VISUAL ANALYTICS PROJECT

MASTER DEGREE
ENGINEERING IN COMPUTER SCIENCE

SAPIENZA UNIVERSITY OF ROME
A.Y. 2024-2025

Davide Fortunato 1936575 – Davide Mazzatenta 2140428

WHAT WE WILL COVER

1. Brief introduction
2. Related works
3. Dataset
4. Visualizations
5. Conclusions and future work

INTRODUCTION

Why this work?

- Soccer analytics gained significant importance
- Scouting is shifting to computational and analytical approaches

Our solution

- Interactive dashboard for analysis
- Dimensionality reduction and clustering algorithms
- For scouts, coaches, and fans

RELATED WORKS

COTTA, Leonardo, et al. *Using fifa soccer video game data for soccer analytics.* In: Workshop on large scale sports analytics. 2016. p. 1-4.

How does it relate to our work?

Cotta et al. demonstrate FIFA data's utility for team and style analysis, aligning with our dashboard's use of EA Sports data and, like us, they apply clustering and dimensionality reduction techniques on it. However, their work is not focused on providing an interactive dashboard but more on validate some real events through the game data.

Content summary

This paper demonstrates how **FIFA video game data** can provide empirical insights into historical soccer phenomena. It validates two iconic events: **Germany's 2014 World Cup Victory Over Brazil**: By analyzing attribute trends (2007–2014) for both national teams using **linear regression**, the authors show that German midfielders improved more in critical *offensive attributes* (e.g., finishing, shot power), while Brazilian defenders prioritized non-defensive skills (e.g., crossing), exposing tactical vulnerabilities. **FC Barcelona's Tiki-Taka Dominance**: Using **PCA and k-means clustering** on midfielders' attributes from the 2012/13 season, Barcelona's players formed a *unique tactical cluster* distinct from other teams, characterized by exceptional vision, passing, and ball control.

RELATED WORKS

A. Cao, X. Xie, R. Zhang, Y. Tian, M. Fan, H. Zhang, and Y. Wu, *Team-Scouter: Simulative Visual Analytics of Soccer Player Scouting*, IEEE Transactions on Visualization and Computer Graphics, vol. 31, no. 1, pp. 1-11, 2025.

How does it relates to our work?

- **Visual Analytics for Soccer Intelligence**:
  Both systems aim to **transform complex soccer data into actionable insights** through interactive visualizations. Your dashboard enables exploration of player attributes and league trends, while Team-Scouter simulates tactical compatibility.

- **Interactive Exploration**:
  Both tools prioritize **user-driven interaction** (e.g., filtering, brushing, linked views). However, while our *dashboard* focuses on **statistical exploration** (e.g., t-SNE clustering, temporal trends), *Team-Scouter*: Emphasizes **predictive simulation** (e.g., modeling how players adapt to new team tactics).

Content summary

Team-Scouter is a visual analytics system designed to improve soccer player scouting by simulating how players adapt to new teams. It employs a **two-level framework**:

- **Match Result Prediction**: Estimates how a player's addition impacts a team's win probability.

- **Behavior Simulation**: Predicts detailed actions (e.g., passes, tackles) in new tactical contexts.

The system provides **interactive tools** to compare a player's historical performance with simulated outcomes and explore tactical fits (e.g., visualizing how a midfielder's role changes in a high-pressing system). Validated through case studies—such as identifying replacements for Cristiano Ronaldo at Real Madrid—and expert feedback, Team-Scouter reduces transfer risks by bridging predictive modeling with visual exploration. It aids clubs in avoiding costly mismatches and optimizing squad-building strategies.

RELATED WORKS

C. Soto-Valero, *A Gaussian Mixture Clustering Model for Characterizing Football Players Using the EA Sports' FIFA Video Game System*, 2017.

How does it relates to our work?

- Similar dataset, based on FIFA video game
- Same methodology framework, which consists in applying a dimensionality reduction technique and then clustering

Content summary

This paper proposes a Gaussian mixture model-based clustering approach to classify 7,705 European football players into distinct roles using 40 attributes from EA Sports' FIFA dataset. Principal Component Analysis (PCA) reduced dimensionality, revealing two dominant components explaining 61.1% variance. A four-cluster Gaussian mixture model identified roles: defenders, midfielders, forwards, and goalkeepers, validated via Bayesian Information Criterion. A gradient boosting model ranked attributes, identifying dribbling as the most discriminative skill. Results confirmed position-specific performance indicators, with goalkeepers separable via unique attributes (e.g., reflexes) and outfield roles differentiated by technical skills. The framework demonstrates how machine learning can extract tactical insights from sports data, aiding player scouting and team strategy design.

DATASET

Main ➡ EA Sports FC 25

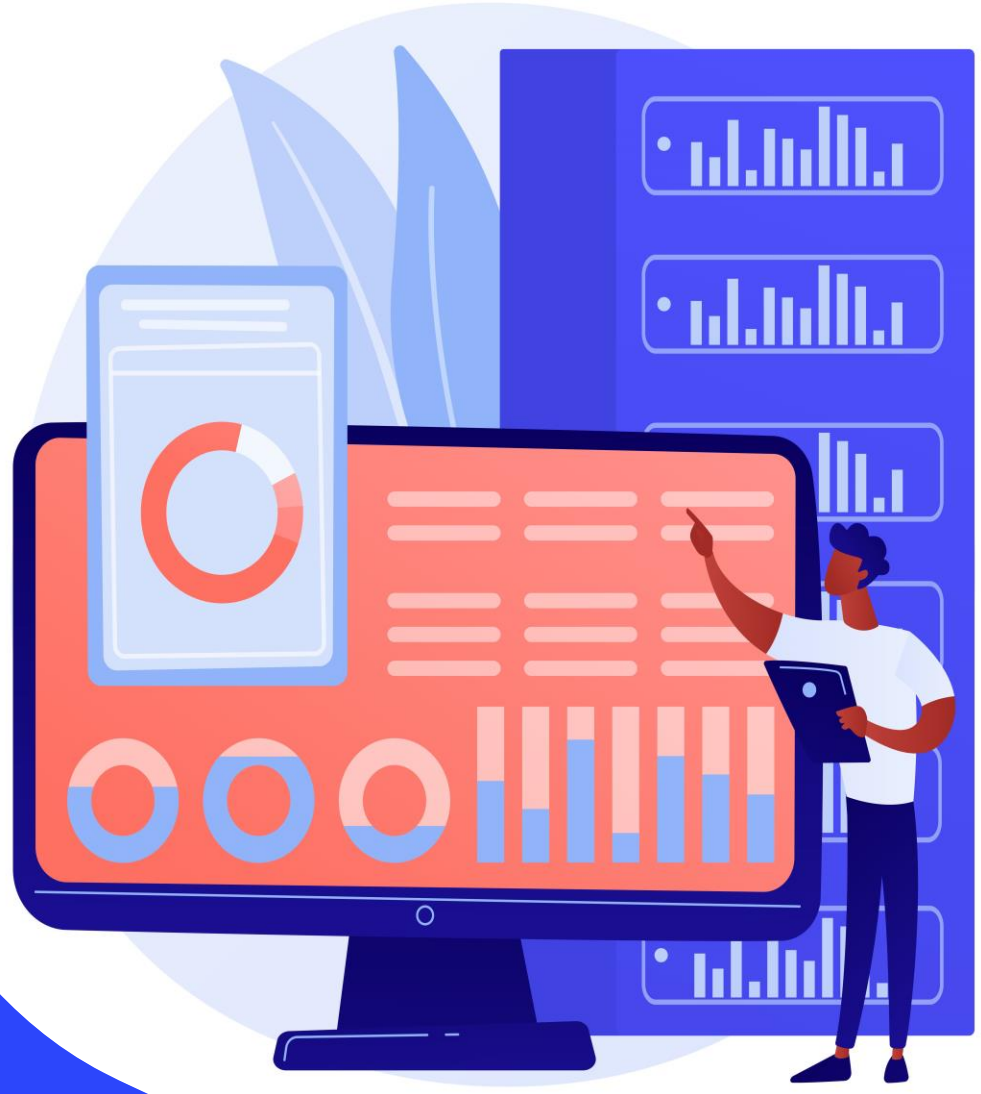
- contains data of football players in 2025
- Personal information, characteristics ratings, market values and images
- over 18000 players

Other ➡ EA Sports FC 24 to FIFA 15

- For recent versions of the game
- Used for visualization over time

DATA PRE-PROCESSING

- **Filter** players by overall bigger or equal than 65
- **Drop** 15 columns useless for future work
- Handle **missing values** by dropping or replacing
- Choose 29 **main attributes** for clustering
- **Scale** them with a standard scaler
- End up with 10878 players and 61 attributes with AS=663741



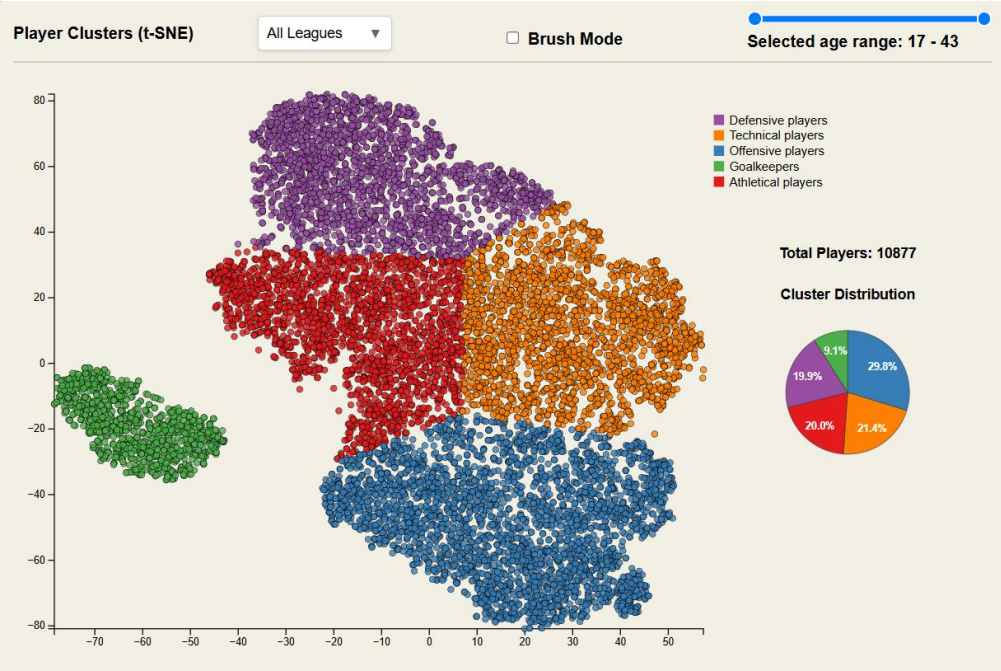
VISUALIZATIONS

- Soccer Field line up
- Scatterplot
- Player profile
- Bar Chart
- Radar Chart
- Line Chart

SOCCER FIELD LINE UP

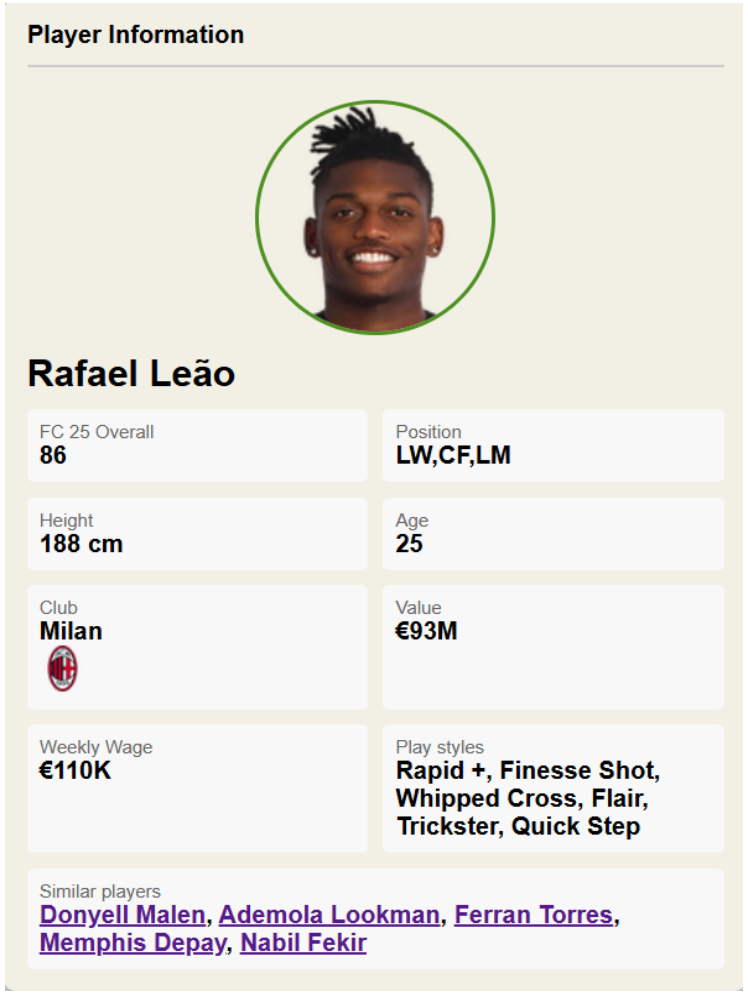
- Represents every possible role in FC 25 (15)
- Uses a qualitative colorBrewer2 palette for each major role
- Allows to filter scatterplot based on role selection
- "Reset Filter" button clears all visualization and filters applied

SCATTERPLOT



- Represent player arrangement based on dimensionality reduction and clustering
- Proximity indicates similarity among players
- Clusters are color-coded by a standard colorbrewer2 qualitative palette
- It is interactive with all the others visualizations
- 3 kinds of filters: league, age, brush
- Pie chart shows distributions

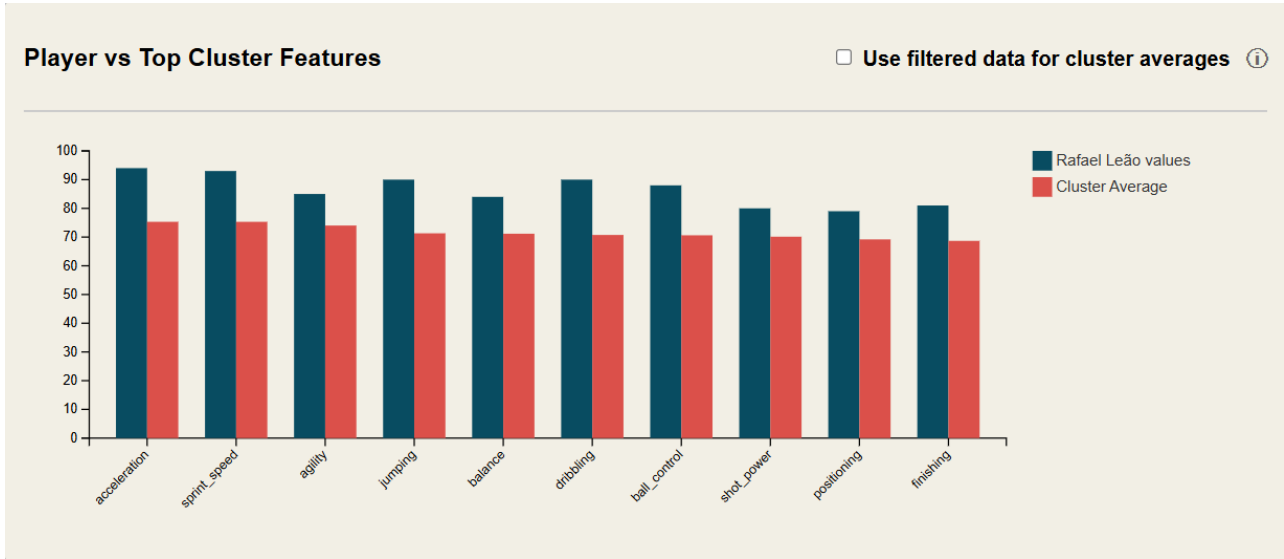
PLAYER PROFILE



- Displays a complete player profile
- Overall rating, positions, club with its logo
- Personal and financial info
- Play styles
- Top 5 most similar players

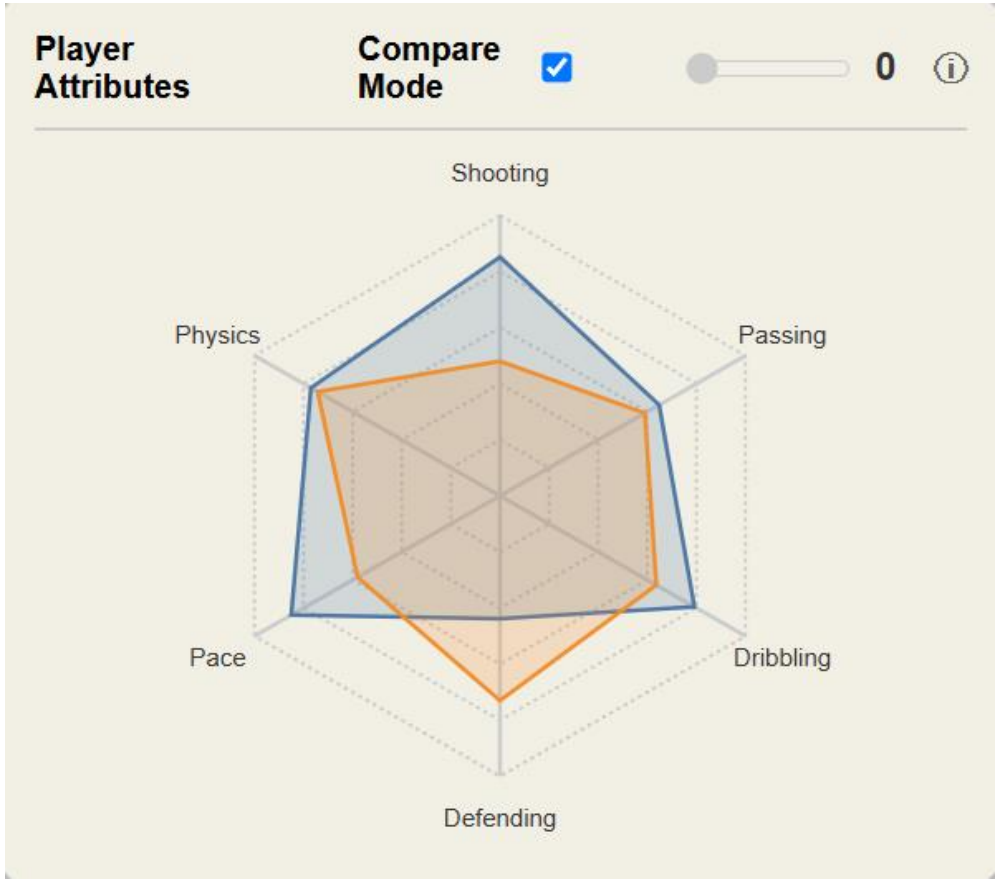
BAR CHART

- Compare a player with a cluster
- Shows top 10 attributes with highest cluster averages
- X-axis has a scale from 0 to 100, Y-axis is dynamic
- 2 color palette from Coolors website
- checkbox allows to visualize averages of filtered players
- Tooltip and infobox



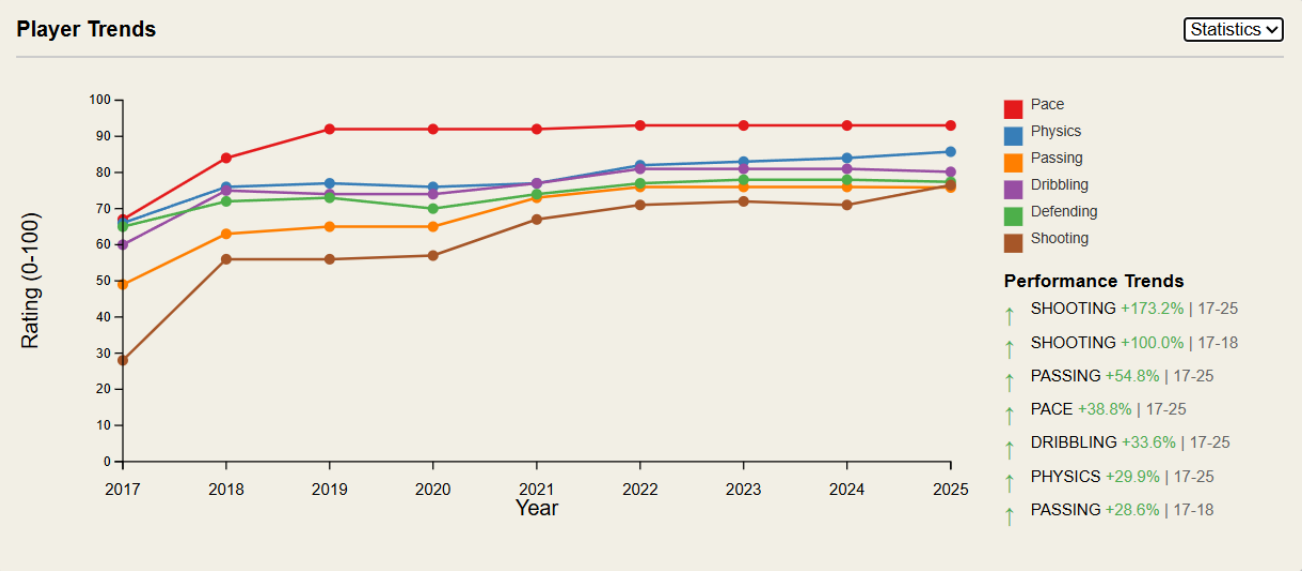
RADAR CHART

- Represents areas of player main attributes
- 5 for GK and 6 for others
- Uses d3.js standard palette
- Tooltip and infobox
- Compare Mode with another player
- Top 5 nearest players with slider



LINE CHART

- Tracks player financial and performance evolution over time, from 2015 to 2025
- Menu with 3 metrics
- X and Y axis are dynamic
- Colors same as scatterplot
- Trend explainer
- Tooltip and legend



CONCLUSIONS

- Data driven analysis with intuitive interface
- Support to soccer analytics
- Tools for different evaluations

FUTURE WORKS

- Integrate advanced stats from real matches
- Line-up simulator
- Predict characteristics with ML
- Team comparisons
- Advanced clustering techniques

THANKS FOR YOUR ATTENTION

GitHub

https://github.com/davidef24/VisualAnalytics