CISB 60 – ML and DL (Fall, 2024)

Final Project: Predicting FIFA Player Potential

Introduction

This project aims to predict the potential of FIFA players using machine learning and deep learning techniques. By analyzing key player attributes such as physical characteristics, skill metrics, and overall ratings, the project provides a data-driven approach to identifying high-potential players.

Objective

The primary goal of this project is to build predictive models for player potential using:

1. A **Random Forest Regressor** (Machine Learning).

2. A **Deep Neural Network** (Deep Learning).

These models will analyze patterns in the data to assist football clubs in making informed scouting decisions, thereby optimizing resource allocation and reducing risk.

Dataset Description

- **Source**: FIFA dataset containing detailed player statistics.

- **Structure**:

    o Total Records: 17,954

    o Total Columns: 51

- **Key Features**:

    o Player name, age, height, weight, overall rating, potential, skill metrics, and preferred foot.

- **Preprocessing Steps**:

    o Missing values were handled.

    o Categorical variables were encoded.

    o Numerical features were scaled.

Models Used

Model 1: Random Forest Regressor

- Justification: Robust to overfitting and effective for non-linear regression tasks.

- Metrics: Mean Squared Error (MSE), R^2 Score.

- Results: The model achieved an R^2 score of ~0.94, indicating excellent performance.

Model 2: Deep Neural Network

- Architecture:
    - Input Layer: Includes all player features.
    - Hidden Layers: Fully connected layers with ReLU activation and Dropout regularization.
    - Output Layer: Linear activation for regression.

- Metrics: Mean Absolute Error (MAE), Loss.

- Results: The network effectively modeled player potential with a low MAE of ~2.5.

Conclusion

The project successfully demonstrates the application of machine learning and deep learning techniques in predicting FIFA player potential. It highlights the value of data-driven decision-making in the sports industry. Key findings include:

- Player attributes such as value_euro, ball_control, and dribbling have significant influence on potential.

- Both Random Forest and Deep Neural Networks are effective in this domain, with high accuracy and low error rates.