INTRODUCTION

REYGA FERDIANSYAH PUTRA

LINKEDIN : https://www.linkedin.com/in/reyga-ferdiansyah/

MEDIUM : https://medium.com/@reygafp

GITHUB : https://github.com/reygaferdiansyah

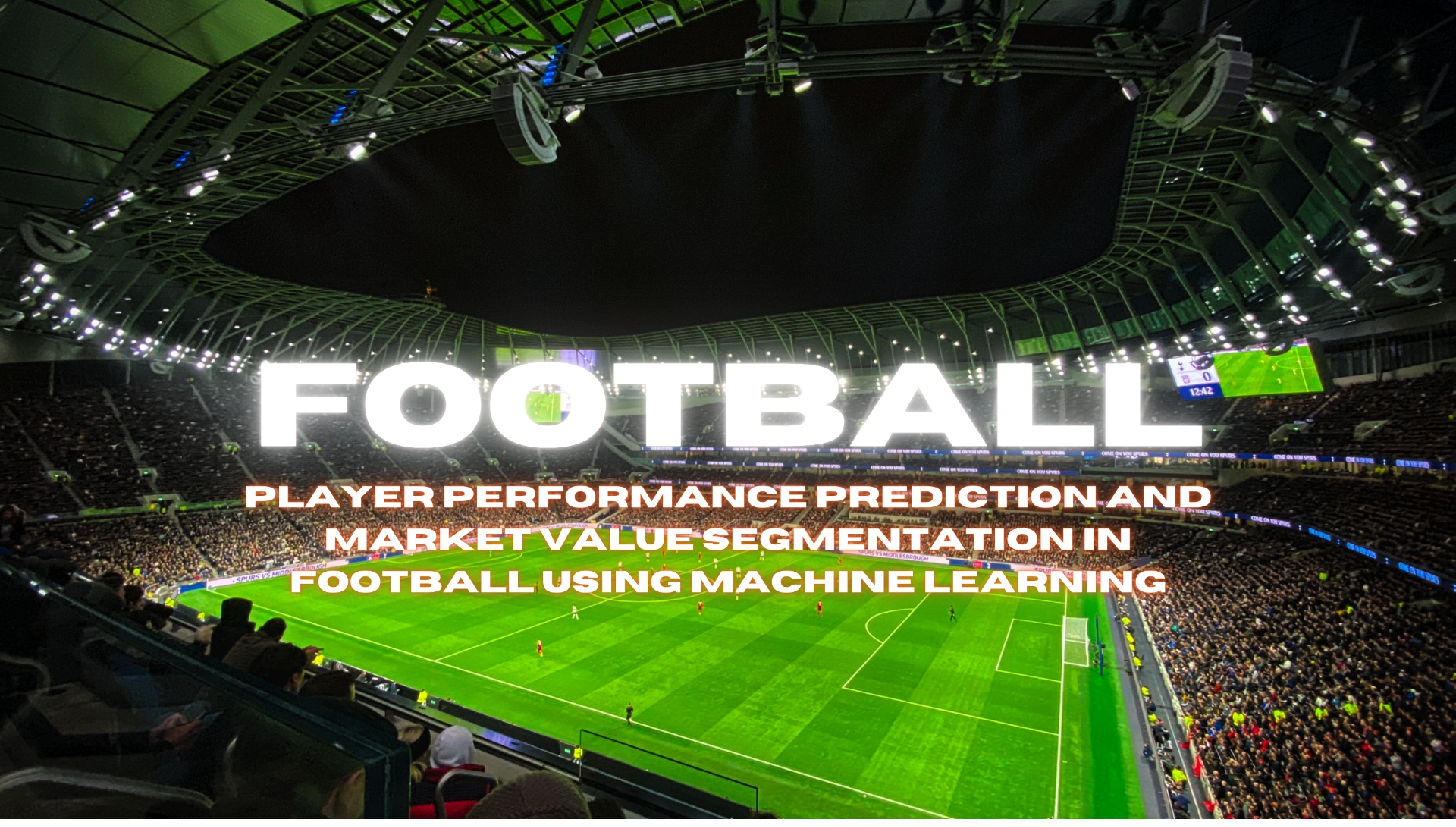
EDUCATION & SERTIFICATION

EDUCATION

NOV 2024 - NOW	BINUS UNIVERSITY
NOV 2024 - NOW	DIBIMBING
2019 - 2023	STIKOM CKI

SERTIFICATION

2024	AZURE AI ENGINEER - MICROSOFT
2024	INNOVATION CHALLENG - SKILVUL
2023	MACHINE LEARNING - DEEP LEARNING. AI


FOOTBALL
PLAYER PERFORMANCE PREDICTION AND
MARKET VALUE SEGMENTATION IN
FOOTBALL USING MACHINE LEARNING

BACKGROUND PROJECT

In modern football, data analysis plays a crucial role in evaluating player performance and market value. This project develops machine learning models to predict player performance metrics such as top scorers, assists, defensive actions, and saves based on match statistics. In addition, clustering techniques are employed to group players according to their market value, providing strategic insights to support objective and efficient player scouting and transfer decisions.

BUSINESS PROBLEM & MAIN PROJECT

Football clubs face significant challenges in objectively evaluating player performance and efficiently managing player market value. Assessments based solely on observation or basic statistics are often insufficient to support strategic decisions such as player acquisition, sales, scouting, or squad rotation.

Therefore, a machine learning–based solution is required to:

- Predict the best-performing players based on their position and key performance indicators (goals, assists, defensive actions, saves).
- Cluster players by market value to provide club management with a strategic overview of player investment potential.

DATA UNDERSTANDING

This dataset contains information on 1,748 football players from top European leagues such as La Liga, Premier League, and Serie A. Each row represents an individual player and includes 27 metrics covering performance statistics, market value, and profile metadata. This dataset reflects player statistics and market data collected during the 2024–2025 football season

Data Sources and Features:

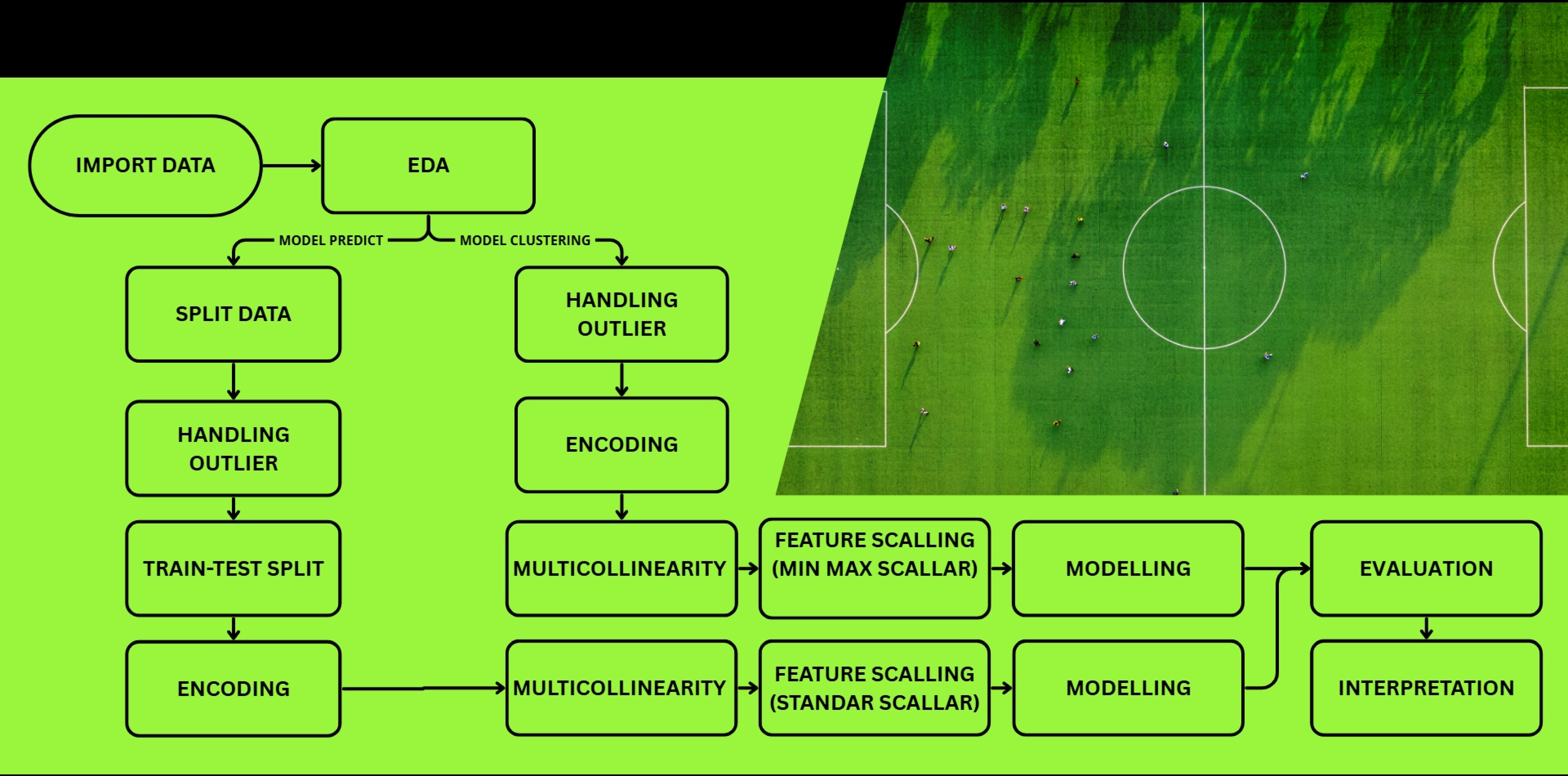
From SofaScore (Match Statistics):

- Offensive Performance: Goals, Expected Goals (xG), Big Chances Missed, Big Chances Created, Total Shots, Goal Conversion %.
- Playmaking Contribution: Assists, Key Passes, Accurate Passes %, Accurate Passes, Successful Dribbles.
- Defensive Performance: Tackles, Interceptions, Clearances, Errors Leading to Goal.
- Goalkeeper Statistics: Saves, Clean Sheets, Penalties Saved, Saves from Inside the Box, Runs Out.
- Overall Performance: Average SofaScore.

From Transfermarkt (Profile & Market Data):

- League, Team, Player Name, Age, Position, Market Value.

FOOTBALL PLAYER PERFORMANCE PREDICTION & MARKET VALUE CLUSTERING WORKFLOW


IMPORT DATA
EDA
MODEL PREDICT
MODEL CLUSTERING
SPLIT DATA
HANDLING OUTLIER
TRAIN-TEST SPLIT
ENCODING
HANDLING OUTLIER
ENCODING
MULTICOLLINEARITY
FEATURE SCALLING (MIN MAX SCALLAR)
MODELLING
MULTICOLLINEARITY
FEATURE SCALLING (STANDAR SCALLAR)
MODELLING
EVALUATION
INTERPRETATION

DATA PREPROCESSING

Import Libraries

Handling missing values and removing irrelevant columns.

Load Dataset (FIFA.csv)

Market value conversion, player position encoding, and key feature extraction.

Check structure and missing values.

dtypes, shape, isna

Handling Missing & Cleaning

Convert 'Market value' and drop the original column.

Feature Engineering

- Goal_Contribution
- Offensive_Impact
- Defensive_Efficiency
- xG_Efficiency (Handling Inf/NaN)

Standardize the position column.

Map positions into 11 categories and correct manual entries.

Correct age data.

Update age = 0 to the correct value.

SPLIT DATA

TOP SCORE

Drop Columns:
- Goals (target)
- Name (Identity features are not needed by the model)
- Goalkeeper statistics

TOP ASSISTS

Drop Columns:
- Assists (target)
- Name
- Goalkeeper statistics

TOP DEFENSIVE

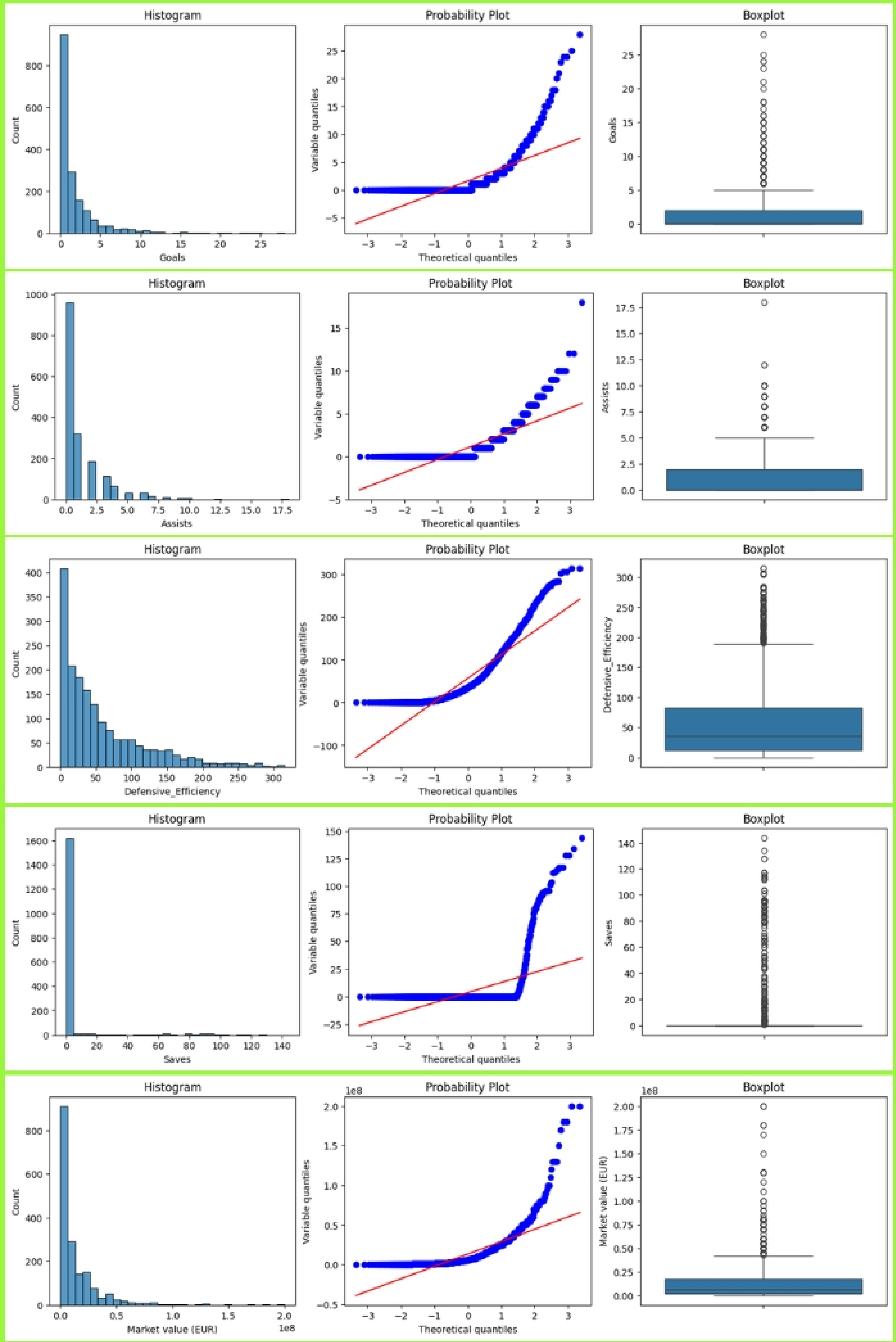
Drop Columns:
- Defensive_Efficiency (target)
- Name
- Goalkeeper statistics

TOP SAVES

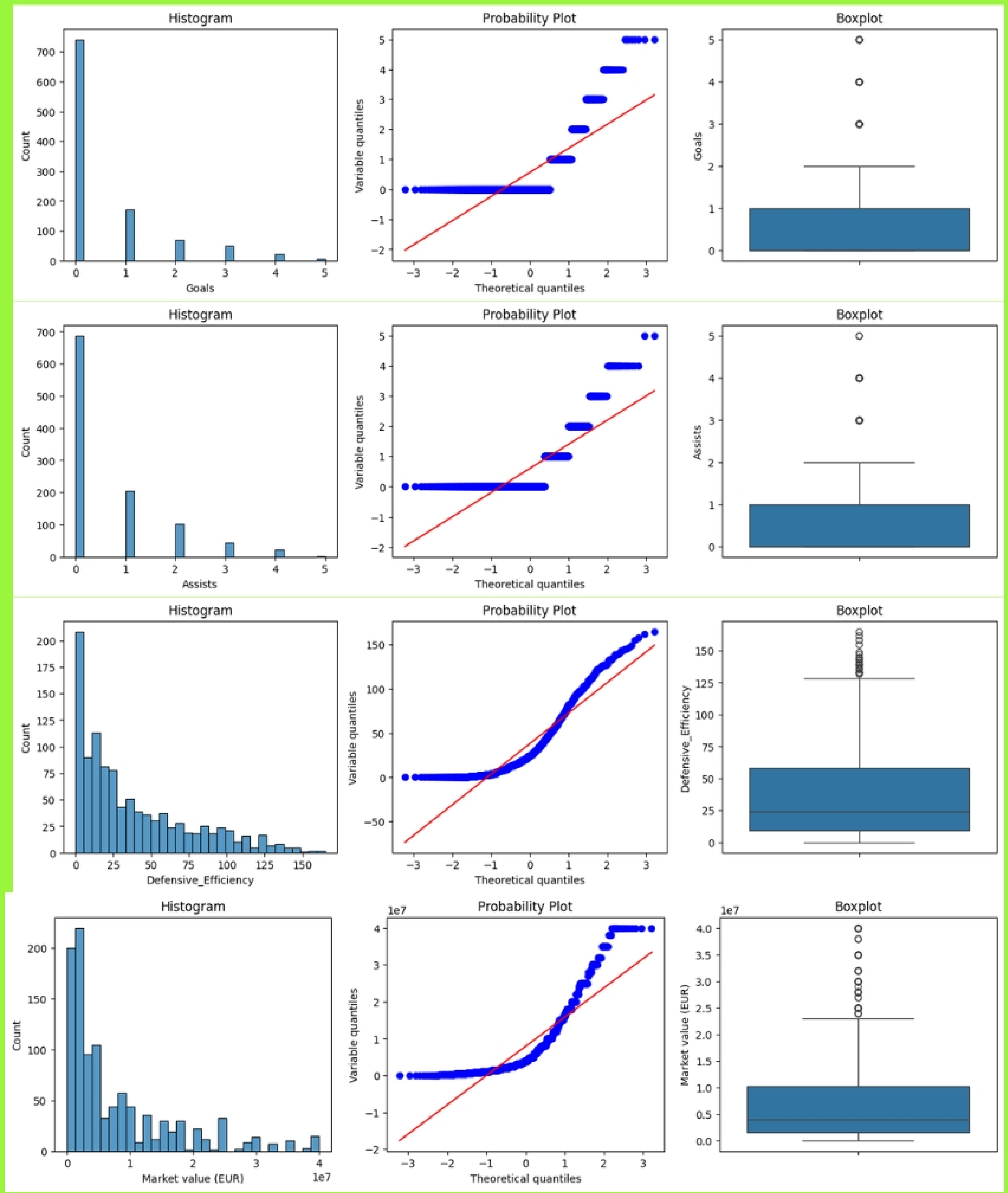
- Target (y): Saves
- Features (X): Selected specifically, not dropped in bulk, but only relevant features are included.

HANDLING OUTLIER

OUTLIER


Histogram
Probability Plot
Boxplot
Count
Goals
Variable quantiles
Theoretical quantiles
Assists
Defensive_Efficiency
Saves
Market value (EUR)
1e8

HANDLING OUTLIER WITH IQR



TRAIN-TEST SPLIT

- 80% training data (X_train, y_train, name_train)

- 20% testing data (X_test, y_test, name_test)

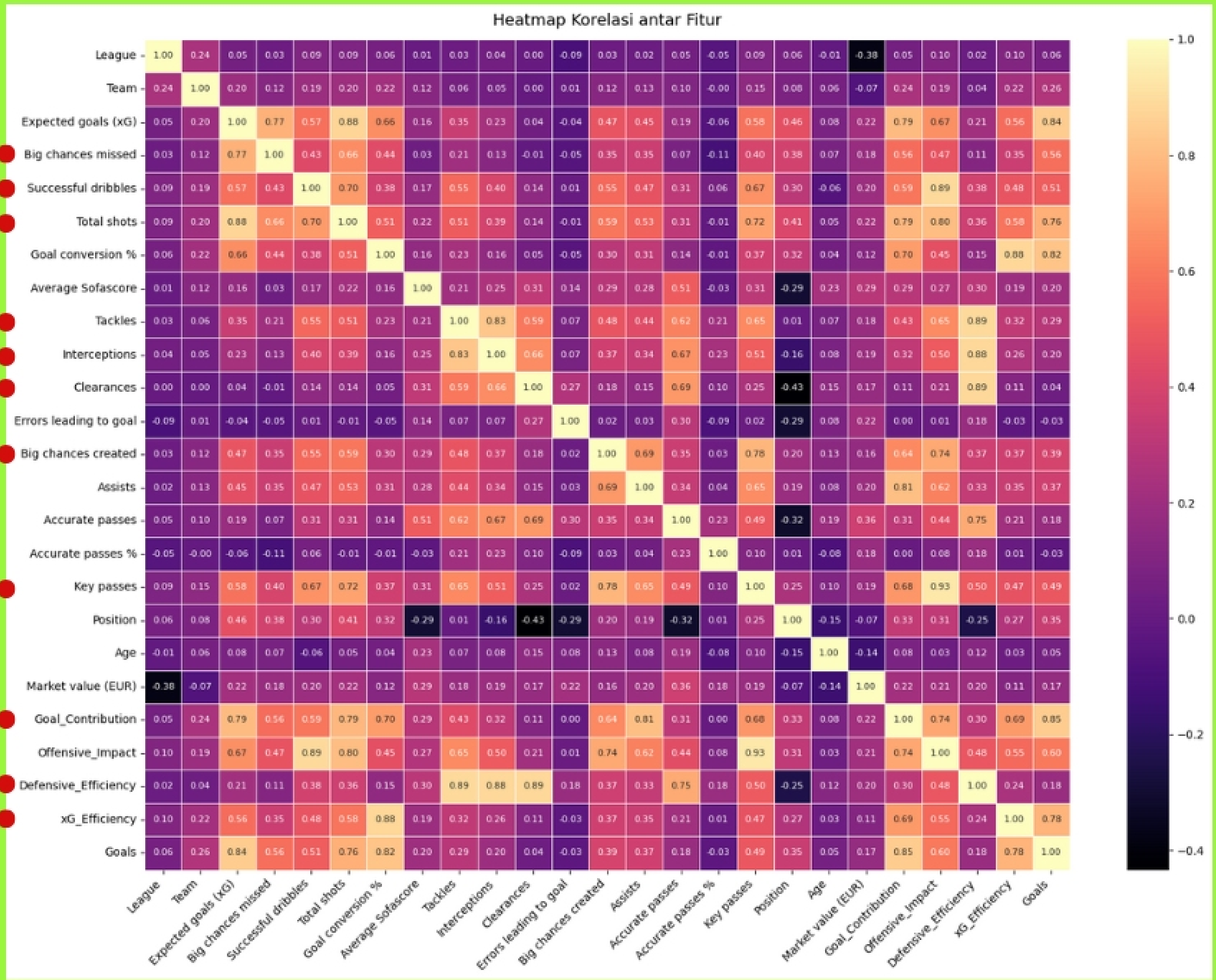
- The parameter random_state=1000 is used to ensure consistent data splitting.

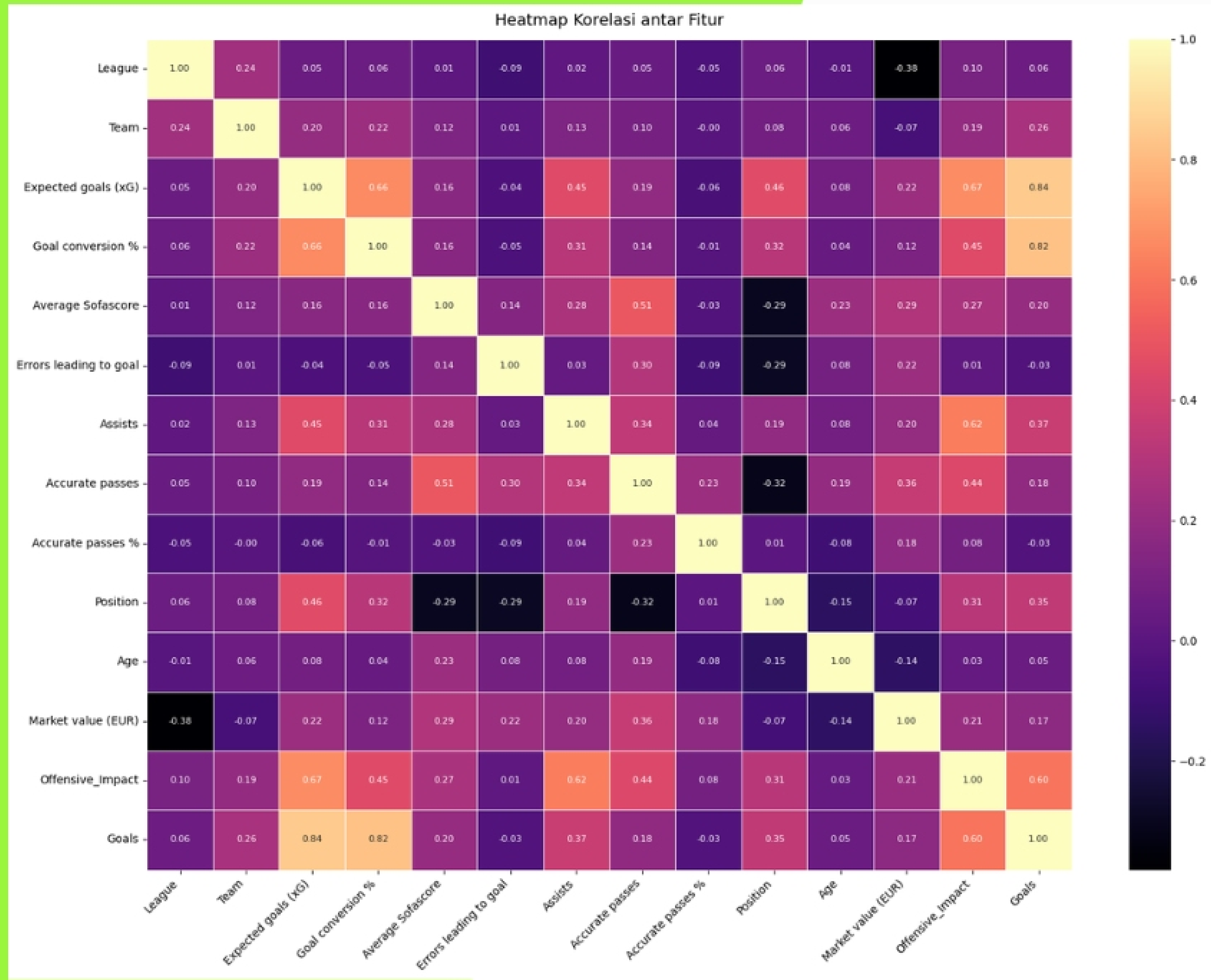
- The variable name_team_clean is included to analyze player identity during evaluation.

MULTICOLLINEARITY

PREDICT TOP SCORE

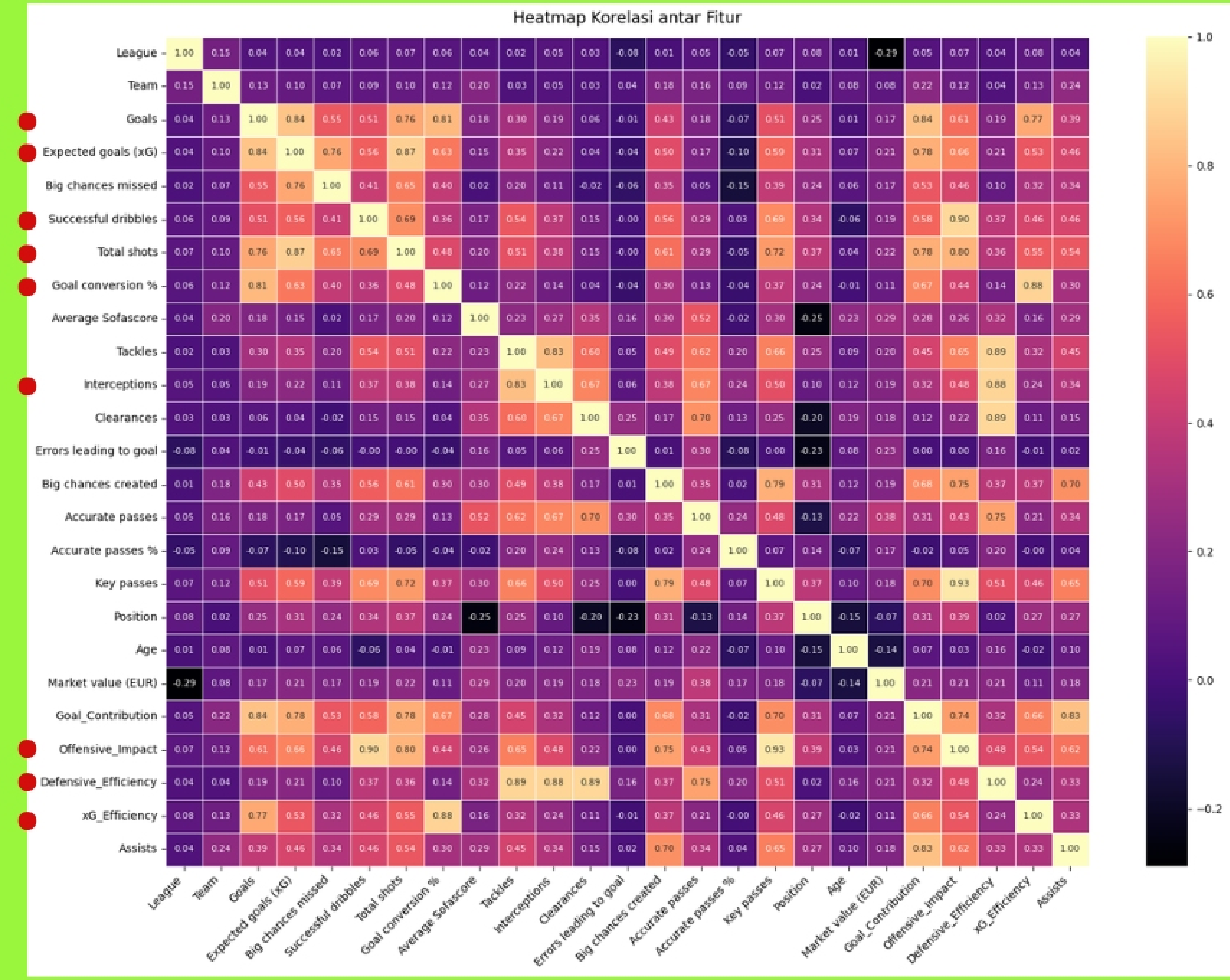
RED MARKS INDICATE DROPPED COLUMNS

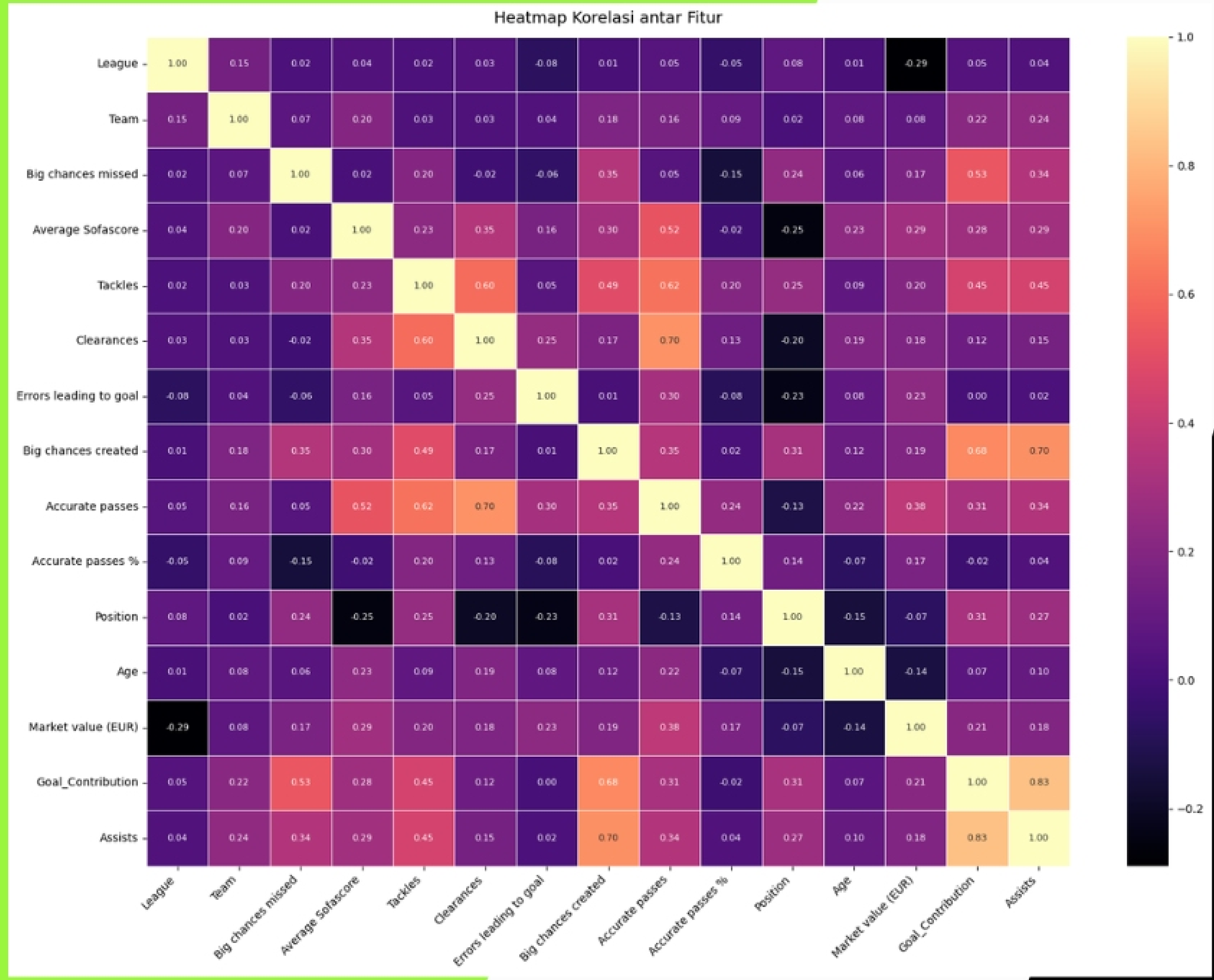




PREDICT TOP ASSISTS

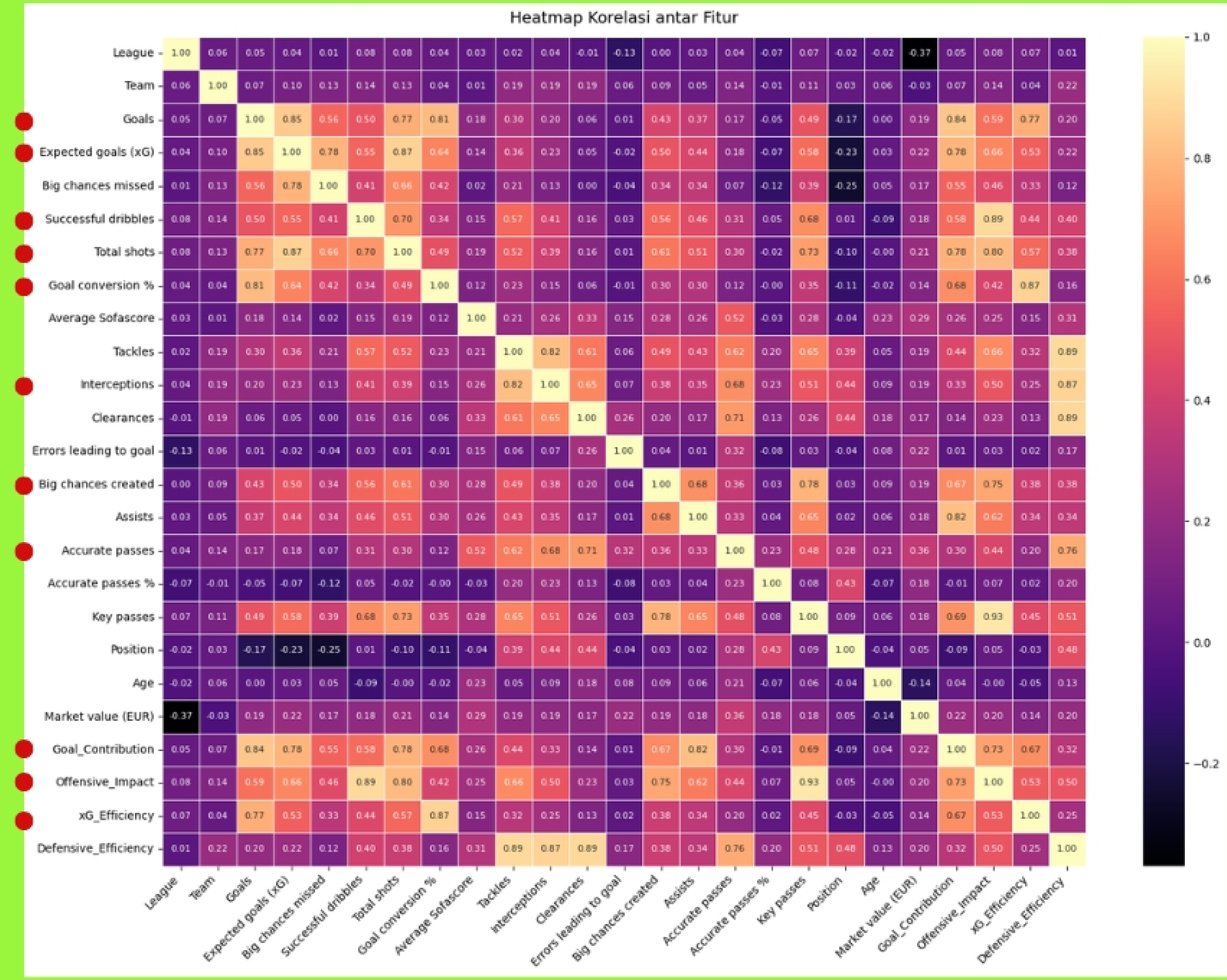
RED MARKS INDICATE DROPPED COLUMNS

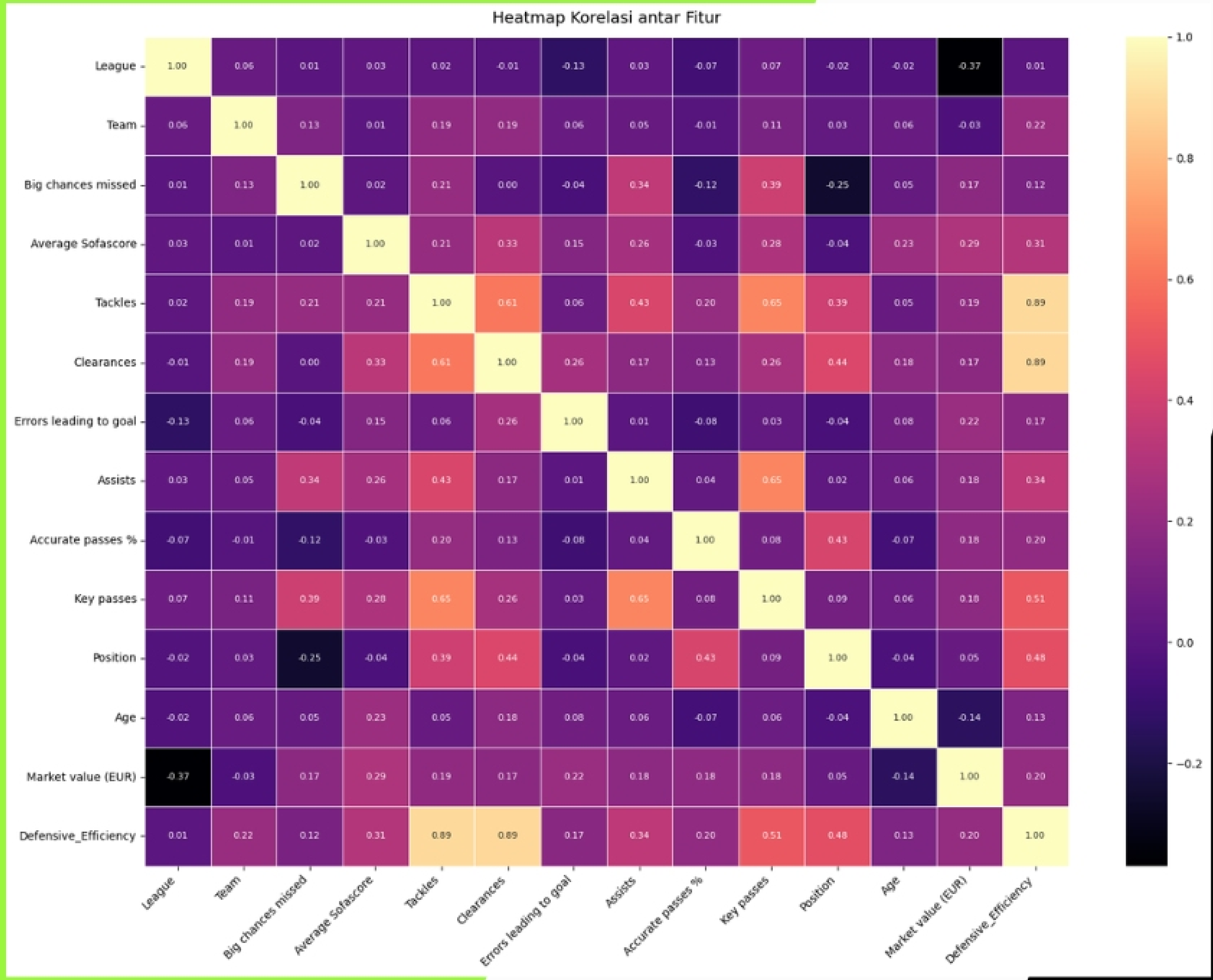




PREDICT TOP DEFFEND

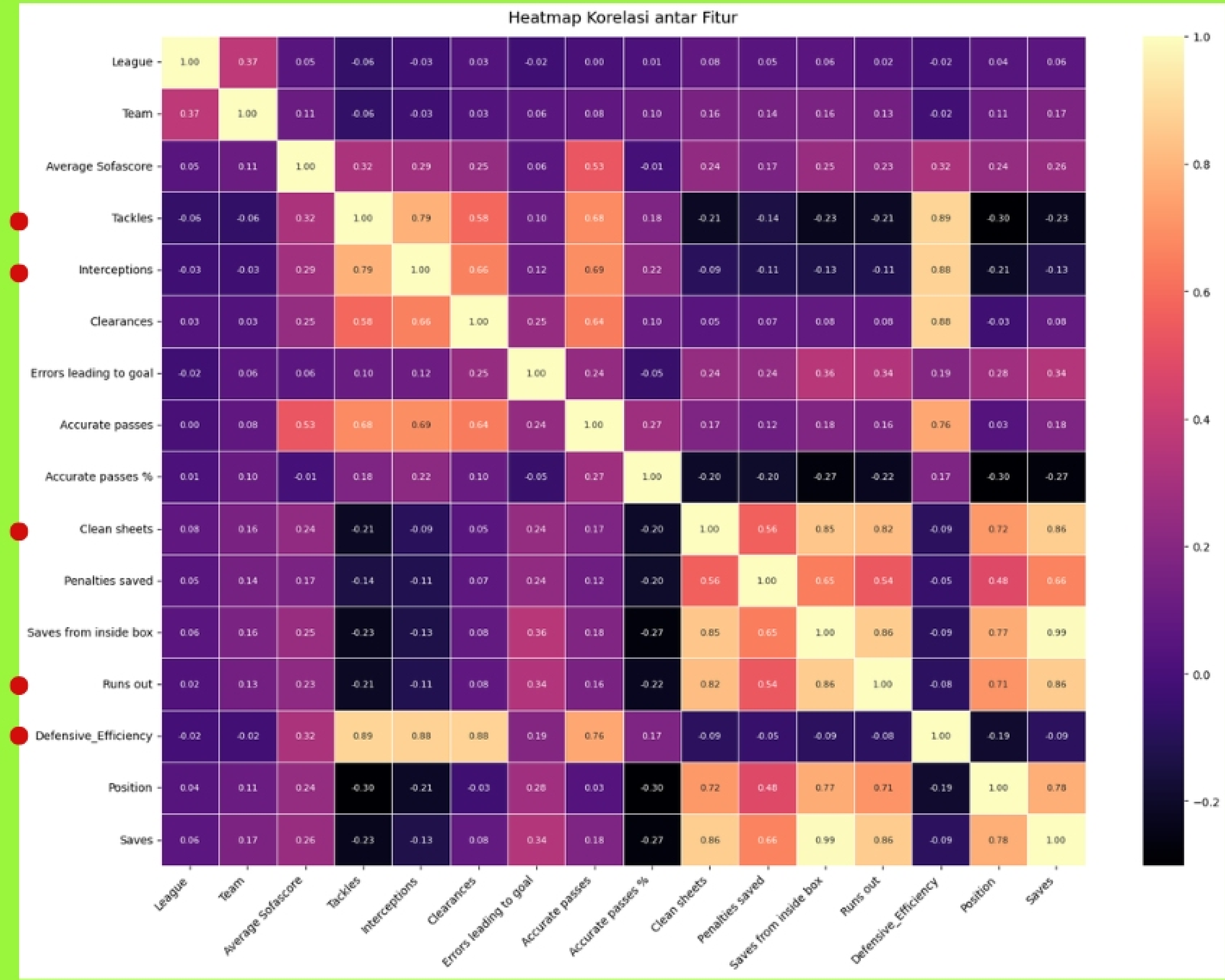
RED MARKS INDICATE DROPPED COLUMNS

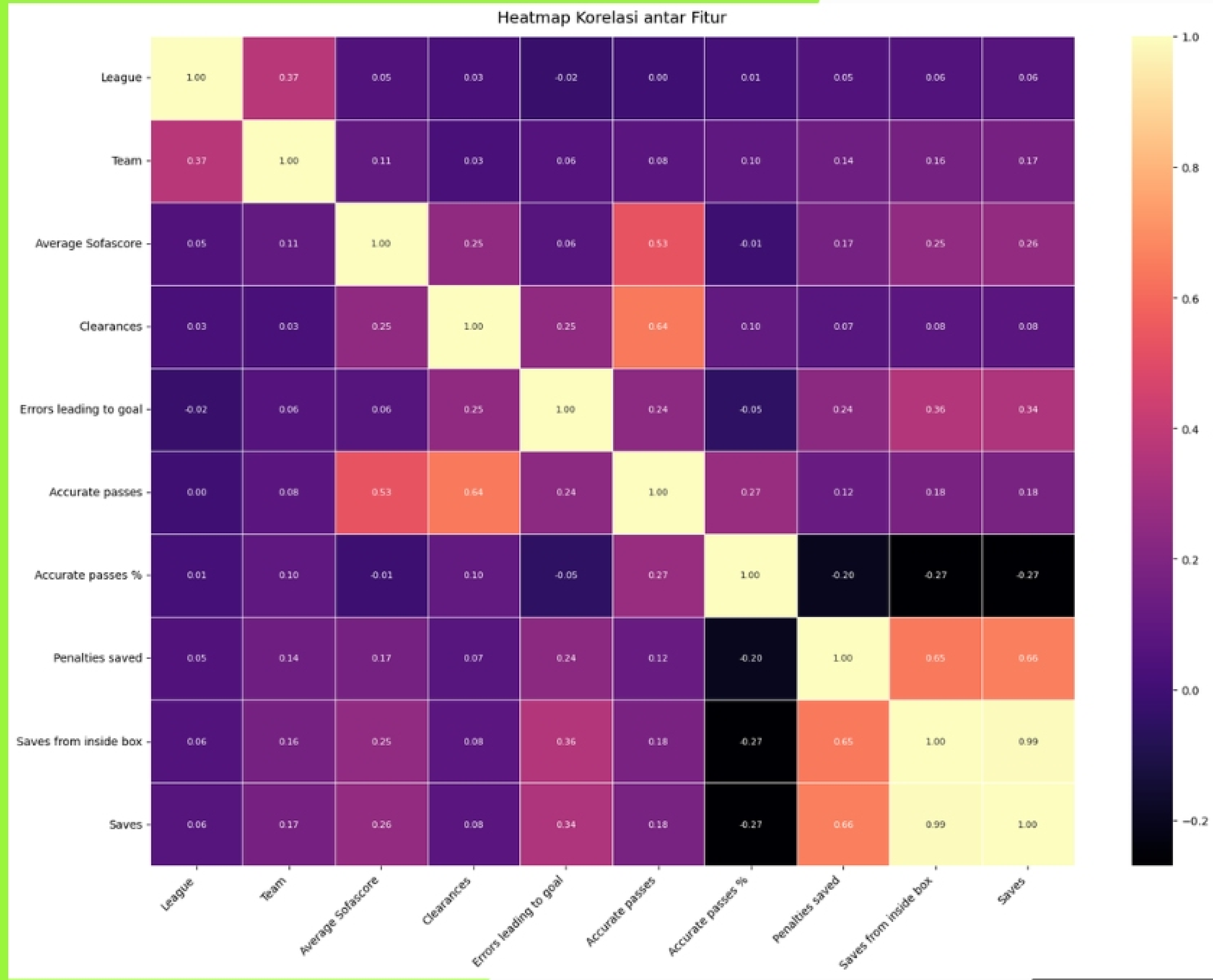




PREDICT TOP SAVES

RED MARKS INDICATE DROPPED COLUMNS





FEATURE SCALLING

MODEL PREDICT

TOP SCORE

	League	Team	Expected goals (xG)	Goal conversion %	Average Sofascore	Errors leading to goal	Assists	Accurate passes	Accurate passes %	Position	Age	Market value (EUR)	Offensive_Impact
count	852.000000	852.000000	852.000000	852.000000	852.000000	852.000000	852.000000	852.000000	852.000000	852.000000	852.000000	8.520000e+02	852.000000
mean	0.592723	0.592723	0.781338	3.397887	6.756796	0.198357	0.607981	325.800469	80.788568	0.592723	26.084507	8.311561e+06	13.382629
std	0.070750	0.292291	1.000897	5.886239	0.196541	0.534890	0.990594	309.757921	8.723065	0.382016	4.556463	9.277122e+06	14.233790
min	0.488372	0.000000	0.000000	0.000000	6.200000	0.000000	0.000000	1.000000	57.280000	0.000000	15.000000	0.000000e+00	0.000000
25%	0.488372	0.388889	0.020000	0.000000	6.630000	0.000000	0.000000	72.750000	76.000000	0.236364	23.000000	1.800000e+06	2.000000
50%	0.615970	0.600000	0.350000	0.000000	6.740000	0.000000	0.000000	222.000000	81.950000	0.773810	26.000000	4.000000e+06	8.000000
75%	0.655589	0.789474	1.150000	6.250000	6.880000	0.000000	1.000000	514.750000	86.852500	0.951613	29.000000	1.200000e+07	22.000000
max	0.655589	1.454545	4.640000	26.670000	7.420000	5.000000	5.000000	1507.000000	100.000000	1.100000	38.000000	4.000000e+07	67.000000

	League	Team	Expected goals (xG)	Goal conversion %	Average Sofascore	Errors leading to goal	Assists	Accurate passes	Accurate passes %	Position	Age	Market value (EUR)	Offensive_Impact
count	852.000000	8.520000e+02	8.520000e+02	8.520000e+02	8.520000e+02	852.000000	8.520000e+02	8.520000e+02	8.520000e+02	8.520000e+02	8.520000e+02	8.520000e+02	8.520000e+02
mean	0.592723	1.250956e-16	-1.250956e-17	-2.293418e-17	-5.427062e-15	0.000000	-3.335881e-17	1.250956e-17	3.429703e-16	1.292654e-16	2.319480e-16	2.501911e-17	5.212315e-17
std	0.070750	1.000587e+00	1.000587e+00	1.000587e+00	1.000587e+00	1.000587	1.000587e+00	1.000587e+00	1.000587e+00	1.000587e+00	1.000587e+00	1.000587e+00	1.000587e+00
min	0.488372	-2.029047e+00	-7.810960e-01	-5.775985e-01	-2.834639e+00	-0.371054	-6.141149e-01	-1.049178e+00	-2.696572e+00	-1.552476e+00	-2.434129e+00	-8.964465e-01	-9.407536e-01
25%	0.488372	-6.977777e-01	-7.611022e-01	-5.775985e-01	-6.455153e-01	-0.371054	-6.141149e-01	-8.174096e-01	-5.492772e-01	-9.333861e-01	-6.773496e-01	-7.023068e-01	-8.001604e-01
50%	0.615970	2.491107e-02	-4.312044e-01	-5.775985e-01	-8.550703e-02	-0.371054	-6.141149e-01	-3.352987e-01	1.332231e-01	4.743067e-01	-1.855752e-02	-4.650250e-01	-3.783806e-01
75%	0.655589	6.735293e-01	3.685478e-01	4.848236e-01	6.272308e-01	-0.371054	3.959737e-01	6.103492e-01	6.955690e-01	9.400142e-01	6.402346e-01	3.978179e-01	6.057721e-01
max	0.655589	2.950245e+00	3.857467e+00	3.955969e+00	3.376362e+00	8.982146	4.436328e+00	3.815539e+00	2.203665e+00	1.328674e+00	2.616611e+00	3.417768e+00	3.769120e+00

MODEL PREDICT

TOP ASSISTS

	League	Team	Big chances missed	Average Sofascore	Tackles	Clearances	Errors leading to goal	Big chances created	Accurate passes	Accurate passes %	Position	Age	Market value (EUR)	Goal_Contribution
count	851.000000	851.000000	851.000000	851.000000	851.000000	851.000000	851.000000	851.000000	851.000000	851.000000	851.000000	851.000000	8.510000e+02	851.000000
mean	0.603995	0.603995	0.808461	6.756005	14.470035	16.920094	0.184489	1.095182	323.494712	80.710517	0.603995	26.104583	8.160034e+06	1.169213
std	0.036018	0.244175	1.257345	0.195803	15.640144	18.844833	0.522477	1.558767	310.766087	8.683992	0.278660	4.591769	9.206495e+06	1.712801
min	0.560000	0.000000	0.000000	6.200000	0.000000	0.000000	0.000000	0.000000	1.000000	57.280000	0.127660	15.000000	0.000000e+00	0.000000
25%	0.560000	0.454545	0.000000	6.620000	2.000000	2.000000	0.000000	0.000000	71.000000	75.720000	0.349112	23.000000	1.800000e+06	0.000000
50%	0.593407	0.600000	0.000000	6.730000	9.000000	10.000000	0.000000	0.000000	216.000000	81.840000	0.611765	26.000000	4.000000e+06	0.000000
75%	0.646341	0.777778	1.000000	6.880000	23.000000	25.000000	0.000000	2.000000	497.500000	86.835000	0.784615	29.000000	1.150000e+07	2.000000
max	0.646341	1.230769	5.000000	7.420000	69.000000	81.000000	5.000000	7.000000	1507.000000	100.000000	2.000000	38.000000	4.000000e+07	9.000000

	League	Team	Big chances missed	Average Sofascore	Tackles	Clearances	Errors leading to goal	Big chances created	Accurate passes	Accurate passes %	Position	Age	Market value (EUR)	Goal_Contribution
count	851.000000	8.510000e+02	8.510000e+02	8.510000e+02	8.510000e+02	8.510000e+02	8.510000e+02	8.510000e+02	8.510000e+02	8.510000e+02	8.510000e+02	8.510000e+02	8.510000e+02	8.510000e+02
mean	0.603995	1.043688e-17	-4.592227e-17	1.753396e-15	3.757277e-17	-3.757277e-17	1.669901e-17	4.174752e-17	-7.097078e-17	8.349503e-18	2.504851e-17	1.795143e-16	1.252426e-17	-7.097078e-17
std	0.036018	1.000588e+00	1.000588e+00	1.000588e+00	1.000588e+00	1.000588e+00	1.000588e+00	1.000588e+00	1.000588e+00	1.000588e+00	1.000588e+00	1.000588e+00	1.000588e+00	1.000588e+00
min	0.560000	-2.475067e+00	-6.433684e-01	-2.841282e+00	-9.257296e-01	-8.983918e-01	-3.533116e-01	-7.030085e-01	-1.038351e+00	-2.699714e+00	-1.710385e+00	-2.419789e+00	-8.868557e-01	-6.830335e-01
25%	0.560000	-6.124193e-01	-6.433684e-01	-6.950079e-01	-7.977784e-01	-7.921995e-01	-3.533116e-01	-7.030085e-01	-8.129690e-01	-5.750180e-01	-9.152112e-01	-6.765167e-01	-6.912266e-01	-6.830335e-01
50%	0.593407	-1.637204e-02	-6.433684e-01	-1.328886e-01	-3.499490e-01	-3.674303e-01	-3.533116e-01	-7.030085e-01	-3.461057e-01	1.301414e-01	2.789771e-02	-2.278955e-02	-4.521243e-01	-6.830335e-01
75%	0.646341	7.121301e-01	1.524259e-01	6.336377e-01	5.457099e-01	4.290119e-01	-3.533116e-01	5.808118e-01	5.602529e-01	7.056759e-01	6.485548e-01	6.309376e-01	3.629970e-01	4.853313e-01
max	0.646341	2.568410e+00	3.335603e+00	3.393133e+00	3.488589e+00	3.402396e+00	9.222108e+00	3.790362e+00	3.810587e+00	2.222575e+00	5.012652e+00	2.592119e+00	3.460458e+00	4.574608e+00

MODEL PREDICT

TOP DEFENSIVE

	League	Team	Big chances missed	Average Sofascore	Tackles	Clearances	Errors leading to goal	Assists	Accurate passes %	Key passes	Position	Age	Market value (EUR)
count	848.000000	848.000000	848.000000	848.000000	848.000000	848.000000	848.000000	848.000000	848.00000	848.000000	848.000000	848.000000	8.480000e+02
mean	38.803066	38.803066	0.811321	6.760413	14.772406	17.385613	0.195755	0.604953	80.71250	7.185142	38.803066	26.142689	8.095341e+06
std	0.476791	8.202774	1.267061	0.199759	15.986235	18.900438	0.536182	0.986724	8.87836	8.279926	17.937964	4.584988	9.066670e+06
min	38.087649	20.230769	0.000000	6.200000	0.000000	0.000000	0.000000	0.000000	57.93000	0.000000	11.408163	15.000000	0.000000e+00
25%	38.087649	34.285714	0.000000	6.630000	2.000000	3.000000	0.000000	0.000000	75.28250	1.000000	19.707071	23.000000	1.700000e+06
50%	38.958491	39.133333	0.000000	6.740000	9.000000	10.000000	0.000000	0.000000	81.96000	4.000000	46.041916	26.000000	4.000000e+06
75%	39.219880	43.461538	1.000000	6.890000	24.000000	25.250000	0.000000	1.000000	87.24500	11.000000	56.523810	29.000000	1.200000e+07
max	39.219880	55.277778	5.000000	7.420000	69.000000	81.000000	5.000000	5.000000	100.00000	41.000000	58.739645	38.000000	4.000000e+07

	League	Team	Big chances missed	Average Sofascore	Tackles	Clearances	Errors leading to goal	Assists	Accurate passes %	Key passes	Position	Age	Market value (EUR)
count	8.480000e+02	8.480000e+02	8.480000e+02	8.480000e+02	8.480000e+02	8.480000e+02	8.480000e+02	8.480000e+02	8.480000e+02	8.480000e+02	8.480000e+02	8.480000e+02	8.480000e+02
mean	1.009884e-14	3.812464e-16	-3.561093e-17	-1.460048e-15	-3.665831e-17	1.256856e-17	-3.351617e-17	-3.718200e-17	-2.230920e-15	8.379042e-18	2.398501e-16	2.304236e-16	4.189521e-18
std	1.000590e+00	1.000590e+00	1.000590e+00	1.000590e+00	1.000590e+00	1.000590e+00	1.000590e+00	1.000590e+00	1.000590e+00	1.000590e+00	1.000590e+00	1.000590e+00	1.000590e+00
min	-1.501369e+00	-2.265485e+00	-6.406948e-01	-2.807093e+00	-9.246157e-01	-9.203952e-01	-3.653057e-01	-6.134543e-01	-2.567585e+00	-8.682905e-01	-1.528104e+00	-2.431689e+00	-8.933950e-01
25%	-1.501369e+00	-5.510353e-01	-6.406948e-01	-6.532341e-01	-7.994342e-01	-7.615751e-01	-3.653057e-01	-6.134543e-01	-6.119604e-01	-7.474452e-01	-1.065186e+00	-6.858345e-01	-7.057845e-01
50%	3.261729e-01	4.028664e-02	-6.406948e-01	-1.022469e-01	-3.612991e-01	-3.909947e-01	-3.653057e-01	-6.134543e-01	1.405931e-01	-3.849094e-01	4.037873e-01	-3.113920e-02	-4.519584e-01
75%	8.747220e-01	5.682494e-01	1.489988e-01	6.490994e-01	5.775619e-01	4.163410e-01	-3.653057e-01	4.005988e-01	7.362120e-01	4.610077e-01	9.884734e-01	6.235561e-01	4.309149e-01
max	8.747220e-01	2.009617e+00	3.307773e+00	3.303856e+00	3.394145e+00	3.367749e+00	8.965395e+00	4.456811e+00	2.173699e+00	4.086367e+00	1.112074e+00	2.587642e+00	3.520972e+00

MODEL PREDICT

TOP SAVES

	League	Team	Average Sofascore	Clearances	Errors leading to goal	Accurate passes	Accurate passes %	Penalties saved	Saves from inside box
count	1146.000000	1146.000000	1146.000000	1146.000000	1146.000000	1146.000000	1146.000000	1146.000000	1146.000000
mean	4.007853	4.007853	6.801099	18.155323	0.175393	373.604712	80.781859	0.036649	2.632635
std	1.108528	2.960001	0.213846	18.821440	0.481760	323.779633	8.577831	0.258414	11.532007
min	2.756614	0.000000	6.200000	0.000000	0.000000	1.000000	57.280000	0.000000	0.000000
25%	2.756614	1.190476	6.650000	4.000000	0.000000	107.250000	75.475000	0.000000	0.000000
50%	3.771505	4.115385	6.790000	12.000000	0.000000	287.500000	81.930000	0.000000	0.000000
75%	5.424242	5.941176	6.940000	26.000000	0.000000	576.750000	86.887500	0.000000	0.000000
max	5.424242	12.315789	7.420000	81.000000	5.000000	1486.000000	100.000000	4.000000	100.000000

	League	Team	Average Sofascore	Clearances	Errors leading to goal	Accurate passes	Accurate passes %	Penalties saved	Saves from inside box
count	1.146000e+03	1.146000e+03	1.146000e+03	1.146000e+03	1.146000e+03	1.146000e+03	1.146000e+03	1.146000e+03	1.146000e+03
mean	3.875124e-17	6.897721e-17	2.526581e-16	1.092785e-16	-2.480079e-17	6.200198e-17	-2.681586e-16	1.240040e-17	-1.472547e-17
std	1.000437e+00	1.000437e+00	1.000437e+00	1.000437e+00	1.000437e+00	1.000437e+00	1.000437e+00	1.000437e+00	1.000437e+00
min	-1.129232e+00	-1.354595e+00	-2.812120e+00	-9.650297e-01	-3.642255e-01	-1.151300e+00	-2.741033e+00	-1.418854e-01	-2.283891e-01
25%	-1.129232e+00	-9.522320e-01	-7.068877e-01	-7.524134e-01	-3.642255e-01	-8.230011e-01	-6.189415e-01	-1.418854e-01	-2.283891e-01
50%	-2.133020e-01	3.634396e-02	-5.192661e-02	-3.271806e-01	-3.642255e-01	-2.660523e-01	1.339083e-01	-1.418854e-01	-2.283891e-01
75%	1.278278e+00	6.534347e-01	6.498174e-01	4.169767e-01	-3.642255e-01	6.276923e-01	7.121039e-01	-1.418854e-01	-2.283891e-01
max	1.278278e+00	2.807960e+00	2.895398e+00	3.340452e+00	1.001892e+01	3.437155e+00	2.241421e+00	1.534389e+01	8.446914e+00

MODEL CLUSTERING

MARKET VALUE

	League	Team	Average Sofascore	Clearances	Errors leading to goal	Accurate passes	Accurate passes %	Penalties saved	Saves from inside box
count	1146.000000	1146.000000	1146.000000	1146.000000	1146.000000	1146.000000	1146.000000	1146.000000	1146.000000
mean	4.007853	4.007853	6.801099	18.155323	0.175393	373.604712	80.781859	0.036649	2.632635
std	1.108528	2.960001	0.213846	18.821440	0.481760	323.779633	8.577831	0.258414	11.532007
min	2.756614	0.000000	6.200000	0.000000	0.000000	1.000000	57.280000	0.000000	0.000000
25%	2.756614	1.190476	6.650000	4.000000	0.000000	107.250000	75.475000	0.000000	0.000000
50%	3.771505	4.115385	6.790000	12.000000	0.000000	287.500000	81.930000	0.000000	0.000000
75%	5.424242	5.941176	6.940000	26.000000	0.000000	576.750000	86.887500	0.000000	0.000000
max	5.424242	12.315789	7.420000	81.000000	5.000000	1486.000000	100.000000	4.000000	100.000000

	League	Team	Average Sofascore	Clearances	Errors leading to goal	Accurate passes	Accurate passes %	Penalties saved	Saves from inside box
count	1.146000e+03	1.146000e+03	1.146000e+03	1.146000e+03	1.146000e+03	1.146000e+03	1.146000e+03	1.146000e+03	1.146000e+03
mean	3.875124e-17	6.897721e-17	2.526581e-16	1.092785e-16	-2.480079e-17	6.200198e-17	-2.681586e-16	1.240040e-17	-1.472547e-17
std	1.000437e+00	1.000437e+00	1.000437e+00	1.000437e+00	1.000437e+00	1.000437e+00	1.000437e+00	1.000437e+00	1.000437e+00
min	-1.129232e+00	-1.354595e+00	-2.812120e+00	-9.650297e-01	-3.642255e-01	-1.151300e+00	-2.741033e+00	-1.418854e-01	-2.283891e-01
25%	-1.129232e+00	-9.522320e-01	-7.068877e-01	-7.524134e-01	-3.642255e-01	-8.230011e-01	-6.189415e-01	-1.418854e-01	-2.283891e-01
50%	-2.133020e-01	3.634396e-02	-5.192661e-02	-3.271806e-01	-3.642255e-01	-2.660523e-01	1.339083e-01	-1.418854e-01	-2.283891e-01
75%	1.278278e+00	6.534347e-01	6.498174e-01	4.169767e-01	-3.642255e-01	6.276923e-01	7.121039e-01	-1.418854e-01	-2.283891e-01
max	1.278278e+00	2.807960e+00	2.895398e+00	3.340452e+00	1.001892e+01	3.437155e+00	2.241421e+00	1.534389e+01	8.446914e+00

MODELLING

MODEL PREDICT

Linear Regression

- Baseline model without regularization.
- Used as an initial benchmark to compare the performance of other models.

Ridge Regression

- Adds a penalty to large coefficients to prevent overfitting.
- Tuning parameter: alpha (using GridSearchCV).

Lasso Regression

- Shrinks some coefficients to zero → automatic feature selection.
- Tuning parameter: alpha (using GridSearchCV).

MLP Regressor

A neural network model capable of capturing non-linear relationships between features.
Tuned hyperparameters:

- hidden_layer_sizes → Layer structure
- alpha → Regularization
- learning_rate_init → Learning rate
- max_iter → Maximum iterations

MODEL CLUSTERING

K-Means Clustering

Optimal k is selected using:

- **Elbow Method → Based on inertia value (amount of variance within clusters)**
- **Silhouette Score → Measures how well each data point fits within its assigned cluster**

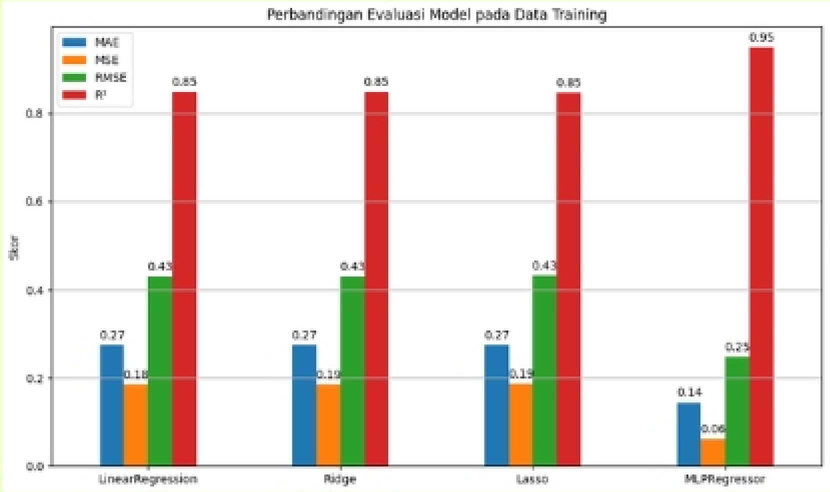
Agglomerative Clustering

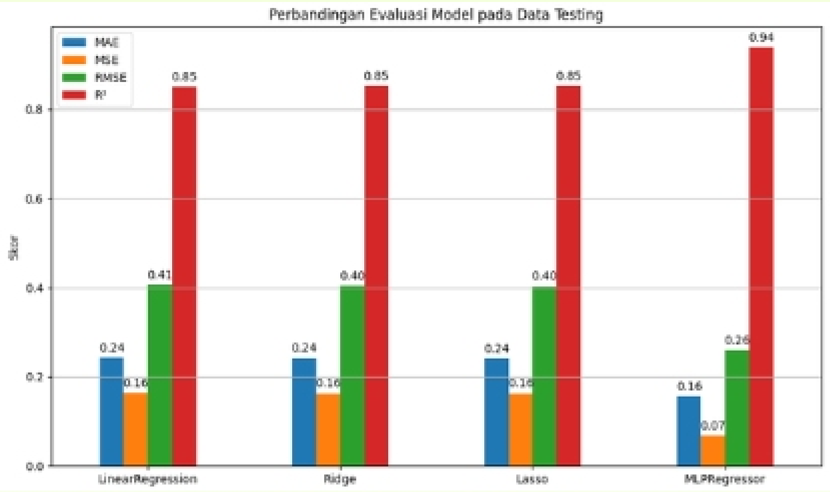
Optimal k is selected using:

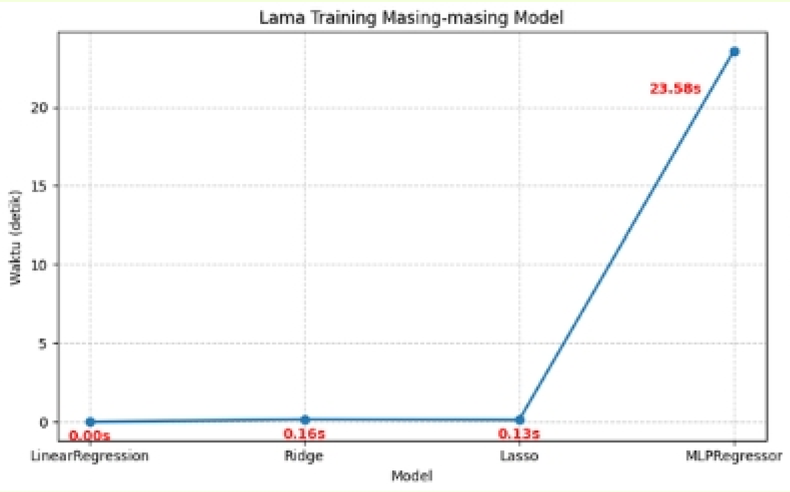
- **Silhouette Score → Measures how well each data point fits within its assigned cluster**

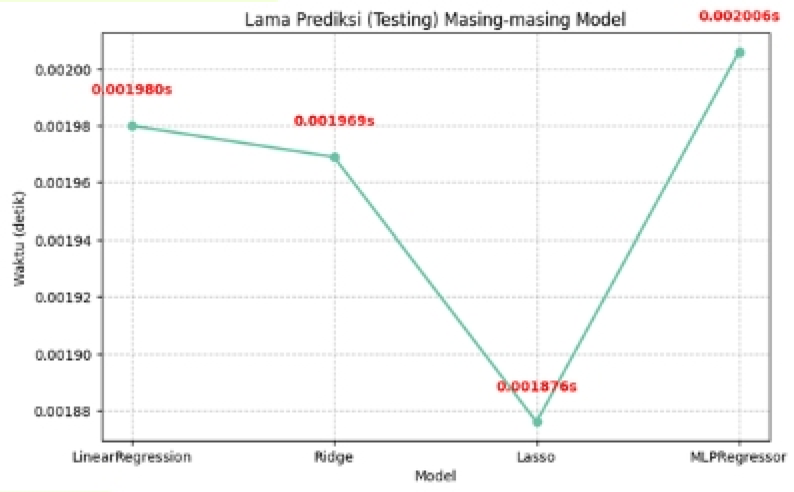
EVALUATION & INTERPRETATION

MODEL PREDICT GOAL









Model	MAE (Train)	MSE (Train)	RMSE (Train)	R² (Train)	MAE (Test)	MSE (Test)	RMSE (Test)	R² (Test)	Training Time (s)	Prediction Time (s)
LinearRegression	0.27495	0.184894	0.429993	0.848773	0.243596	0.164778	0.405929	0.850717	0	0.00198
Ridge	0.274103	0.185113	0.430248	0.848594	0.242104	0.163215	0.403999	0.852133	0.16	0.001969
Lasso	0.27412	0.187485	0.432995	0.846653	0.241232	0.163101	0.403858	0.852236	0.13	0.001876
MLPRegressor	0.143978	0.061645	0.248285	0.949579	0.155699	0.067806	0.260397	0.93857	23.58	0.002006

Conclusion and Recommendation

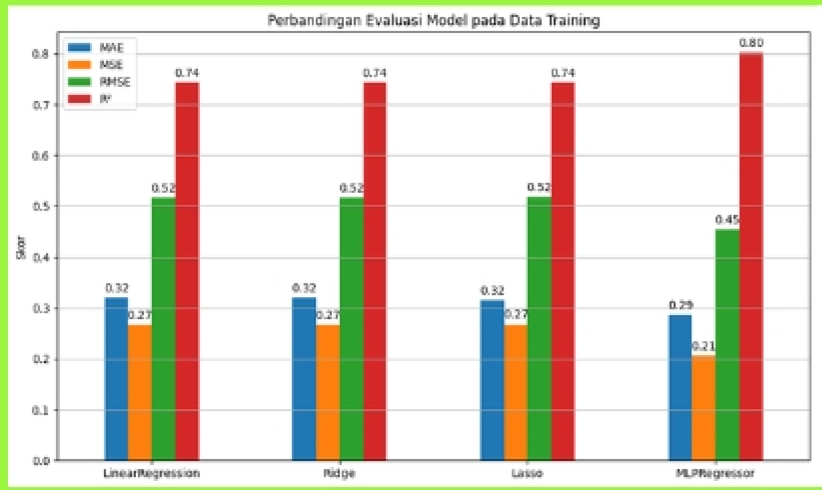
Conclusion

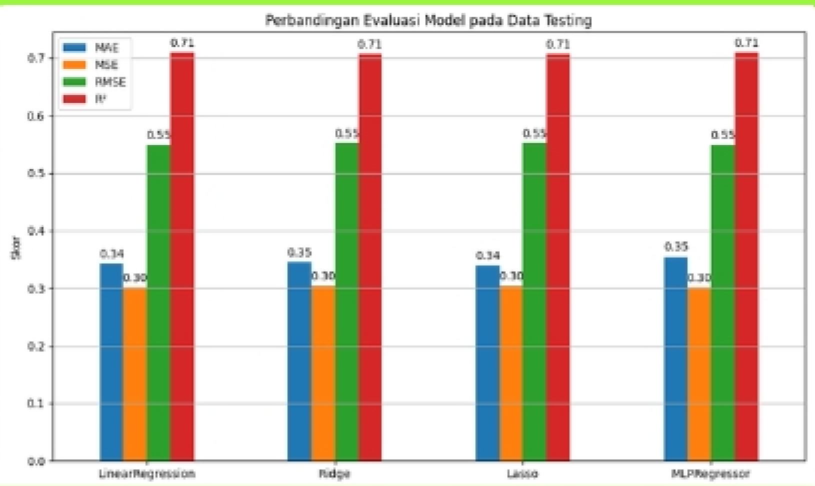
- **MLPRegressor provides the highest accuracy ($R^2 \approx 0.94$), but requires the longest training time (23.58 seconds).**
- **Linear, Ridge, and Lasso are much faster to train (< 0.2 seconds) while still maintaining solid accuracy ($R^2 \approx 0.85$).**
- **All models deliver very fast prediction times (< 0.003 seconds).**

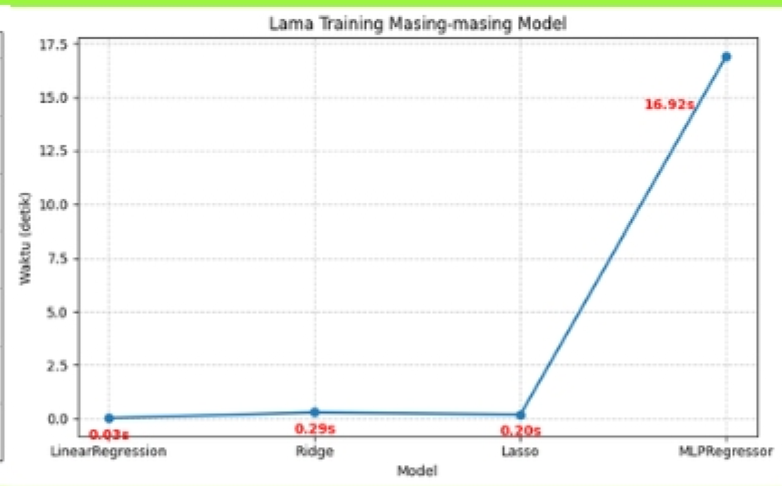
Recommendation

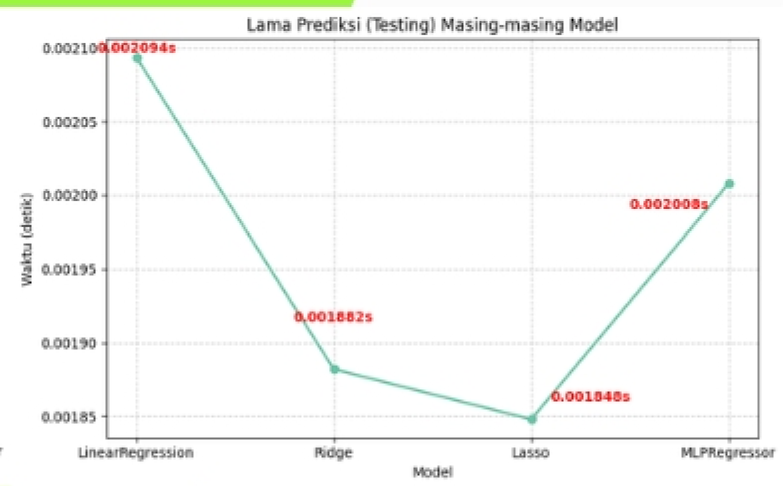
- **Use MLPRegressor for maximum accuracy if training time is not a constraint.**

MODEL PREDICT ASSISTS









Model	MAE (Train)	MSE (Train)	RMSE (Train)	R^2 (Train)	MAE (Test)	MSE (Test)	RMSE (Test)	R^2 (Test)	Training Time (s)	Prediction Time (s)
LinearRegression	0.320234	0.267057	0.516776	0.743939	0.343397	0.301862	0.549419	0.709504	0.03	0.002094
Ridge	0.320687	0.267248	0.51696	0.743756	0.346259	0.304411	0.551734	0.707051	0.29	0.001882
Lasso	0.31586	0.268025	0.517711	0.743011	0.340528	0.304382	0.551708	0.707078	0.2	0.001848
MLPRegressor	0.286522	0.20581	0.453662	0.802665	0.354931	0.301977	0.549524	0.709393	16.92	0.002008

Conclusion and Recommendation

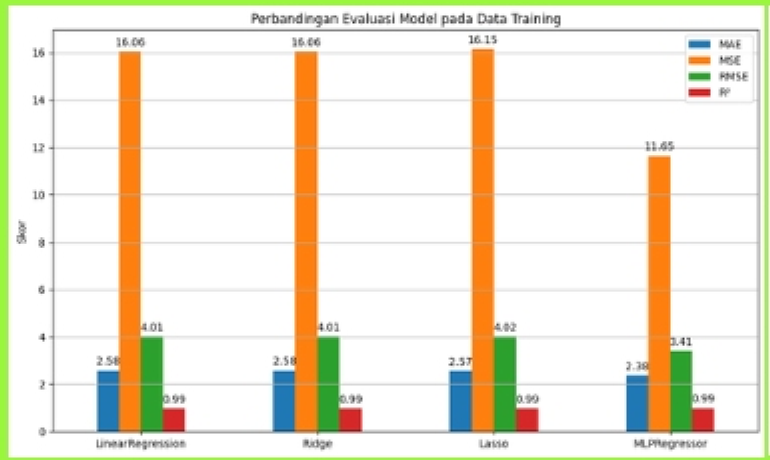
Conclusion

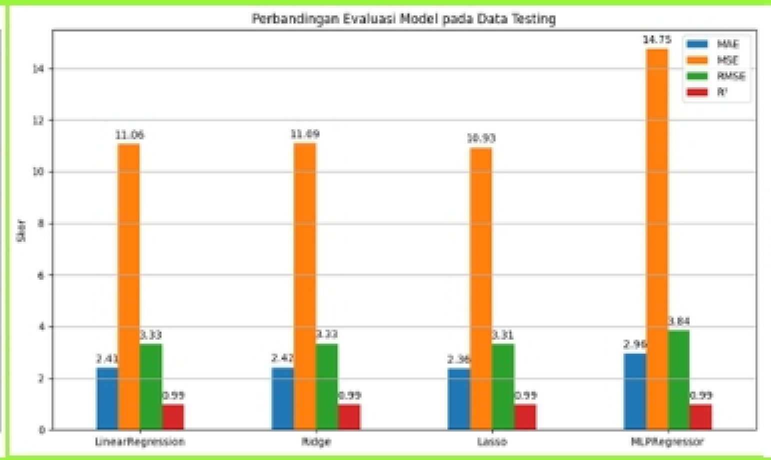
- MLPRegressor performs best on the training data ($R^2 = 0.80$), but offers only slight improvement on the testing data ($R^2 = 0.709$).
- Linear, Ridge, and Lasso show relatively similar performance on the testing data ($R^2 \approx 0.71$) with significantly faster training and prediction times.
- All models provide highly efficient prediction times (< 0.003 seconds), making them suitable for real-time systems.

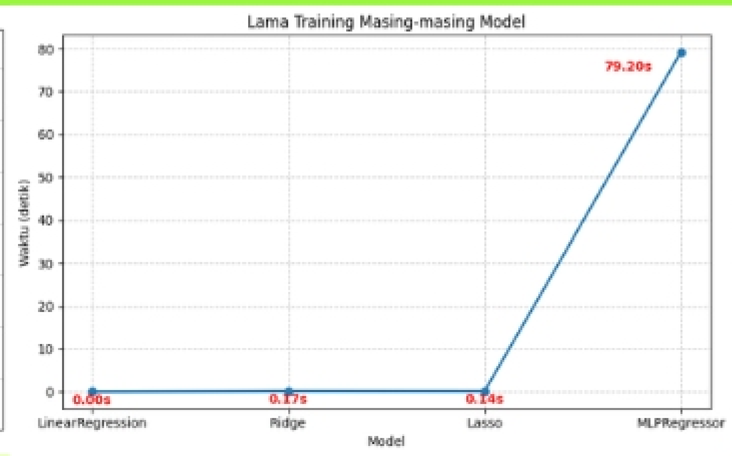
Recommendation

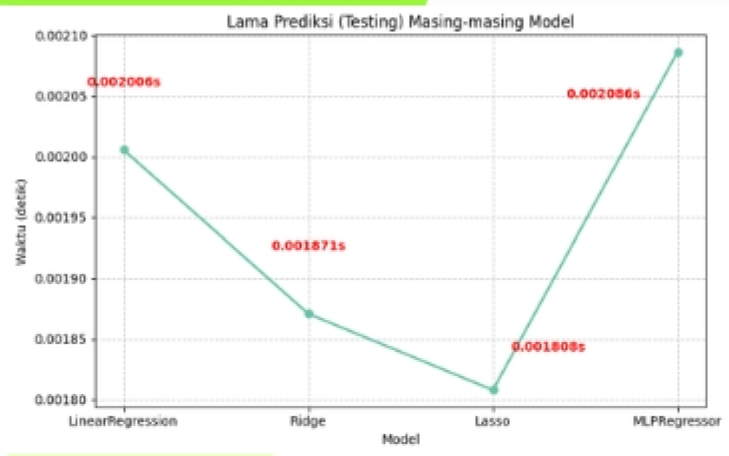
- Use Linear Regression or Lasso if efficiency and speed are a priority, as they deliver nearly equivalent performance on testing data with much shorter runtime.

MODEL PREDICT DEFENSIVE









Model	MAE (Train)	MSE (Train)	RMSE (Train)	R² (Train)	MAE (Test)	MSE (Test)	RMSE (Test)	R² (Test)	Training Time (s)	Prediction Time (s)
LinearRegression	2.577891	16.060334	4.007535	0.988724	2.409295	11.064631	3.326354	0.990538	0	0.002006
Ridge	2.580933	16.061858	4.007725	0.988723	2.415389	11.091995	3.330465	0.990514	0.17	0.001871
Lasso	2.571597	16.151378	4.018878	0.988661	2.360466	10.925631	3.305394	0.990657	0.14	0.001808
MLPRegressor	2.378031	11.645101	3.412492	0.991824	2.957274	14.745653	3.840007	0.98739	79.2	0.002086

Conclusion and Recommendation

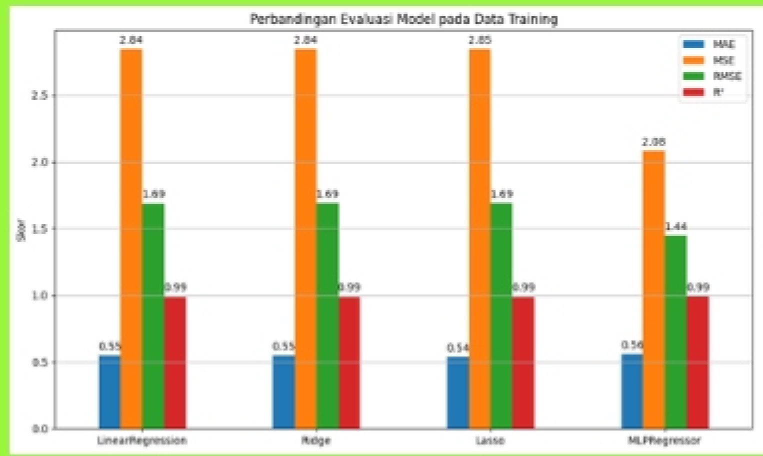
Conclusion

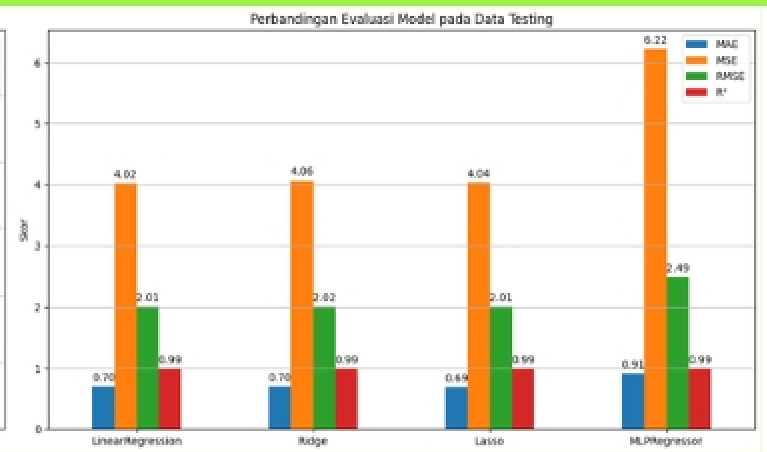
- MLPRegressor excels on the training data ($R^2 = 0.9918$), but shows a slight drop in testing performance ($R^2 = 0.9874$) and requires the longest training time (79.2 seconds).
- Lasso Regression delivers the best performance on testing data ($R^2 = 0.9907$) with highly efficient training and prediction times.
- Linear and Ridge Regression remain consistent, fast, and accurate ($R^2 \approx 0.9905$), making them suitable for systems that prioritize speed and stability.

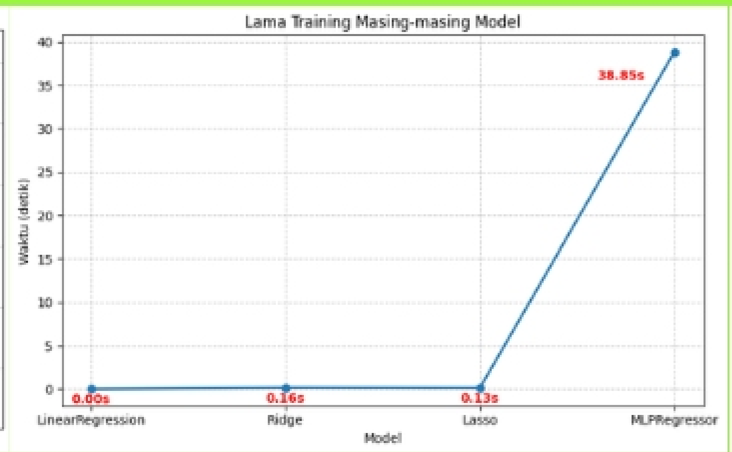
Recommendation

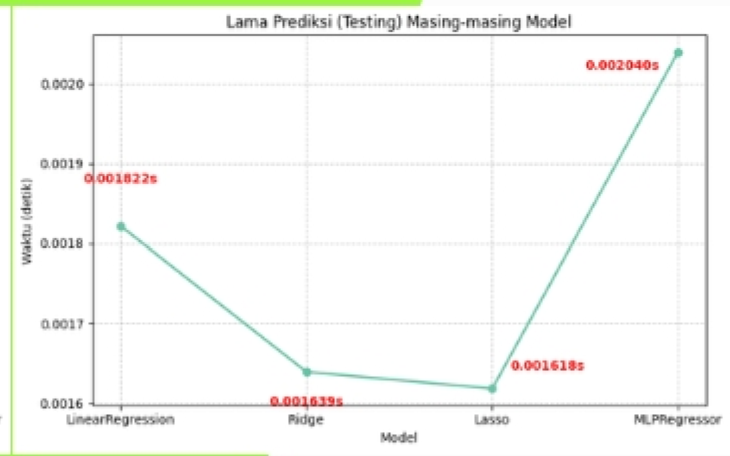
- Use Linear Regression for real-time systems that demand extreme speed and low latency.

MODEL PREDICT SAVES









Model	MAE (Train)	MSE (Train)	RMSE (Train)	R² (Train)	MAE (Test)	MSE (Test)	RMSE (Test)	R² (Test)	Training Time (s)	Prediction Time (s)
LinearRegression	0.547807	2.8445	1.686564	0.990676	0.700844	4.023981	2.005986	0.990468	0	0.001822
Ridge	0.549315	2.84494	1.686695	0.990675	0.704274	4.062787	2.015636	0.990376	0.16	0.001639
Lasso	0.53533	2.845411	1.686834	0.990673	0.688623	4.035778	2.008925	0.99044	0.13	0.001618
MLPRegressor	0.559593	2.081409	1.442709	0.993178	0.914031	6.224135	2.494822	0.985256	38.85	0.00204

Conclusion and Recommendation

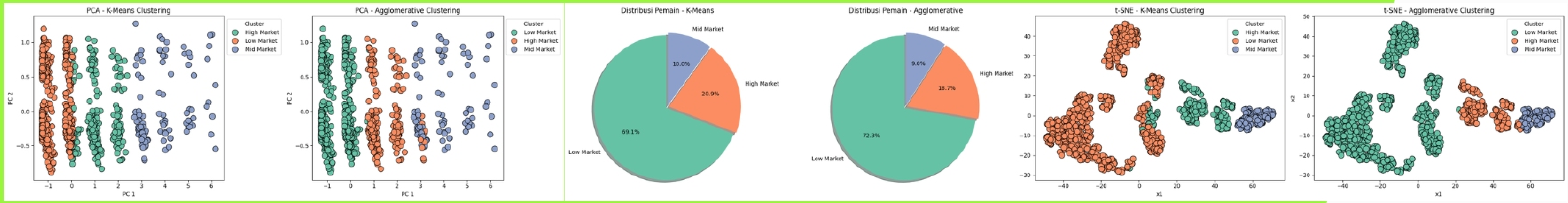
Conclusion

- MLPRegressor demonstrates the best performance on training data ($R^2 = 0.9932$) but drops on testing data ($R^2 = 0.9853$) and requires long training time (38.85 seconds).
- Linear, Ridge, and Lasso Regression show highly consistent performance on testing data ($R^2 \approx 0.990$) with very low latency.
- Lasso Regression achieves the lowest MAE and MSE on the testing set among the linear models.

Recommendation

- Use Linear Regression if prediction speed is the top priority.

MODEL CLUSTERING MARKET VALUE



Visualization Aspect	K-Means Clustering	Agglomerative Clustering
PCA (Linear)	The PCA visualization for K-Means shows fairly clear cluster separation and even distribution. The clusters appear symmetrical and are easier to interpret visually.	PCA visualization for Agglomerative reveals overlapping between clusters, especially between Low and High Market segments. The distribution is less distinct.
t-SNE (Nonlinear)	In the t-SNE visualization, K-Means shows strong dominance of the Low Market color, with significant overlaps. The clusters look less organic.	t-SNE for Agglomerative shows better-separated clusters, more natural spatial structure, and clearer distinction of High and Mid Market segments.
Pie Chart Distribution	K-Means shows a balanced distribution: 69.1% Low Market, 20.9% High Market, 10% Mid Market, indicating good economic variation capture.	Agglomerative skews towards Low Market (72.3%), with Mid and High Market forming a smaller portion — showing more conservative segmentation.

Conclusion and Recommendation

Conclusion:
K-Means Clustering generates more balanced, distinct, and well-separated clusters, making it highly suitable for clean and structured player market segmentation that can be readily implemented in systems. In contrast, Agglomerative Clustering is better at capturing the natural structure of the data, but its results tend to be less balanced and show noticeable overlaps between market segments, particularly between Low and High Market groups.

Recommendation:
It is recommended to use K-Means Clustering for player market segmentation, as it offers a more practical, efficient, and interpretable solution for real-world applications.

INSIGHT DAN REKOMENDASI BISNIS

Use Case	Model Recommendation	Business Recommendation	Implementation Strategy
Top Scorer Prediction	MLPRegressor ($R^2 \approx 0.94$)	Focus on players with the highest goal contributions using a high-accuracy model. Suitable for transfer decisions.	Applied in weekly scouting reports or performance-based player ranking systems. Training can be done offline.
Top Assist Prediction	Linear Regression / Lasso ($R^2 \approx 0.71$)	Use fast and efficient models to track key pass contributors. Supports midfield rotation and attacking strategies.	Integrated into real-time dashboards for coaches or analysts throughout the season.
Top Defensive Player Prediction	Linear Regression ($R^2 \approx 0.9905$)	Evaluate defensive performance objectively and regularly. Supports tactical decisions and starting XI selection.	Used in real-time defensive analytics systems to support match decisions and tactical planning.
Top Goalkeeper Saves Prediction	Linear Regression ($R^2 \approx 0.9904$)	Automatically monitor goalkeeper save performance. Supports weekly evaluations and defensive strategies.	Implemented in automated goalkeeper performance monitoring tools and weekly efficiency reports.
Player Market Value Segmentation	K-Means Clustering	Cluster players objectively by market value to optimize transfer strategy and investment planning.	Used in scouting and transfer frameworks to identify undervalued/overpriced players and manage budget efficiently.

STRATEGIC VALUE

Strategic Aspect	Explanation	Business Impact
Player Performance Prediction	Machine learning models predict player performance (scorer, assist, defense, saves) based on match statistics.	• More accurate recruitment • Early talent identification • Reduced risk of subjective decisions
Player Market Value Segmentation	Clustering (e.g., K-Means) groups players based on economic value into segments (Low, Mid, High Market).	• Identify undervalued players • Optimize transfer spending • Improve transfer strategy
Data-Driven Decision Making	Integration of prediction and segmentation results in scouting, transfers, and team lineup decisions.	• Sharper and more targeted team strategies • More efficient budget management • Higher player ROI

THANK YOU

The beauty of football lies not just in the game, but in how we understand it. Enjoy the game.

link project : https://github.com/reygaferdiansyah