Football Player Transfer Prediction

Group No:1
Nafis Islam (14.02.04.014)
Hafizul Islam Himel (14.02.04.024)

February 14, 2018

1 Introduction

In professional football, a transfer is the action taken whenever a player under contract moves between clubs. It refers to the transferring of a player's registration from one association football club to another.Players are only allowed to move form one club to another club only during transfer window and also by the law. When players transfer form one club to another club some sort of compensation is paid which is called transfer fee. When a player moves from one club to another club he terminates his old club contract and sign another contract with the new club. In some cases, however, transfers can function in a similar manner to player trades, as teams can offer another player on their squad as part of the compensation.

We all know that every year there are two transfer windows. One is summer transfer window and the other one is winter transfer window.So, our model gives us prediction whether a transfer happened or not in this winter of 2017/2018. We all know that transfer had been done of a player depending on his performance. We collect the player's six months (June-December) labeled data for training. Through this data we predict whether he moved to a new club or not.

2 Dataset

2.1 Abstract

This dataset is build to predict whether a transfer is happened or not. Usually transfer is done by the performance of a player. This dataset contains two classes, where 0 is for no transfer 1 is for transfer. Each classes contains 50 instances. In this dataset transfer is predicted of a player based on the performance of six months (June - January) 2017/2018.In this dataset there are thirteen features including the predictive class [1].

Table 1: Attributes

No	**Attributes**
1.	Player
2.	Age
3.	Matches
4.	Goals
5.	Own_Goals
6.	Assists
7.	Yellow_cards
8.	Second_Yellow
9.	Red_Cards
10.	Subs_on
11.	Subs_off
12.	Price
13.	Transfer(predictive attribute)

2.2 Selected Attributes

For getting higher accuracy we reduced some features from the dataset. The selected features are:

Table 2: Selected Features

	Attributes
1.	Player
2.	Matches
3.	Goals
4.	Assists
5.	Yellow_Cards
6.	Subs_Off
7.	Price
8.	Transfer(predictive attribute)

3 Models

We've used 5 models in this dataset to see the which one gives better accuracy & result.As we used K-Fold Cross Validation with k=5, we divided our data set into 5 slices and ran classifier on different slices exactly 5 times and we kept lists to keep the results of each classifier.

3.1 Random Forest Classifier

Random forest classifier is an ensemble learning method for classification, regression and other tasks, that operate by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees. Random decision forests correct for decision trees' habit of overfitting to their training set [2].

3.2 Decsion Tree Classifier

Decision Trees (DTs) are a non-parametric supervised learning method used for classification and regression. The goal is to create a model that predicts the value of a target variable by learning simple decision rules inferred from the data features.For instance, decision trees learn from data to approximate a sine curve with a set of if-then-else decision rules. The deeper the tree, the more complex the decision rules and the fitter the model [3].

3.3 Support Vector Classification

In machine learning, support vector machines (SVMs, also support vector networks) are supervised learning models with associated learning algorithms that analyze data used for classification and regression analysis. Given a set of training examples, each marked as belonging to one or the other of two categories, an SVM training algorithm builds a model that assigns new examples to one category or the other, making it a non-probabilistic binary linear classifier [4].

3.4 K-Nearest Neighbors Classifier

K-nearest neighbors algorithm (k-NN) is a non-parametric method used for classification and regression.[1] In both cases, the input consists of the k closest training examples in the feature space. The output depends on whether k-NN is used for classification or regression.In k-NN classification, the output is a class membership. An object is classified by a majority vote of its neighbors, with the object being assigned to the class most common among its k nearest neighbors (k is a positive integer, typically small). If k = 1, then the object is simply assigned to the class of that single nearest neighbor.In k-NN regression, the output is the property value for the object. This value is the average of the values of its k nearest neighbors [5].

3.5 Gradient Boosting Classifier

Gradient boosting is a machine learning technique for regression and classification problems, which produces a prediction model in the form of an ensemble of weak prediction models, typically decision trees. It builds the model in a stage-wise fashion like other boosting methods do, and it generalizes them by allowing optimization of an arbitrary differentiable loss function [6].

4 Performance Scores of Models

4.1 Random Forest Classifier

Avg Accuracy	90%
Avg Precision	92%
Avg Recall	90%
Avg F1 Score	89%

4.2 Decision Tree Classifier

Avg Accuracy	82%
Avg Precision	84%
Avg Recall	82%
Avg F1 Score	81%

4.3 Support Vector Classifier

Avg Accuracy	89%
Avg Precision	91%
Avg Recall	88%
Avg F1 Score	88%

4.4 K-NN Classifier

Avg Accuracy	91%
Avg Precision	92%
Avg Recall	91%
Avg F1 Score	90%

4.5 Gradiant Boosting Classifier

Avg Accuracy	86%
Avg Precision	88%
Avg Recall	86%
Avg F1 Score	85%

5 Discussion

As we can see that the KNN classifier shows most accuracy regarding this dataset, as KNN classifier can capture much more complex relationships between our data points. After KNN classifier Random Forest classifier also gives good accuracy for this dataset. On the other classifiers, we get low accuracy like Decision tree classifier, as we think our feature engineering was not top notch and we have limited amount of data.

References

[1] Transfermarkt. Transfermarkt, 2017.

[2] Wikipedia. Random forest — wikipedia, the free encyclopedia, 2017.

[3] Scikit Learn. Decision tree classifier, 2017.

[4] Wikipedia. Support vector machine — wikipedia, the free encyclopedia, 2017.

[5] Wikipedia. K-nearist neighbour — wikipedia, the free encyclopedia, 2017.

[6] Wikipedia. Gradient boosting classifier — wikipedia, the free encyclopedia, 2017.