

## Introduction

There has been research done on ODI and Test match cricket but very few on T20 cricket, which is currently more favorite than its older brothers. And that's why we decided to do research on this format of the game. The result of a T20 cricket match depends on lots of in game and pre-game attributes. Pre-game attributes like condition, venue, pitch, team strength etc. and in game attributes like wickets in hand, run rate, total run, strike rate etc. influence a match result predominantly. We gave more emphasis on in game attributes as our prediction will be when match is in progress. Our intentions would be to finding out the attributes which is most affecting the result in different phases of the game. We broke an innings into three phases: Power-play (1-6 overs), Mid-overs (7- 16) and final overs (17-20). Prediction will be active till the last over of mid overs phase. We consider an entire cycle of process of data mining, decision making and preparing a model to predict. Mining the data according to the attributes and different phases we have divided important to construct meaningful statistics. Modeling a problem for prediction requires several intelligent assumptions and molding the problem with collected data-sets. As we already mentioned cricket is a game of uncertainty and T20 format is the most unpredictable format rather than the other two format because it is the shortest format of the game and one over can change the result of a game

In this project we tried to design a prediction model which can go with this unpredictability and try to make a result prediction.

## Dataset

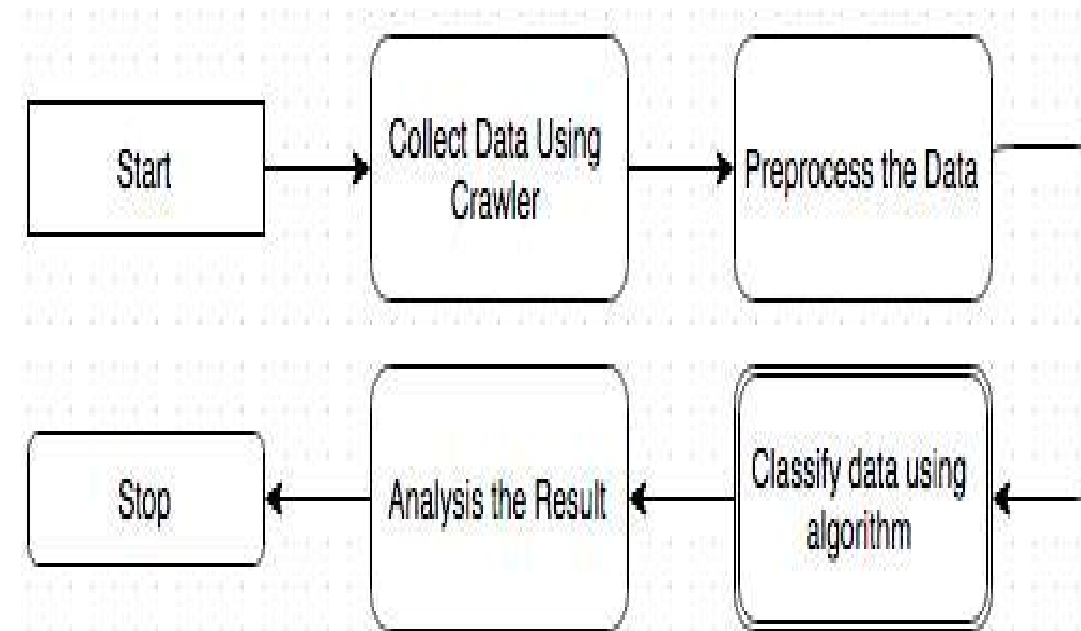
For our project, Data is collected from Statsguru. It is ESPN Cricinfo's cricket statistics maintenance database, where all the data relating to all the matches of cricket are saved.

The attributes are:

- Venue
- runs of home team in a segment (MR)
- runs of opposition team in a segment (OR)
- run rate of home team in a segment (MRN)
- run rate of opposition team in a segment (ORN)
- teams batting order (which team batted first, which team second)
- number of wickets fallen for home team (MW)
- number of wickets fallen for opposition team (OW)
- the result of the match

## Problem Definition

The aim is to prepare a model which will predict the result of a T20 cricket game while the match is in progress. Our main objective is to combine pre-game data and in-game data in order to design a good predictive model



workflow

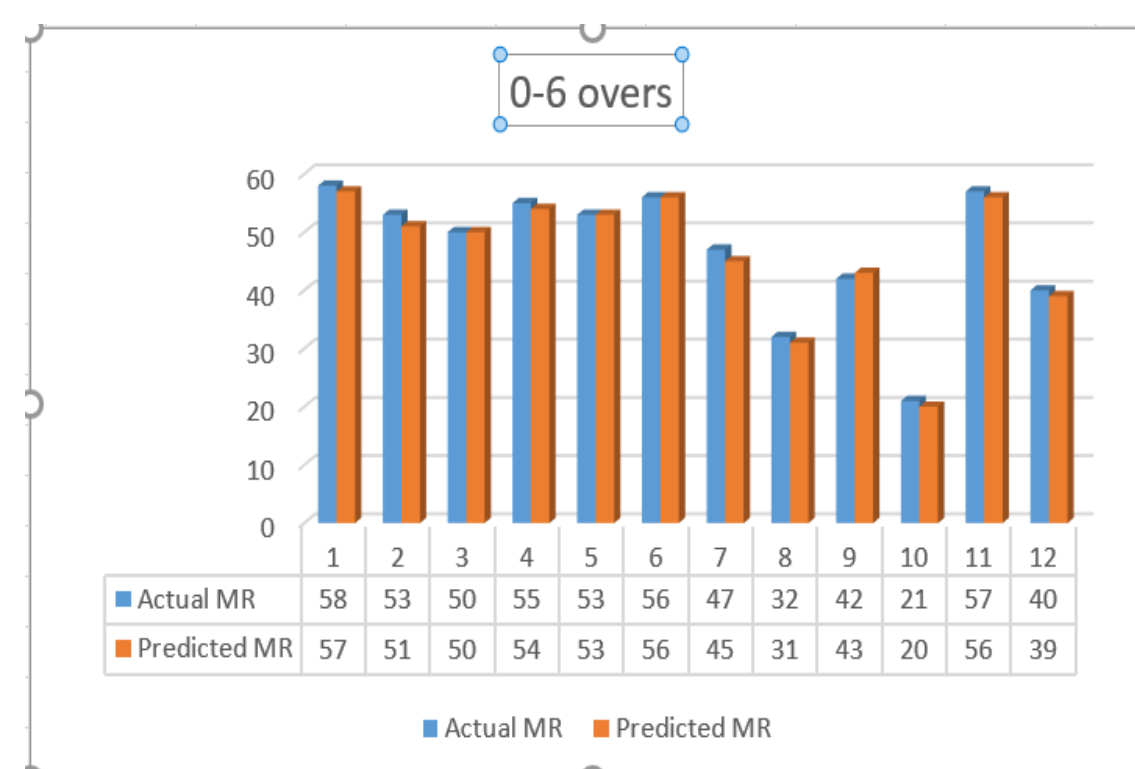


Fig1: first segment MR prediction

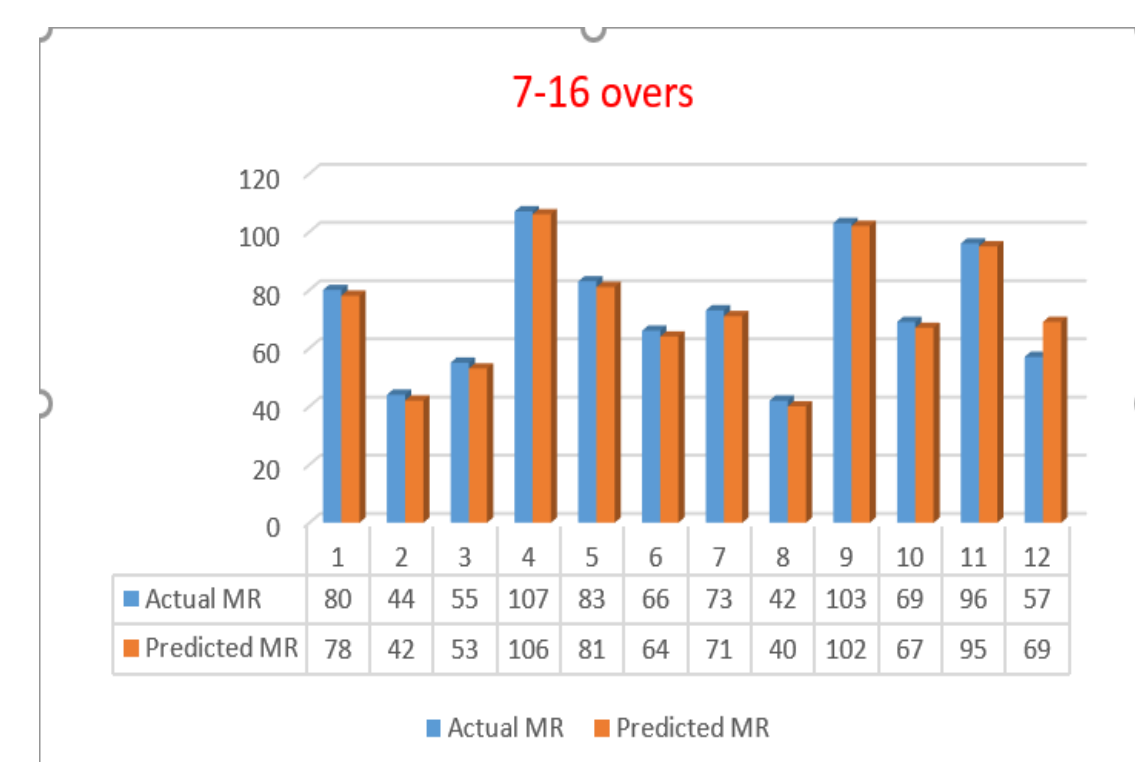


Fig2: second segment MR prediction

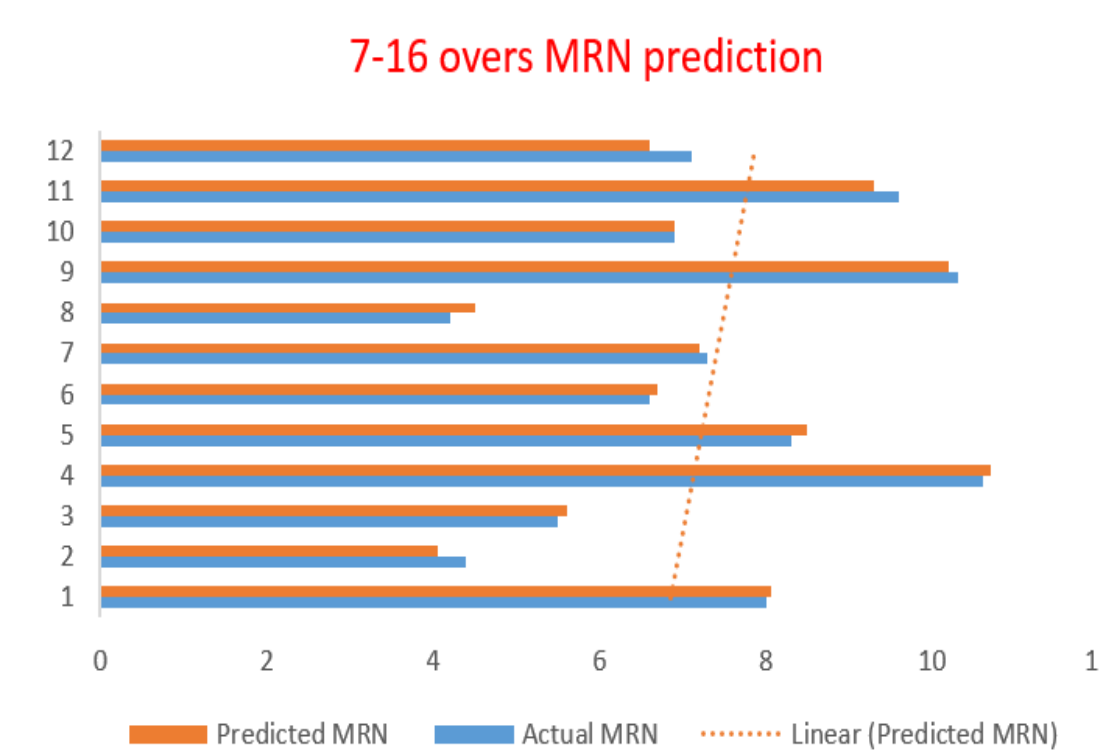


Fig3: second segment MRN prediction



Fig4: Final Result Prediction

## Analysis and Result

As we have divided our total work into 3 segments. First of all, we selected our first 6 over data set and we have implemented the algorithm over it. Then we have analyzed the data visually and figure out the distribution of values. The following metrics were used to determine the performance of our model: Time taken to build the model, Kappa statistics, mean absolute Error, Root Mean Squared Error, Relative Absolute Error, Prediction accuracy.

**1.First Segment:** The prediction of first segment was quite difficult as there is no instantaneous attributes as we mentioned before. That is why we have just tried to find out some average run scored by the team and the result was not so bad, the average run which we predicted using multiple linear regression it was just based on home or away attributes.

**2.Second Segment:** In order to predict the second segment run we have analyzed which attributes should take and which attribute will impact strongly. We have selected teams wicket, teams run rate, teams run to give a prediction about probable run can be scored in second segment. The figures 2 and 3 shows the run prediction and run rate prediction of second segment

**3.Third Segment:** Third segment prediction is the result of first and second segment prediction. Based on the result of first two segment, third segment prediction give its output.

## Conclusion

Our main goal in this project to develop a model to predict the outcome of a T20 cricket match while the game is in progress. We used the data of previous matches played between the team in order to design our model. We have used decision tree algorithm and weka to design this model. Along with decision tree, we also used multiple linear regression to predict match result with each segment. This knowledge will help us in the future to design a better prediction model.

### Reference:

- <http://stats.espncricinfo.com/ci/engine/stats/index.html>
- <http://www.icc-cricket.com/world-t20/stats>