**Predictive Modeling of a Chess Player’s Style using Machine Learning**

Aniketh Kalagara

Indus International School, Bangalore

[aniketh.kalagara@gmail.com](about:blank)

Acknowledgements

Thank you for the guidance of Dashiell Young-Saver from Harvard University in the development of this research paper.

***Abstract* - For many novice and professional chess players, the emergence of ‘unbeatable’ artificial intelligence (AI) chess models has had a dehumanizing effect on the sport. In this paper, I propose a ‘rehumanizing’ use of chess AI.**

**Specifically, I built AI models that assist chess players in categorizing their own personal human playing styles,**

**based on the moves they make during games. Logistic regression and neural network models were trained using**

**4500 games from five of the most famous chess players in the world: Magnus Carlsen, Hikaru Nakamura, Robert**

**James Fischer, Vishwanathan Anand, and Garry Kasparov. When tasked with associating unlabeled game moves**

**with one of the 5 famous players, the best model (a neural network) performed 33.5% better than random guessing.**

**This model can be deployed to help players learn how closely their own moves can be associated with each of these**

**famous players, which gives them a new sense of their own playing style and whose moves they should study to**

**advance within their own style of play.**





**1. Introduction**

Over the last few decades, chess has been hit with an increased influence of technology to assist players in the sport. The world championship is a prime example, months of strenuous memorization of every single possible line provided by computers, the two participants memorize hundreds of games provided. In this computer-assisted environment, many assume that Artificial Intelligence (AI) models will soon completely take over the sport. However, it must be noted that there is no set way to play a chess game. After the first move is played, 400 possible board setups exist. After the second move, there are 197,472 possible setups. After the third: 121 million. AI can cut down those possibilities to a few 100 of the “best” moves (in terms of equally high winning probably). But how does a player identify which move of those equally “best” moves to take? This is where chess becomes human. The influence of a playing style on the moves and ideas one creates in a game is what defines why chess is so competitive. In this project, I propose a new use of chess AI: to help players identify their own playing styles. Then, when they’re faced with the decision of their next move, they’ll know which move is most aligned with their personal style of play.

The availability of top level chess games is vast, with thousands of games available in PGN (Portable Game Notation) formats throughout chess databases. Chess players of the recent past have different playing styles, yet, they have achieved greatness in the sport. An example can be of the following players: Magnus Carlsen, Robert James Fischer, Garry Kasparov, Vishwanathan Anand, and Hikaru Nakamura. Within this elite list contains four world champions and a renowned world number 2. The success these players have obtained are in different eras. Hence, the development of their playing styles will obviously differ, showcased by the vast amount of analysis which top level analysts and chess authors have highlighted in many books of these five greats. Using this difference as a baseline to explore playing styles can also be used in the effectiveness of AI in identifying the unique qualities these five contain in their approach in a chess game. Hence, I defined my research question as: ***How accurately can artificial intelligence match specific chess games to the top level chess player who played them?***

A successful model will accurately match chess games – based solely on their moves – to the players who played them. Then, this model can be deployed to analyze moves of novice chess players, such that they can identify which of the five “greats” their own playing style matches the most.

A notable study worth highlighting, conducting similar research is *Classification of Chess Games and Players by Styles Using Game Data* by M. G. P. B. Jayasekara in 2018 [1]. The present study differs in terms of the size of our datasets and the types of models used (I use logistics regression & neural networks, they used other algorithms). In addition, the feature extract we employ also differs.

**2. Methodology**

**2.1 Dataset Used**

This study’s original dataset contains games of twelve extremely strong players of different eras. The dataset contained the player name, color they played as in the game, the opponent and opponent FIDE classical rating, the result for the player, among other game data, but most importantly, the game given in standard chess notation, PGN.

To boost the interpretability of the results for novice chess players and to lower computational barriers, I decided to cut the dataset down from 12 players to five.

**2.2 Feature Extraction**

Due to the massive amount of possibilities a chess game has, machine learning models cannot use the raw moves of a game as its features. Otherwise, every game is like a new game, and there are no generalizable patterns to learn from. Hence, I had to extract common features from games, such that we could use them for prediction on future games.

There are eight distinct features that I extracted…

**2.2.1 Game Length**

Players with more conservative playing styles might get into longer games. Hence, this was an important feature to measure.

**2.2.2 Number of Trades**

Trades represent pieces getting taken off the board in succession. When considering the number of trades, equal and unequal trades were considered as trades. Even if a bishop was traded for a rook (a bishop being of lower value than the rook), we considered it as a trade. This was done because trades represent game simplification, indicating that the larger the trades, the more the player likes playing with less pieces on the board.

**2.2.3 Queen Lifetime**

The queen lifetime as the name says represents how long the queen has stayed on the board. The queen, being the strongest piece on the board, has a strong presence. Hence, some players tend to want to remove it quickly to ensure stretched out and positional games in comparison to players who want to keep it on the board for longer representing an attacking playing style.

**2.2.4 Number of Central Pawns**

A chess board is divided into eight columns or in chess terms, files, represented by letters “a” to “h”. The pawns located on the “c”, “d”, “e” and “f” files are considered to be pawns which can occupy the center. Along with the eight files, there are eight rows or in chess terms, ranks, represented by numbers “1” to “8”. The central ranks are considered to be ranks “4” and “5”. The number of central pawns represent how closed or open the game is, highlighting a player’s style depending on the positions which arise out of these pawns and their positions, or otherwise known as pawn structure.

**2.2.5 Piece Advancement**

This metric measures how many times each player goes on to the “opponent’s side of the board”. A chess board as mentioned has 8 ranks. The “white side of the board” is considered to be ranks “1” to “4” and of course black has ranks “5” to “8”. The more advanced one’s position is, the more attacking it is considered to be, with less advancements correlating to a more positional style of play.

**2.2.6 Queen Entry**

Again, using the queen as a metric is important due to the value which the piece provides in the game. In this scenario, the queen entry is simply defined by the amount of moves each side’s queen takes to make its first move. The faster the queen enters the game, the more aggressive that player intends to play.

**2.2.7 Castling**

****

Figure 1: Castling

Castling is a special move which involves the subsequent movement of the king and rook. Castling can only occur once in the game given that the king or rook involved in the castling has not moved prior. The process of castling is an extremely common practice associated with king safety. There are two types of castling as shown in figure 1, the king moving to the right for white or left for black is called “king-side castling” and the other is known as “queen-side castling”. A player who switches up the castling they do, whether it’s not castling at all throughout the game or between the two types, often correlates to a dynamic, creative style of play.

**2.2.8 Piece Moves**

This metric simply checks the number of piece moves done by white’s and black’s knights, bishops, rooks, queens and kings. Again, using the first letter of PGN notation, every instance a specific letter is present, the counter for the specific white or black piece counter is appended. The number of piece moves will of course differentiate between player to player depending on the openings and strategies one uses in the game.

**2.3 Cleanup**

After extracting the features, I finalized the dataset with the columns correlating to the name of the grandmaster, the metrics and the color. The color they played influences their playing style, because black is considered to be more defensive compared to white as they play second. Also, the color the player of focus played will come with any PGN notation inputted, hence, we decided to take it directly from the dataset and use it as a ninth metric.

To prepare for modeling, I one-hot encoded the categorical variables ( the castling metric and color metric). This means that the categorical variables were re-classified in terms of numerical, preferably binary outputs. For example, if the player is white, there can be a white metric in which the value is 1, representing true, and the black metric where black is 0, representing false. This can be achieved for any categorical variable type through “one-hot encoding”.

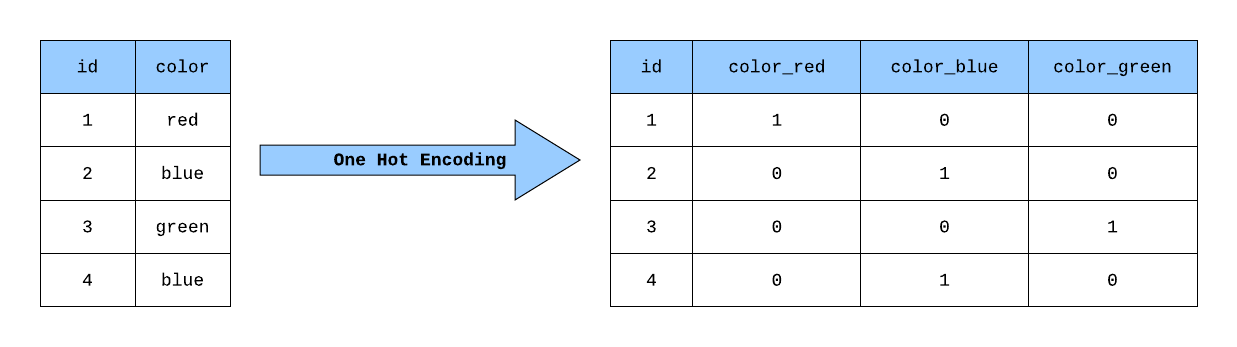
Figure 2: One-Hot Encoding

Figure 2 shows one-hot encoding, we use one-hot encoding to make data into these binary values such that it can be used in AI model creation. Once the one-hot encoded data is placed back into the dataset, it is ready to be placed into a model.

**2.4 Logistic Regression Model**

Prior to feature extraction, the model wasn’t split and the twelve players still remained. Hence, I decided prior to splitting to the five chosen grandmasters, I wanted to test the model using just two players to see how it performed on a logistic regression model to see if it is ready for genuine model creation.

Firstly, a logistic regression model is my baseline model of choice to see how much better the model can perform compared to random guessing. Using a baseline model helps give an idea on how the main, ideal model would perform in comparison as it ideally should be better and given it is not, we know that there are bugs in the code. Logistic regression is a model used to predict binary outputs of 0 and 1, or true and false. In this scenario, the logistic regression model will predict the occurrence of whether the game is of one player or not. The model will predict on a scale from 0 to 1 and whichever the model predicts is closer to the true value, it will predict that as the player.

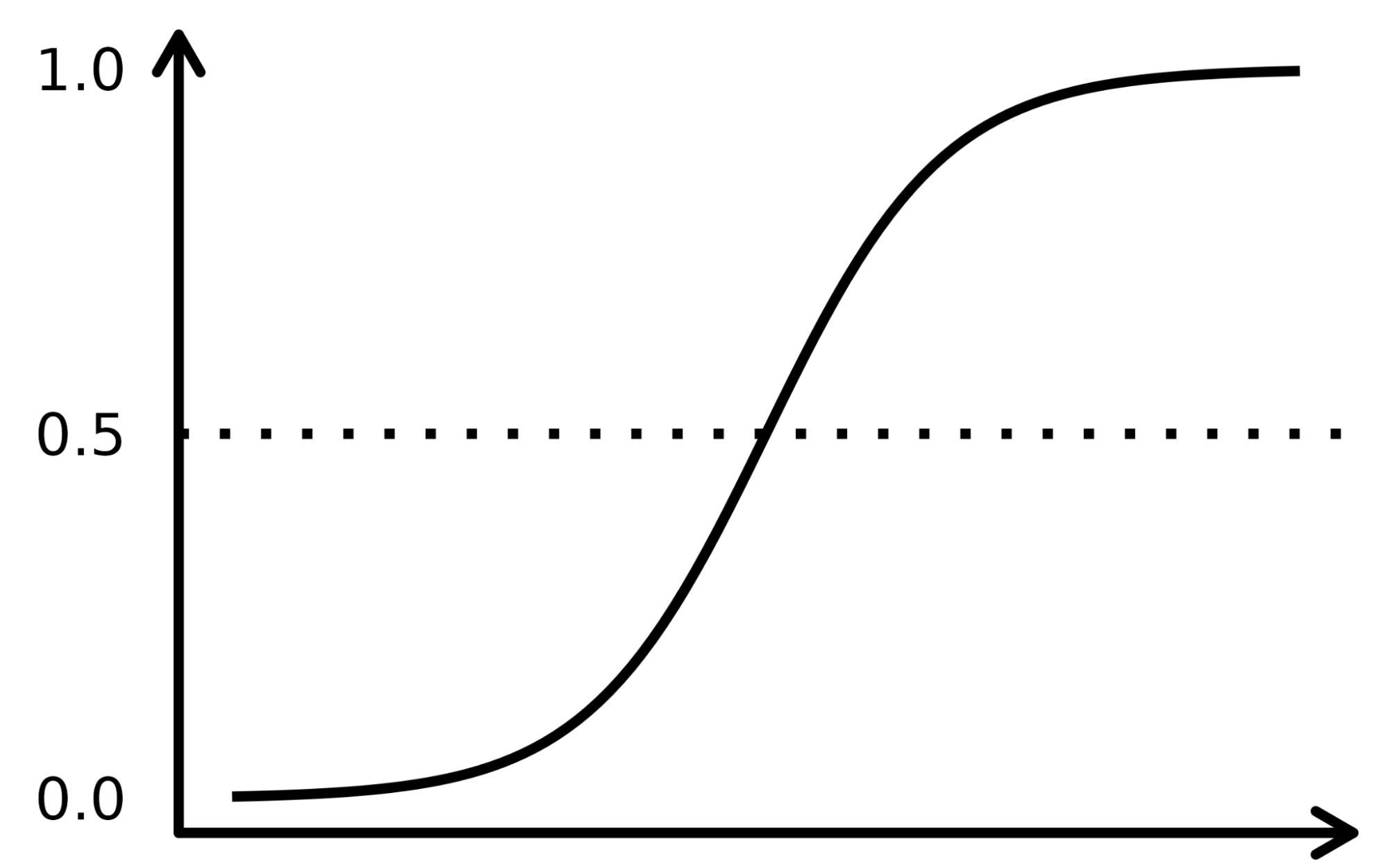


Figure 3: Logistic Regression Diagram

A diagrammatic representation of logistic regression is displayed with a sigmoid or S-shaped curve structure. As said earlier, the model will base its result on such a scale and accordingly predict between the two players. If the model successfully predicts above random guessing, it can be applied to the chosen five players as well. The training to testing split, which involves the dataset being split randomly into data which the model will be trained by and a test set for the model to predict results on, is a 70 to 30 train to test split. Every time any model is run, there will be a random state factor which means every time, there will be a random split occurring, which means there won’t be the same data in a train set and a test set ever.

Prior to creation of any model, due to the dataset being of a random amount of games per each player, it is important that the amount of games per player remains equal and the players chosen are known to be of different styles such that the best accuracy can be obtained and the model can be accurately measured to see if the feature extraction did not have any bugs. Hence, for all models created, the baseline amount of games will be 900. Along with that, for the two player logistic regression, we took two players of fairly different playing styles. A sacrificial, extremely dubious and aggressive playing style of the 8th world champion, Mikhail Tal and the grinding, positional and extremely accurate endgame playing style of the 16th world champion and one of our five players as well, Magnus Carlsen.

Once the model has been used on two players, we can move on to providing a baseline model for five players as we now know the model functions accurately. Using the logistic regression model will enable us to check whether the neural network model will perform better or not and see how much better it performs in comparison to random guessing.

**2.5 Neural Network Model**

Here is the actual model which will be used to check for five players. This model will be intended to answer the research question to the best of its ability. Prior to creation of the model as well, 900 games was the baseline to measure the results of the neural network.

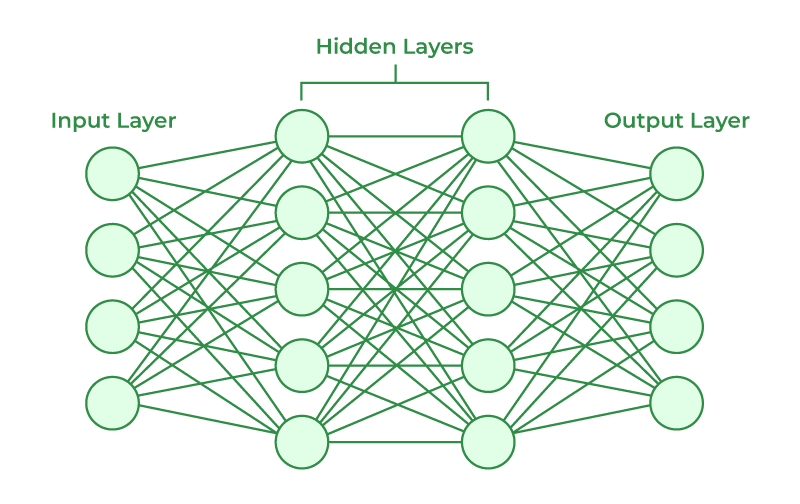


Figure 4: Neural Network Diagram

A neural network is a method of AI which is inspired from the human brain, using layers connected by neurons to help computers identify and predict results. Neural networks use different methods such as the number of times a model should be trained known as epochs, the amount of layers and the amount of weights or nodes which connect to each neuron, etc. The output of this neural network will be a five node layer for the five players which the neural network should be tested on. Throughout the process, the neural network will be tuned in the form of the variety of metrics which can be provided by a neural network. The process of tuning will undergo multiple trial and error processes to identify the best performing model with the highest accuracy.

Different model architectures were tested on the basis of training and validation datasets derived from the 70% split. The trial and error process involved looping through different model architectures, changing the number of hidden layers and the number of nodes per layer while keeping epochs constant. Finally, as per a multitude of trials, the final model architecture had 3 hidden layers with 40 neurons per layer.

**2.6 Calculations**

Once the best accuracy is obtained from the neural network model, a simple calculation should be made to see how well it performed above random guessing. Often when AI models provide an accuracy, it is seemingly quite low, but this low accuracy is not a true showcase of how well the model has performed.

First, let’s consider the two player model. If the model is randomly guessing every player for every game, there will be a 50% accuracy. For the five player model, the random guess accuracy would be 20%. Taking the actual accuracy of either model trained, we can identify how well the model is performing above random guessing. For example, the accuracy of the logistic regression model on two players is 55%. To identify how well it has performed above random guessing, we subtract 55 from 50, giving 5% and divide that according to the random guessing accuracy, being 50%. This results in a value of 0.1 or a 10% better performance than random guessing. Although a 55% accuracy seems really bad, the model is actually predicting something, which means it can be applied given further testing. Hence, using this metric as a measure of accuracy enables a better understanding on the performance of an AI model which will be used for all three models and their best performing accuracy which will be shown in the following results table.

**3. Results**

The following table presents the results of all three models:

| **Model Type** | **Accuracy** | **Percent improvement over random guessing** |
| --- | --- | --- |
| Logistic Regression (2 Players) | 61.6% | 23.2% |
| Logistic Regression (5 Players) | 24.5% | 22.5% |
| Neural Network (5 Players) | 26.7% | 33.5% |

The results clearly highlight that, initially, the model did function given two players and post that, had a good improvement from the logistic regression model to the neural network by a 11% percent improvement over random guessing.

**4. Discussion And Conclusion**

With more experimentation and even larger neural network models, along with stronger processing power, the model can be maximized given the features extracted. Along with that, we can see that the number of features extracted are only nine, with so many factors affecting a player’s playing style, ranging from piece sacrifices to openings played, there are factors which can affect a player’s style not being fed into the model. Hence, using these two as possible improvements, we can maximize the use of PGN notation to bring about new features with even better performing models.

The application of this research can be applied to help newer chess players identify their playing styles. What I wish to do with this research would be creating an application in which players input their own games and according to the features extracted, help lead themselves towards these top grandmasters, whether it is these five or anyone else. This provides an opportunity to help rehumanize the effect AI can have on chess, a goal which I hope this research will aim to address.

**References**

1. Jayasekara, M. G. P. B. “Classification of Chess Games and Players by Styles Using Game Data.” Classification of Chess Games and Players by Styles Using Game Data, 2018. https://dl.ucsc.cmb.ac.lk/jspui/bitstream/123456789/4219/1/2014MCS033.pdf.
2. Renyu, Liu. “Do Chess Masters Have Different Playing Styles?” Kaggle, January 15, 2020. https://www.kaggle.com/code/liury123/do-chess-masters-have-different-playing-styles.