



Ahsanullah University of Science and Technology (AUST)

Department of Computer Science and Engineering

Match Result Prediction

Project Report

Course No.: CSE4108

Course Title: Artificial Intelligence Lab

Submitted To-

Mr. Md. Siam Ansary

Lecturer, Department of CSE, AUST.

Submitted By-

Kaho Fardin Hasib	170204063
Md. Farhan Fuad	180104082
Abdullah AlMohaimen	180104098

Introduction

Match Result Prediction is a process where we are analysing the datas of football matches played by all the teams of English premier league and predicting results by using ML techniques.

Brief Description of the dataset:

Name	Match Result Prediction
File format	.CSV
Dimension	366*10
No. of Columns	10
No. of Rows	366
No of feature columns	9
Name of features	H Team, A Team, FTHG, FTAG, HTGS, HTFormPts, ATFormPts, FTR, Date
Target	Win Prediction

Description of the features:

H Team: A team which is playing on it's own ground.

A Team: A team playing away from it's own ground.

FTHG: It is the goal scored by a home team in a full match.

FTAG: Goal scored by the away team in a full match.

HTGS: Half time goal shoot.

HTFormPts: All the points a team acquired as a home team.

ATFormPts: All the points a team acquired as a away team.

FTR: Full time result.

Date: Date of the matchday.

Description of the used ML models:

- **Decision Tree Classifier** – It is a tree structured classifier, where internal nodes represent the features of a dataset, branches represent the decision rules and each leaf node represents the outcome. In a decision tree, there are two nodes, which are the Decision node and leaf node.
- **Naive Bayes** - a simple technique for constructing classifiers: models that assign class labels to problem instances, represented as vectors of feature values, where the class labels are drawn from some finite set.
- **KNN Classifier** - an algorithm that uses data and classify new data points based on similarity measures.

- **Support Vector Machine** - Support vector machines (SVMs) are a set of supervised learning methods used for classification, regression and outliers detection. The advantages of support vector machines are: Effective in high dimensional spaces.
- **Random Forest** - Random forest is a *Supervised Machine Learning Algorithm* that is *used widely in Classification and Regression problems*. It builds decision trees on different samples and takes their majority vote for classification and average in case of regression.

Performance scores of the models

The accuracy of the model is given below:

Mode	Accuracy Score	Precision Score	Recall Score
Decision Tree Classifier	100%	Home: 1 Away: 1	Home: 1 Away: 1
Naive Bayes	86.3%	Home: 0.94 Away: 0.80	Home: 0.79 Away: 0.94
KNN Classifier	71.23%	Home: 0.73 Away: 0.7	Home: 0.63 Away: 0.79
Support Vector Machine	0.64	Home: 0.76 Away: 0.83	Home: 0.91 Away: 0.39
Random Forest	98.63	Home: 1 Away: 0.97	Home: 0.97 Away: 1

Conclusion

We can see that the models performed better on the model of Random Forest and giving near to accurate prediction for different level of random state, though for Naive Bayes, output value fully accurate, maybe because of bug. Though the recall score is questionable for some models as it given output value as , other scores are pretty much good. So, we can say Classification models are most suitable for this dataset. Random Forest, KNN and Decision Tree Classifier are almost similar when comes to accuracy.

References:

1. <https://www.premierleague.com/tables>