opening moves I do is e4, Nf3,Bc4 and d5,Nc6,Bf5 for white and black. Italian like openings are handled this

way.

(appendix 8)

Two sampled T-test resulting in a p value=0.207, between Italian like and its opposite when null hypothesis is that "There is no difference in mean between them" is failed to be rejected.

Another interesting result is observed when the difference between the opponent and I was greater or lower than at most 50. Even when I'm 50 points at most lower the average accuracy of the opponent is 73.1 while the other one is 74.5. Seeing that even though they played slightly worse I lost 65/105 games whereas against 74.5 averaged accuracy opponent, I won 58/98.
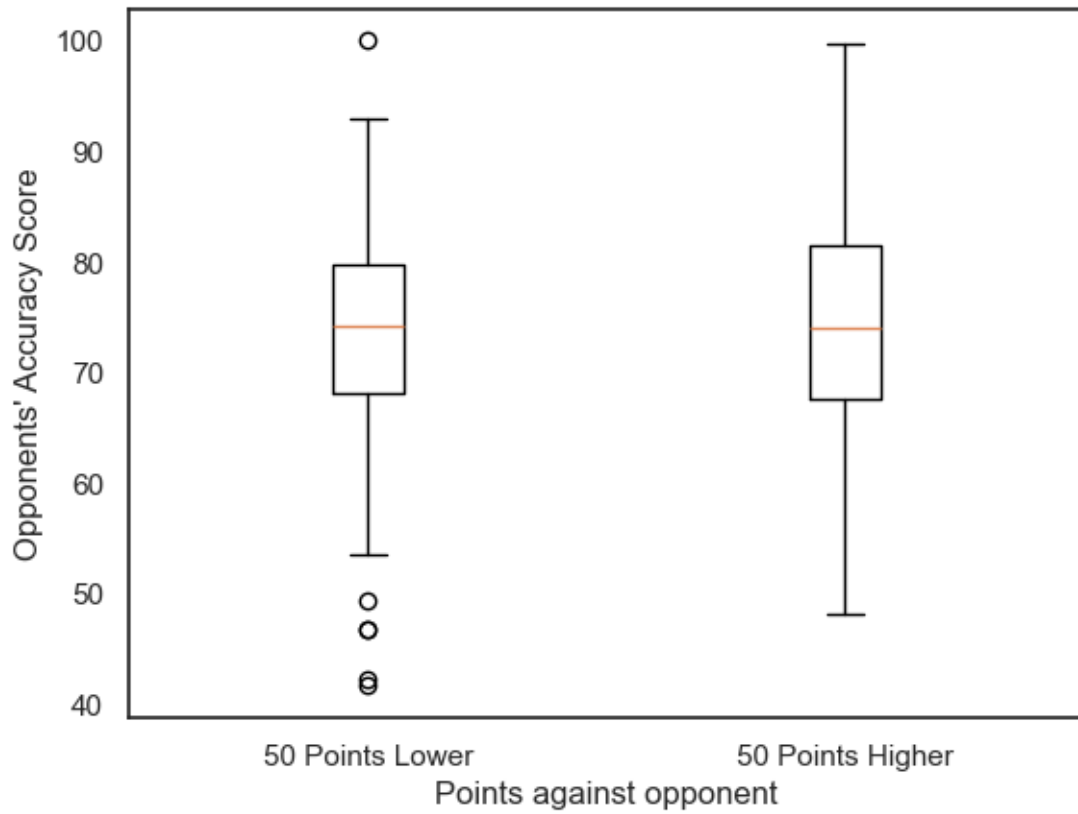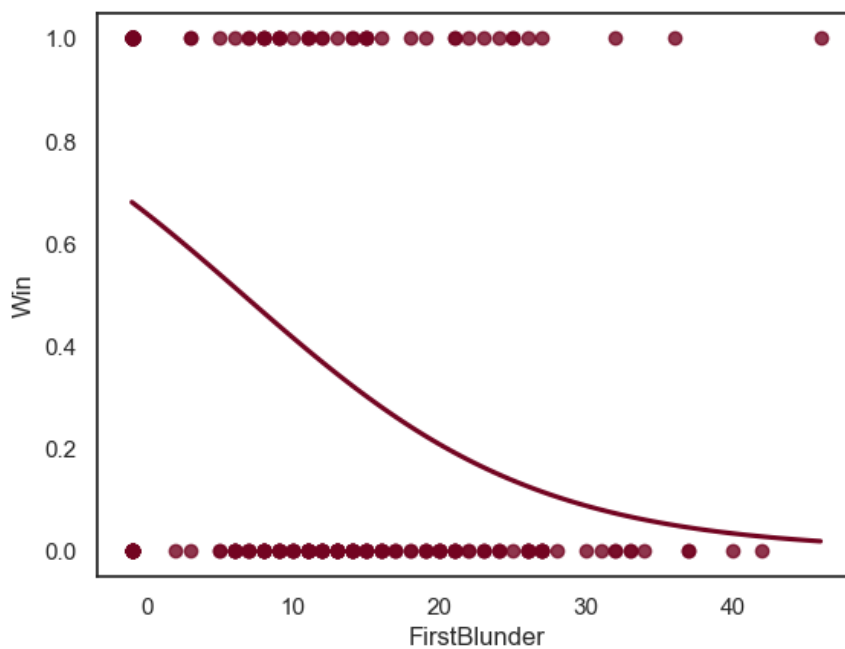
(appendix 9)
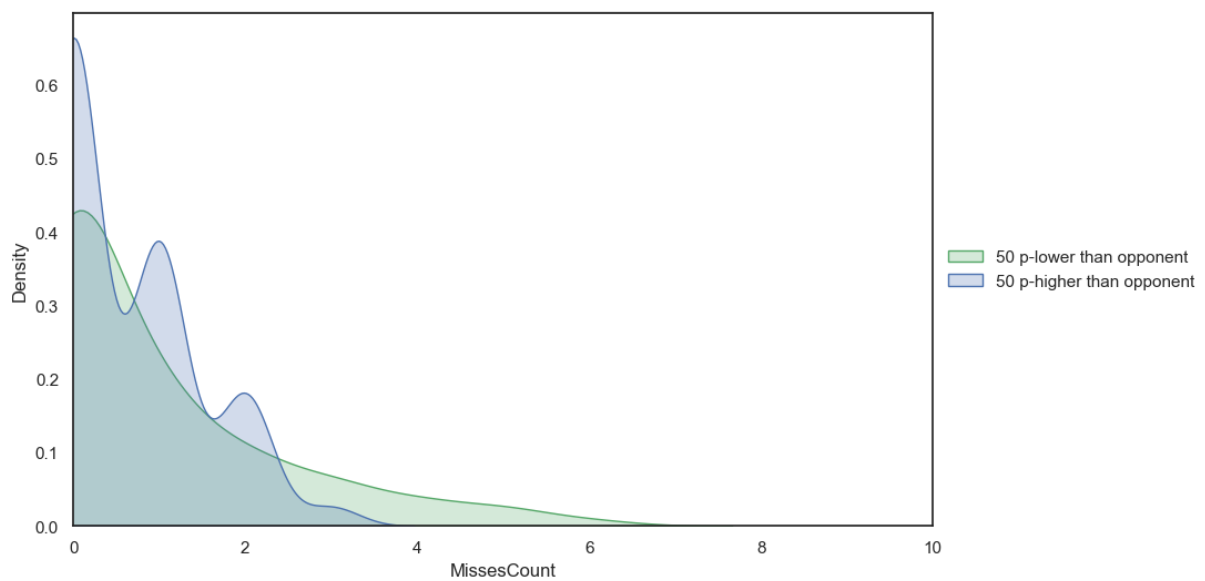
(appendix 10)

Done a T-test, where Null hypothesis is "There is no difference in mean between two samples". P value is 0.054 where significance level is set to 0.06. So we rejected and showed that there is a significance of overlooking the moves.

Logistic regression between the first blundered move and win is to be predicting the outcome with 0.68 accuracy which is not bad with itself.

(appendix 11)

Lastly, I've tuned hyperparameters with decision tree classifier using max depth and min samples and with the given cols = ["FirstBlunder","Opening_map","AccuracyO","AccuracyM","MoveCount"," RatingDiff","TimeOfDay_map","Player_map"], its classification accuracy is 86%. It mostly mistakes the Draw class for Lost class.

Confusion Matrix

	Draw	Lost	Won
Draw	0	1	1
Lost	0	31	4
Won	2	0	18
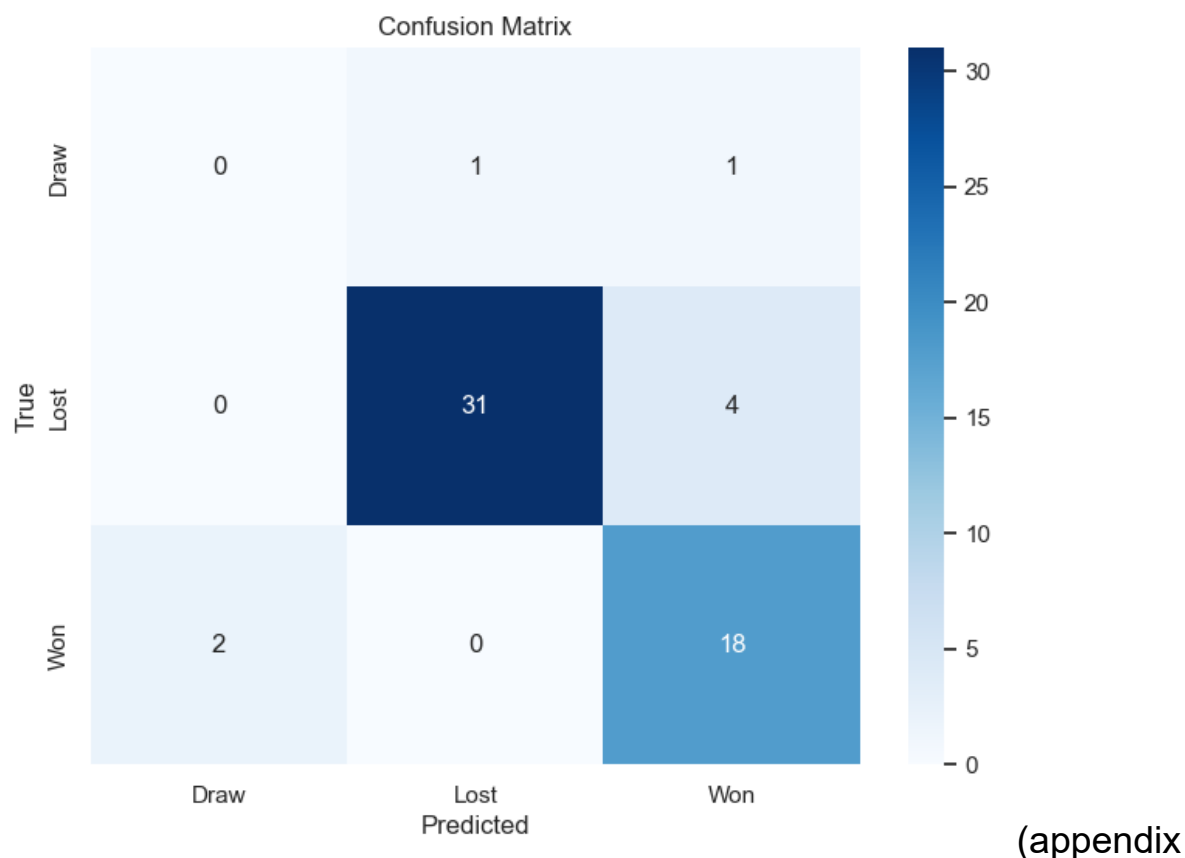
(appendix 12)

Results:

No significant causation is found between Italian like openings that I frequently play with the results of my games. The null hypothesis is not rejected. However, the time of the day and a slight difference in rating between me and my opponent had a greater effect on the failure than any other. Other than the weaknesses stemming from lack of knowledge upon some openings, most of the failure of the games presumably comes from biological time and psychological pressure I am on.