Checkmate by Algorithm: Analyzing Online Chess Games through the Lens of Machine Learning

Thanya Nguyen
Mourtalla Diop

Our Data We Chose

And Why we chose it

- Chess Dataset
- Originally choosing running related
- Both have previous chess background
- Both in Chess Club at UTK

Our Data's features and what we chose from it

- Game ID;
- Rated (T/F);
- Start Time;
- End Time;
- Number of Turns;
- Game Status;
- Winner;
- Time Increment;
- White Player ID;
- White Player Rating;
- Black Player ID;
- Black Player Rating;
- All Moves in Standard Chess Notation;
- Opening Eco (Standardised Code for any given opening, list here);
- Opening Name;
- Opening Ply (Number of moves in the opening phase)

- Number of turns
- Game status
- Winner
- Time increment
- White player rating
- Black player rating
- Opening ECO

Thanya’s Work

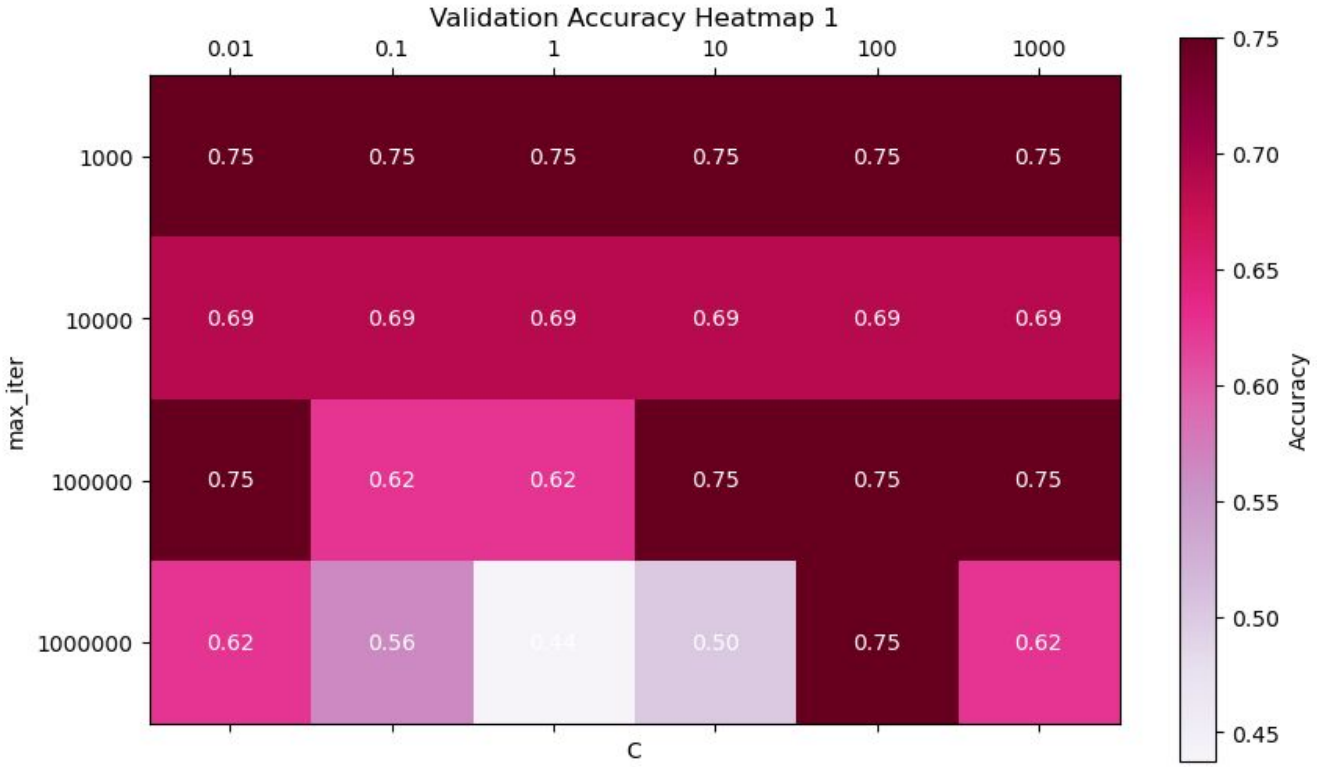
Dataset Visualization

- Separated ratings into bins
- Wanted to see the categories and if there was a color that won more than the other
- As the ratings inc, there is not much difference in winner color

The Steps

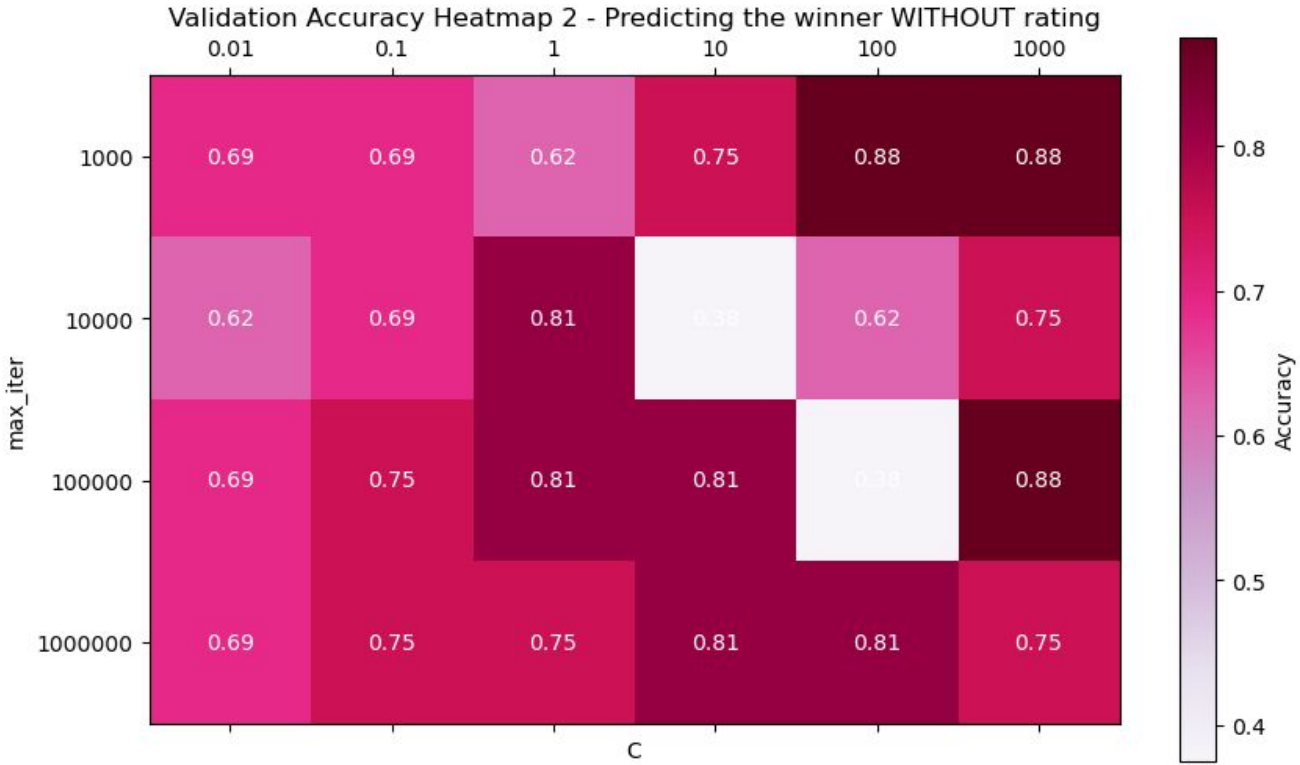
- Down-select the features for the Vienna Opening
- Calculate the rating difference between white and black (use for later)
- Collected features accordingly
- Build heatmap with max_iter and C values

Results - Heatmap 1



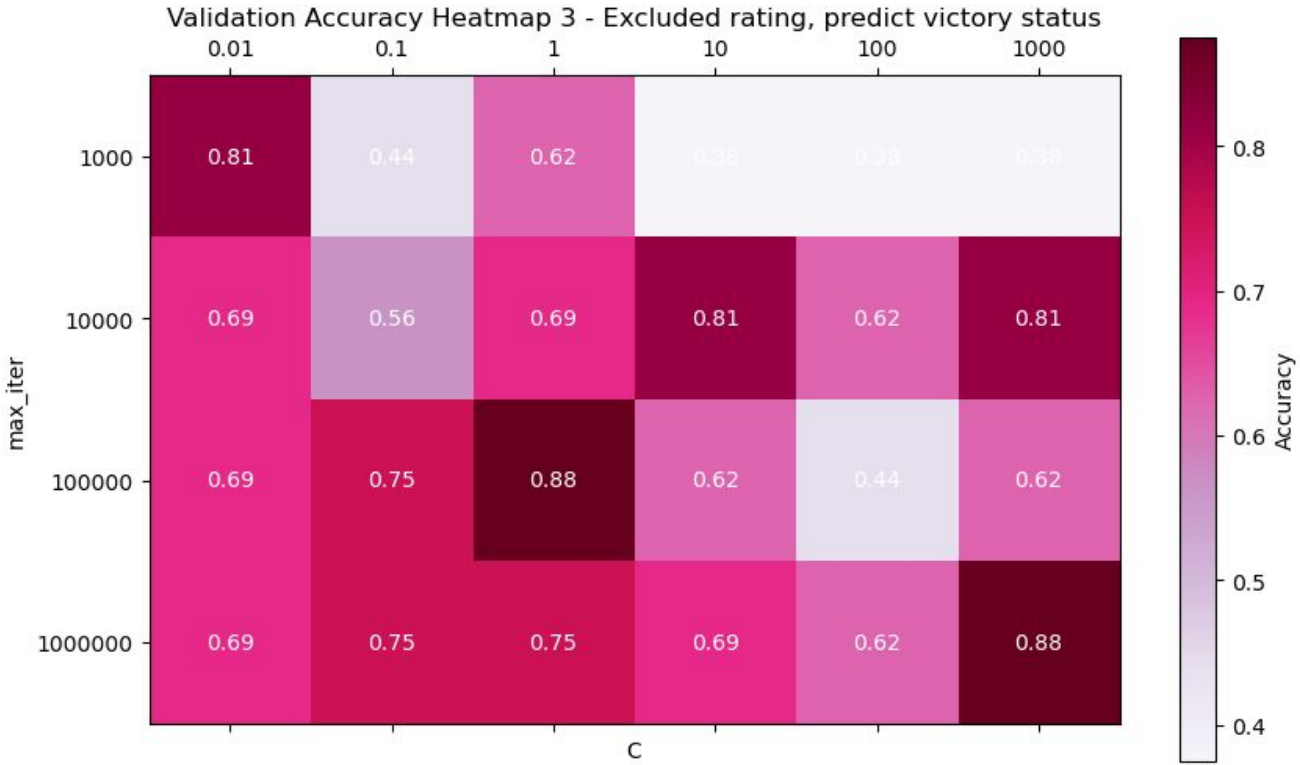
- Average validation accuracy: 0.6822
- Tried to predict winner color given each players ratings
- Many repeats, looks a bit concerning

Results - Heatmap 2



- Average validation accuracy - 0.7188
- Very diverse colors, very pink
- Highest average validation accuracy
- Tried to predict the winner color WITHOUT black/white rating given

Results - Heatmap 3



- Average validation accuracy: 0.6484
- Tried to predict the victory status (checkmate, draw, etc.) without the ratings given.
- Lighter colored heatmap than the other ones

Conclusion

- The best heatmap was heatmap 2, when the ratings were not given
- Could come to the conclusion that it may be potentially due to overfitting when the rating is included. Too much data provided

Common Mistakes

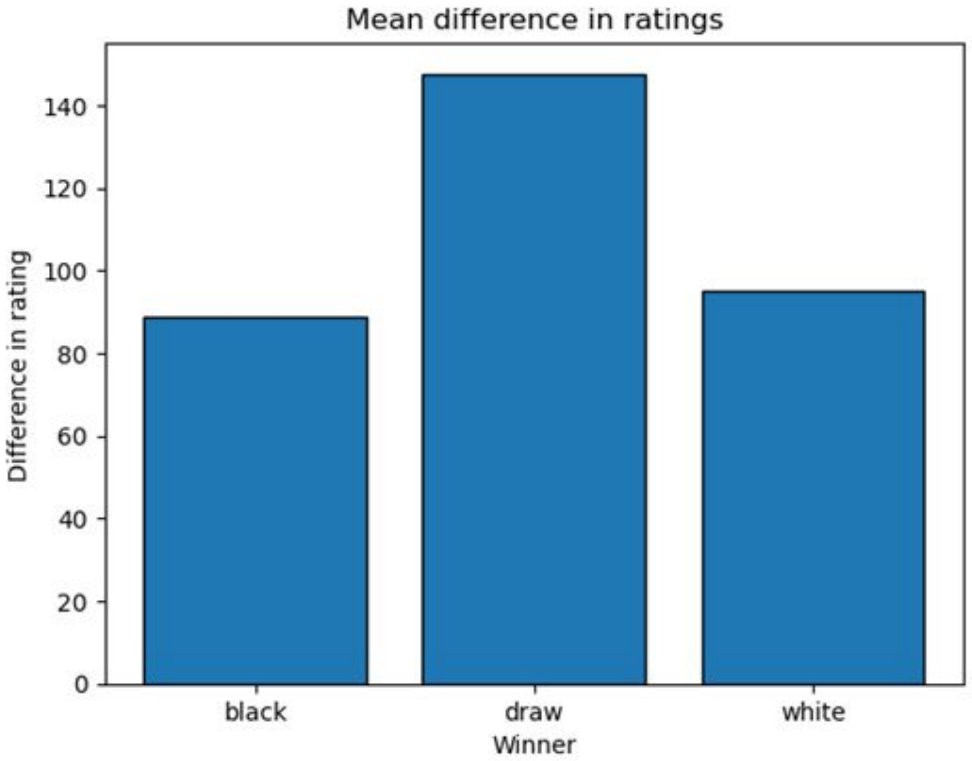
- Had different keys that may affect the map
  - Checkmate: 1
  - Resign: 2
  - Draw: 3
- Different encoding may have skewed the heatmap
- Did not pursue decision trees because scared of free will

Future work

- What would this look like with different openings like Queen’s Gambit? Caro Kann?
- Pursue decision trees
- Use a heatmap to determine “hot spots” for pieces on the chess board

Talla’s Work

Dataset Visualization

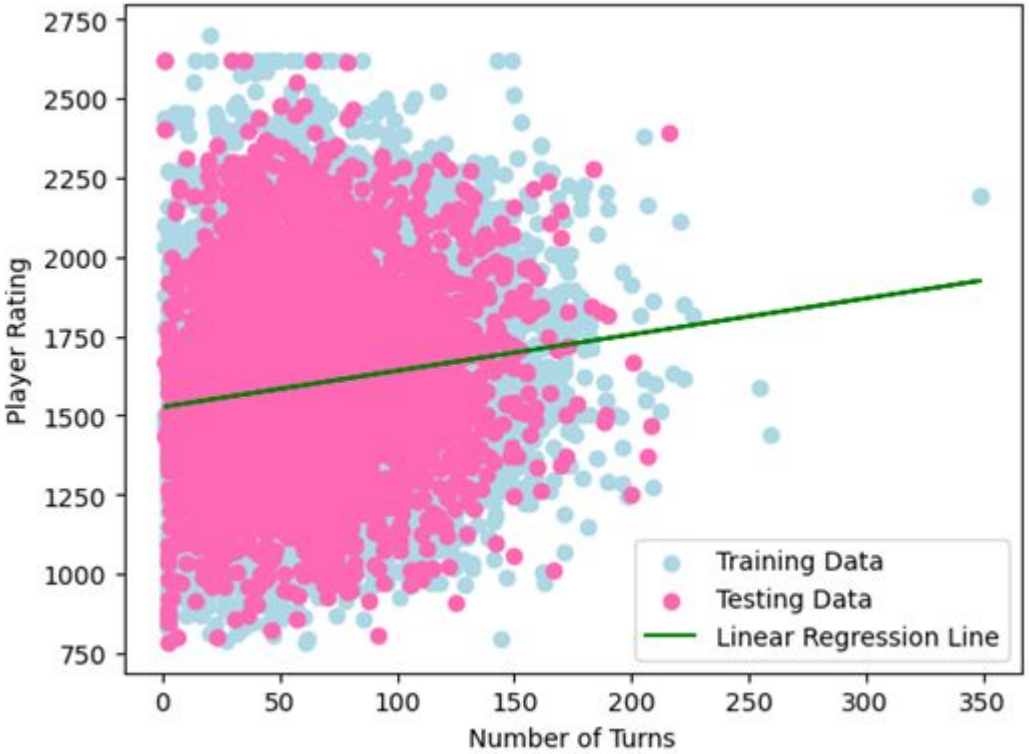


The histogram shows the difference in rating as white players have higher winning rate than black players

The Steps

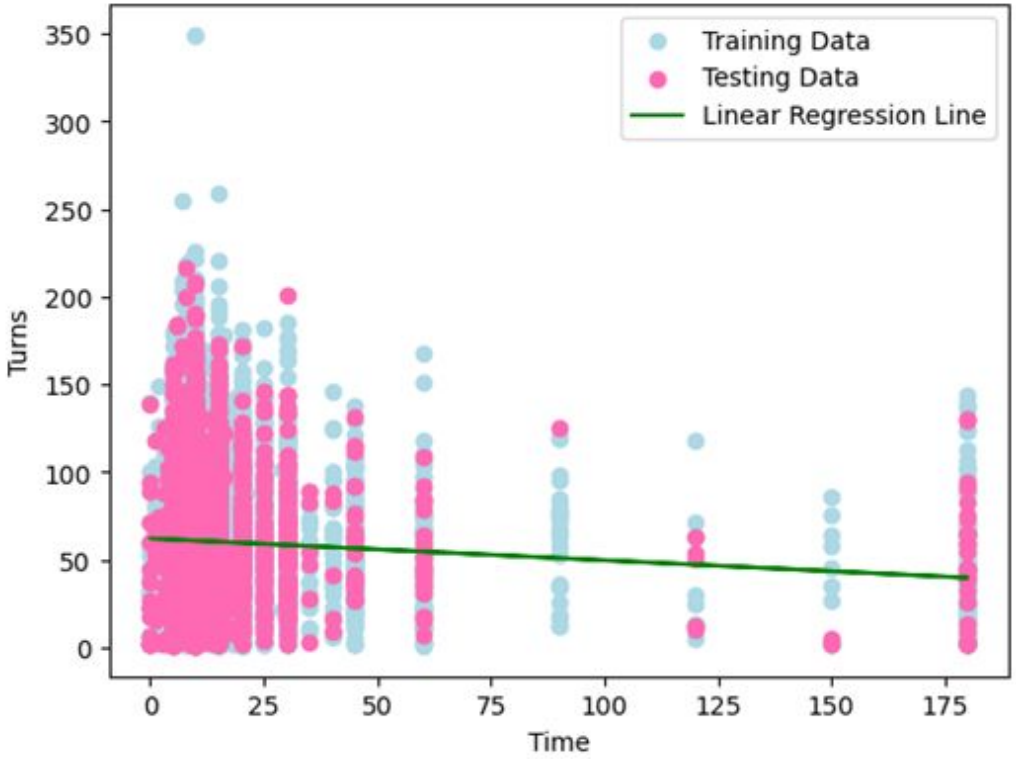
- Read the data
- Split data into training and testing sets
- Create a linear regression model and train it with training data
- Use trained model to make predictions
- Get the coefficients
- Create a scatter plot

Results



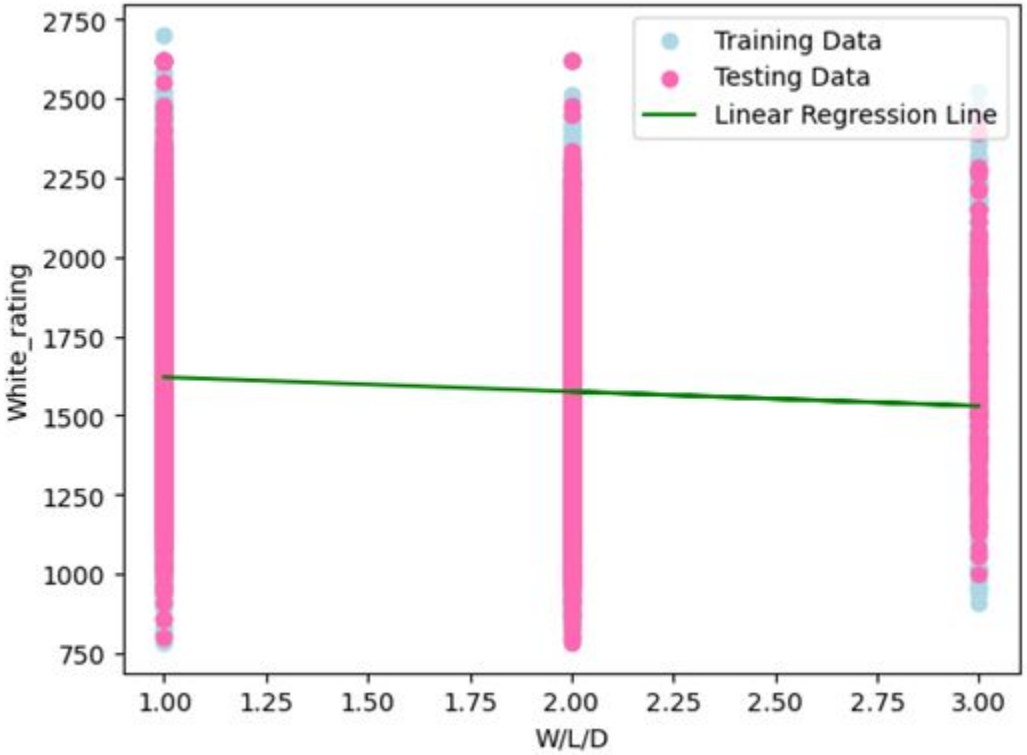
Linear regression between number of turns and player ratings,the plot shows that a higher numbers of turns correlates with higher player ratings as the linear regression captures the linear relationship in the data

Results



It visualizes the relationship between the time increment and the corresponding number of turns,it shows that the change in the predicted number of turns increase in the time increment

Results



This learn regression model shed light on how modification on winner encoding affects the anticipated white player rating, as for every unit increase in the winner encoding the expected white player rating changes

Conclusion

- The best linear regression which is the one that provides insights into the relationship between the winner encoding and white player ratings in chess games. Further exploration and refinement of the model could enhance its predictive capabilities and provide a more nuanced understanding of the factors influencing white player ratings.

Common Mistakes

- The linear model assumes a linear relationship, which may not capture more patterns in the data
- The model’s performance on unseen data (test set) is crucial for assessing it’s real world utility.

Future Work

- would consider exploring more additional features or more complex models if the linear relationship seems insufficient
- Analyze residuals to identify areas where the model may not accurately capture the variance in the number of turns

Thank you