Ahsanullah University of Science and Technology (AUST)

Department of Computer Science and Engineering

Bangladesh Cricket Match Result Analysis before and after Covid-19 Period and Win Prediction

Project Report

Course No.: CSE4108

Course Title: Artificial Intelligence Lab

Submitted To-

Mr. Md. Siam Ansary

Lecturer, Department of CSE, AUST.

Submitted By-

Ishmam Faruki : 180104026

Nipun Paul: 180104042

Mutakabbirul Islam: 180104044

Introduction

Bangladesh Cricket Match Result Analysis before and after Covid-19 Period and Win Prediction is a process where we are analysing the datas of cricket games played by Bangladesh Cricket Team all over the world and predicting win by using ML techniques.

Brief Description of the dataset:

Name	Bangladesh-national-cricket-team-win-by-margin-against-opponents
File format	.CSV
Dimension	617x8
No. of Columns	8
No. of Rows	617
No of feature columns	7
Name of features	Winner, Result, Margin, BR, Match, Ground, Start Date, Covid
Target	Win Prediction

Description of the features:

Winner : Name of the winning team.

Result: Match result, it could be win, lose or tie.
Margin: It defines the win margin, whether a team win by runs or wickets
.
BR: Balls reamaining while chasing totals or bowling out opponents.

Match: Declares the game format and participating team name.

Ground: Name of the cricket ground of venue.

Start Date: Date of the matchday.

Covid: Defines if the matchday was pre covid or post covid days. Takes two parameters, yes for post covid and no for pre covid match days.

Description of the used ML models:

- **Logistic Regression -** a statistical model that in its basic form uses a logistic function to model a binary dependent variable.

- **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.

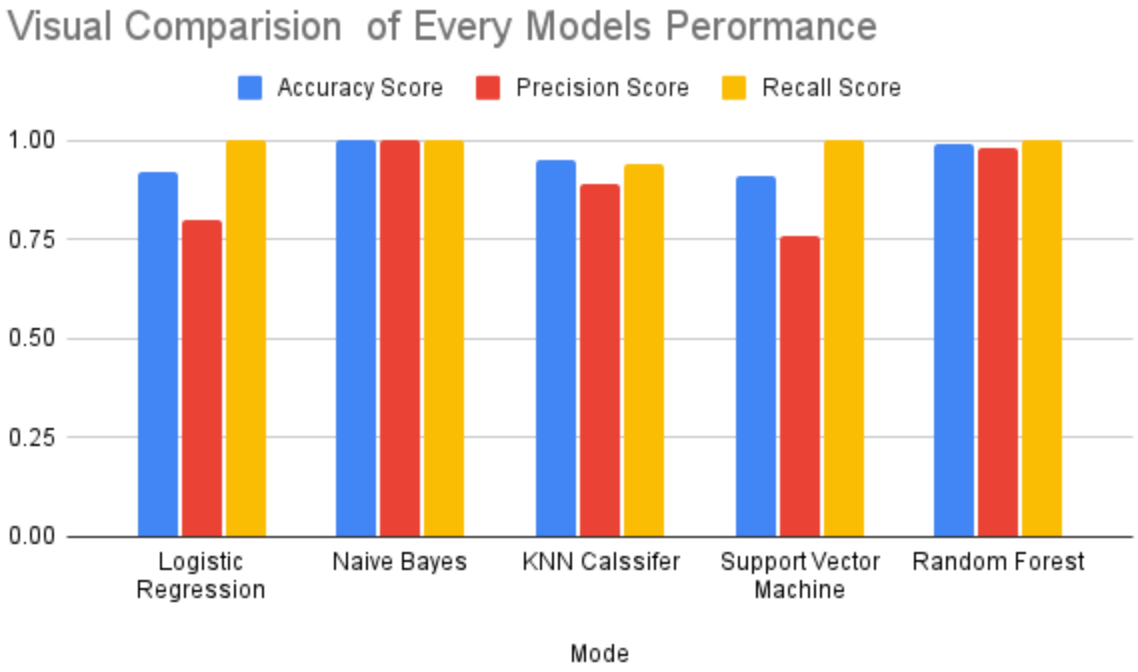
- **Support Vector Machine** - Support vector machines (SVMs) are a set of supervised learning methods used for classification, regression and outliers detection. The advantages of support vector machines are: Effective in high dimensional spaces.

- **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
Logistic Regression	0.921	0.80	1
Naive Bayes	1	1	1
KNN Calssifer	0.95	0.89	0.94
Support Vector Machine	0.91	0.76	1
Random Forest	0.99	0.98	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 Logistic Regression are almost similar when comes to accuracy.

References:

1. https://stats.espncricinfo.com/ci/engine/stats/index.html