



University of New Haven

TAGLIATELA COLLEGE OF ENGINEERING

Electrical & Computer Engineering and Computer Science

Electrical & Computer Engineering & Computer Science (ECECS)

TECHNICAL REPORT TEMPLATE



SPRING 22

CONTENTS

Project Name.Error! Bookmark not defined.

Executive SummaryError! Bookmark not defined.

Technical Report.....3

Highlights of Project3

Submitted on:.....3

Abstract.....4

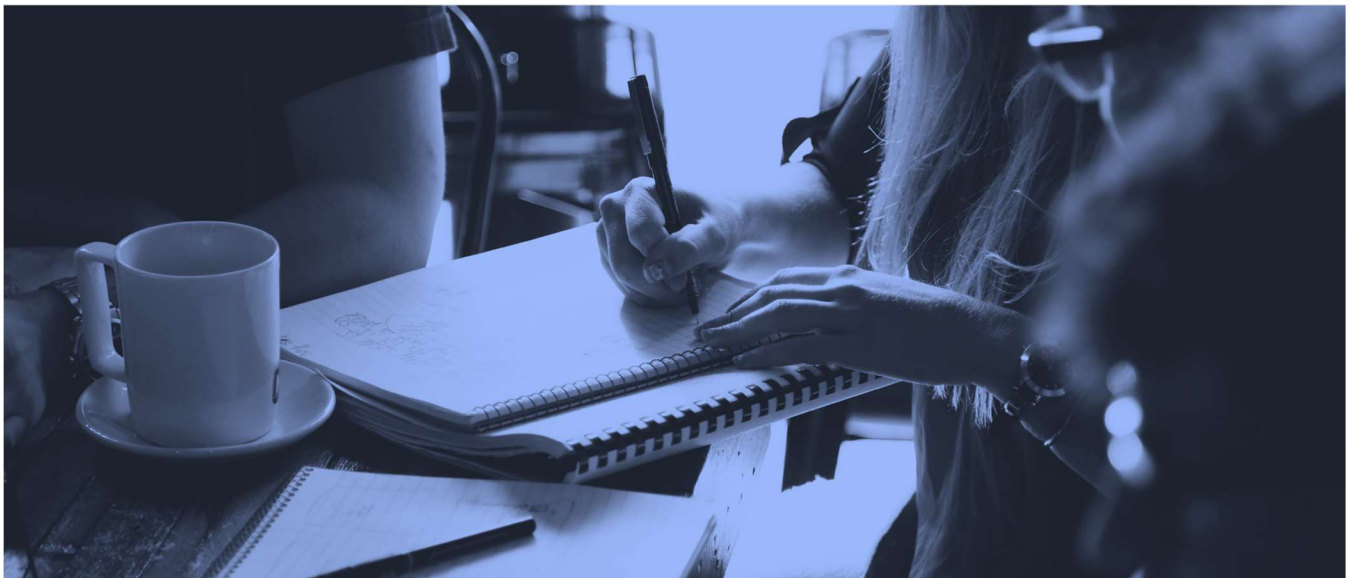
Methodology6

Results Section8

Discussion.....Error! Bookmark not defined.

ConclusionError! Bookmark not defined.

Contributions/ReferencesError! Bookmark not defined.



Team Members:

Name 1 Khalil Malik Muhammad

Name 2 Yegneshwar Rao Ginjupalli

Name 3 Srividhya Pasam

Name 4 Sreeja Garlapati

Questions?

Contact : yginj1@unh.newhaven.edu

Technical Report

Data Engineering and Analysis of Cricket

Highlights of Project

This project aimed at forming a team of 11 members from across all the cricket teams, maximizing the chance of winning the match.



**Submitted on:
04/30/2023**

Abstract

The T-20 world cup is a major international cricket tournament, featuring all the cricket teams present across the world compete in the shortest format of the game. In this study, we conducted a comprehensive analysis of recently held T-20 world cup covering all the aspects of the tournament, including player performance, team strategies and match outcomes with of forming a new team of 11 from all the available teams.

The data was collected from Espn, Cricinfo and official cricket governing bodies. We used visualization tools to understand the player performance, team performance and metrics that impact them. Further our analysis also helped in identifying the strengths and weakness of players, helping us to go with a analytics based decision.

The analysis in this notebook also presents the analysis like Exploratory Data analysis, individual player performance, team performance and insights related to the strengths and weakness of the player and team.

Executive Summary

Introductory Section

Cricket is a game that has always relied on data and statistics to make informed decisions. With increased availability of cricket data, it has become easy to analyze player and team performance and use data driven approaches to improve team strategies.

One of the most popular formats of cricket is T-20, where teams play a fast-paced game with a limited number of overs. In this context data analytics has become an increasingly important tool for teams looking to create the best possible squad.

In this study we will explore the recent T-20 cricket world cup data to form a team of 11 from the teams across the world to score a minimum of 180 runs and can defend a maximum score of 150 runs. In this study we used extensive data engineering techniques to generate the data that is useful in making data driven decisions.

Methodology

Following CRISP-DM methodology for Data Science Process



Business understanding - To form a new cricket team of 11 best players for the Indian Premier League IPL.



Data understanding - Gather, assess and explore the relevant data from various sources like *ESPN*, Cricinfo website to understand its structure and quality.



Data preparation - Clean, transform and prepare the data to make it ready for modeling by using web-scraping, Excel and python.



Modeling - Choose appropriate modeling techniques like SQL and PowerBI reports to derive insights from the data.

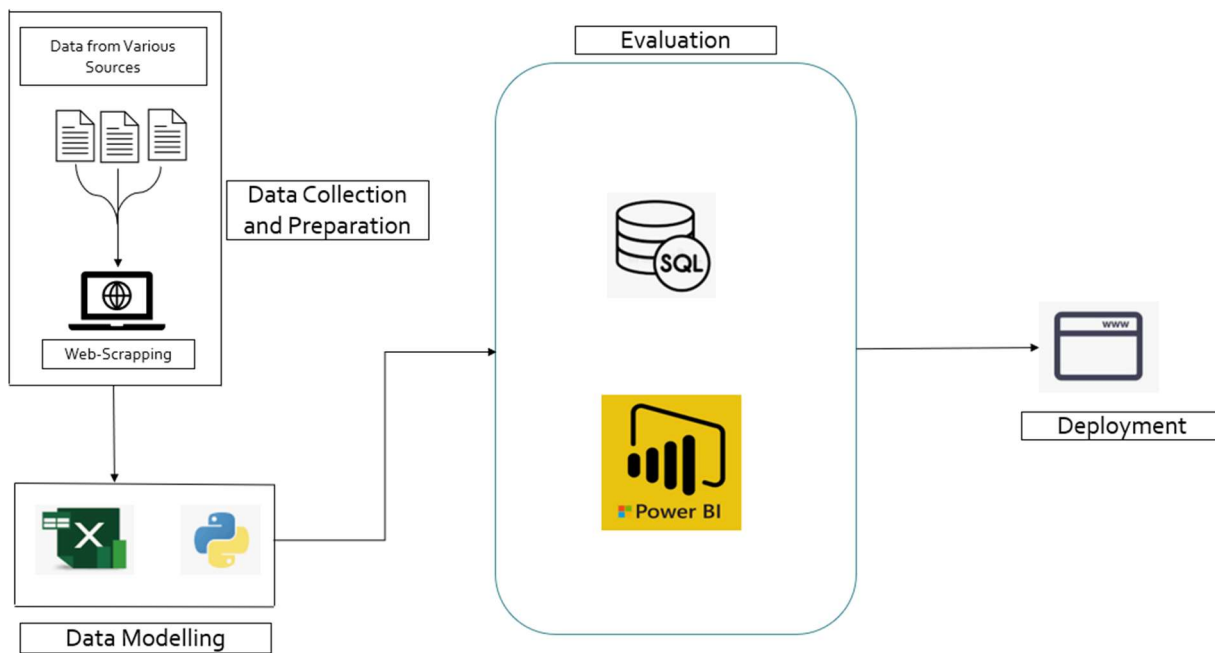


Evaluation - The performance of the players is evaluated and then visualized using the graphs and histograms to finalize the team of 11 members.



Deployment - Communicate the results to the Board members and deploy the models in a web page for ongoing use.

Project Layout:



Detailed Flow Chart

Filters:

Middle Order Anchors:

PARAMETERS	DESCRIPTION	CRITERIA
Batting Average	Average Runs Scored in an Innings	>0
Strike Rate	No of runs scored per 100 balls	>125
Innings Batted	Total Innings Batted	>3
Average Balls Faced	Average of Faced Balls in an Innings	>52
Batting Position	Batting order	<2

Wicket Keepers/Lower Order Anchors:

PARAMETERS	DESCRIPTION	CRITERIA
Batting Average	Average Runs Scored in an Innings	>25
Strike Rate	No of runs scored per 100 balls	>130
Innings Batted	Total Innings Batted	>3
Average Balls Faced	Average of Faced Balls in an Innings	>12
Innings Bowled	Innings Bowled by Bowler	<1

All Rounders :

PARAMETERS	DESCRIPTION	CRITERIA
Batting Average	Average Runs Scored in an Innings	>15
Strike Rate	No of runs scored per 100 balls	>140
Innings Batted	Total Innings Batted	>3
Average Balls Faced	Average of Faced Balls in an Innings	>30
Innings Bowled	Innings Bowled by Bowler	<2
Bowling Economy	Runs Allowed Per Over	<7
Bowling Strike Rate	Minimum Balls to take Wicket	<20

Bowlers:

PARAMETERS	DESCRIPTION	CRITERIA
Batting Average	Average Runs Scored in an Innings	>15
Bowling Average	No of runs per wicket	<20
Dot Balls %	% of Dot Balls Bowled	>40
Bowling Style	Fast or Spinner	>30
Innings Bowled	Innings Bowled by Bowler	<5
Bowling Economy	Runs Allowed Per Over	<7
Bowling Strike Rate	Minimum Balls to take Wicket	<16

Conclusion:

With the above set guidelines, a team of 11 members were listed from all the teams available with the aim of scoring 180 runs and defending 150 runs in the hardship. And the team is.



