Extending Adila to Professional Football Data: A Case Study with FIFA23

Javiera López, Javiera Azócar Oliva, Pablo Poblete, Gabriel Acevedo

Motivation

The chances to win a football match increase when the right tactics are chosen and the opponent's behavior is well anticipated. Similarly, automatically generating teams with complementary skills is valuable for complex tasks, yet the huge number of possible attribute combinations challenges traditional algorithms, and even deep learning models may overlook fairness (Dashti et al., 2022). To address this, we developed a structured, scalable dataset that captures multiple participant attributes, supporting fair and effective team generation in data-driven environments.

Introduction

Team formation can be seen as a kind of recommendation problem: instead of suggesting items to a user, we recommend a group of people with the right skills to accomplish a task. Thus, teams need members with varied and complementary skills. However, most past methods have been domain-specific with their own datasets, making it difficult to compare or extend results (Dashti et al., 2022). To enable large-scale team recommendations, frameworks like OpeNTF have been recently developed. OpeNTF is an open-source system that trains neural networks on large parallel datasets, using standard metrics for evaluation. It uses abstract "team" classes to model different scenarios, such as research teams, movie casts, or patent inventor groups. Our main contribution is extending these tools with new data and practical validations in a non-traditional domain. We applied the Adila and OpeNTF models to a sports context by creating an adapted dataset and testing team recommendation algorithms under realistic conditions.

Dataset

Adila uses the file teams.pkl that organizes teams as structured lists of instances, containing player information, skills, and metadata relevant to use with Adila. This structure facilitates experimentation with team formation models and performance simulation in research environments.

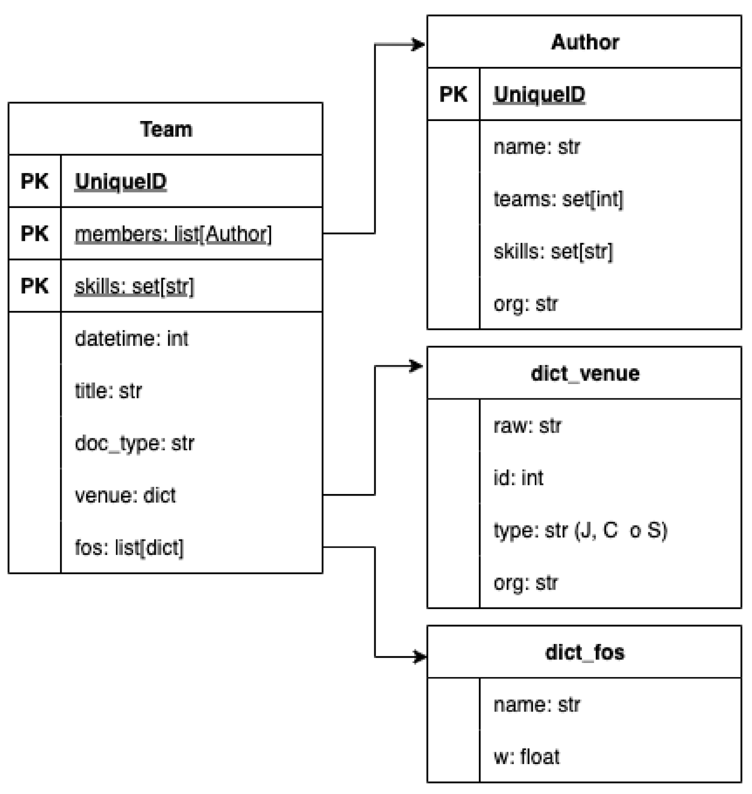

Figure 1: Diagram of the data in teams.pkl

Model & Results

Main Pipeline Description: The pipeline operates as follows for each `.pred` file (base model predictions): (1) Loads base team predictions from the file `fpred`, generated by an initial recommendation model. (2) Loads team-related information from `fteamsvecs`, such as vector-based team representations. (3) Loads data split definitions (training, validation, and test sets) from `fsplit`. (4) Applies a fairness-aware reranking algorithm (e.g., `det_cons`, a deterministic method with fairness constraints) to adjust the team recommendations. (5)Evaluates the results before and after reranking in terms of both utility and fairness metrics.

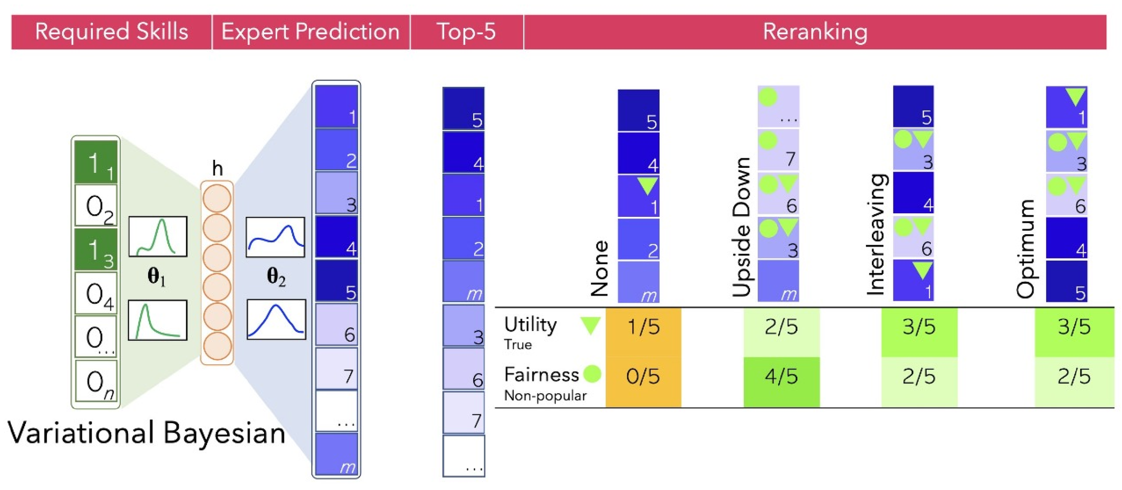

Figure 2: Adila Pipeline

Statistic	Value
#Publications (teams)	4,877,383
#Unique Authors (members)	5,022,955
#Unique Fields of Study	89,504
Avg. #Authors per Publication	3.06
Avg. #FOS per Publication	8.57
Avg. #Publications per Author	2.97
Avg. #FOS per Author	16.73
#Publications without FOS	?
#Publications with Single Author	768,956
#Publications with Single Skill	5,569

Table 1: Overall statistics of the Adila dataset

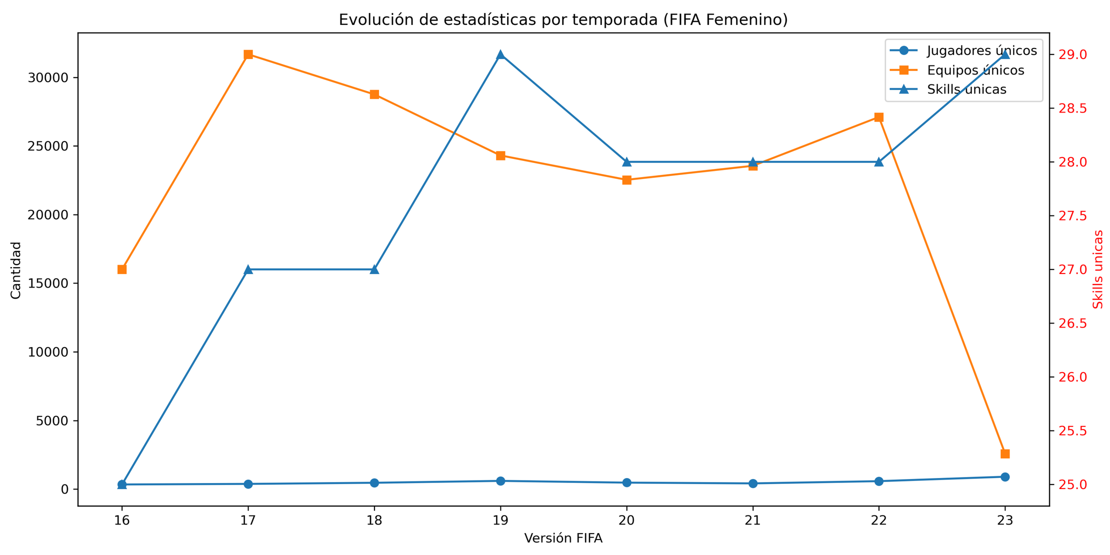

Figure 3: Seasonal trends for female players

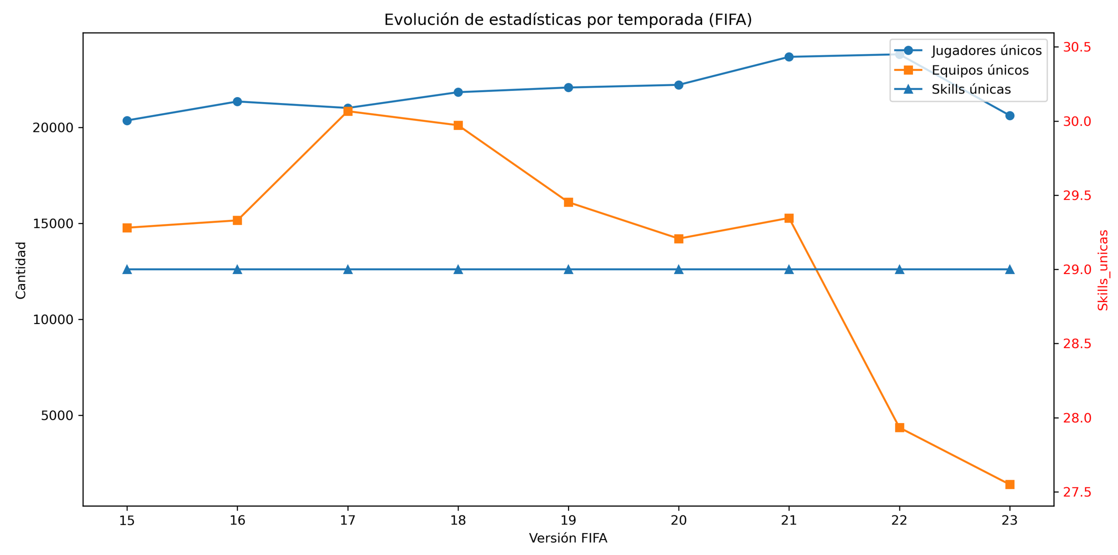

Figure 4: Seasonal trends for male players

Despite the high number of unique players per season, the number of represented teams drops sharply in recent versions. This reinforces the methodological decision to work only with the latest versions (FIFA 22–23), which present a lower data volume but a more consolidated structure.

FIFA Version	#Teams	#Players	#Skills
15	14,774	20,361	29
16	15,156	21,347	29
17	20,842	21,010	29
18	20,111	21,831	29
19	16,101	22,074	29
20	14,206	22,214	29
21	15,281	23,672	29
22	4,365	23,803	29
23	1,403	20,621	29

Table 2: Statistics Male Players

FIFA Version	#Teams	#Players	#Skills
16	16,003	336	25
17	31,676	377	27
18	28,756	460	27
19	24,309	597	29
20	22,533	470	28
21	23,552	415	28
22	27,107	577	28
23	2,584	899	29

Table 3: Statistics Female Players

Conclusions and Future Work

In this project we have worked on two main lines we have adapted Adila, a team building system just originally designed for experts, to a sports context: the creation of soccer teams. Hand in hand with this comes the second line that consisted in the elaboration of a dataset of soccer players, this second part was the main contribution, since it lays the foundations for future applications and research, it is interesting the applicability of the models in this dataset, since it can be evaluated how it behaves when gender bias may conflict with technical performance. So, the next steps may be to:

- Train Adila and OpeNTF model with the elaborated dataset
- Explore fairness in other attributes (age, nationality, position).

References

[1] Dashti, A., Saxena, K., Patel, D., Fani, H. OpeNTF: A Benchmark Library for Neural Team Formation. International Conference on Information and Knowledge Management, Proceedings, 3913–3917, 2025. https://doi.org/10.1145/3511808.3557526

[2] Higuera Cabañes, C., Creatura, M., Bercowsky, M. BBVA AI Factory | Fairness in AI: How can we address bias to build equitable systems? l, 2025. https://www.bbvaaifactory.com/es/fairness-in-ai-how-can-we-address-bias-to-build-equitable-systems/

[3] Fani Lab. OpeNTF Tutorial – UMAP24 Dataset, 2024. Presentado en ACM UMAP 2024. Recuperado de https://fani-lab.github.io/OpeNTF/tutorial/umap24/

[4] Fani, H. Tutorial on Collaborative Team Recommendation for Skilled Users: Objectives, Techniques, and New Perspectives. Fani's Lab!, School of Computer Science, University of Windsor, 2024. Presentado en ACM UMAP 2024. Recuperado de SlideShare.

[5] Loghmani H., Fani H., Rueda G., Paul E., Lakshmi Y. Moasses R. Adila: Fairness-Aware Team Formation. WorkShop Collaborative Team Recommendations for Skilled Users, 2022 https://github.com/fani-lab/Adila.git