Predicting the Outcome of a Chess Game Using the Random Forest and Support Vector Machine Algorithms

Ryne Swift

University of Manitoba

December 6, 2023

Héctor Pulido's thesis

- Héctor Pulido's thesis was crucial in the construction of this project.[1]

[1] Predicting the Outcome of a Chess Game by Statistical and Machine Learning Techniques.

Project startup

- Downloaded a database of more than 6.5 million chess games
- Applied the "so" filter to reduce the database's size into approximately 498,000 games
- Returned games which had at least one player containing "so" in their first or last name
- Name independent of player strength
- Elo rating is an approximation of a player's strength
- Remaining players held Elo ratings between 1 to 3000
- Converted the remaining games into PGN format
- Necessary to clean the dataset

Data cleaning and processing

- Removed games which presented errors using Python
- Approximately 17,000 games were removed
- Appended additional data to the dataset

Source: https://www.obviously.ai/post/data-cleaning-in-machine-learning

Dataset variables

fen_eval	white
sum_eval	white_win
white_elo	white_draw
black_elo	white_lose
black_avg_elo	black_win
black	black_draw
result	black_lose
white_avg_elo	

Table 3.2: Variables of the dataset

fen_eval: A list of the numerical evaluations of the game's board positions. Used the 40^{th} fen_eval for predictions.
sum_eval: The summation of the numerical values of the fen_eval list for each game.
white_elo: The Elo of the white player in the game.
white_avg_elo: The average Elo of the white player across every game they played as white within the dataset.
white: The name of the white player.
white_win: The cumulative frequency of wins as white for the white player across the dataset.
result: The outcome of the game.

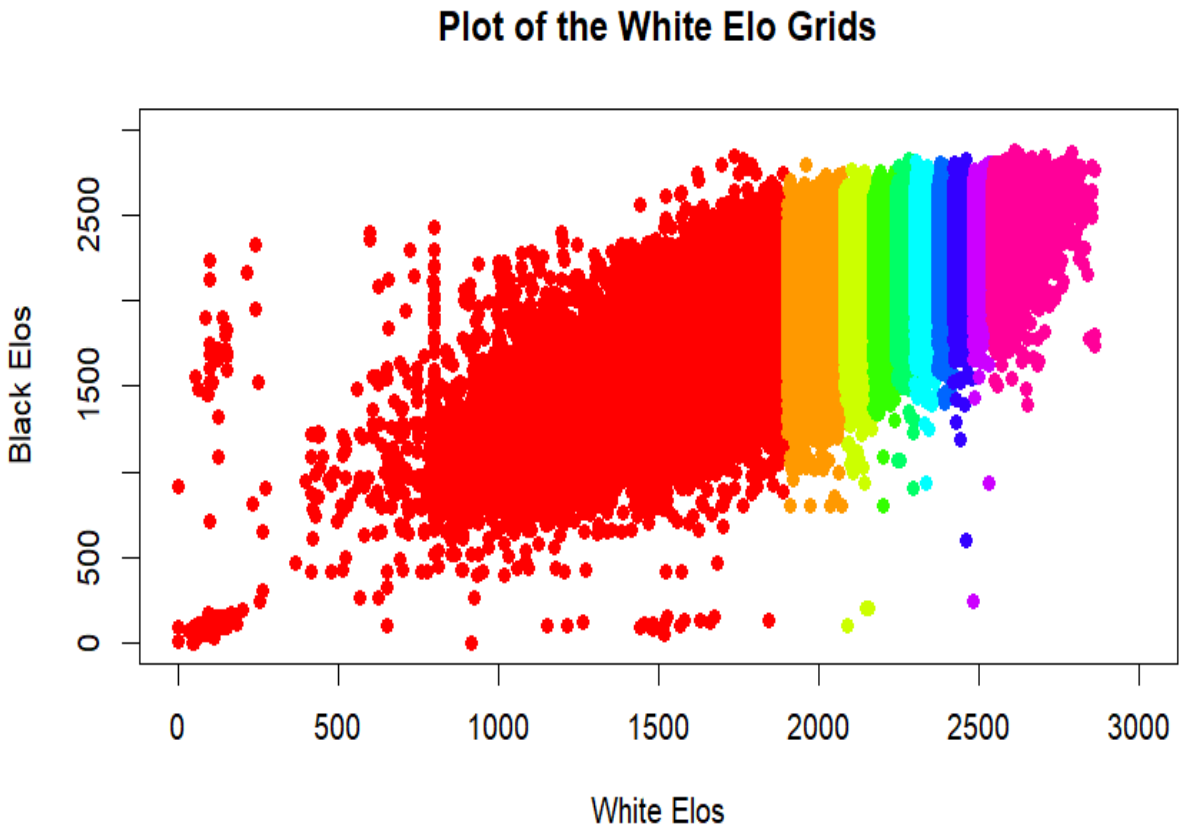
Limitations of the data

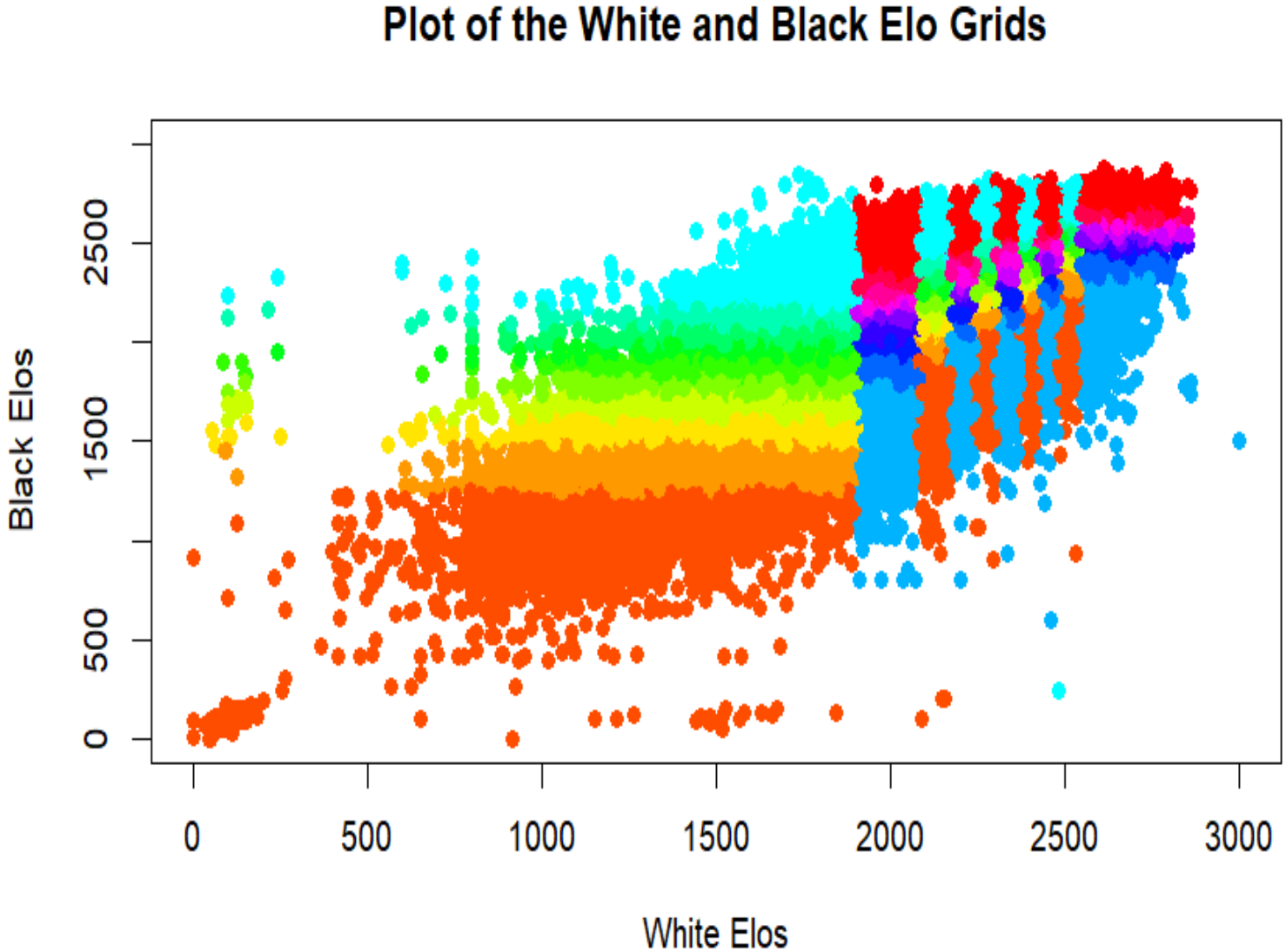
- Fictional games
- Assumes Elos are accurate indicators of a player's strength
- Cheating

Source: https://www.dailymail.co.uk/news/article-8830323/Chess-grandmaster-Igors-Rausis-caught-cheating-phone-loo-banned-competing-again.html

Dataset consolidation

- Removed short games, and games where at least one player name was missing
- Approximately 233,000 games left in dataset
- Separated the dataset into 100 segments



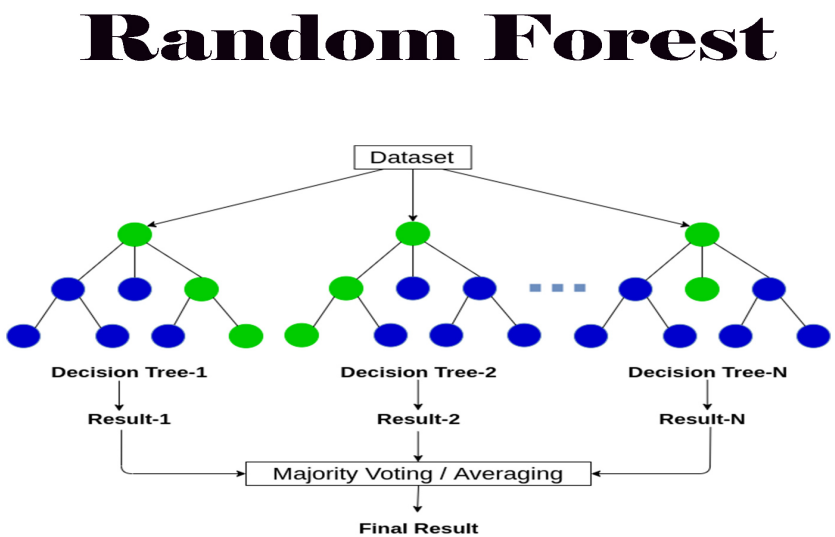

Plot of the White and Black Elo Grids
Black Elos
0
500
1500
2500
0
500
1000
1500
2000
2500
3000
White Elos

Model types

- Win and loss models
- Win, draw and loss models
- Performed to determine algorithms' performance given the absence of draws as an outcome

Random Forest mechanics

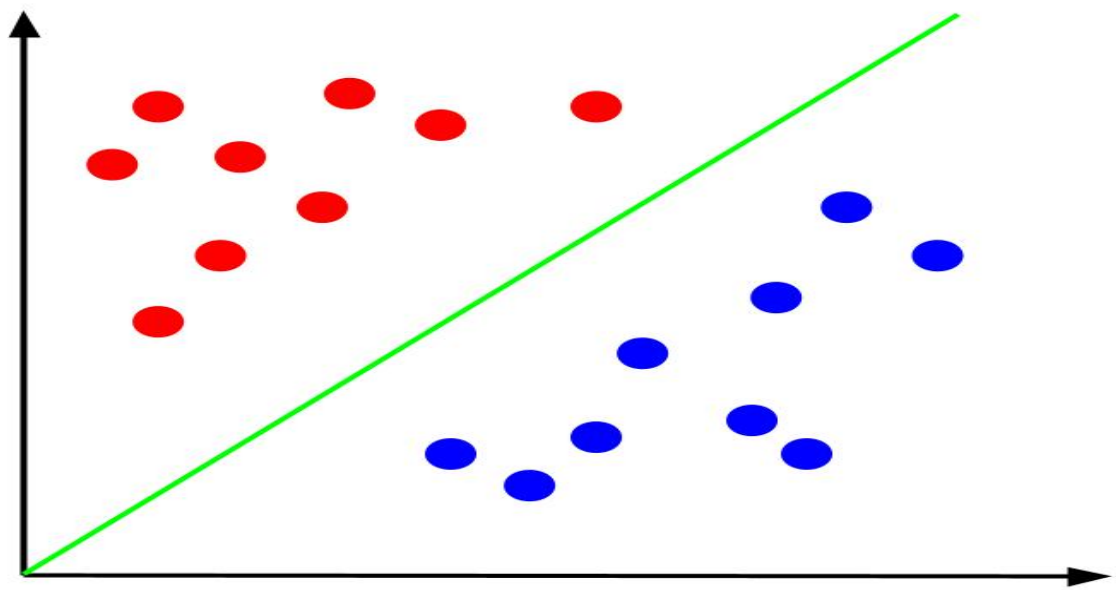
- Each tree trained using bootstrapped training data
- Random subset of predictors chosen for each tree
- Each tree produces a predicted result
- Trees' predictions used to make the model



Source:
https://anasbrital98.github.io/blog/2021/Random-Forest/

SVM mechanics

- Line created to optimally separate the data points
- Placement of the data point relative to the line decides its predicted outcome
- 3 pairs of classifiers for win, draw and loss models



Source: https://geekflare.com/support-vector-machine/

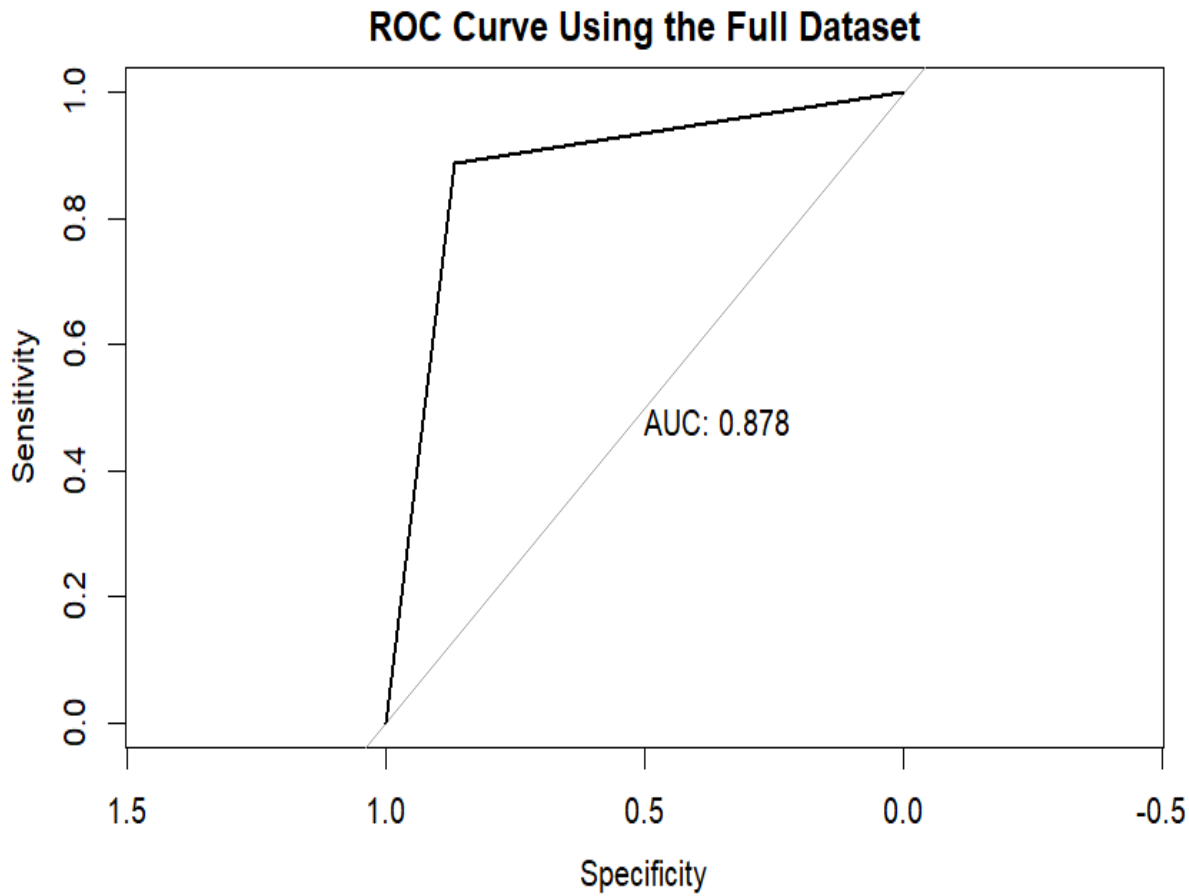
Setup of the models' implementation

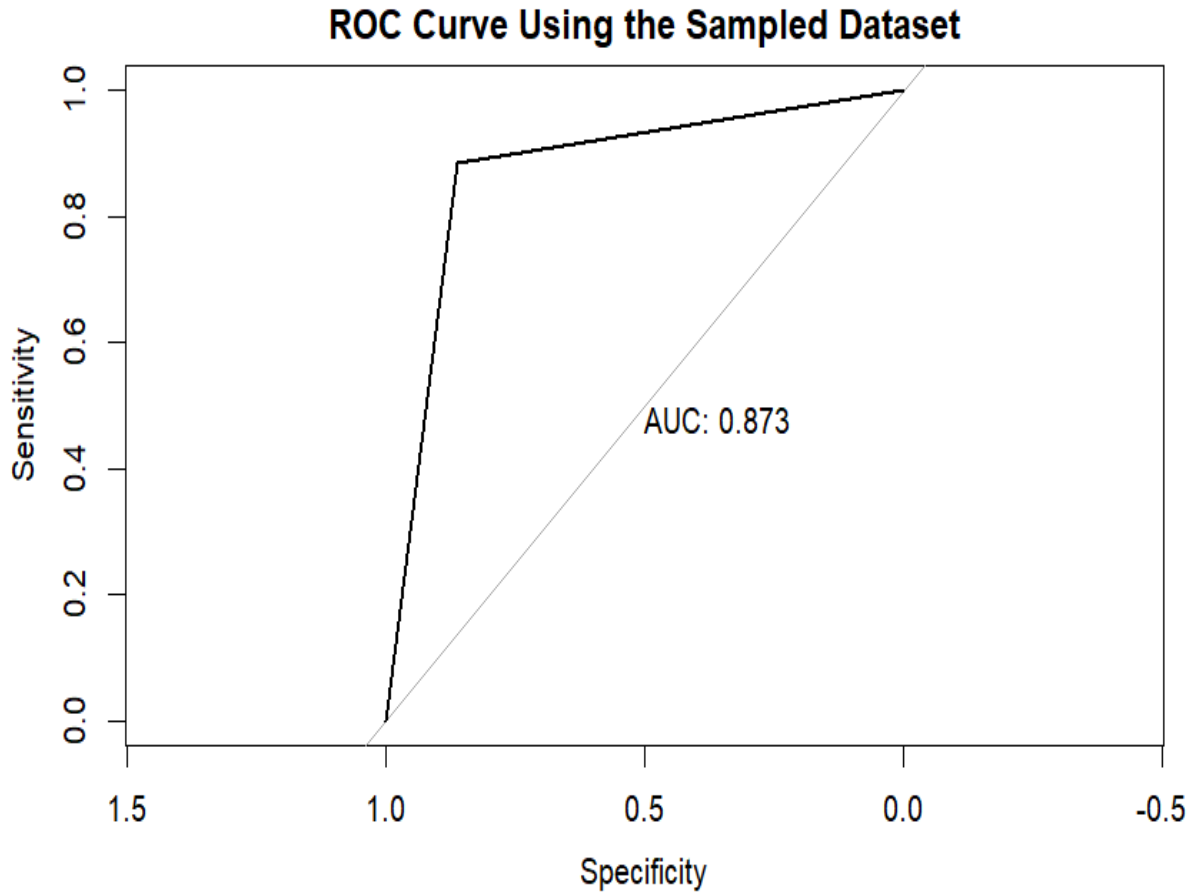
- Used seed 427 for Random Forest and SVM
- Divided the datasets into training and testing data
- Appended additional columns

Columns added for both models	Description
Delta Average Elo	Difference of the white and black average Elos
Percentage Results	Percentage of results for fen, white, and black
Delta Percentage	The percentage difference between win, lose, and draw for white and black
Columns Added for Win and Loss Only	**Description**
Delta White	Difference between wins and losses for white
Delta Black	Difference between wins and losses for black

ROC curve

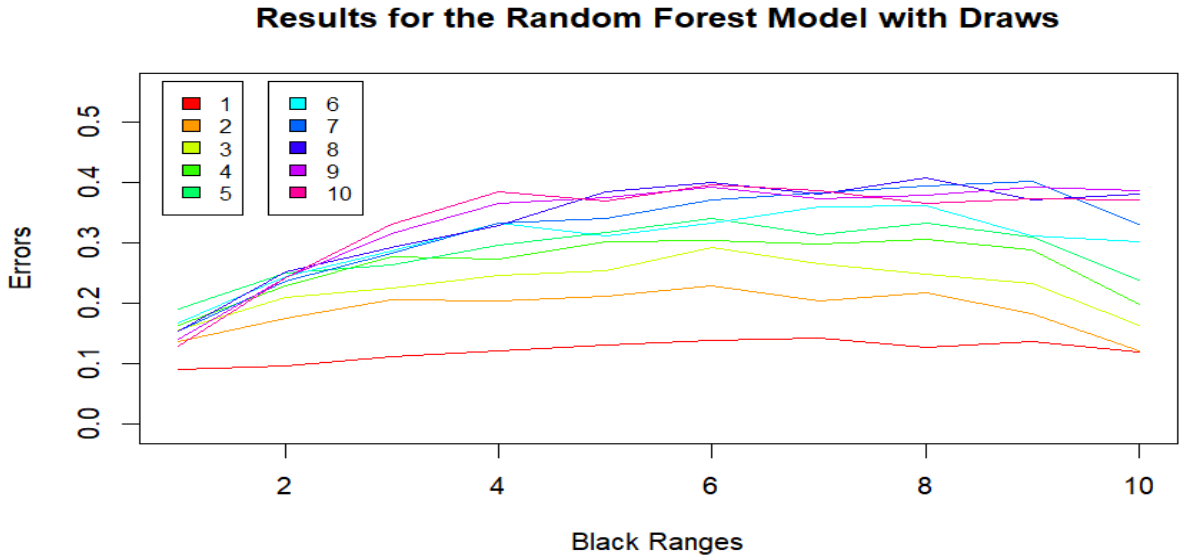
- Needed to justify Random Forest
- Values between full and sampled dataset are similar

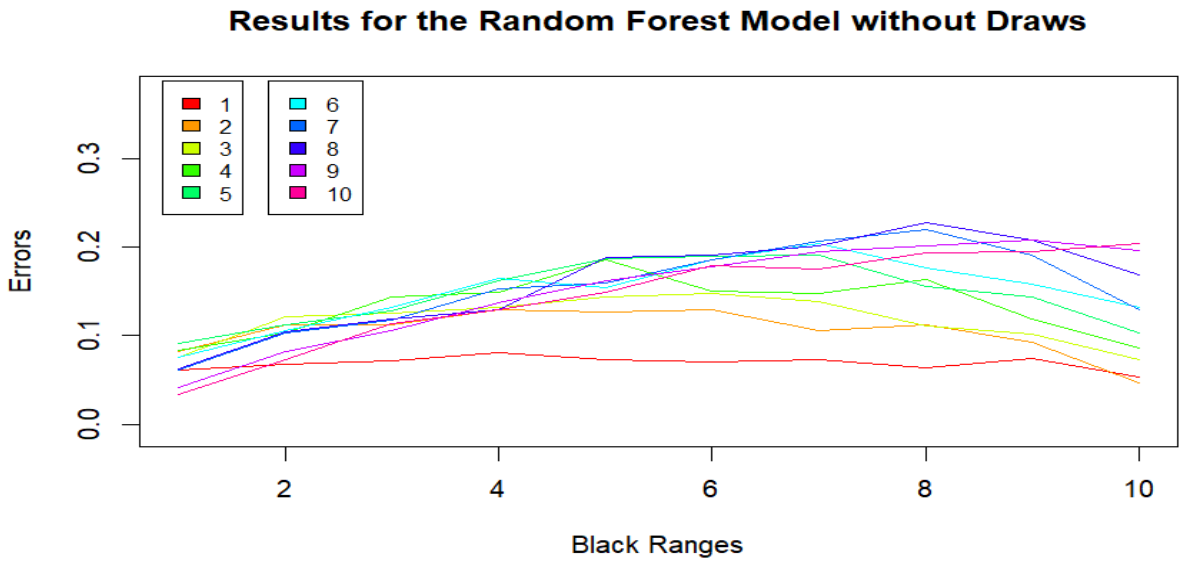


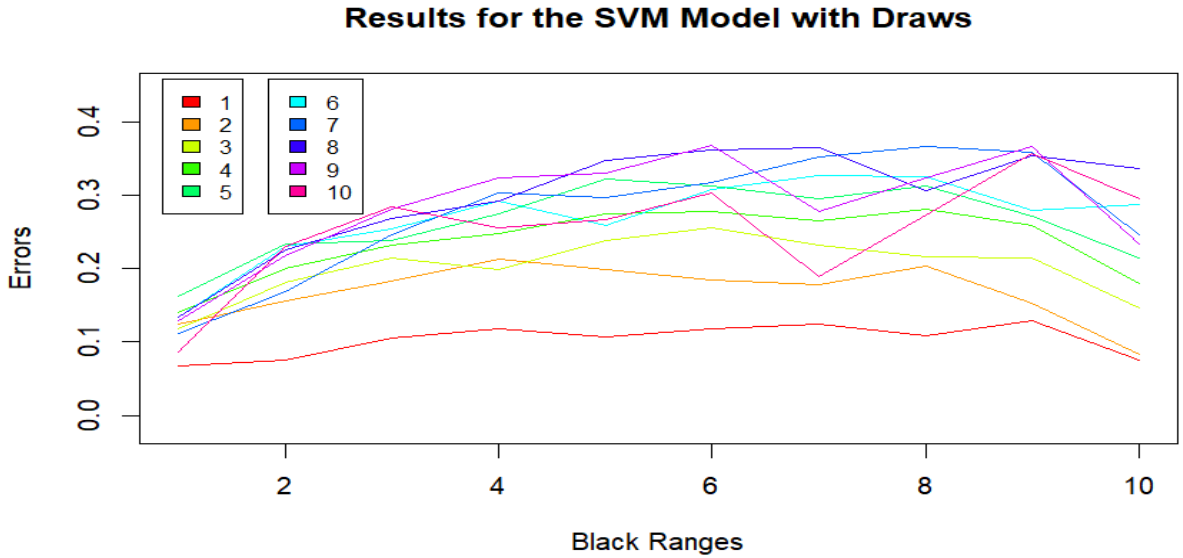

ROC Curve Using the Sampled Dataset
Sensitivity
0.0
0.2
0.4
0.6
0.8
1.0
AUC: 0.873
1.5
1.0
0.5
0.0
-0.5
Specificity

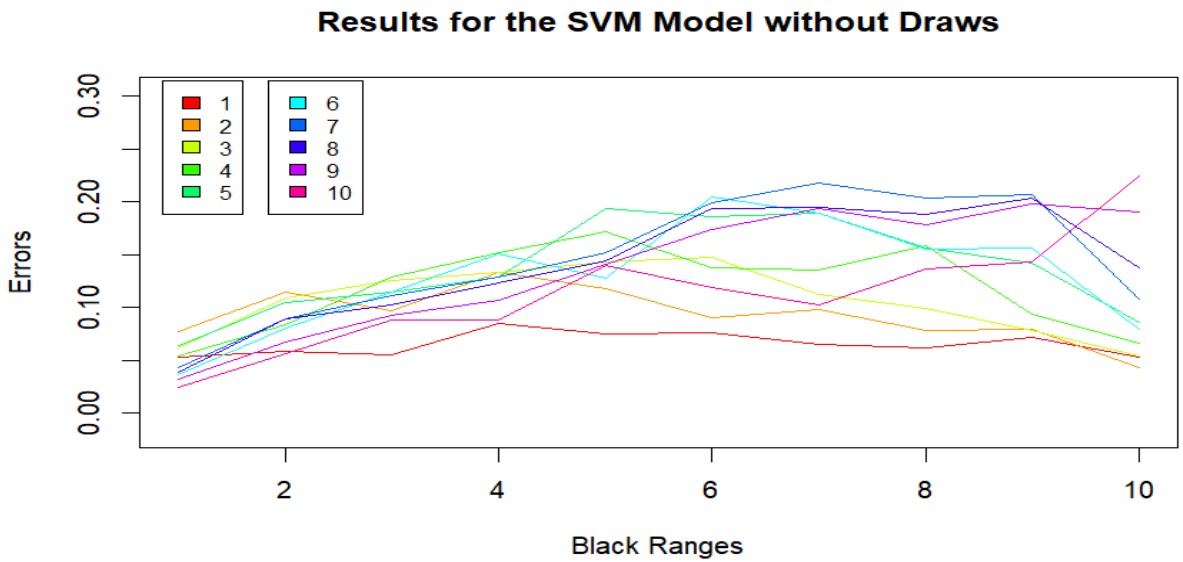
Results

- Plots' colours are the errors for each white grid across the 10 black subgrids
- Error is the proportion of time that a model's prediction for a game is incorrect
- Sample rows of the tables displayed


Results for the Random Forest Model with Draws
Errors
0.0
0.1
0.2
0.3
0.4
0.5
1
2
3
4
5
6
7
8
9
10
2
4
6
8
10
Black Ranges


Results for the Random Forest Model without Draws
Errors
0.0
0.1
0.2
0.3
1
2
3
4
5
6
7
8
9
10
2
4
6
8
10
Black Ranges


Results for the SVM Model with Draws
Errors
0.0
0.1
0.2
0.3
0.4
1
2
3
4
5
6
7
8
9
10
2
4
6
8
10
Black Ranges


Results for the SVM Model without Draws
Errors
0.00
0.10
0.20
0.30
1
2
3
4
5
6
7
8
9
10
2
4
6
8
10
Black Ranges

Grid	errors	% Black Correct	% Draw Correct	% White Correct	Tree Used
2	0.10	89.52	92.31	87.46	63
3	0.11	86.48	90.09	84.51	67
33	0.28	71.90	67.61	73.04	200
41	0.19	89.29	92.91	78.75	96
61	0.15	100.00	93.94	82.96	98
70	0.33	67.76	67.53	100.00	222
75	0.39	58.00	53.92	62.57	116
85	0.38	72.34	55.86	63.37	117

Random Forest model results with draws

Grid	errors	% Black Correct	% White Correct	Tree Used
2	0.07	93.48	92.42	84
3	0.07	91.24	92.92	100
33	0.14	91.41	84.20	148
41	0.09	75.61	91.70	77
61	0.06	100.00	93.66	58
70	0.13	84.58	70.00	87
75	0.19	81.82	85.75	252
85	0.16	80.33	85.57	207

Random Forest model results without draws

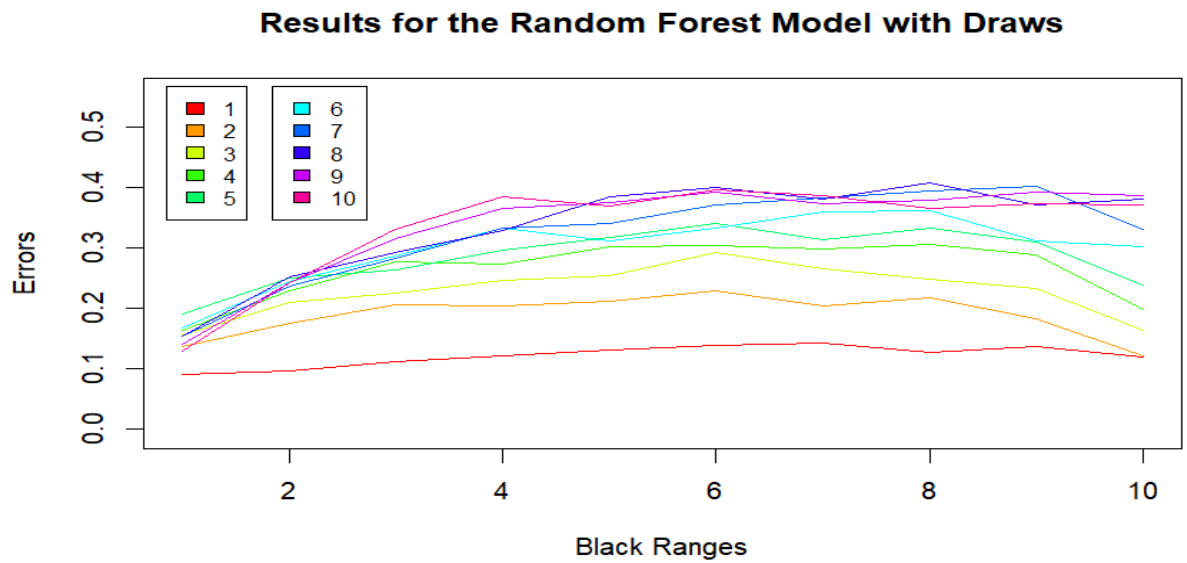
Grid	errors	% Black Correct	% Draw Correct	% White Correct	Cost Used
2	0.08	85.31	90.79	87.70	10
3	0.11	84.40	88.84	82.67	1
33	0.23	73.53	68.60	72.43	1
41	0.16	75.00	91.03	78.47	1
61	0.11	62.50	84.00	86.53	10
70	0.25	69.33	64.10	36.73	10
75	0.35	61.33	56.44	61.89	1
85	0.33	60.87	59.48	63.67	1

SVM model results with draws

Grid	errors	% Black Correct	% White Correct	Cost Used
2	0.06	89.16	94.96	10
3	0.06	90.73	88.05	10
33	0.13	83.76	85.19	1
41	0.06	81.40	93.46	10
61	0.04	89.47	93.76	1
70	0.11	89.20	66.67	10
75	0.14	69.35	80.00	10
85	0.14	80.30	83.37	1

SVM model results without draws

- Errors generally increase as the black grid number increases
- First white grid remains stationary for every model
- Initial three white grid ranges contain the least errors, and final two consist of the most
- Models without draws possess less error than models with draws
- Random Forest and SVM performed similarly



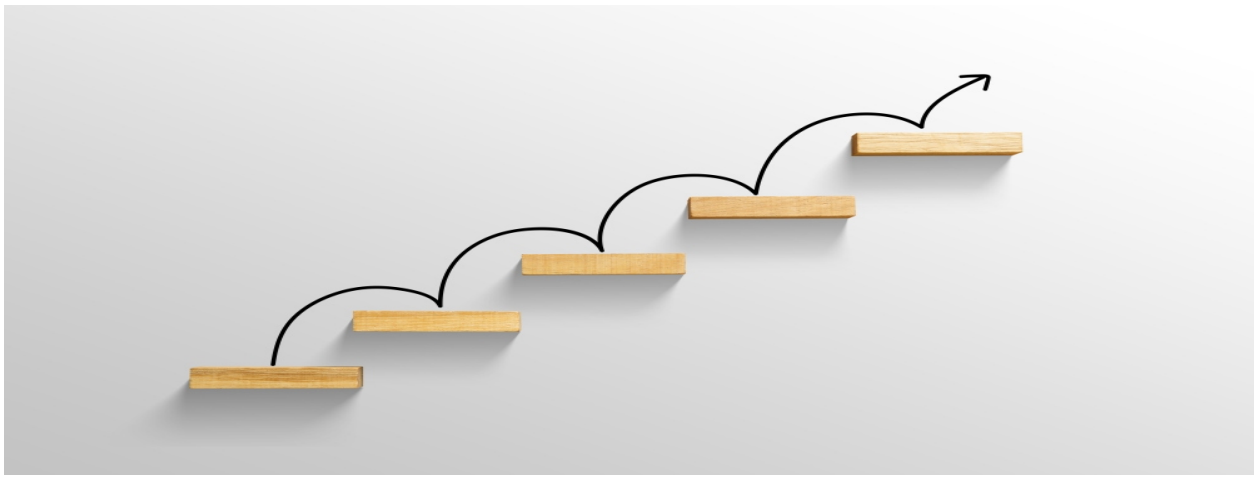
Explanations

- sum_eval likely most useful for predictions in the first white grid
- Change in the importance of predictor variables
- Varying nature of skill levels
- Absence of draws may simplify the algorithms' classifications

Source: https://www.facultyfocus.com/articles/teaching-and-learning/five-strategies-for-mastering-the-art-of-answering-questions-when-teaching-and-presenting/

Improvements

- A larger and more reliable dataset
- Use of other machine learning algorithms
- Use of chess engines



Source: https://www.teachermagazine.com/au_en/articles/change-leading-to-improvement

References

Pulido, Héctor, "Predicting the Outcome of a Chess Game by Statistical and Machine Learning Techniques".
https://www.obviously.ai/post/data-cleaning-in-machine-learning
https://www.dailymail.co.uk/news/article-8830323/Chess-grandmaster-Igors-Rausis-caught-cheating-phone-loo-banned-competing-again.html
https://anasbrital98.github.io/blog/2021/Random-Forest/
https://geekflare.com/support-vector-machine/
https://www.facultyfocus.com/articles/teaching-and-learning/five-strategies-for-mastering-the-art-of-answering-questions-when-teaching-and-presenting/
https://www.teachermagazine.com/au_en/articles/change-leading-to-improvement