

Usecase 7.1 - (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 7.1

Step 1: Defining the Problem Statement

We are interested in developing a robust predictive model that can accurately estimate the category of football players based on a comprehensive set of variables. This model aims to leverage both basic player information and detailed professional statistics to determine a player's market value category when they are transferred between clubs.

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.
 - **Modify target variable to train the second model**
- Model Training: Train **the selected** models 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 the models like Accuracy, Precision, Recall, F1 Score, and the Confusion Matrix.
- Hyperparameter Tuning: Fine-tune the hyperparameters of each model to optimize performance. Utilize techniques like grid search.
- Model Validation: Validate the model's performance using the test set to ensure that the model generalizes well to unseen data.
- Overfitting Check: Ensure that models are not overfitting by comparing training and validation performance.
- Underfitting Check: Ensure that models are not underfitting by assessing if the performance is significantly poor both on training and validation sets.

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

Not applicable

Step 9 : Model Performance Maintenance in Production

Not applicable