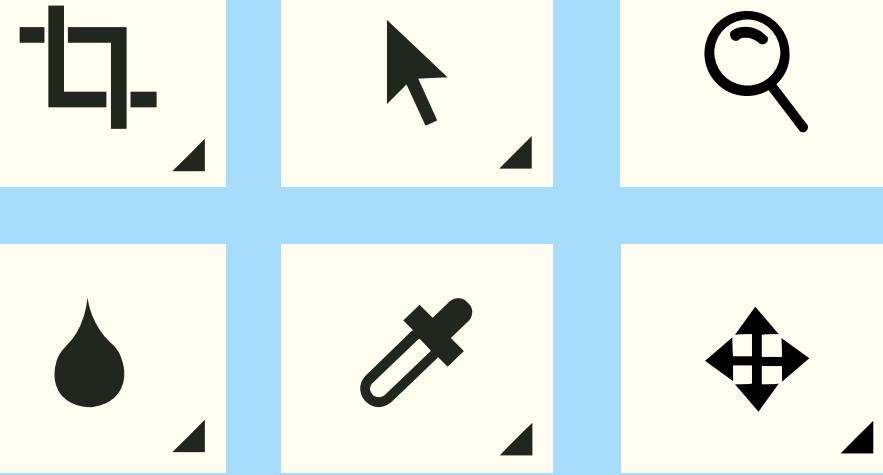


**LET'S
PLAY
CHESS**

BUT FAIRLY



THE PRO- BLEM



CHESS-PLAYING MACHINES . . .



1996 - Deep Blue - IBM



2004 - Stockfish



2017 - AlphaZero

and

. . . CHESS-PLAYER CHEATERS



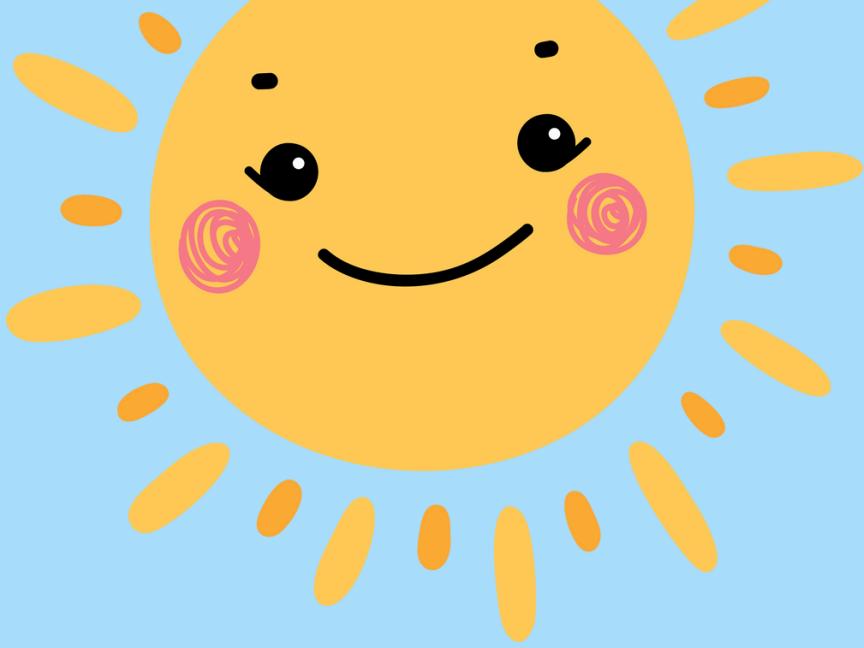
ON BOARD



V S

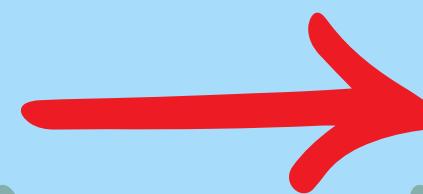
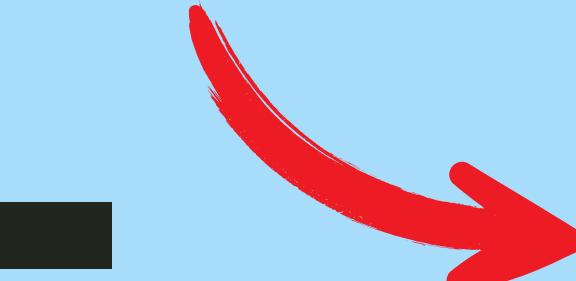


OUR SOLU- TION



WE CREATED A RANDOM FOREST MODEL CAPABLE OF ANALYSING A CHESS GAME AND RECOGNISING POTENTIAL CHEATERS... AND AN APP TO USE THE MODEL.

THIS IS HOW
WE DID IT



of course
MACHINE LEARNING!



FIRST OF
ALL WE HAD
A PROBLEM:

WE COULDN'T
FIND DATASET
OF CHEATERS-
LABELLED
CHESS GAMES

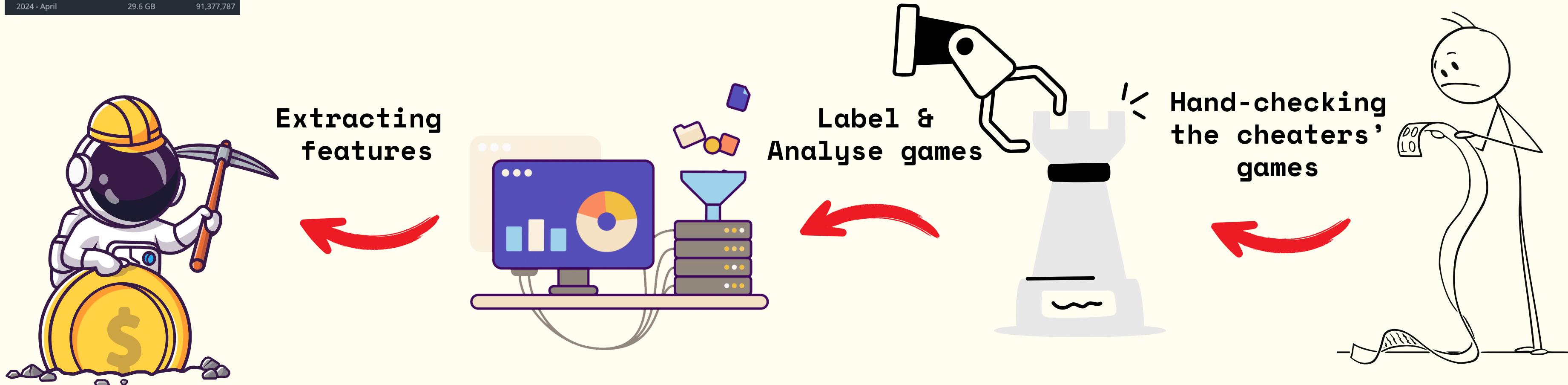
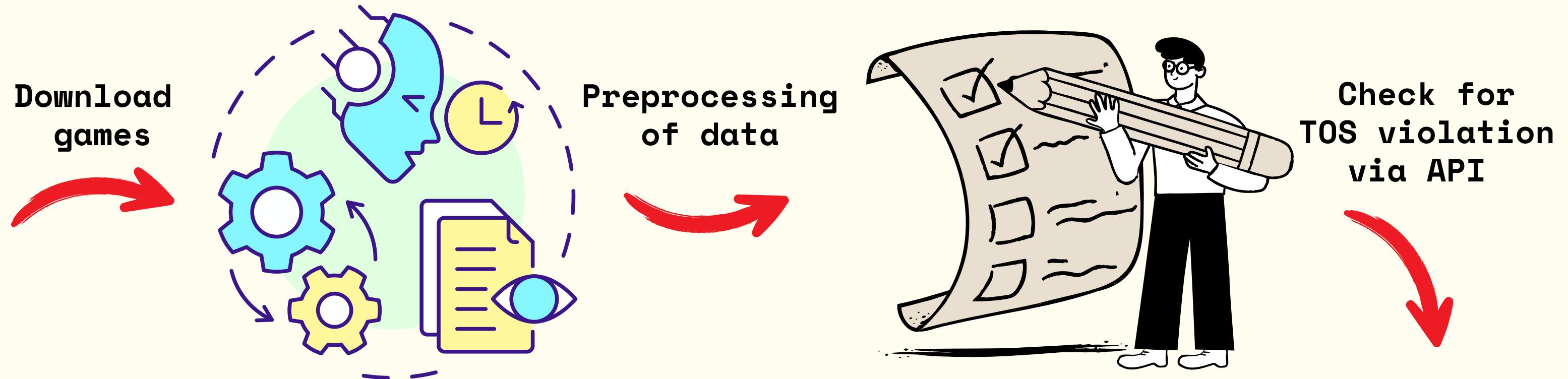
SO WE NEEDED TO
CREATE IT

BUT HOW?

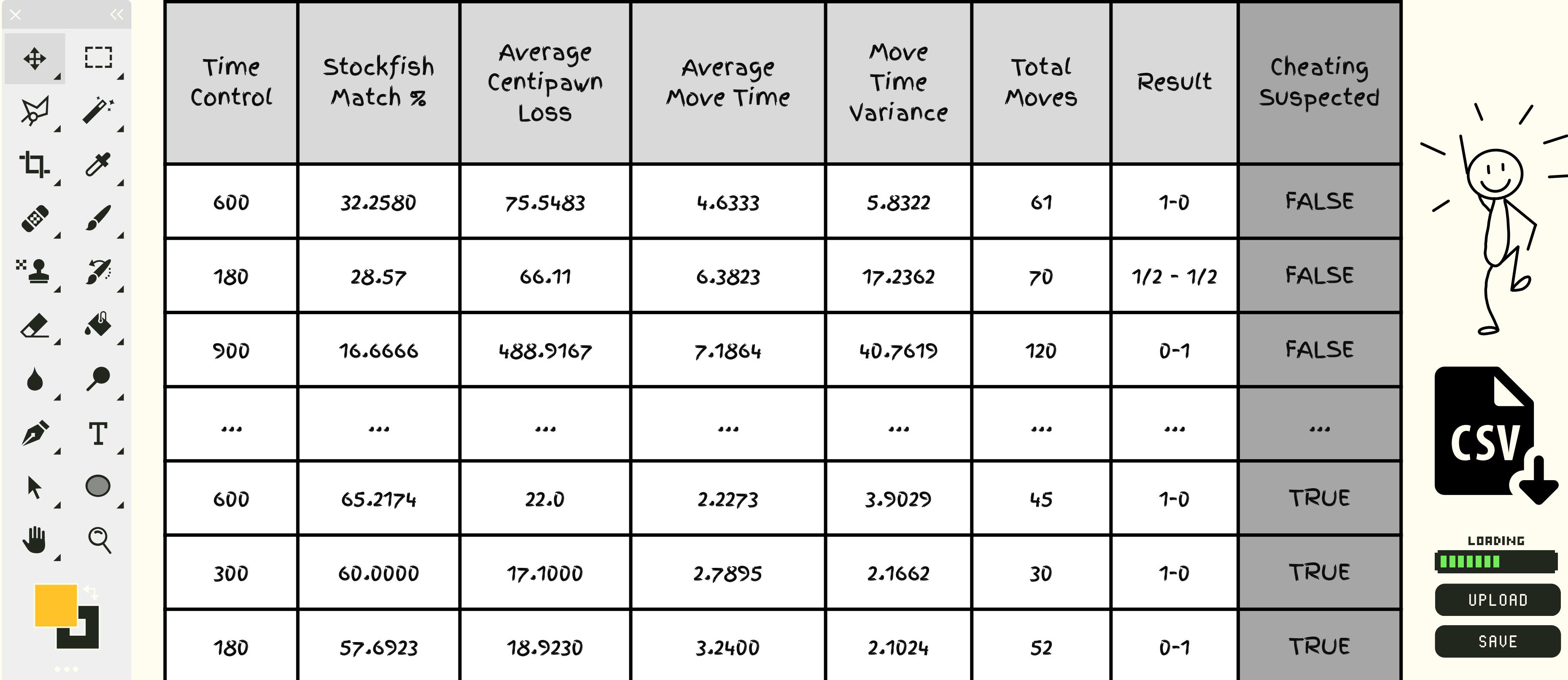


HOW WE MADE UP THE DATABASE:

Month	Size	Games
2025 - April	29.8 GB	91,757,350
2025 - March	31.8 GB	97,512,351
2025 - February	29.2 GB	89,430,612
2025 - January	32.9 GB	100,412,379
2024 - December	31.6 GB	96,587,411
2024 - November	29.6 GB	90,847,982
2024 - October	30.7 GB	94,254,891
2024 - September	28.6 GB	87,713,219
2024 - August	30 GB	92,198,878
2024 - July	29.3 GB	90,106,180
2024 - June	29.1 GB	89,342,529
2024 - May	30.7 GB	94,400,051
2024 - April	29.6 GB	91,377,787



THE FEATURES WE USED:



The screenshot shows a software interface with a toolbar on the left containing various icons for file operations like opening, saving, and filtering. The main area displays a table with the following data:

Time Control	Stockfish Match %	Average Centipawn Loss	Average Move Time	Move Time Variance	Total Moves	Result	Cheating Suspected
600	32.2580	75.5483	4.6333	5.8322	61	1-0	FALSE
180	28.57	66.11	6.3823	17.2362	70	1/2 - 1/2	FALSE
900	16.6666	488.9167	7.1864	40.7619	120	0-1	FALSE
...
600	65.2174	22.0	2.2273	3.9029	45	1-0	TRUE
300	60.0000	17.1000	2.7895	2.1662	30	1-0	TRUE
180	57.6923	18.9230	3.2400	2.1024	52	0-1	TRUE

On the right side of the table, there is a CSV file icon with a download arrow, and below it are three buttons labeled "LOADING", "UPLOAD", and "SAVE".

FINAL LY WE COULD TRAIN THE MODEL

RANDOM FOREST CLASSIFIER

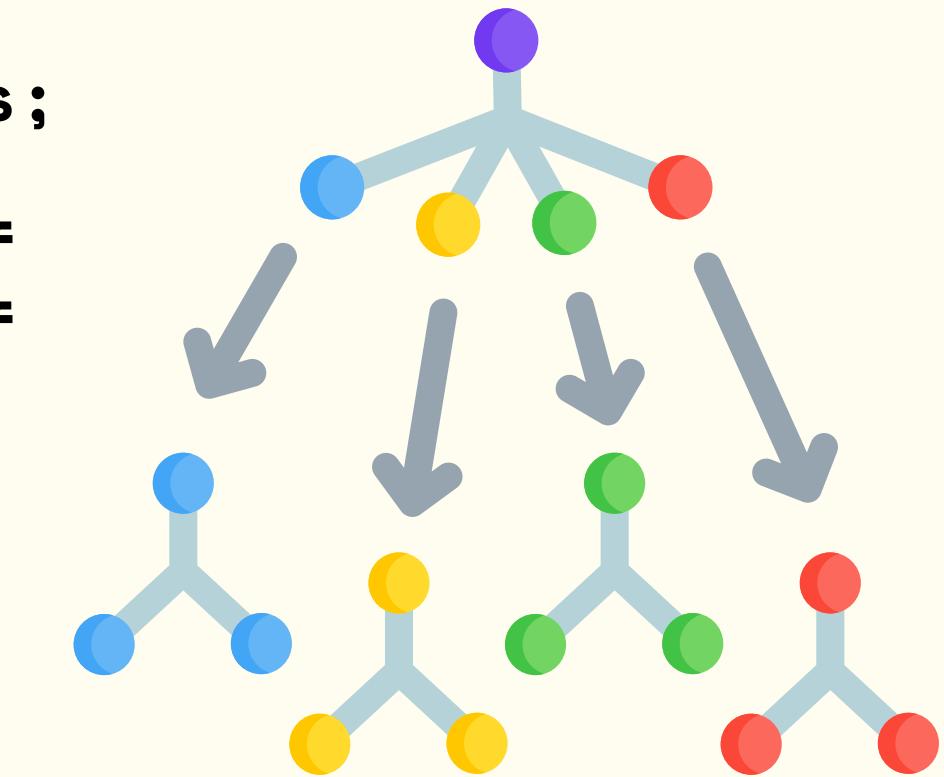
- Using a grid search function, we obtained the following best parameters:

Max Depth	Min. sample leaf	Min. sample split	N. estimators	Max. features
None	1	2	200	4

- We divided the dataset into TRAINING SET and TEST SET with a respective percentage of

70% - 30%

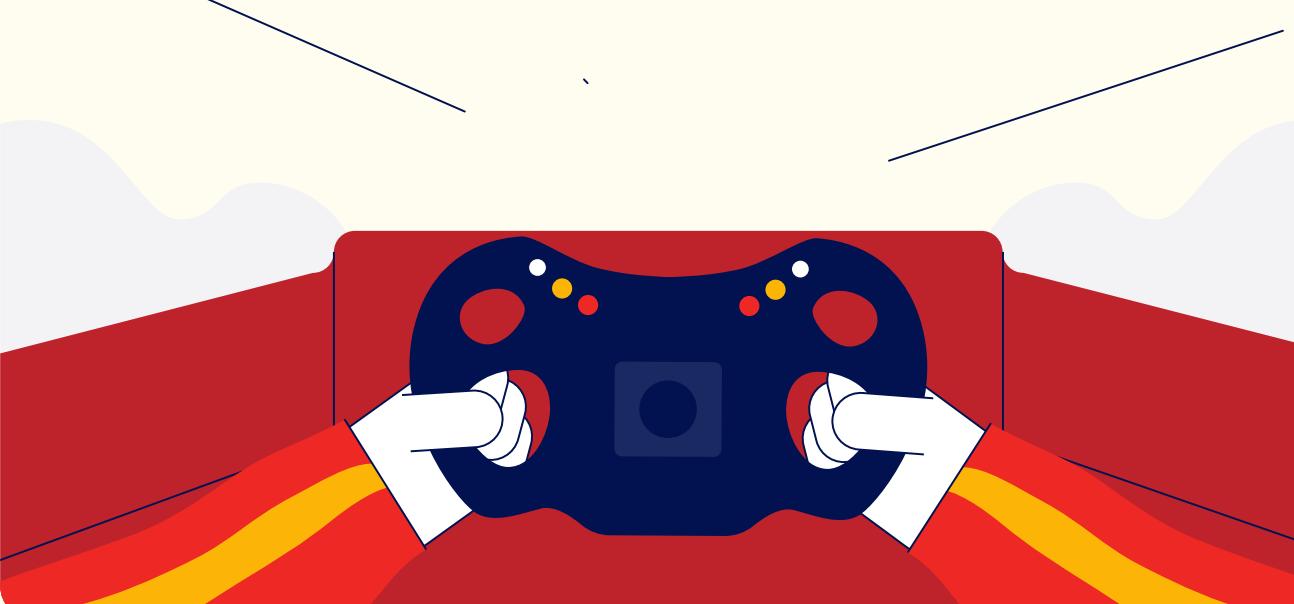
- We used One Hot Encoder for preprocessing of categorical attributes;
- We imposed a stratified distribution of statistical units given the minority of cheaters out of the total;
- WE ASKED THE MODEL TO PRODUCE:
 - CONFUSION MATRIX
 - VARIABLE IMPORTANCE
 - STATISTICS AS:
Accuracy, Precision
Recall and F1 Score



MODEL PERFORMANCE

Statistics:

ACCURACY	0.9343
PRECISION	0.9722
RECALL	0.7447
F1 SCORE	0.8434

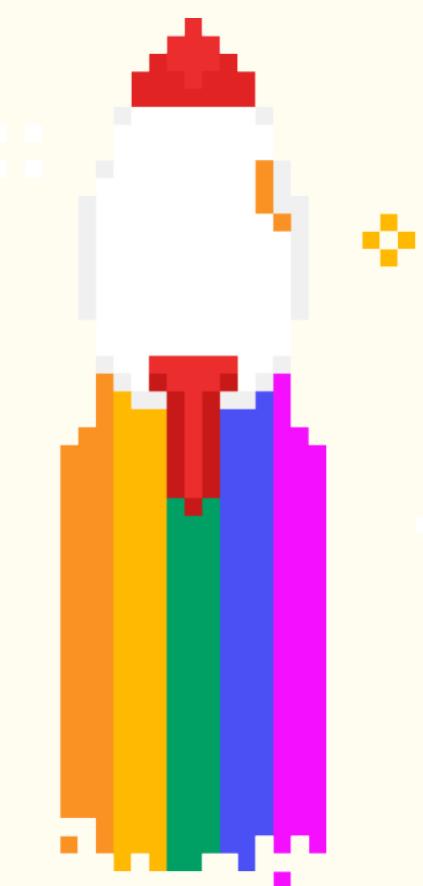
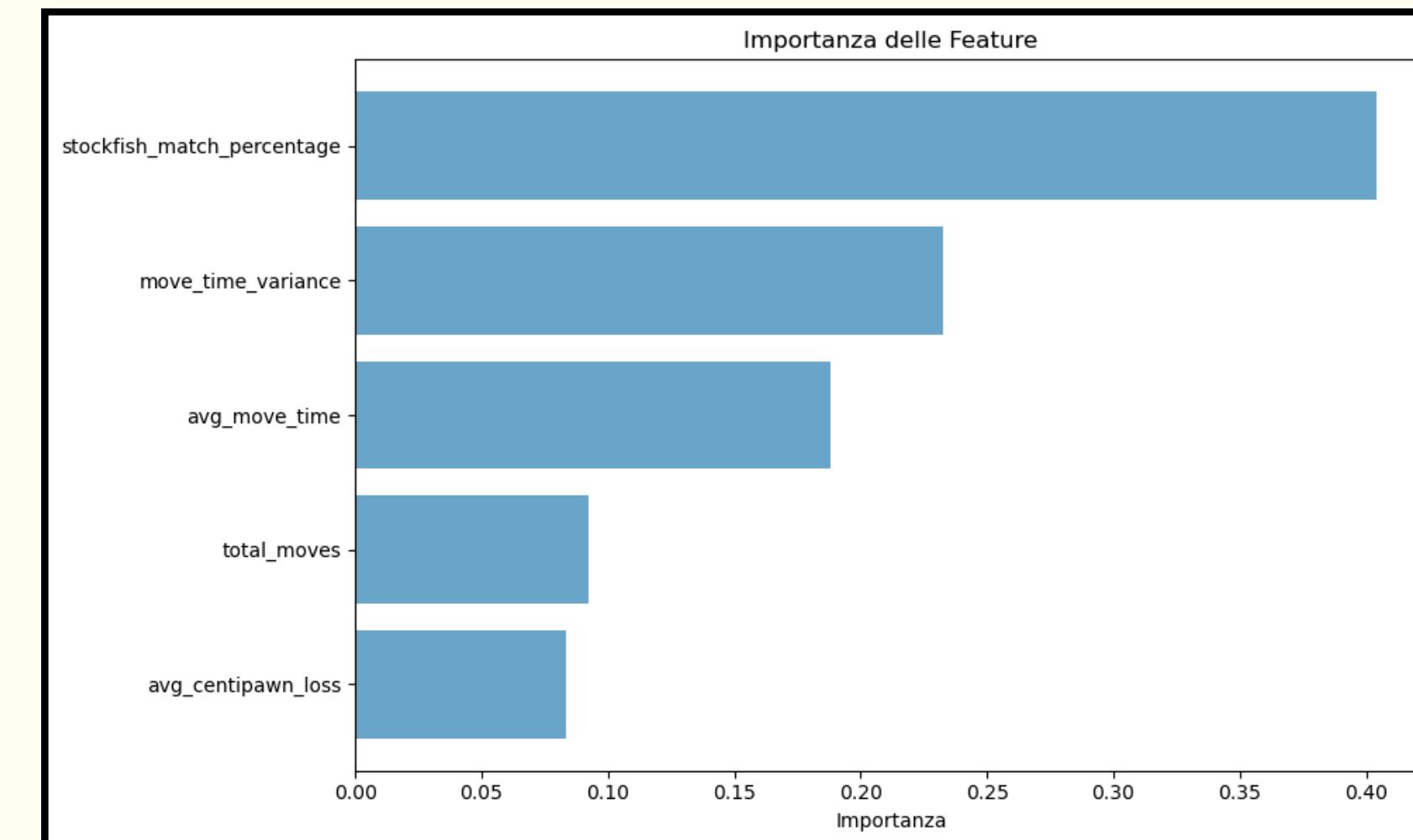


Confusion Matrix:

149	2
10	37



Variable Importance:



THE APP: the process



- 1) Search the games via players' names;**
- 2) Analyse the game with Stockfish;**
- 3) Extract the feature for the model;**
- 4) Query to the Random Forest Model;**
- 5) Output the prediction with features values;**
- 6) Show the game on the chessboard.**

THE APP



q3k2r/1R3Q1p/2p4P/5p2/8/P1p1P3/1pP2PP1/1K1R4 b k - 0 25

Analisi PGN

Analisi Lichess

Inserisci i nomi dei giocatori per analizzare l'ultima partita giocata su Lichess. L'analisi verrà effettuata sul tuo avversario per verificare eventuali segni di cheating.

Il tuo username Lichess:

miziesamle

Username dell'avversario da analizzare:

BernardoBusoni

Analizza Partita

Risultato dell'analisi:

✓ Nessun segno di cheating rilevato

Probabilità di cheating: 1.00%

Feature estratte:

avg_centipawn_loss: 442.08

avg_move_time: 7.61

move_time_variance: 57.02

stockfish_match_percentage: 29.17

time_control: 180.00

total_moves: 24.00



8/6k1/8/r7/8/1p3P2/p3K1P1/R7 w - - 0 53

Analisi PGN

Analisi Lichess

Inserisci i nomi dei giocatori per analizzare l'ultima partita giocata su Lichess. L'analisi verrà effettuata sul tuo avversario per verificare eventuali segni di cheating.

Il tuo username Lichess:

Kyrgyzstan001

Username dell'avversario da analizzare:

VomeroCorason

Analizza Partita

Risultato dell'analisi:

⚠️ Rilevati possibili segni di cheating!

Probabilità di cheating: 53.29%

Feature estratte:

avg_centipawn_loss: 5.75

avg_move_time: 3.18

move_time_variance: 9.20

stockfish_match_percentage: 59.62

time_control: 300.00

total_moves: 52.00

○ VomeroCorason

0

0

0

Tournament Points Studies Forum Posts



This account violated the Lichess Terms of Service

Member since May 22, 2025

Active 24 hours ago

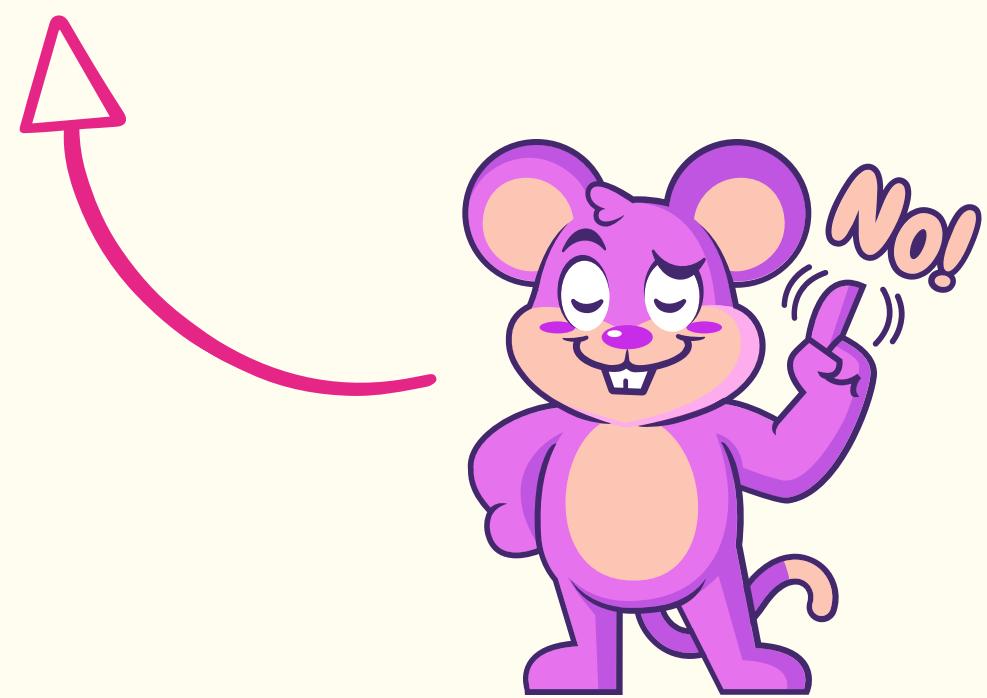
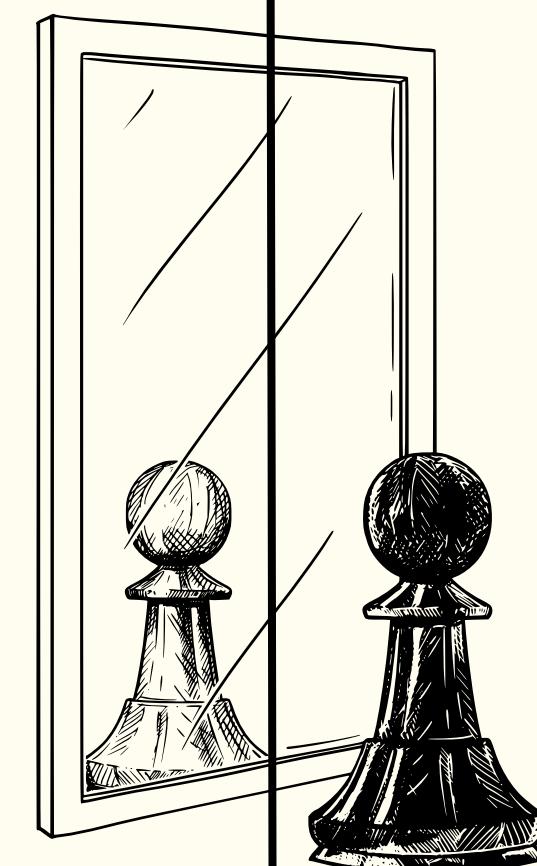
Time spent playing: 54 minutes



mò lo
frego

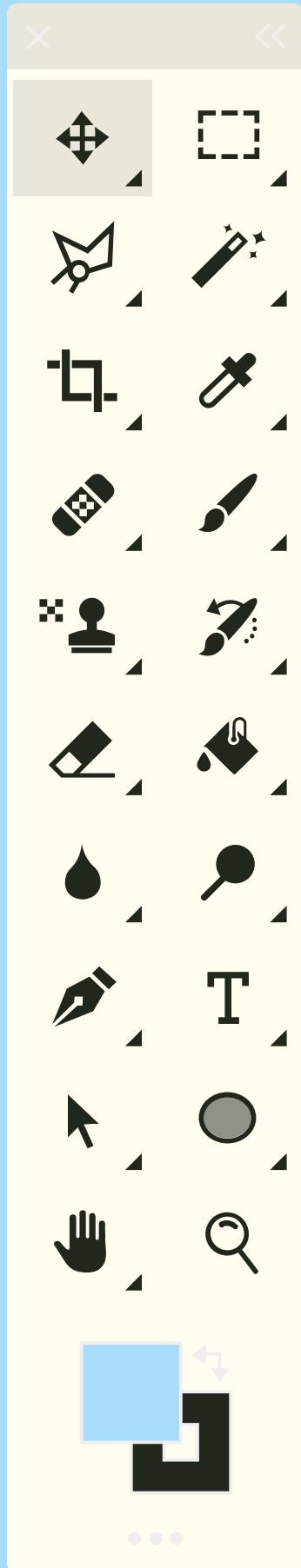
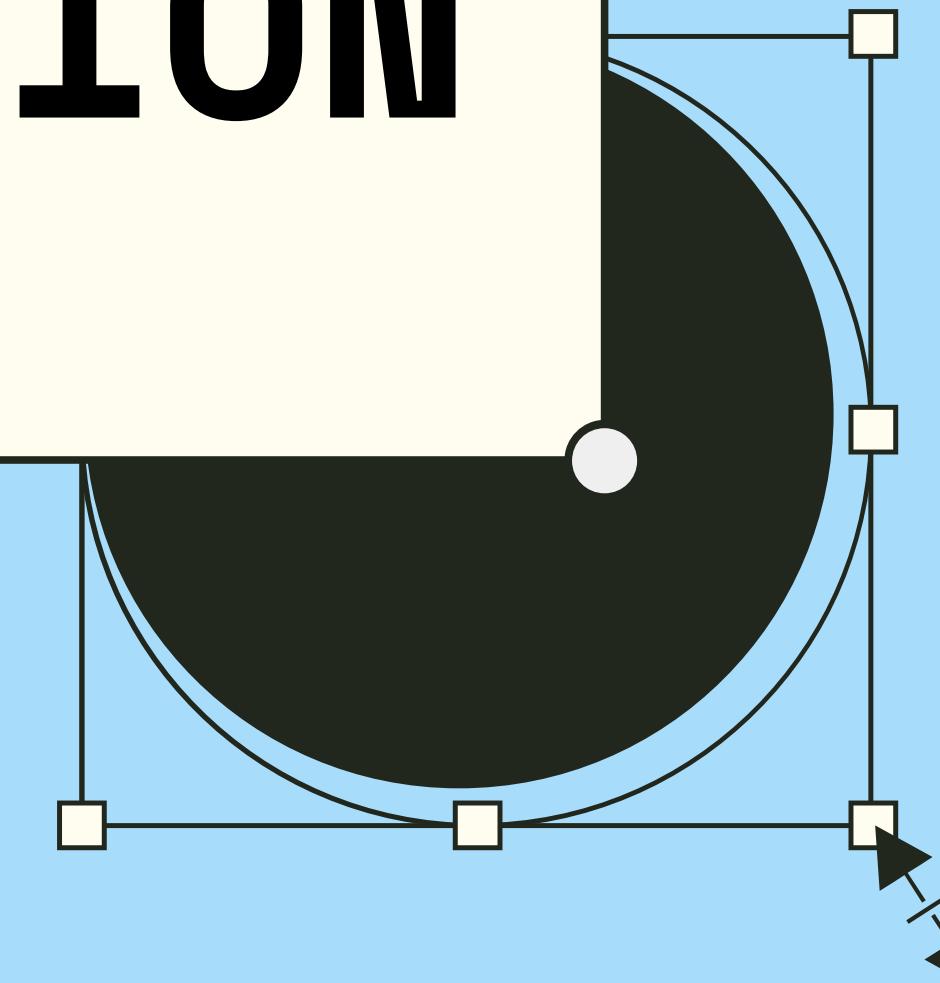


TAKEAWAY MESSAGE



**THANKS FOR
YOUR ATTENTION**

SAVE



REFERENCES



- An Introduction to Statistical Learning 2ed.
- StatQuest: Random Forests
- Stockfish
- Lichess open database
- Cursor
- Chess.com - Articles

