**Development of “Criclytics” (cricket analytics) for predictive analysis of cricket players and teams.**

By

Chinmay R Ranade

Pratyush P Kerhalkar

Mayur B Mahajan

Sameer V Nandanwar

**TABLE OF CONTENTS**

**Introduction 4**

**Architecture 5**

**Data Sources 6**

**Hardware requirements 7**

**Software requirements 8**

**Testing 9**

**Performance numbers 10**

**Conclusion 11**

Introduction

We have gathered ball by ball Datasets of 3995 cricket matches comprising of **405** Test matches, **6** other multi-day matches, **1,522** One-day internationals, **199** other one-day matches, **734** T20 internationals, **157** international T20s, **636** Indian Premier League matches,**190** Big Bash League matches, **121** NatWest T20 Blast matches and **24** Pakistan Super League matches featuring **60** countries, **45** club teams and **2** representative XIs going back as far as **2009** (for women), and **2005** (for men). We have proposed to perform “Criclytics” on datasets comprising of International matches played by Men excluding Indian Premier League, Big Bash League, NatWest T20 Blast and Pakistan Super League Matches.

1. **PROBLEM STATEMENT**

Based upon above acquired data, we have proposed to perform “Criclytics” i.e. Cricket Analysis which has predictions based upon following 2 categories.

1. **For Team:**

Winning and Score predictor (WASP) using predictive analysis to predict the result

of the match based upon various factors like team record against opposition,

venue where match is being played etc.

1. **For Individual Players:**

Predicting final Innings score of respective team based upon average statistics

of respective players impact in average conditions.

**2. SCOPE OF PROBLEM STATEMENT**

Above problem statement is applicable for all the matches that are played on International level under ICC authorization wing including Test, One-day International and T-20 Matches.

**3. WHAT IS THE MARKET NEED OF PROJECT**

Our project can be produce effective value for money in Betting Institutions and Fantasy leagues like Dream11, IPL fantasy leagues etc. It can even assist IPL franchises for comparative analysis to determine which Player should they invest upon and who will be better will be better value for money.

**4. WHY IT CANNOT BE ACHIEVED TRIVIALLY**

Current ICC protocol implements Cricket score predictions based only upon Current Average Scoring rate of match. The trivial Score predictions doesn’t consider fall of Wickets and past history of player against particular opponent and on particular ground, Which makes Score predictions less accurate and less interesting.

**Architecture**

Data Source [cricsheet.com]

Data Transform 

Data Cleaning 

Data Processing

HDFS

Data Dump

Report

Model Planning

Model Building 

Deployment

Model Evaluation 

**Data Sources**

1. Schema

Data is Downloaded from [cricsheet.com](https://cricsheet.org/) which is in the “YAML” format. General sample schema of the data is as follows;

meta:

data\_version: 0.9

created: 2013-02-22

revision: 1

info:

city: Auckland

dates:

- 2005-02-17

gender: male

match\_type: T20

winner: Australia

overs: 20

player\_of\_match:

- RT Ponting

teams:

- New Zealand

- Australia

toss:

decision: bat

winner: Australia

umpires:

- BF Bowden

- AL Hill

venue: Eden Park

innings:

- 1st innings:

team: Australia

deliveries:

- 0.1:

batsman: AC Gilchrist

bowler: DR Tuffey

non\_striker: MJ Clarke

runs:

batsman: 0

extras: 1

total: 1

This data has ball by ball information which we have later converted into “CSV” format using Python for better analytical processing and inference.

Data obtained from [cricsheet.com](https://cricsheet.org/) has volume of 445 MB and Volume can be increased further to have better predictions and fault tolerance.

**Hardware Requirements**

1. Commodity Hardware

**Software Requirements**

1. Anaconda Navigator

2. Apache Spark

3. Tableau Desktop