TalentFinder

**Table of Contents**

1. Introduction

2. Installation

3. Configuration

4. Deployment

5. Directory Structure

# 1. Introduction TalentFinder is a project aimed at analyzing football data and evaluating player talents based on collected information from various sources, such as sofascore and transfermarkt.

# 2. Installation

To run the project locally, follow these steps:  
  
1. Clone the repository:  
*git clone https://github.com/klaudiamarciniakk/TalentFinder.git  
cd TalentFinder*  
  
  
2. Install the required libraries::  
*pip install -r requirements.txt*

# 3. Configuration

Before running the machine learning scripts, ensure that you have the appropriate data in the data directory. If not, use the data fetching and updating scripts. The obtained data should be in the correct format as indicated in the csv files in the flags, players, sofascore, and transfermarkt subdirectories. Also, make sure you can connect to the database. To do this, you must have access to the services of the Faculty of Mathematics and Computer Science at Adam Mickiewicz University via VPN. The instructions can be found below: https://laboratoria.wmi.amu.edu.pl/uslugi/vpn/

# 4. Deployment

Below are examples of how to run the main project scripts.

**Running data collection scripts**

**1. Collecting player information from the transfermarkt portal:**  
*python scrappers/transfermarkt/transfermarkt\_scrapper.py*  
  
**2. Fetching player ratings from the sofascore portal:**  
*python scrappers/sofascore/scrapper\_sofascore.py*

## Data processing

**1. Generating part 1 of flags:**  
*python database/add\_flags.py*

**2. Generating part 2 of flags:**  
*python database/flags\_KM.py*

**3. Merging CSV files with flags:**  
Running jupyter notebook: *database/merge\_two\_csv\_with\_flags.ipynb*

**Model training**

**1. Training using random forest with classification:** Running jupyter notebook: *ml/* *clasification /* *Clasification\_transfer.ipynb*

**2. Training using linear regression:**  
Running jupyter notebook: *ml/* *regresion /* *Regresion\_transfers.ipynb*

**3. Running jupyter notebook:**  
Running jupyter notebook: *ml/* *random\_forest /* *random\_forest.ipynb*

# 5. Directory Structure

TalentFinder/  
├── data/  
│ ├── flags/  
│ ├── players/  
│ ├── sofascore/  
│ └── transfermarkt/  
├──documentation/  
├──editdata/  
├── ml/

│ ├── clasification/ │ ├── random\_forest/ │ ├── regresion/  
├── scrappers/  
│ ├── sofascore/  
│ └── transfermarkt/  
├── readme.md └──requirements.txt