



AI BootCamp

Project 2: UEFA Euro 2024 Tournament Prediction

Maggie Rosen, Mike Szumski, Kurt Andreassen



Executive Summary



Our goal is to predict top finalists in the UEFA Euro 2024 tournament now underway in Germany. The competition began in March 2023 with 54 European national men's teams, from which 20 teams made it into the group stage. Of these just 16 made it into the knockout stage, and the final will be on July 14 in Berlin.

We utilized results from qualifying rounds, from the UEFA league tournament, and other international friendlies. Data was supplemented with published predictions and team ratings.

Project Approach

Data Exploration & Transformation

Machine Learning & Optimization

Predictions

Drawing Conclusions

Data Collection and Cleanup Overview

CSV concatenation :

- [Kaggle](#) :

Matches played including goals, home and away teams, stadium information, referees, and more from 1874 to current.

- [FootyStats](#) :

Expected Goals (xG) rating - *"Not all shots on goal are created equal."*

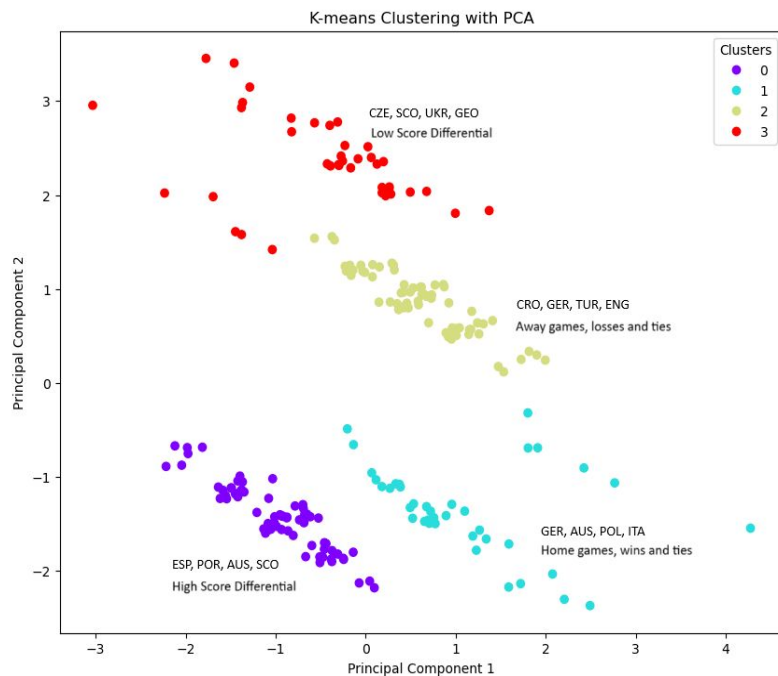
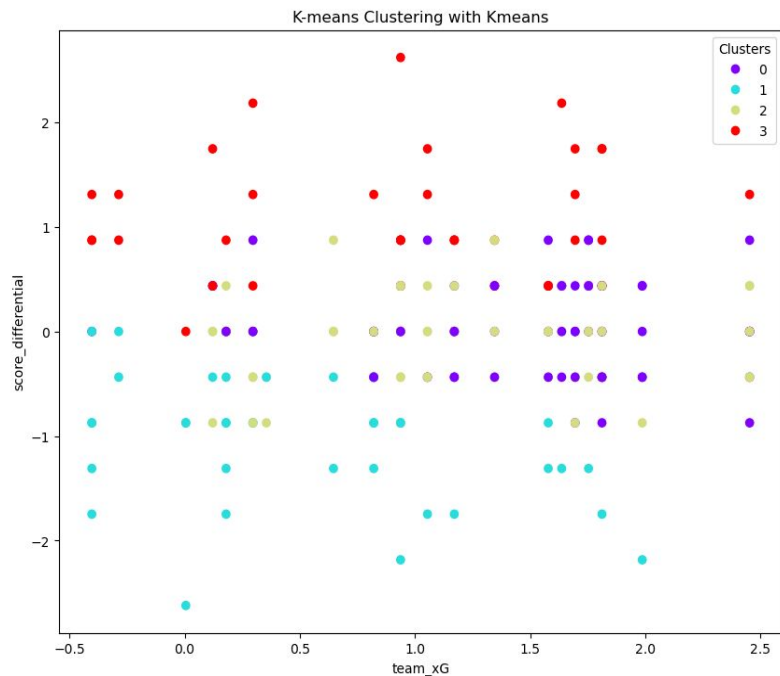
- [Oddsportal](#) :

Each team's blended odds from 10 betting websites of winning the tournament

Data Cleansing (Iterative updates)

- Change match data into team data
- Filter for recent qualifying and friendly matches
- Drop NaN and irrelevant columns
- Replace NaN values with appropriately scaled values (xG, Odds)
- Update Null values
- Create column: target_top_4
 - Initially based on highest goal differential (goals scored - goals allowed) during the qualifiers. After quarterfinals, we switched to the actual top four teams.

Unsupervised Learning



Comparison of Supervised Learning Models

Model	Precisi	Recal	F1-sco	comment
Logistic Regression	0.93	1.00	0.97	
Decision Tree Classifier	0.93	0.93	0.93	
Random Forest Classifier	0.93	0.93	0.93	
C-support Vector Classification (SVC)	0.82	1.00	0.90	<< selected for additional analysis
K-nearest Neighbors	0.78	1.00	0.88	
Gradient Boosting Classifier	1.00	1.00	1.00	Overfitted
Cross Validation (GB Classifier)			0.88	
Gradient Boosting Classifier (with emphasis on opponent_xG)	1.00	1.00	1.00	Overfitted
Stratified K-fold			0.88	

Hypertuning the SVC model

We used **GridSearchCV** to optimize tuning of model parameters:

SVC model	values tested	Optimal value
C	[0.1, 1, 10, 100]	10
kernel	['rbf', 'linear', 'poly', 'sigmoid']	sigmoid
gamma	['auto', 'scale']	scale
shrinking	[True, False]	FALSE

After tuning the SVC model scored 0.97, an increase of 0.07.

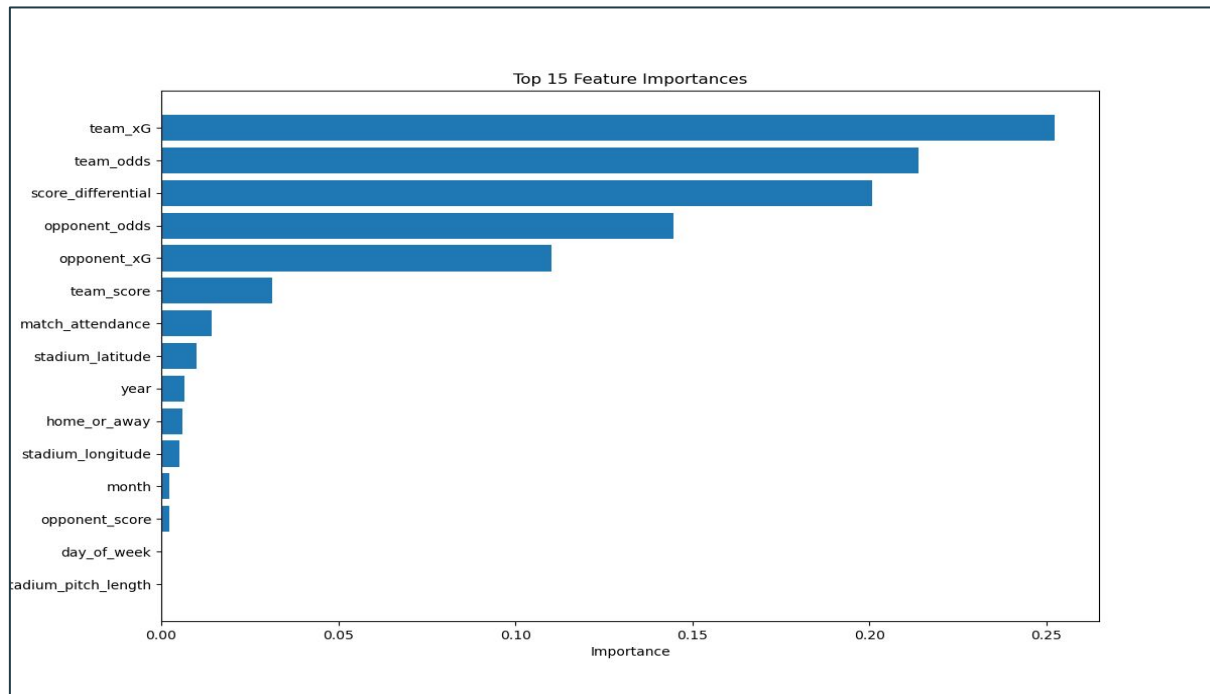
Feature Importance

- GradientBoostingClassifier

- Mean Cross-Validation

Score: 0.94

- "feature_importances_"



Prediction Models

Model	Mean Squared Error	R^2 Score
GradientBoostingRegressor	0.001	0.998
RandomForestRegressor	0.002	0.998
DecisionTreeRegressor	0.003	0.997
SVR	0.207	0.794
LinearRegression	0.120	0.881

- Regression Models vs Classifier Models
- Feature of Importance: "team_xG"
- Mean squared error and R^2 scores

Semi - Finalist

UEFA Euro 2024 Semi - Finalists			
ESP	FRA	NED	ENG

Predicted Semi - Finalists : team_xG			
GER	ITA	POR	ENG

Predicted Semi - Finalists : team_odds			
AUT	ITA	NED	ENG

Questions / Future Plans

Taking the project to the next level:

- Reshape the data to take a view of each team involved by summarizing match level data
- Rebuild the model to treat a team as a collection of individuals and collect information about the individual team members, either starters, or both starters and substitutes.
- Include additional features, such as manager experience, days since last match, a measure of player health, and referee.
- Incorporate additional teams performance data from the tournament itself, as each game is played.

Results & Conclusions

- Best feature in our models for predicting the winners, are team odds. These were collected from betting sites, and we in effect confirmed that they're a good indicator of success.
- Using past performance to predict future success in mens soccer is a difficult business as our work with the models indicated. Good for the fans that outcome is subject to many many factors, but difficult for those of us who try to pick winners.
- In the end, the models were helpful but did not accurately predict the top 4. A soccer game between top teams is full of unpredictable chances, questionable calls, problematic weather, varying levels of spectator energy, and simple human error all making a reliable prediction challenging.

PREDICTED WINNER OF EURO 2024



The team from **Spain** is the odds-on favorite to win the final on July 14.



Q&A