ITEC 4700 – Final Project

Predicting Soccer Players' Value Using FIFA 23 Statistics

Spring 2023

Created by:
Bradley Iversen & Mahmood Shukor

TABLE OF CONTENTS

01

INTRODUCTION

Team members, into to the Project, Dataset, and Technologies used.

About the Project

The objective of this project is to predict soccer players' value based on their FIFA 23 statistics. These statistics include players' information, and players' in-game ratings.
To do this, we test multiple models that we learned in class to determine which model works best with the dataset. In addition, we will also work to find which variable has the most weight when determining soccer players' value.

Data Collections

The data is collected from Kaggle: Fifa 23 Players Dataset.
The raw data consist of 18,539 rows and 89 attributes.
These attributes include players' personal information (Name, Age, Height, Weight, Nationality), players' value and wage, & in-game statistics (overall, position, stats).

	Known As	Full Name	Overall	Potential	Value(in Euro)	Positions Played	Best Position	Nationality	Image Link	Age	Height(in cm)	Weight(k
0	L. Messi	Lionel Messi	91	91	54000000	RW	CAM	Argentina	https://cdn.sofifa.net/players/158/023/23_60.png	35	169	
1	K. Benzema	Karim Benzema	91	91	64000000	CF,ST	CF	France	https://cdn.sofifa.net/players/165/153/23_60.png	34	185	
2	R. Lewandowski	Robert Lewandowski	91	91	84000000	ST	ST	Poland	https://cdn.sofifa.net/players/188/545/23_60.png	33	185	
3	K. De Bruyne	Kevin De Bruyne	91	91	107500000	CM,CAM	CM	Belgium	https://cdn.sofifa.net/players/192/985/23_60.png	31	181	
4	K. Mbappé	Kylian Mbappé	91	95	190500000	ST,LW	ST	France	https://cdn.sofifa.net/players/231/747/23_60.png	23	182	

Technologies

1. Python - The programming language
2. Python Libraries:
    a. Pandas - Load and Manipulate DataFrame
    b. Numpy
    c. MatPlotLib - Generate Visualizations
    d. Scikit-learn - Apply Machine Learning Models
    e. XGBoosting
    f. Neural Network Keras
3. Kaggle - Dataset
4. Jupyter Notebook - Python Notebook
5. GitHub Repository - File sharing & collaboration online

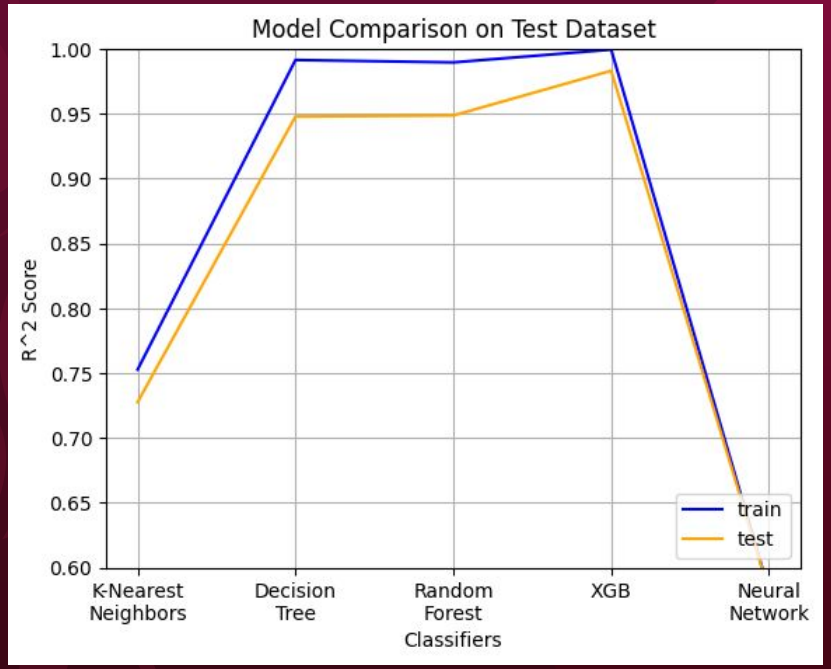
02 DATA MODELING

Process	Description
Data Acquisition and Exploration	The first step is to acquire the data and explore it to gain insights into the data's structure, quality, and characteristics.
Data Preprocessing	Once the data has been acquired and explored, it needs to be preprocessed before it can be used for modeling. The preprocessing steps include removing duplicates, handling missing values, scaling the features, and encoding categorical variables.
Feature Selection	The next step is to select the relevant features for the model. This is done by analyzing the correlation between the features and the target variable.
Data Split	And the last step of the data modeling is splitting the data set into a training set (80%) and a test set (20%). The training set is the subset of data used to train the model. The test set is the subset of data used to test (i.e. evaluate) the trained model.

03

ANALYSIS & RESULTS

R^2 Score on All Model



KNN

Train: 0.753
Test: 0.728

Decision Tree

Train: 0.99
Test: 0.947

Random Forest

Train: 0.99
Test: 0.948

XGBoosting

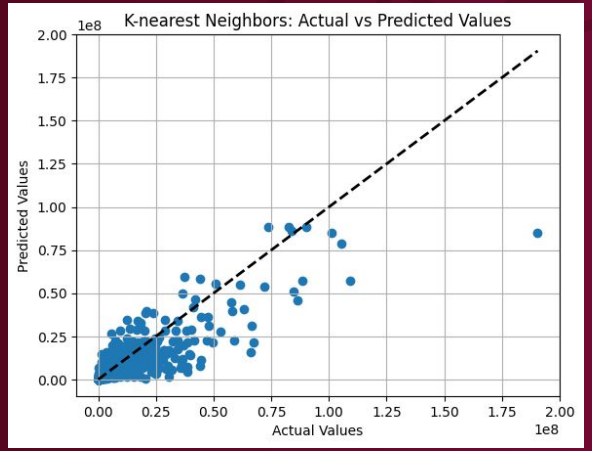
Train: 0.999
Test: 0.983

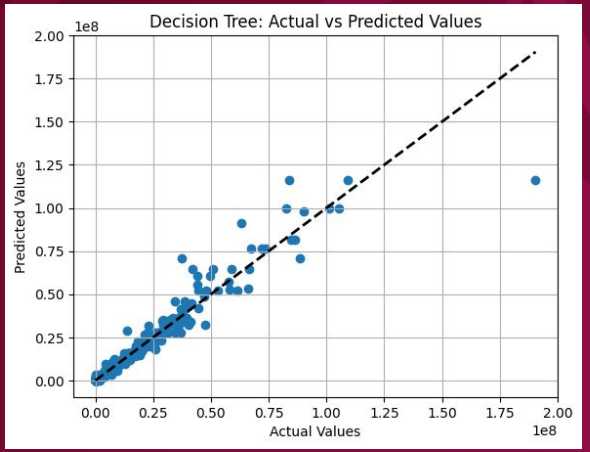
XGBoosting generates the most highest R^2 score.

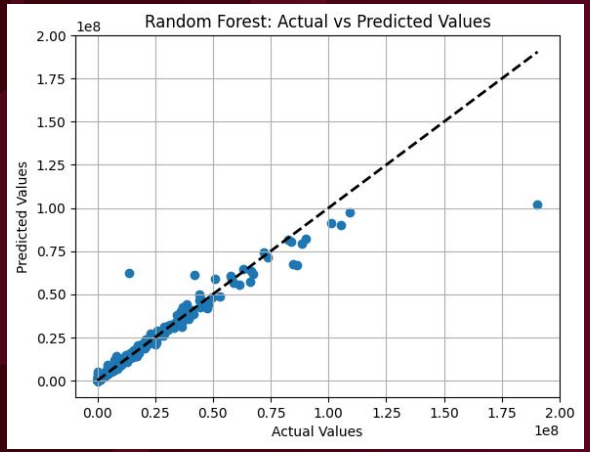
Neural Network

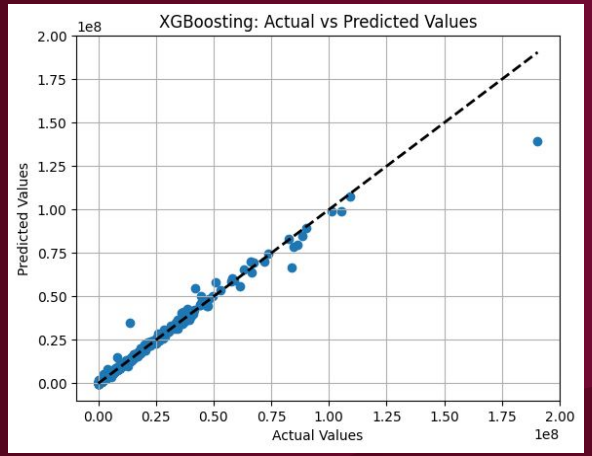
Train: 0.58
Test: 0.58

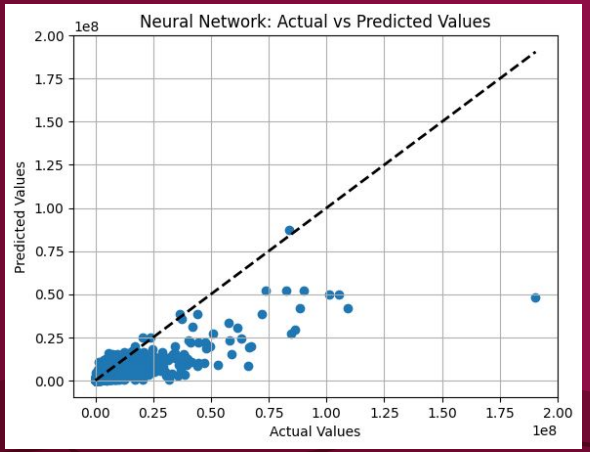
Actual vs Predicted Values Plots



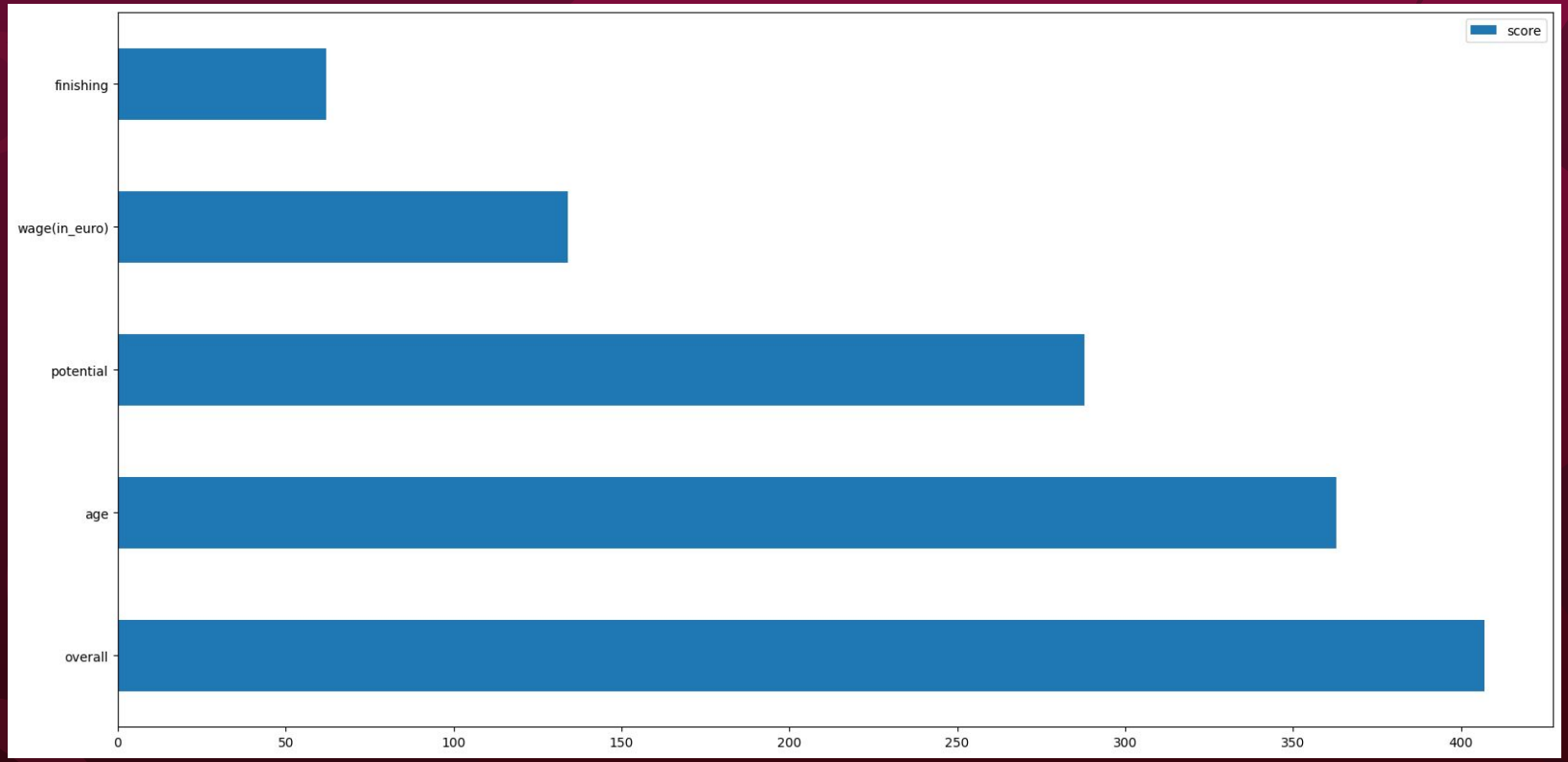








XGBoosting Best Features



Our GitHub Page: GitHub Repository: ITEC4700_FinalProject

References

[1] Singh Naik, S. (2022). FIFA 23 Players Dataset. Retrieved from Kaggle: https://www.kaggle.com/sanjeetsinghnaik/fifa-23-players-dataset

[2] Sklearn documentation. (n.d.). Retrieved from https://scikit-learn.org/stable/documentation.html

[3] Keras documentation. (n.d.). Retrieved from https://keras.io/

[4] Transfermarkt. (n.d.). Retrieved from https://www.transfermarkt.com/

[5] Ding, Yan. ITEC 4700 Lecture Notes.

THANKS!

DO YOU HAVE ANY QUESTIONS?