Usecase 7.2 - (Lab)

By: eng. Esraa Madhi

So we defined data science as: It's the process of asking interesting questions, and then answering those questions using data.

For any **Data project** we will go through these steps:

1. Defining the Problem Statement
2. Collecting Data
3. Data Quality Checking and Remediation
4. Exploratory Data Analysis
5. Building Machine Learning Models
6. Model Evaluation
7. Communicating Results
8. Model Deployment
9. Model Performance Maintenance in Production

Usecase 7.2

Step 1: Defining the Problem Statement

During our previous engagement with this dataset, we constructed a predictive model capable of estimating football players' categories. These categories were the result of feature engineering and statistical analysis to define their variables. Our current objective is to refine these categories or clusters by incorporating both the basic information of the players and professional statistics, particularly focusing on their market value during inter-club transfers.

Step 2: Collecting Data

Football players dataset was collected for this lab encompassing player demographics (age, height, playing position) and performance metrics (goal scoring, assists, injury history) for the seasons 2021-2022 and 2022-2023.

Step 3: Data Quality Checking and Remediation

Step 4: Exploratory Data Analysis

Done in the notebook

Step 5: Building Machine Learning Models

Step 6: Model Evaluation

- Feature Engineering: Apply feature engineering techniques to create new features or modify existing ones such as :
  - Encode categorical variables
  - Normalize or standardize numerical features.
- Model Training: Train at least 2 unsupervised models from your choice on the training dataset. Ensure that you have a separate validation set or employ cross-validation to assess model performance during training.
- Performance Metrics: Use appropriate performance metrics to evaluate your models.

- Hyperparameter Tuning: Fine-tune the hyperparameters of each model to optimize performance.

Step 7: Communicating Results

When conducting the analysis and building the predictive model, it is crucial to maintain clear and comprehensive documentation. This will not only facilitate understanding and reproducibility of the work but also allow others to follow and build upon your methodology. Use markdown cells in your notebook to provide detailed commentary on the following key aspects:

- Feature Engineering: Outline the steps taken to manipulate or create new features to improve model performance.
- Hyperparameter Optimization: Explain the process and methods used to fine-tune model hyperparameters.
- Performance Metric Visuals: Include charts or graphs that illustrate the performance of your model across various metrics.
- Feature and Prediction Insights: Offer an interpretation of how different features influence the model's predictions.

Step 8: Model Deployment

Implement the deployment of our most proficient model as follows:

- Construct a FastAPI endpoint to serve the model
- Develop a Streamlit application that features:
  a. A visualization of the model's clustering results
  b. An interface that interacts with the previously established FastAPI to predict football players' categories.

Step 9 : Model Performance Maintenance in Production

Not applicable