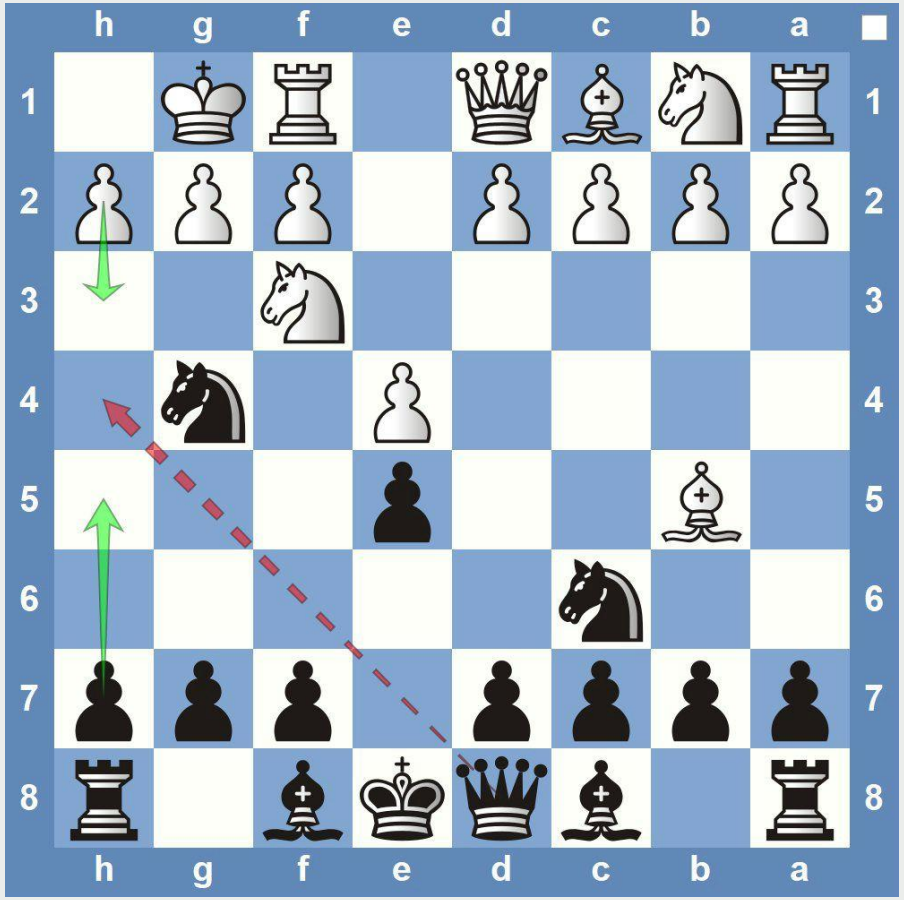
MACHINE LEARNING & CHESS

Diego Izaguirre

WHY CHESS?


h g f e d c b a
1 2 3 4 5 6 7 8

Overview

- Data
- Preprocessing
- Feature selection
- Results

DATA

TARGET

Resignation
Checkmate

FEATURES

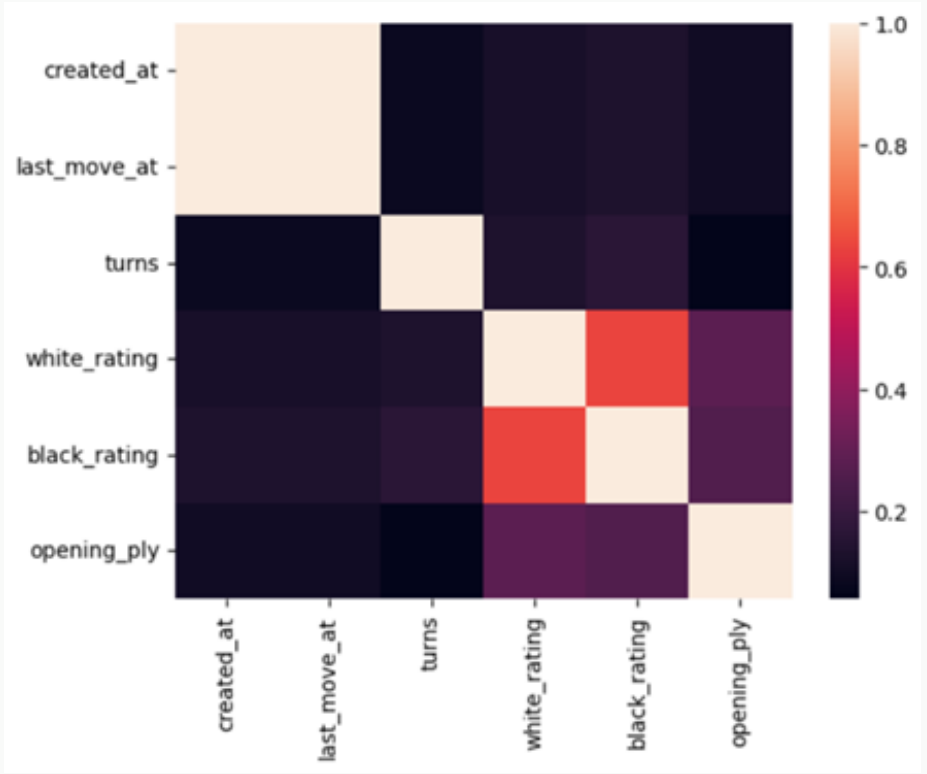
Game ID	Rated	Start Time	End Time	# of turns
Winner	Time Incr.	White ID	White Rating	Black ID
Black Rating	Moves	Opening Eco	Opening Name	Opening Ply

Important Features

- Start Time – End Time
- Moves
- Opening Eco

START & END TIME

- Extremely collinear
- Combined into one feature



OPENING ECO

- Ex) D10
- Letter identifier denotes type
- Binned based on type

A	Flank Opening
B	Semi Open
C	Open / French Defense
D	Closed / Semi Closed
E	Indian Defense

MOVES

- Ex) e4 e5 d3 d6 Be3 c6 Be2…

Q	Queen
K	King
N	Knight
B	Bishop
R	Rook

OTHER PREPROCESSING

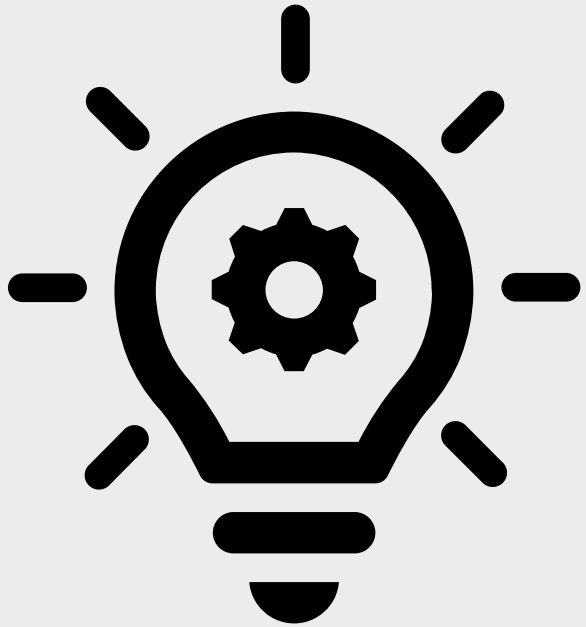
- Min-Max Scaling
- Dummy variables
- Class rebalancing

Resignation	56%
Checkmate	32%
Draw & Out of Time	13%

Feature Selection

- Wrapper via model
- Univariate via mutual info classification

SELECTION TYPE	SELECTED FEATURES
RANDOM FOREST	['turns', 'white_rating', 'black_rating', 'queen_moves', 'opening_ply', 'knight_moves', 'king_moves_moves', 'bishop_moves', 'rook_moves', 'game_time']
SVM	['turns', 'white_rating', 'queen_moves', 'opening_eco_Indian Defense']
GRADIENT BOOSTING	['turns', 'white_rating', 'black_rating', 'queen_moves', 'opening_ply', 'knight_moves', 'king_moves_moves', 'bishop_moves', 'rook_moves', 'game_time']
UNIVARIATE	['rated', 'turns', 'white_rating', 'black_rating', 'queen_moves', 'knight_moves', 'king_moves_moves', 'bishop_moves', 'rook_moves', 'game_time']



Grid Search

- Optimized parameters on all the models
- Ran using typical 5-fold CV

Methods

5-fold cross validation

3 runs of each model (full fit, wrapper, univariate)

Utilizing optimized parameters.

Full Fit

	Accuracy	AUC
RF	0.70	0.74
SVM	0.66	0.69
Grad Boost	0.69	0.73

Wrapper

	Accuracy	AUC
RF	0.69	0.73
SVM	0.64	0.67
Grad Boost	0.68	0.72

Univariate

	Accuracy	AUC
RF	0.69	0.73
SVM	0.65	0.68
Grad Boost	0.68	0.72

Model Results

BEST MODEL?

Limitations

- Move patterns lost
- Heavily imbalanced classes

Conclusions

SELECTED FEATURES

- Moves per piece
- Opening ply
- Time
- Ratings

DO OPENINGS MATTER?

- Maybe?
- Development > Opening Type

¡Questions!