

CricAI

A Smart Prediction Tool for the game of Cricket.

Major Project 1: Group 07

1. Jalaz Kumar
2. Pritish Chauhan
3. Aditya Thakur
4. Aarti Ramoul
5. Rajat Patial
6. Amandeep Prasad

Faculty Supervisor:

Mr. Rajiv Kumar
Assistant Professor
CSED, NIT Hamirpur

GOAL:

To design a smart prediction system which predicts the outcome and scores of future cricket matches using various pre-game attributes.

Phase 1: Outcome Prediction(Classification Problem)

Phase 2: Score Prediction(Regression Problem)

MOTIVATION:

Sports Analytics using Machine Learning is a blooming sector in the field of Computer Science. After Football, Cricket is the second most popular sports with a fan base of around 2.5 billion and mostly popular in South Asia, Australia, The Caribbeans and UK. It has tremendous spectator support and the masses show great interest in predicting the outcome of games.

The result of any fixed-over format cricket match depends on lots of in-game and pre-game attributes. Pre-game attributes like venue(Home/Away/Neutral), past track-records(Past 3 months/Against), Pitch & Ground Records, Innings(First/Second) etc. and in-game attributes like Current Run Rate, Wickets in hand, Strike Rate etc. influence a match result predominantly.

We strive to give emphasis only on the pre-game attributes right now, as our different classifiers strive to provide the best possible outcome of a future cricket match.(Win/Lose)

WHAT WE PROPOSE TO ACHIEVE:

1. Develop a Web based UI Platform for hosting the Predictive Models as well as cricket statistics.
2. Observe the results by doing comparative analysis for cricket datasets using 3 different models:
 - a. Bayesian Models
 - b. SVM Models using Gaussian Kernels
 - c. Neural Networks

PAST WORKS:

From our literature survey, we found that very limited machine learning work has been done on game of cricket. Though cricket shares some attributes with other sports such as baseball, it still remains unique in certain respects and deserves to be analyzed independently.

Most of analyzing studies on cricket so far have been conducted using statistical methods.

1. Bailey and Clarke conducted a study to predict the outcome in one day international cricket while the game is in progress. This study was performed using statistical models.
Journal of Sports Science and Medicine, 2006, Vol. 5, pp. 480-487.
2. WASP(Winning and Score Predictor), 2012 is product of some extensive research of Dr. Scott Brooker and Dr. Seamus Hogan at University of Canterbury in New Zealand. The WASP System is grounded on the theory of Dynamic Programming.

TECH-STACK

Python & it's Libraries:

1. Scikit-learn - Bayesian and SVM Classifiers
2. Tensor Flow - Neural Networks
3. Pandas
4. Scrapy(Data Scraping)
5. Django - Web Framework

PostgreSQL - Database Engine

HOW WE PROPOSE TO ACHIEVE PHASE(1):

- 1) Data Collection & Pruning of datasets
- 2) Using Sklearn, design & training of Bayesian Models
- 3) Using Sklearn, design & training of SVM Classifier.
- 4) Comparative Analysis of the 2 classifiers for some test data.
- 5) Development of the Hosting Platform
- 6) Using Tensor Flow, design & training of backpropagation Neural Nets.
- 7) Comparative Analysis of the 3 classifiers for test data

If accomplished all these aims well in time, Preparing foundations for accomplishing the Phase(2).

DATASETS:

- CricSheets Datasets from www.cricsheets.org available in yaml format.
- ESPN CricInfo Datasets from www.stats.espncricinfo.com needs scraping.

Both of these datasets need lots of pruning as well as boosting to make them suitable as the training data for our Classifiers.

CONCLUSION:

This tool if successfully implemented will be based on our results of detailed analysis with real data and moreover, can be substantially checked for future live matches.

By analysing different pre-game attributes of the game, we will be able to predict the winning criteria formulated using the datasets for future cricket matches in a much more impactful manner.

Thank You...