

School of Computing and Mathematics

PUSL3119
Computing Individual Project

BSc (Hons) *Computer Science*

Mahindu Bandaranayake

Statistical and Predictive Bowling Analysis of New Zealand Cricket

2022/2023

Acknowledgement

I would like to make this an opportunity to thank all the colleagues and lecturers from the 1st and 2nd year to show my utmost sincerity to them, for guiding and helping me throughout my undergraduate journey.

The journey of course has not been easy, during a tough time period with all the crisis that was present in the country, working through and dedicating time despite these obstacles was an immense amount of pressure on each and every student had to undergo and move forward in order to complete the projects in due time. Also, I would like to show my appreciation to the university's academic staff members and support stuff for providing and facilitating me with all the essential needs that helped me to coordinate my studies methodically.

Another person that I would like to show my gratitude is the Dean of the Computing Faculty Mr. Rasika Ranaweera, who has been with us and showing up as our mentor and motivating us to reach our full potential at all times, thank you sir for your companionship.

And last but not least a special vote of thanks to my supervisor Mr. Pramudya Tilakaratne who has been guiding me since day one of the project module as well as the previous year modules which he conducted. "Thank you sir for your great service as a lecturer and for your kind nurturing words which helped me a lot to grow under your supervision.

Abstract

Cricket is one of the biggest cultured sporting- events in Sri Lanka, Millions have poured their hearts to witness the Sri Lankan Cricket Team performing well in ICC events against other cricketing nations in the world. In here, discussions are taken place as to why the team has been losing matches constantly against the least encountered cricketing nation (New Zealand) in One-day Internationals and providing a solution with the interference of ICT to help solve the issue of selecting ‘Ideal’ and ‘Optimum Players’ to counterattack through thorough analysis of prescriptive data and predictive data of the opponent team, calculated using Machine Learning techniques namely; K-NN and Multiple-Linear Regression, which are vital components of performance analysis, mandatorily required to be conducted by coaching stuff and captains.

KEYWORDS : K-NN, Multiple-Linear Regression, ICT, Predictive Data

Table of Content

Acknowledgement -----	2
Abstract -----	3
Chapter 1 : Introduction -----	7
1.1 Stating the Project Domain -----	7
1.2 Stating the Rationality of the Project-----	8
Chapter 2 : Background-----	11
2.1 Stating the Solution-----	11
2.2 Business Objectives-----	12
2.3 Project Objectives-----	12
Chapter 3 : Scope of the Project -----	14
3.1 Setting up a boundary -----	14
3.2 Identified Components -----	14
3.3 Functions -----	15
3.4 Representation Criteria-----	15
Chapter 4 : Literature Review -----	16
4.1 Domain Overview-----	16
4.2 Existing Systems -----	16
4.3 Algorithms and Calculation Concepts -----	17
4.3 Technical Analysis-----	19
4.4 Highlighted Technology -----	19
4.5 Reflection -----	20
Chapter 5 : Development Methodology -----	21
5.1 Step 1: Data Source -----	21
5.2 Step 2 : Evaluation of the Dataset -----	21
5.3 Step 3 : Connecting MongoDB to the Development Environment -----	21
5.4 Step 4 : Code Implementation using Dash Logics -----	22
5.4.1 ML Algorithm Implementation (Multiple Linear Regression)-----	23
5.4.2 ML Algorithm Implementation (K-NN) -----	24
Chapter 6 : Technology Identification-----	25
6.1 Web Application Technology -----	25

6.2	Database Technology -----	26
6.3	Development Environment (IDE) -----	27
6.4	Dependencies and Libraries -----	27
Chapter 7 : UML Diagrams	-----	28
7.1	Use Case Diagram -----	28
7.2	Class Diagram -----	29
Chapter 8 : Development	-----	30
8.1	Web Application -----	30
8.1.1	Starting Page -----	30
8.1.2	History Section-----	30
8.1.3	Entries Page -----	31
8.1.4	Records Page -----	32
8.1.5	Preference Page -----	33
8.1.6	Prediction Page-----	34
8.1.7	Modernize Page-----	35
8.1.8	Administration Page-----	35
Chapter 9 : Requirements	-----	36
9.1	Functional Requirement-----	36
9.2	Non-Functional Requirement -----	39
End Project Report	-----	40
Project Post - Mortem	-----	42
Conclusion	-----	43
Reference List	-----	44
Appendices	-----	46
User Guide	-----	46
PID	-----	49
Interim	-----	65
Unit Test Cases	-----	73
Project Management	-----	75
Supervisor Meeting Records	-----	76

Table of Figures

<i>Figure 1: Batting.....</i>	7
<i>Figure 2 : Bowling</i>	8
<i>Figure 3 : History between SL vs NZ</i>	8
<i>Figure 4 : ODI Summary 2023 SL vs NZ</i>	9
<i>Figure 5 : Social Media Comments 1.....</i>	9
<i>Figure 6 : Social Media Comments 2.....</i>	10
<i>Figure 7 : Low Totals in the Scorecard.....</i>	11
<i>Figure 8:Regression Analysis Steps</i>	18
<i>Figure 9 : MongoDB Connection.....</i>	21
<i>Figure 10 : Dash Layout.....</i>	22
<i>Figure 11 : Dash Callback Functions.....</i>	22
<i>Figure 12 : Dash Plotly Logo</i>	25
<i>Figure 13 : Python Logo.....</i>	25
<i>Figure 14 : Flask Logo</i>	26
<i>Figure 15 : React Logo</i>	26
<i>Figure 16 : MongoDB Logo</i>	26
<i>Figure 17 : Visual Studio Code Logo.....</i>	27
<i>Figure 18 : Use Case Diagram</i>	28
<i>Figure 19 : Class Diagram</i>	29
<i>Figure 20 : Application Screenshot : Starting Page.....</i>	30
<i>Figure 21 : Application Screenshot: History Section</i>	31
<i>Figure 22 : Application Screenshot : Entries Page</i>	32
<i>Figure 23 : Application Screenshot : Records Page</i>	32
<i>Figure 24 : Application Screenshot : Preference Page</i>	33
<i>Figure 25 : Application Screenshot : Prediction Page</i>	34
<i>Figure 26 : Application Screenshot : Modernize Page.....</i>	35
<i>Figure 27 : Application Screenshot : Administration Page</i>	35
<i>Figure 28 : Overall Idea of the Project</i>	43

Chapter 1 : Introduction

1.1 Stating the Project Domain

Cricket is a famous outdoor sport invented and played in England with the second largest fan base in the world, also embraced by the other nations in the Indian subcontinent as well countries such as Australia, New Zealand, South Africa and the Caribbean etc. In the present day, the ICC(International Cricket Council) have ensured that the sport is emerged in the European continent as well. Countries such as Netherlands, Denmark, United States of America are already a part of the ICC body and competing in tournaments to qualify for the major ICC event in the future.

Cricket is a bat and ball game played between two teams, to be exact between two nations consisting of ‘11 players’ from each side. Each team will be involving with batting and balling, which will be done first, either batting or bowling will be decided by the team winning the toss before the game starts.

Cricket is played in three different formats, namely; Test, ODI (One-Day Internationals) and T20I (Twenty-Twenty Internationals). Where there will be two ‘Innings’ for the ODI’s and T20I’s matches and 4 Innings for the Test matches. An Inning will be either batting or bowling by one particular team, where teams will be swapping their roles after an innings is over. In general, one team will be batting while the other team will be bowling. ODI’s and T20I’s are considered as shorter format of the game where 50 overs are dedicated in the ODI’s and 20 Overs in the T20I’s for each team. In test cricket, countless number of balls can be bowled within the span of five days.

A delivered ball in cricket is the action performed by the bowler, where the bowler is bowling the ball at a batsman. An over consists of 6 deliveries where no consistent overs are allowed to be bowled by one particular bowler in any format of the game. In an ODI the bowlers are allowed to bowl up to maximum of 10 overs while only 4 overs are allowed in the T20I’s.



Figure 1: Batting



Figure 2 : Bowling

1.2 Stating the Rationality of the Project

The Sri Lanka Cricket Team have been underperforming in the recent past in all the three formats, losing matches in a constant basis in bilateral series and in World Cups as well against tough opponents such as Australia, India, South Africa and New Zealand. Out of those teams, the New Zealand team is the least encountered team and out of these encounters majority of the wins lie under the name of New Zealand.

The below table supports the above-mentioned statement.

New Zealand in Sri Lanka ODI Series	2013/14	drawn	(3)	1-1
Sri Lanka in New Zealand ODI Series	2014/15	New Zealand	(7)	4-2
ICC Cricket World Cup (in Australia/New Zealand)	2014/15	Australia		
Sri Lanka in New Zealand ODI Series	2015/16	New Zealand	(5)	3-1
ICC Champions Trophy (in England)	2017	Pakistan		
Sri Lanka in New Zealand ODI Series	2018/19	New Zealand	(3)	3-0
ICC Cricket World Cup (in England)	2019	England		
ICC Men's Cricket World Cup Super League	2020-2022/23	-		

Figure 3 : History between SL vs NZ

Within the span of 5 years only two bilateral series have been played between the two sides. Also, a major dis-advantage is that the Sri Lanka team haven't toured New Zealand for quite some time in a regular basis. The last tour was back in 2019 and the latest tour was in 2023 and even then the performance outcome of the series can be considered as below average.

The following figure show the summary of the ODI series which was held during the month of March 2023

ESPNcricinfo		Live Scores	Series	Teams	News	Features	Videos	Stats	IPL 2023
Sri Lanka tour of New Zealand	▼		Home	Fixtures and Results	MVP	Videos	Stats	Squads	Photo
New Zealand won by an innings and 58 runs									
Sat, 25 Mar	RESULT	1st ODI (D/N), Auckland, March 25, 2023, Sri Lanka tour of New Zealand	New Zealand	274 (19.5/50 ov, T:275)	76				
			Flag: New Zealand						
			Flag: Sri Lanka						
		New Zealand won by 198 runs							
Tue, 28 Mar	ABANDONED	2nd ODI (D/N), Christchurch, March 28, 2023, Sri Lanka tour of New Zealand	New Zealand						
			Flag: New Zealand						
			Flag: Sri Lanka						
		Match abandoned without a ball bowled							
Fri, 31 Mar	RESULT	3rd ODI (D/N), Hamilton, March 31, 2023, Sri Lanka tour of New Zealand	Sri Lanka	157 (32.5/50 ov, T:158)	159/4				
			Flag: Sri Lanka						
			Flag: New Zealand						
		New Zealand won by 6 wickets (with 103 balls remaining)							

Figure 4 : ODI Summary 2023 SL vs NZ

Stating the obvious, the Sri Lanka Cricket Team have been receiving negative and hate comments from the fans as well cricket expertise's, directing certain predictions as to where the team would see themselves in the future ICC events. The fingers are pointed towards the selection committee as for the failure of selecting and preparing a strong challenging team to face the opponent team.

The following figures will support the above stated issue

Rajitha Abeysinghe @RajithaSL · Jan 6, 2021

Sri Lanka cricket has become a real joke to the world since **last few years** 😢 #missing the glory days

Angad@holi @Angadholi1 · Sep 7, 2022

Replying to @Shoaib_Jatt

Last few months has been bad for Indin **cricket** but doesn't mean you foul our superiority , Pakistan and **Sri Lanka** suffered a lot for 5 to 6 **years** for quality **cricket**.

Sushant Adhikari @Sushant_adh18 · Aug 25, 2021

Replying to @No_gooly

Yap. I wanted him to travel to **Sri Lanka** because I didn't want to see him in the bench & at the same time missing the limited overs **cricket**(where he's doing pretty good in **last few years**).

Sakun @Sakun_SD · Mar 14, 2021

Maiden ODI half century for Wanindu Hasaranga!

He is the best thing that happened to **Sri Lanka cricket** in **last few years**.

#WIVSL

Figure 5 : Social Media Comments 1



Figure 6 : Social Media Comments 2

- ❖ Therefore, this proves the fact that necessary game plans and strategies were not developed when it comes down to playing crucial series against different teams in major ICC events. The main reason for that would be the lack of pre-game plan developed by the relevant authorities to counter attack. "***One major finding is that not every player's performance is low or high at a given ground or opposition.***"(Attygalle and Sunethra, 2014). Therefore, an optimum game plan can be determined by considering the past player records and analyzing them in a suitable way to extract useful details regarding the players in order to utilize the pool resources and talent in players by taking calculated risks with the hope of creating winning chances.

Chapter 2 : Background

2.1 Stating the Solution

According to this a specific article, it states that past records or referred to as historical data collections are subjected for analysis to depict the game facts and draw line with predictions (Pramanik *et al.*, 2022). Therefore, this proves the fact that certain factors can be decided and developed using past records in order to create scientific measurements in the action plan before the game starts. Moreover, the above statement can be clarified by particular referenced lines from an article, where it states significant performances analysis in both bowling and batting department will bring out emphasis of the statistical data (Saikia, Bhattacharjee and Lemmer, 2012).

The solution idea will be developed around the bowling department as the importance of analyzing statistical data for the above identified problem statement in the previous chapter, which will support the cause. Since according to statistical data, during matches between Sri Lanka and New Zealand, New Zealand Team have managed to restrict the Sri Lankan Team for a moderate low score at most of the occasion.

MATCHES					
	Team 1	Team 2	Date	Venue	Result
1	New Zealand 274 (49.3 ov)	Sri Lanka 76 (19.5 ov)	Mar 25, 2023	Auckland	New Zealand
2	New Zealand	Sri Lanka	Mar 28, 2023	Christchurch	ABANDONED
3	Sri Lanka 157 (41.3 ov)	New Zealand 159/4 (32.5 ov)	Mar 31, 2023	Hamilton	New Zealand

Figure 7 : Low Totals in the Scorecard

According to various articles New Zealand Team being a ‘Sena’ nation, possesses a quality bowling lineup consisting of equally good seam and spin bowlers to control the run rate of the opponent team by not letting the batting team to score freely in their home grounds. Generally, even the pitches in New Zealand are in the favor of the bowlers, therefore with the weather conditions and other factors, the winning chance are pretty much in the hands of the bowling unit.

As previously mentioned the problem lies in the hands of the selection committee being unable to select and preparing a team with similar performance rate in order to counter attack. The selection committee is accountable for utilizing the resources and pool of players to obtain the maximum output.

As per this article, performance of a bowler is measured in a different way rather than sticking to the traditional way comparing the average, economy and the strike-rate of bowlers; ***“to quantify the performances of bowlers by combining the three traditional measures viz bowling average, economy rate and bowling strike rate”*** (Saikia, Bhattacharjee and Lemmer, 2012). In order to achieve this task the basic requirement is to gather the fundamental components to make the calculated predictions and use this data to calculate the individual performance as well

Evaluation of past data is mandatory when it comes to 21st century, due to the evolution of digital systems and also the game of cricket generating a high volume of performance statistics compared to other sports. So, to make predictions and make tough decisions in a competitive environment a system to retain and evaluate history of records and make some necessary prediction in advance to oversee foreseeable future will benefit the team to overcome the current struggle.

Solution Summary

- ❖ **A system to calculate individual player performances and predict useful performance attributes to help develop strategies and game plans through statistical analysis.**

2.2 Business Objectives

Development of such a system can be extended further by including all the other teams as well, since the initial scope is considering only one particular team. Based on an analytical point of view the requirements can be satisfied whether it is the insight of the batting unit of the opponent team, or the fielding insight can be provided by using similar concepts and techniques.

With the evolution of the game and modernization of the rules and regulation of the sport the performance level will be measured from different levels and might not be the same as before therefore, adaptation to these constraints are significantly important. Nowadays franchise cricket leagues being a major part of a player's career is a factor to be highlighted. Indication and Evaluation of these statistics might play a huge role in the selection phase for the main ICC events.

2.3 Project Objectives

In the game of cricket quantifying player performance is vital when selecting a team in the international level (Mukherjee, 2014). Therefore, analysing prescriptive and predictive data is an essential part in doing well and receiving a win in difficult conditions.

- To keep a track of important player performance details of other teams
- Significant ground records relating to venues in New Zealand
- Visualization of statistical information in form of charts
- Performing important calculations to ease the process (Average, Economy and Strike Rate)
- Predicting the ‘Average’ of active players based on the increment in number of matches
- Calculating the ‘Bowling Performance’ of a bowler to evaluate among the other bowlers
- Depiction of player attributes
- Updating current databases with new player/performance details
- Indicating records from previous series (Match Summaries)
- Comparison selection against benchmark attributes
- Ability to alter previous data (Add or Remove player profile details)

Other Objectives beyond the Current Scope

- Continuity of the analyzes while embedding new features in to the system in the future
- A path to create similar format for other the teams with the current intellectual circumstances.

Chapter 3 : Scope of the Project

3.1 Setting up a boundary

The specific team taken in to consideration is the New Zealand Cricket Team, where player's data with the priority of bowling are considered as the core to lay the foundation to kick start the project

3.2 Identified Components

Bowler's Attributes taken in to consideration (Series based)

- Matches Played (Innings)
- Wickets taken
- Overs Bowled
- Runs conceded
- Average
- Economy
- Strike Rate
- Dot Balls Bowled

Series Attributes

- Total runs conceded by Seam Bowlers
- Total runs conceded by Spin Bowlers
- Total Wickets taken by Seam Bowlers
- Total Wickets taken by Spin Bowlers
- Average of Spin Bowlers
- Average of Seam Bowlers

Match Summaries

- Score in the 1st Innings
- Score in the 2nd Innings
- Winning Team

Player List

- Player Name
- Bowling Style

3.3 Functions

Calculations

- Average calculations based on series data
- Economy calculations based on series data
- Bowling Strike Rate calculation based on series data
- Lemmer's Bowling Performance value calculations based on series data
- Predictive Average of a Bowler calculations done using Multiple Linear Regression
- Player comparison calculations done using K-NN method

3.4 Representation Criteria

Charts

- Past Records (Series based) – Average
- Past Records (Series based) – Economy
- Past Records (Series based) – Strike Rate
- Past Records (Series based) – Wickets
- Past Records (Series based) – Overs
- Past Records (Series based) – Runs Conceded
- Past Records (Series based) – Dot Balls
- Latest Records (Player Performance) – Average
- Latest Records (Player Performance) – Economy
- Latest Records (Player Performance) – Strike Rate
- Predicted Average (For each Player)
- Accuracy Output (Predicted Average Confirmation)
- Player Performance Comparison (All)

Chapter 4 : Literature Review

4.1 Domain Overview

Cricket is a team sport played with a bat and a ball, where two teams take turns to bowl and bat according to the toss. The team who wins the toss has the opportunity to decide whether to bat or ball during the 1st innings of any format of the game out of the three formats; Test matches, ODI and T20 (Krishnamohan, 2021). When it comes down to Sri Lankan Cricket, one of the authors has conducted a qualitative research based on fan opinions on why SLC Team is failing to produce match winning results and one of the factors that this article highlights is the ‘poor management strategies’ where suitable decisions have not been made during critical conditions of the match time. (Dananjaya Weerasinghe, no date)

4.2 Existing Systems

Author states; obtaining quantified measurements can be used for critical team selection proceeding by considering the individual performance level , furthermore the author states that collection of stat numbers of bowling and batting is always important as a whole. In this article, research was conducted based on the number of dismissals of batsmen during a considerable period to determine the most successful bowler of the history (Mukherjee, 2014). Another author states the development of a system to compute the number of runs need to be scored every over in an ODI match (Krishnamohan, 2021). One of the authors mentions “*Sports analytics and data visualization can play a crucial role in selecting the best players for a team*”, and a toss related data visualizing project was carried in order to support the decision making when selecting players for a particular team (Kanungo and Tulasi, 2019). Another very important study which was conducted on IPL teams was to measure the overall bowling performance of a team by using the CBR (Combined Bowling Rate) technique to understand the post-match success rate. The particular equation was developed on top of the initial equation found by H.Lemmer mentioned in one of his articles which is known by the name LBP (Lemmer’s Bowling Performance), where the rate of performance is calculated for a single player (Saikia, Bhattacharjee and Lemmer, 2012). Furthermore, an approach was casted to determine the team score using linear regression without the involvement of the concept of big data and a big data framework (Awan *et al.*, 2021). A comprehensive statistical analysis in bowling performance was conducted to understand the vulnerability of wickets taken by Fast bowlers and Medium Fast bowlers (Malhotra and Krishna, no date)

4.3 Algorithms and Calculation Concepts

Certain initial calculations need to be made before the full equation is situated. The following are the essential sub equations are part of the main equation

- (i) **Bowling Average:** Number of Runs Conceded by the Bowler per Wicket taken.

$$\text{Bowling Average} = \frac{\text{Total Runs Conceded}}{\text{Total Wickets taken}}$$

- (ii) **Bowling Economy:** Number of Runs Conceded by the Bowler per Over Bowled

$$\text{Bowling Economy} = \frac{\text{Total Runs Conceded}}{\text{Total Overs Bowled}}$$

- (iii) **Bowling Strike Rate:** Number of Balls that Bowler has to Bowl to get a Wicket

$$\text{Bowling Strike Rate} = \frac{\text{Total Balls Bowled}}{\text{Total Wickets taken}}$$

Therefore, according to the article, the Lemmer's Bowling Performance equation which calculates the individual player performance is derived as follows;

$$\text{Lemmer's Bowling Performance} = (\text{Wickets Taken}) \times 25 - (\text{Runs Conceded} \times \text{Economy})$$

And in order to calculate the team performance using Combined Bowling Rate the above equation can be altered

$$\text{Combined Bowling Rate} = \frac{3}{(\text{1/Average} + \text{1/Economy} + \text{1/Strike Rate})}$$

(Bhattacharjee, Pahinkar and Pahinkar, 2012)

Based on an article it has stated that regression has the best fitted models to predict a variable using other external variables.

These variables can be categorized in two main types

- Independent Variables
- Dependent Variables

Regression analysis is when a relationship is built between these variables to define the unknown variable. The first step is to identify the specific dependent variable and the independent variables in order to create the model using a suitable framework. The dependent variable will be considered as the unknown variable and other variables will be forming patterns and combination to find the missing term in the equation.

The following steps can be followed in an orderly manner to receive a desired predictive output

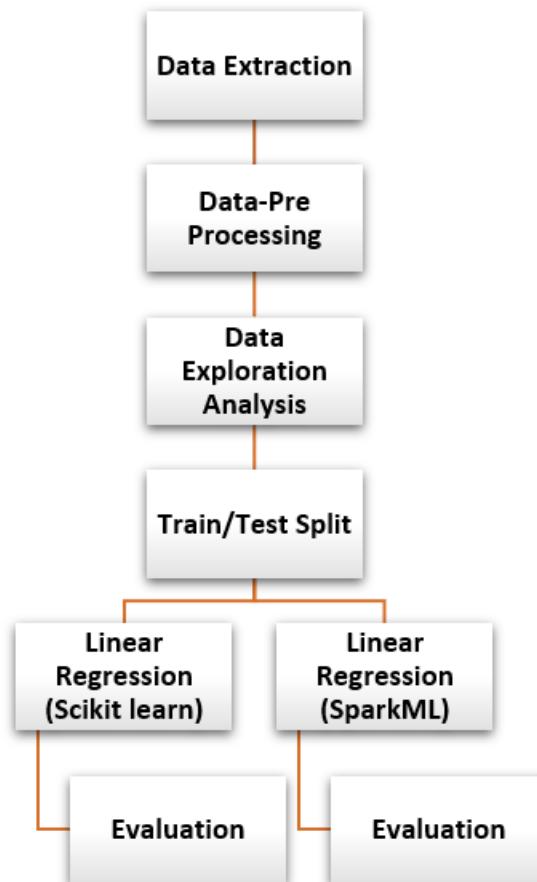


Figure 8:Regression Analysis Steps

Furthermore, K-Nearest Neighbor (K-NN) algorithm can be used to categorize an unknown variable in to an existing dataset. Once after the necessary calculations are performed (distance calculation), the labelled data point will be assigned to a target class. This algorithm is a simple concept of classifying and recognizing a data variable

(Awan *et al.*, 2021)

4.3 Technical Analysis

Python libraries such as **Pandas** for data analysis and **Plotly** (“*it is easy to learn no matter what your background or experience is and has a huge community to help in case of errors in almost all domains*”) were identified as involved technology to create responsive charts when analyzing a data visualizing project developed to track Covid patients in certain states in India. Impotency of Matplotlib library was also identified through various articles [(Sayeed, 2020),(McKinney, no date)]. Multiple Linear Regression is considered as a supervised learning technique, which has the capacity of forming one or more than one relationship between independent and dependent variables. Also due to robust nature, these relationships can be formed in a suitable manner to predict something (Awan *et al.*, 2021).

4.4 Highlighted Technology

Dash is a framework containing unique features to minimize the code content, building app using high level languages such python, F# etc . Many more important useful functions can be obtained from the official website of Dash-Plotly to develop an interactive web application as mentioned in the article (Kudale *et al.*, 2093).

4.5 Reflection

In a summary, predictive and statistical project can be developed using many technologies such as Python, Julia or R but the optimum and the most reliable language would be Python, due to availability of upgradable libraries and full documentation. Many Machine Learning projects relating to Cricket were identified and the main aspect is to identify a quality dataset to train in order to retrieve an automatic output through the raw data. Visualizing is also a part of these projects since charts and illustration would depict better understanding than values. Therefore, decisions which has to be made regarding cricket team selections will be more much easier with the help of from statistical and predictive application along with the other application automating other tasks in cricket by providing IT solutions to develop strategies for the game plan. With the involvement of both concepts of Machine Learning and Statistical Calculations helps to bring out a full picture of prescriptive data to understand the resources from a different manner.

Chapter 5 : Development Methodology

5.1 Step 1: Data Source

Identifying a reliable data source to extract relevant player details within a timeframe of 5 years is essential, since the system will be containing informative and descriptive data from 2018 to 2022. As mentioned in this article unavailability of ball-by-ball data is the down side to conduct more investigative research in depth (Krishnamohan, 2021)

Fortunately, a suitable data source to extract identified project components was found

<http://www.cricmetric.com/index.py>

5.2 Step 2 : Evaluation of the Dataset

In order to proceed forward, the next step was to separate the identified data in to different csv files. The three main components that will be present in the system are;

- Player
- Series
- Ground

Then data will be formed in a manner that is relevant for each and every player with the above identified constraints, vice versa the same procedure will be carried out for the other two components as well.

5.3 Step 3 : Connecting MongoDB to the Development Environment

A NoSQL Database will be used to store the relevant csv files in the cloud and retrieve the items to the development environment using a connection

The following is a screenshot from the development code connecting the database string in to the python files.

```
10
11
12 client = MongoClient('mongodb://localhost:27017/')
13 db = client['AnalyzingDataset']
14 collection = db['Player_Average_Prediction']
15
16 futuredata_collection=db["FUTUREDATA"]
```

Figure 9 : MongoDB Connection

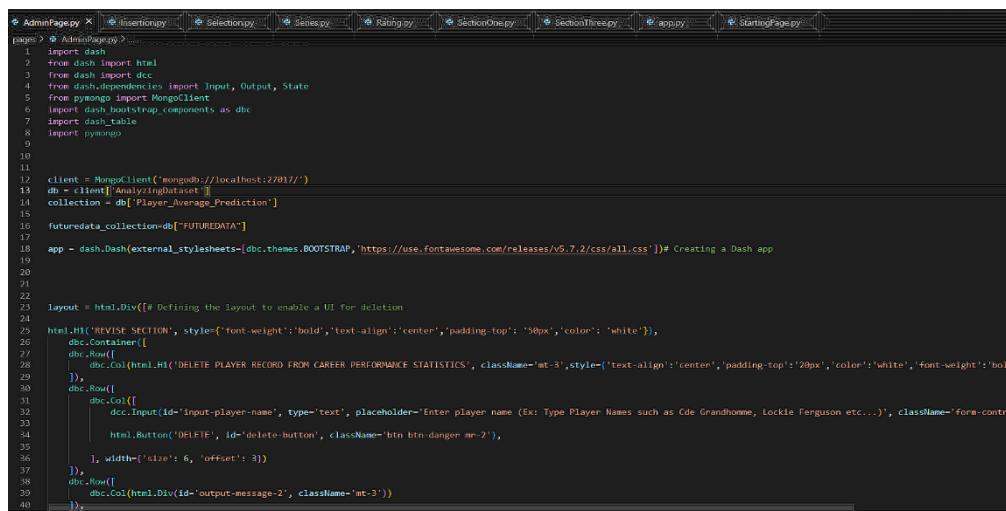
5.4 Step 4 : Code Implementation using Dash Logics

With the help of the Dash Plotly documentation <https://dash.plotly.com/> necessary logics and concepts can be extracted and can be used to create instances to develop the User Interface as well as the Machine Learning Models.

Three main steps which are required to be followed:

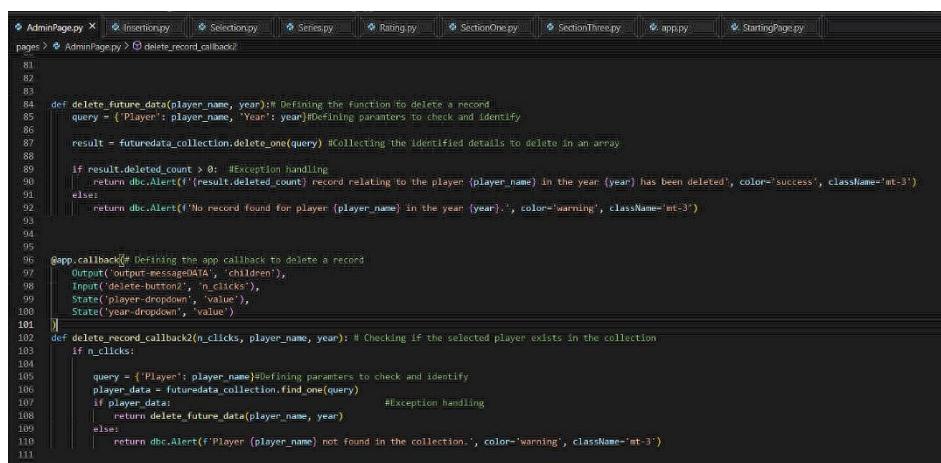
- Implementing the Code of the Layout
- Creating the Callback functions
- Linking the Callback function in the main file (app.py)

The following are some screenshots from the development environment creating the layout of Administration UI and the Functions in it.



```
AdminPage.py
pages > AdminPage.py
1 import dash
2 from dash import html
3 from dash import dcc
4 from dash.dependencies import Input, Output, State
5 from pymongo import MongoClient
6 import dash_bootstrap_components as dbc
7 import dash_table
8
9 import pymongo
10
11
12 client = MongoClient('mongodb://localhost:27017/')
13 db = client['Analyzingdataset']
14 collection = db['Player_Average_Prediction']
15
16 futuredata_collection=db["FUTUREDATA"]
17
18 app = dash.Dash(external_stylesheets=[dbc.themes.BOOTSTRAP,'https://use.fontawesome.com/releases/v5.7.2/css/all.css'])# Creating a Dash app
19
20
21
22
23 layout = html.Div([# Defining the layout to enable a UI for deletion
24
25     html.H1('DELETE SECTION', style={'font-weight':'bold','text-align':'center','padding-top': '50px','color': 'white'}),
26     dbc.Container([
27         dbc.Row([
28             dbc.Col([
29                 dbc.Row([
30                     dbc.Col([
31                         dbc.Input(id='input-player-name', type='text', placeholder='Enter player name (Ex: Type Player Names such as Cde Grandhomme, Lockie Ferguson etc...) ', className='form-control'),
32                         html.Button('DELETE!', id='delete-button', className='btn btn-danger mr-2'),
33                     ], width={'size': 6, 'offset': 3})
34                 ]),
35                 dbc.Row([
36                     dbc.Col(html.Div(id='output-message-2', className='mt-3'))
37                 ])
38             ])
39         ])
40     ])
41 
```

Figure 10 : Dash Layout



```
AdminPage.py
pages > AdminPage.py > delete_record_callback
81
82
83
84 def delete_future_data(player_name, year):# Defining the function to delete a record
85     query = { 'Player': player_name, 'Year': year}#Defining parameters to check and identify
86
87     result = futuredata_collection.delete_one(query) #Collecting the identified details to delete in an array
88
89     if result.deleted_count > 0: #Exception handling
90         return dbc.Alert(f'{result.deleted_count} record relating to the player {player_name} in the year {year} has been deleted', color='success', className='mt-3')
91     else:
92         return dbc.Alert(f'No record found for player {player_name} in the year {year}.', color='warning', className='mt-3')
93
94
95
96 @app.callback(# Defining the app callback to delete a record
97     Output('output-message2', 'children'),
98     Input('delete-button2', 'n_clicks'),
99     State('player-dropdown', 'value'),
100     State('year-dropdown', 'value')
101 )
102 def delete_record_callback2(n_clicks, player_name, year): # Checking if the selected player exists in the collection
103     if n_clicks:
104
105         query = {'Player': player_name}#Defining parameters to check and identify
106         player_data = futuredata_collection.find_one(query) #Exception handling
107
108         if player_data:
109             return delete_future_data(player_name, year)
110         else:
111             return dbc.Alert(f'Player {player_name} not found in the collection.', color='warning', className='mt-3')
112
113 
```

Figure 11 : Dash Callback Functions

5.4.1 ML Algorithm Implementation (Multiple Linear Regression)

Based on the facts mentioned in the conducted [Literature Review](#), a particular framework needs to be selected in order to create and train the models relevant to the dataset. Out of the two, the prominent and effective method of approach is to proceed with the '**Scikit Library**'. With the help of the documentation provided, an empty model can be created which satisfy the prediction need.

https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LinearRegression.html

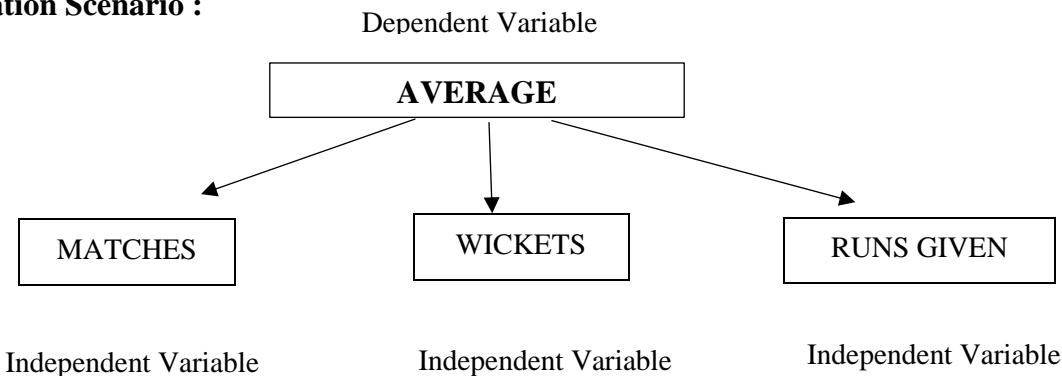
<https://plotly.com/python/ml-regression/>

The identified independent and dependent variables to implement the model are as follows

Multiple Linear Regression

To find the relationship between many independent variables with a specific dependent variable

Application Scenario :



Outcome :

Predicting the Average of the Bowler based on numbers Matches Increased

Models :

Spin Model (For Spin Bowlers)

Seam Model (For Seam Bowlers)

Output :

Graphs and Values

Predicted Average – Gauge Chart

Correlation Matrix – Heatmap

Correlation Values

Player Stats – 3D Scatter Plot

Based on an article, it has mentioned that, runs, wickets, overs, runs_last_5, wickets_last_5, striker, non-striker, and total runs are considered as independent variable in defining the final ‘team score’ (Dependent Variable) (Awan *et al.*, 2021)

So, using that particular idea it can be decided that whatever the dependent variable that is required to be predicted, the requirement is to have direct and indirect relationships with the independent variable. As discussed previously in the Literature Review the initial calculations have direct links on each other. For example, the strike rate cannot be calculated without considering the overs bowled since the number balls bowled is calculated by multiplying the number of overs by 6 (An Over consist of 6 balls)

In order to predict the bowling average , the direct linking variables are the number of wickets taken by the bowler and runs conceded by the bowler.

So, the average will be predicted accordingly based on the increment in number of matches of a player.

5.4.2 ML Algorithm Implementation (K-NN)

Another feature that has been implemented in the system is the ‘**Player Selection**’. Where a particular player name along with the relevant of match details will be depicted based on the input parameters from the user by running a scan through the collection of data to detect similar values in the database.

Based on the Dash Plotly documentation regarding K-NN <https://plotly.com/python/knn-classification/>, the specific task was much easier to be implemented. Idea behind the code is to create an instance to calculate the distance between the known variable (Average and Matches User Input) and the unknown variable (Possible Closer Values to User Input) and display the point with the least distance with assigning the values to the unknown variable.

Chapter 6 : Technology Identification

6.1 Web Application Technology

Dash Framework



Figure 12 : Dash Plotly Logo

Dash framework is built using three main components;

- Python
- Flask
- React

Python



Figure 13 : Python Logo

Due to its robust nature and readable language the most popular programming language is used for the development process. Python language will be used to proceed with data analysis using the pandas library, which usually is carried out in separately Jupyter Notebook and brought through an integrating API, but since the evolution of dash the process can be easily conducted in an IDE which helps to connect the machine learning models as well visualized charts with the frontend UI.

<https://www.python.org/doc/>

Flask



Figure 14 : Flask Logo

With the help of ‘Werkzeug’ routing system, the framework is capable of handling incoming API calls by intaking the links in to the Jinja2 template engine in order to connect with the programming language functions. Simply what it means is the gap between the User interface and Backend machine learning/data analysis charts and graphs techniques are linked together.

<https://flask.palletsprojects.com/en/2.2.x/>

React



Figure 15 : React Logo

React is a JavaScript library built by the Facebook company to develop dynamic User Interfaces using robust UI components. With the help of this library and combination of bootstrap more attractive frontend finishing can be obtained.

<https://react.dev/learn>

6.2 Database Technology



Figure 16 : MongoDB Logo

A NoSQL Database Management System (DBMS) is used to store and retrieve data in the form of documents in the development environment, furthermore with the cloud technology integrated, the integrity of the data store can be ensured. Also, due to the availability of the UI version of the database, the tasks are being simplified and clarified to easily embed in the application development. <https://www.mongodb.com/products/compass>

6.3 Development Environment (IDE)

Visual Studio Code



Figure 17 : Visual Studio Code Logo

The relevant codes are written in files with the .py extension (python files). All the User Interface codes, and Machine Learning model implementation can be organized under a particular file thanks to the huge amount of work load carried out by flask integration to the dash framework. All the necessary extensions can be installed in to the project environment using the terminal by entering the relevant key commands.

6.4 Dependencies and Libraries

After installing Python (Python 3.10.5) and MongoDB (db version v6.0.2) in the environment variables the following can be done

Using the VSCode terminal, the following dependencies and libraries were downloaded and installed in to the project environment under the relevant folder

Local folder Path : C:\Users\USER\Documents\PUSL3119 Computing Project (22AYAUM)\DEVELOPMENT PHASE - FINAL TOUCHES

- ✓ pip install dash (Version: 2.9.1)
- ✓ pip install pandas
- ✓ pip install dash-bootstrap-components 1.4.1
- ✓ pip install dash-core-components 2.0.0
- ✓ pip install dash-html-components 2.0.0
- ✓ pip install dash-table 5.0.0
- ✓ pip install flask 2.2.2
- ✓ pip install plotly 5.14.0
- ✓ pip install pymongo 4.3.2
- ✓ pip install http 7.2.2
- ✓ pip install pytest

7.1 Use Case Diagram

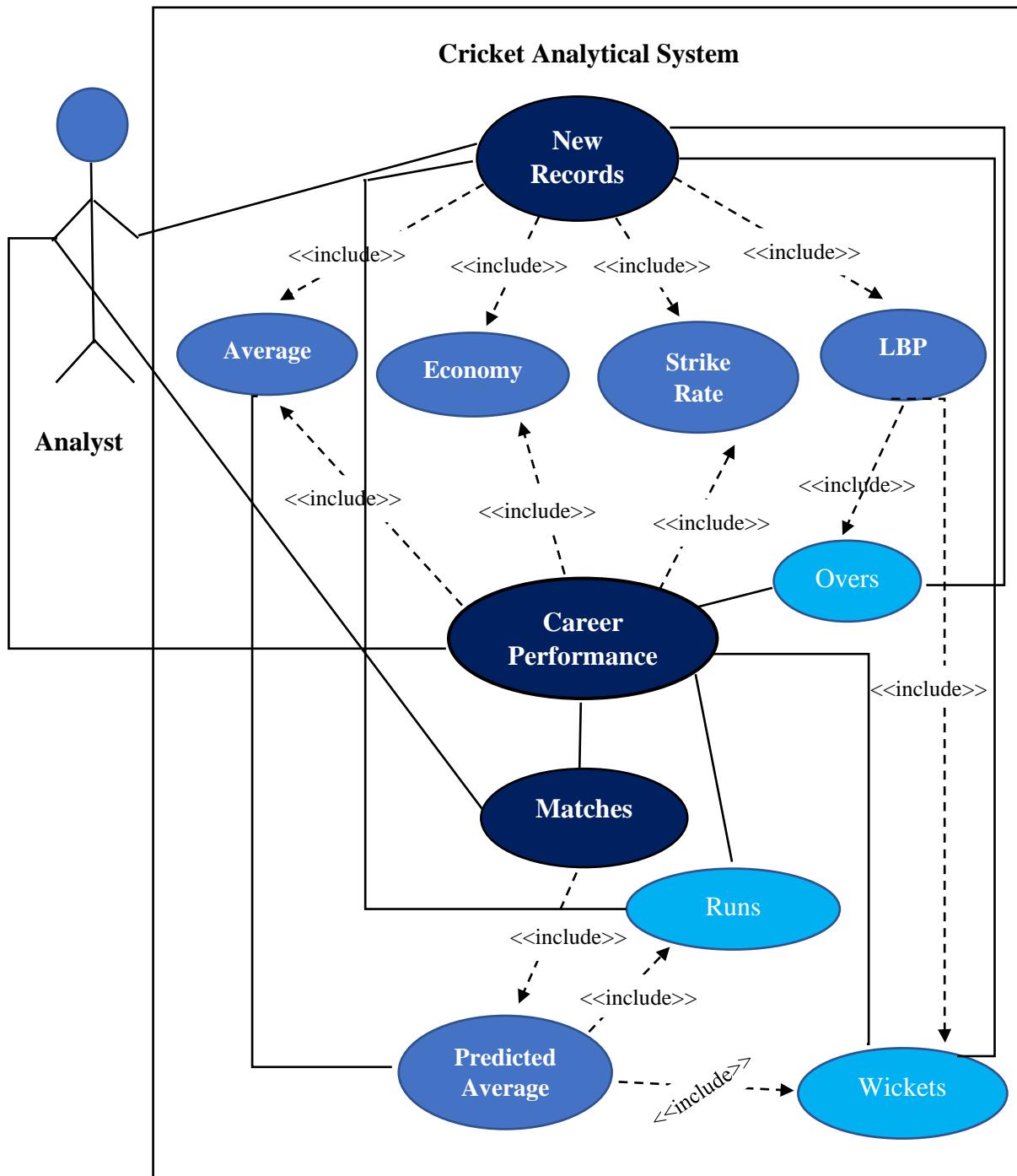


Figure 18 : Use Case Diagram

7.2 Class Diagram

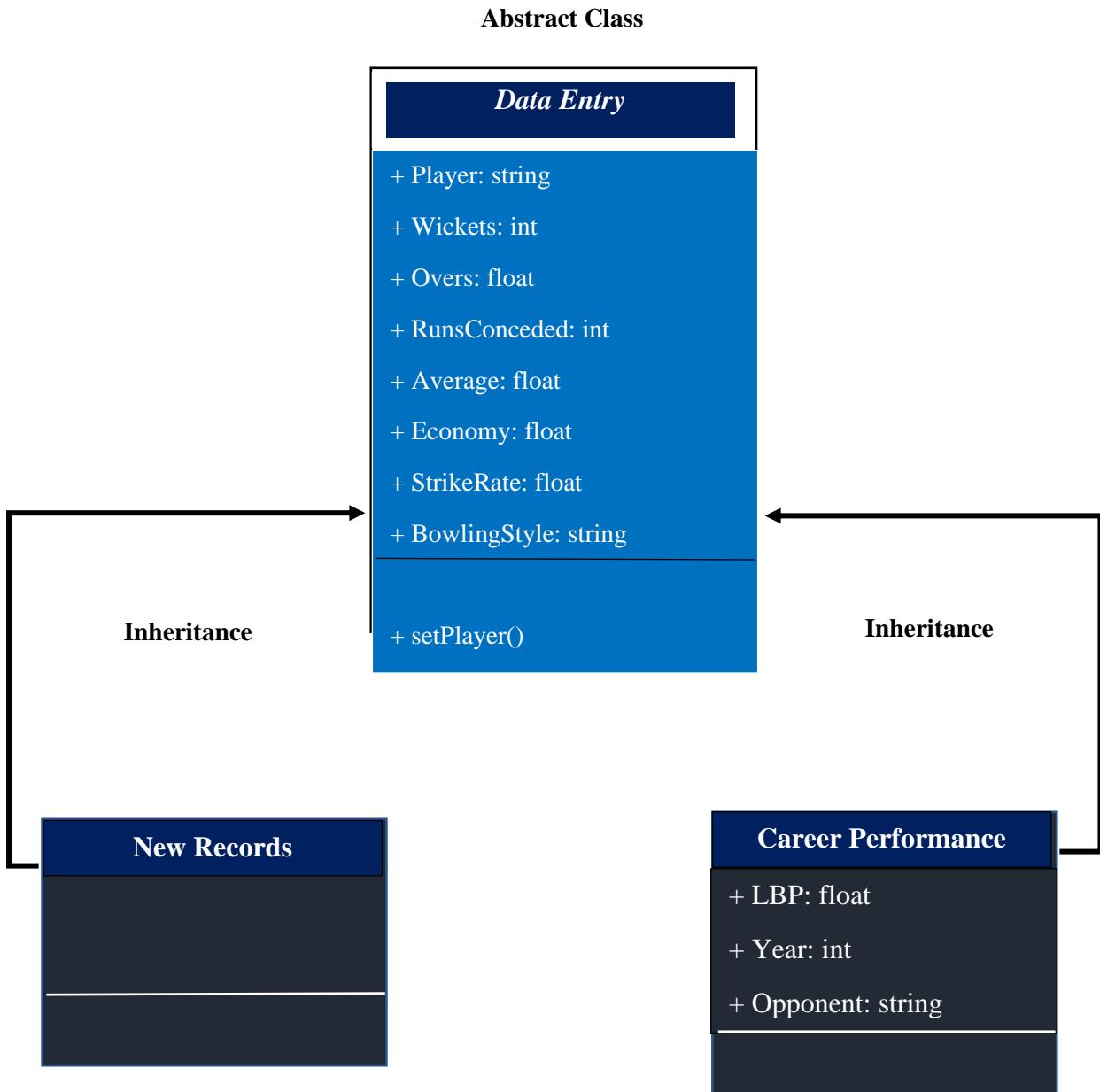


Figure 19 : Class Diagram

Chapter 8 : Development

8.1 Web Application

Using the above-mentioned [technologies](#) the development process was carried in the Visual Studio Code IDE using the previously mentioned [dependencies and libraries](#).

8.1.1 Starting Page

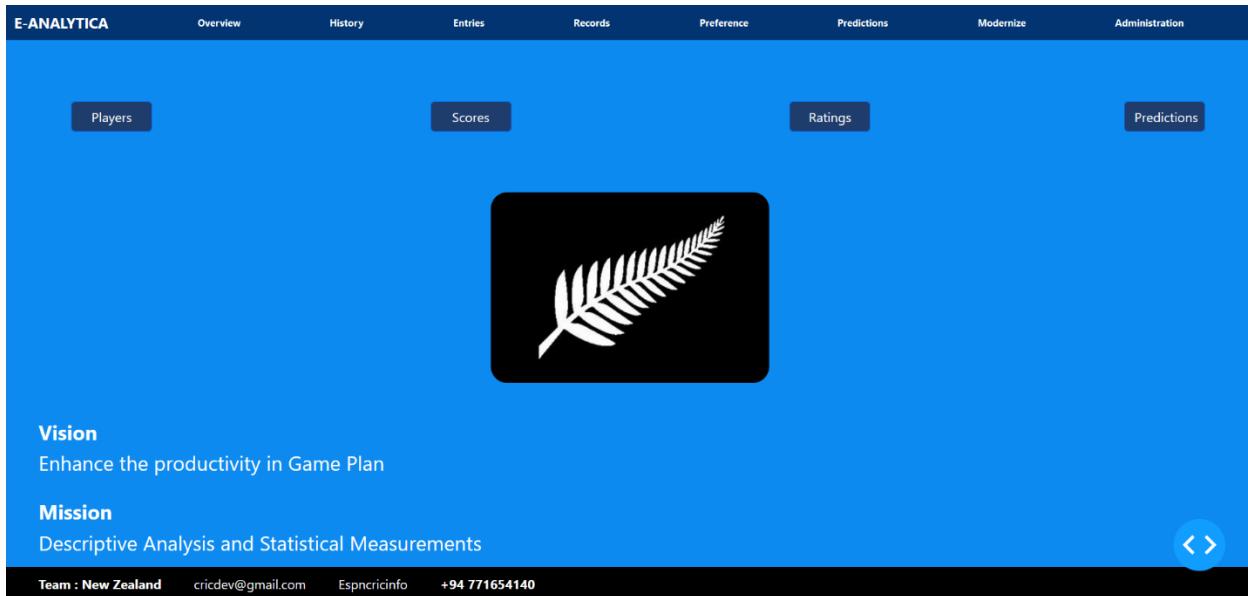


Figure 20 : Application Screenshot : Starting Page

8.1.2 History Section

A screenshot of the E-Analytica web application's History section. The header is identical to the starting page. The main content area is titled "PAST RECORDS" and displays "MATCH SUMMARIES (2018-2022)" with a series of blue buttons for matches: 2018 PAK vs NZ, 2018 ENG vs NZ, 2019 BAN vs NZ, 2019 IND vs NZ, 2019 SL vs NZ, 2021 BAN vs NZ, 2020 IND vs NZ, and 2022 NED vs NZ. Below this, it shows "BOWLERS PLAYER LIST BASED ON THE ABOVE SERIES OF MATCHES (2018-2022)" with another series of blue buttons: 2018 NZ LIST-1, 2018 NZ LIST-2, 2019 NZ LIST-1, 2019 NZ LIST-2, 2019 NZ LIST-3, 2021 NZ LIST, 2020 NZ LIST, and 2022 NZ LIST. A detailed table for the "ENG VS NZ 2018 PLAYER LIST" is shown in a box, listing players like TG Southee, MJ Santner, C Munro, TA Boult, C de Grandhomme, IS Sodhi, and LH Ferguson along with their bowling styles. A blue circular arrow icon is located on the right side of the page.

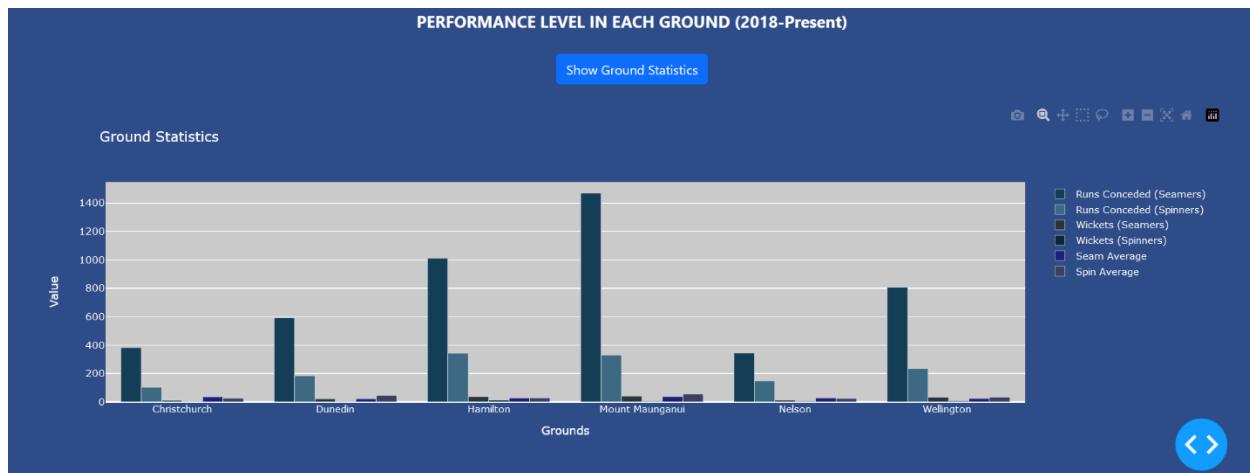


Figure 21 : Application Screenshot: History Section

8.1.3 Entries Page

The figure shows the "NEW PLAYER RECORDS (SERIES STATISTICS)" page. The page features a series of input fields for entering player statistics. Each field has a placeholder text and a red border. A blue circular icon with a double-headed arrow is located to the right of the last two input fields. At the bottom, there is a green button labeled "ACCESS THE COLLECTION".

- ENTER THE SPECIFIC YEAR: Year
- ENTER THE NUMBER OF MATCHES PLAYED: Matches Played
- ENTER THE NAME OF THE OPPONENT TEAM: Country Name
- ENTER THE PLAYER NAME: Player Name...
- ENTER THE BOWLING STYLE OF THE BOWLER: Ex: Spin or Seam
- ENTER THE NUMBER OF RUNS CONCEDED BY THE BOWLER: Runs Conceded by the Bowler
- ENTER THE NUMBER OF OVERS BOWLED BY THE BOWLER: Overs Bowled
- ENTER THE NUMBER WICKETS TAKEN BY THE BOWLER: Wickets Taken

[ACCESS THE COLLECTION](#)

The screenshot shows a table titled "Overall Bowling Performance" with the following columns: Year, Matches, Opponent, Player, Bowling Style, Runs Conceded, Overs, Wickets, Average, Economy, Strike Rate, and Overall Bowling Performance. The data spans from 2022 to 2023 across multiple international teams.

Year	Matches	Opponent	Player	Bowling Style	Runs Conceded	Overs	Wickets	Average	Economy	Strike Rate	Overall Bowling Performance
2022	3	INDIA	TG SOUTHEE	SEAM	119	21.3	5	23.80	5.59	26.00	-539.84
2022	2	INDIA	AF MILNE	SEAM	124	20	4	31.00	6.20	30.00	-668.80
2022	3	INDIA	LH FERGUSON	SEAM	132	22.5	4	33.00	5.87	35.00	-674.40
2022	1	INDIA	DJ MITCHELL	SEAM	25	7	3	8.33	3.57	14.00	-14.29
2022	3	INDIA	MJ SANTNER	SPIN	80	13	1	80.00	6.15	78.00	-467.31
2022	1	INDIA	MG BRACEWELL	SEAM	18	2	0		9.00		-162.00
2022	3	INDIA	MJ HENRY	SEAM	97	24	1	97.00	4.04	144.00	-367.04
2023	2	SRI LANKA	HB SHIPLEY	SEAM	63	15.3	8	7.88	4.12	11.75	-59.41
2023	2	SRI LANKA	DJ MITCHELL	SEAM	44	10	5	8.80	4.40	12.00	-68.60
2023	2	SRI LANKA	MJ HENRY	SEAM	26	16	3	8.67	1.62	32.00	32.75
2023	2	SRI LANKA	BM TICKNER	SEAM	55	11.5	2	27.50	4.78	37.00	-213.04
2023	1	SRI LANKA	ISH SODHI	SPIN	40	8	0		5.00		-200.00

REFRESH THE COLLECTION



Figure 22 : Application Screenshot : Entries Page

8.1.4 Records Page

The screenshot shows a series of cards for different cricket series, each with a set of tabs for viewing bowling statistics. The series listed are PAK VS NZ 2018, ENG VS NZ 2018, BAN VS NZ 2019, IND VS NZ 2019, SL VS NZ 2019, IND VS NZ 2020, BAN VS NZ 2021, and NED VS NZ 2022. Each card has tabs for OVERS, WICKETS, AVERAGE, ECONOMY, STRIKE RATE, DOT BALLS, and RUNS GIVEN.

Figure 23 : Application Screenshot : Records Page

8.1.5 Preference Page

The screenshot shows the E-Analytica application interface. At the top, there is a navigation bar with tabs: Overview, History, Entries, Records, Preference (which is currently selected), Predictions, Modernize, and Administration.

The main content area is titled "PLAYER COMPARISON SECTION". It features a search bar labeled "SEARCH FOR PLAYER PERFORMANCE" with the placeholder "Enter player name" and a "Search" button. Below the search bar are two input fields: "ENTER YOUR PLAYER'S AVERAGE VALUE:" with a value of "0" and "ENTER THE NUMBER OF MATCHES PLAYED BY THE PLAYER:" with a value of "0". A blue "RUN SCAN" button is positioned between these two input fields.

A large central box displays "Performance Statistics" for the player "Matt Henry". The table contains the following data:

Attribute	Value
Player	Matt Henry
Average	22.6
Matches	13

The name "Matt Henry" is also displayed prominently below the statistics table. There are two circular icons with arrows pointing left and right, likely for navigating through multiple player profiles.

Figure 24 : Application Screenshot : Preference Page

8.1.6 Prediction Page

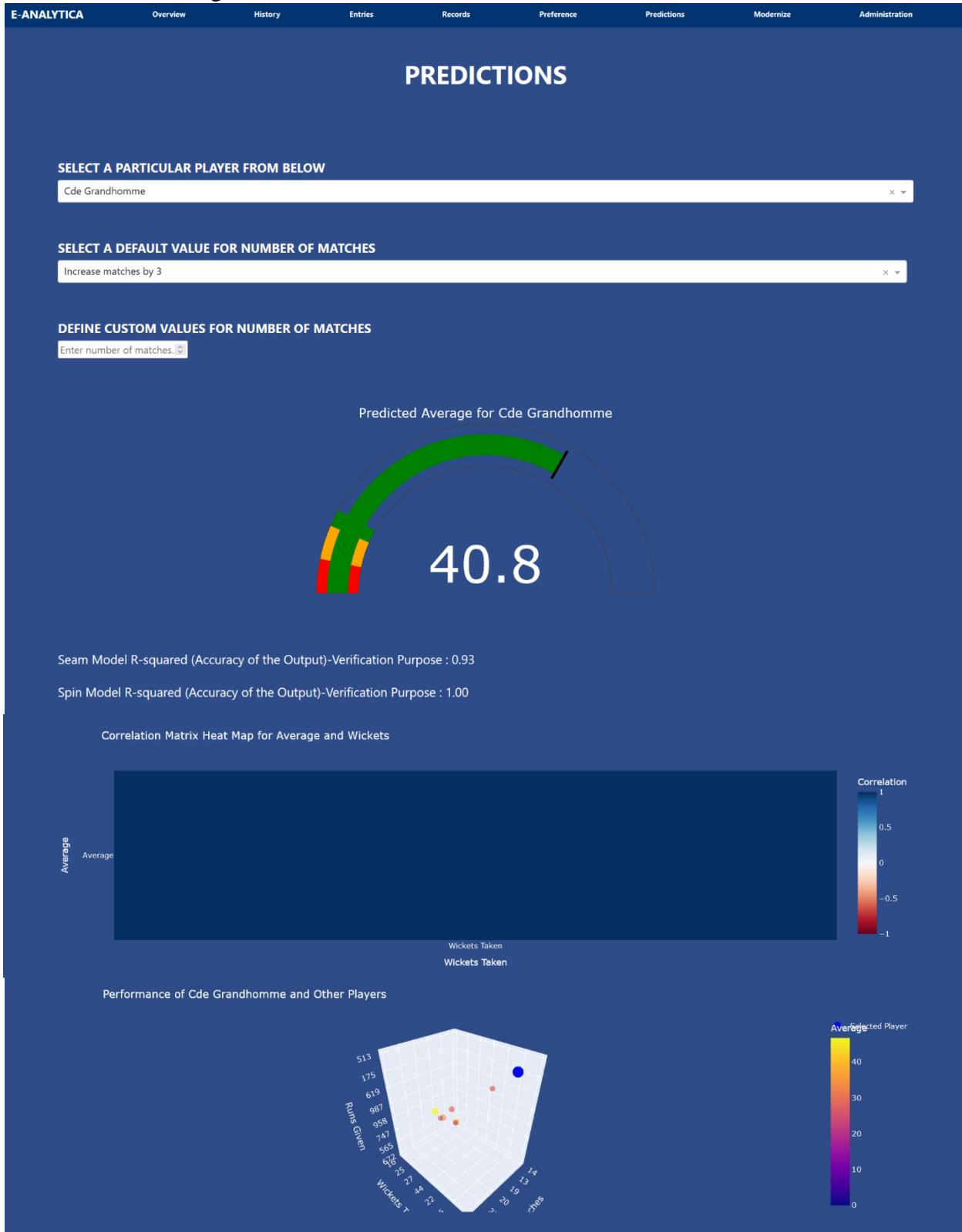


Figure 25 : Application Screenshot : Prediction Page

8.1.7 Modernize Page

The screenshot shows the 'Modernize' section of the application. At the top, there is a header bar with tabs: Overview, History, Entries, Records, Preference, Predictions, Modernize, and Administration. Below the header, the title 'UPDATE SECTION' is displayed. A sub-section titled 'INSERT STATS IN TO THE COLLECTION (CAREER PERFORMANCE STATISTICS)' contains several input fields for player statistics:

- Player Name: Player
- Matches Played: Matches
- Wickets Taken by the Player: Wickets
- Runs Conceded by the Player: Runs Given
- Bowling Type (i.e Spin or Seam): Bowling Style (with a blue double-headed arrow icon)
- Role in the Team (i.e Part-Timer or Main Bowler): Role

Below these fields are two buttons: 'Update the Collection' and 'Clear'. Underneath is a section titled 'View Collection (Player Stats in New Zealand from 2018-Present)' which displays a table with columns: Player, Matches, Wickets, Runs Given, Average, Bowling Type, and Role. A blue double-headed arrow icon is positioned next to the table.

Figure 26 : Application Screenshot : Modernize Page

8.1.8 Administration Page

The screenshot shows the 'Administration' section of the application. At the top, there is a header bar with tabs: Overview, History, Entries, Records, Preference, Predictions, Modernize, and Administration. Below the header, the title 'REVISE SECTION' is displayed. There are two main sections for record management:

- DELETE PLAYER RECORD FROM CAREER PERFORMANCE STATISTICS:** A text input field for entering player names and a red 'DELETE' button.
- DELETE NEW ENTRIES FROM NEW RECORDS:** Two dropdown menus for selecting a specific player and a specific year, followed by a red 'DELETE RECORD' button and a blue double-headed arrow icon.

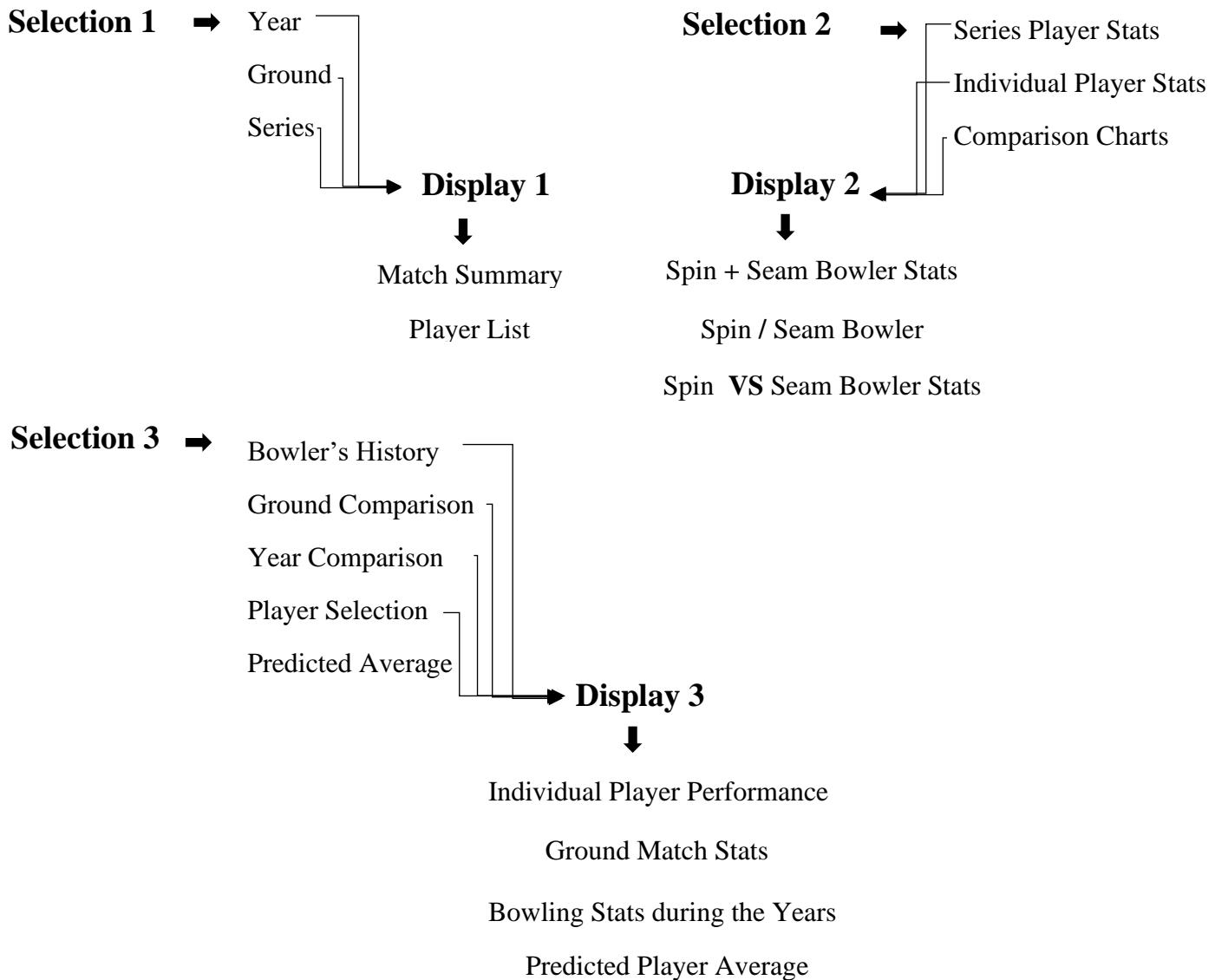
Figure 27 : Application Screenshot : Administration Page

Chapter 9 : Requirements

9.1 Functional Requirement

Exploratory/Analytic statistical application contains the functional option of viewing different categorized data, defining significant values in a graphical and numerical format to explain the logic and the importance of the data displayed

Navigation



Each property contains in detail information under the specific content which elaborates the points in a methodical manner. The graphs indicates the subtle changes, highest and lowest values, assumptions and variation among different statistics which helps the analyst to develop strategies.

Contribution to the Series

- Wickets – Total wickets by each bowler (seam/spin)
- Runs Conceded – Total runs conceded by every bowler
- Overs Bowled – Total Overs bowled by each bowler in the Series
- Dots – No of Balls Bowled by Bowlers without Conceding Runs
- Economy Variations Among all the Bowlers (**Runs Conceded/Overs Bowled=Eco**)
- Average Variation Among all the Bowlers (**Runs Conceded/Wickets Taken**)
- Bowling Strike Rate of every Bowler (**Balls Bowled/Wickets Taken**)

Comparison Charts (Seam Vs Spin)

- Total Wickets for the Spin and Seam Bowlers
- Total Overs Bowled by the Seam and Spin Bowlers
- Average, Bowling Strike Rate and Economy among different type of Bowlers
- Dots balls, Runs Given, Matches Played by Seam and Spin

Depicting The Main Objectives Using Statistical Representation and Calculations

- Bowler Stats over the 5 years at a particular Ground (**Total Wickets, Totals Runs Conceded, Total Overs Bowled and Averages for both spin and seam bowlers at that Venue are considered**)
- Seam Vs Spin Ground Stats (**Total wickets, Runs Conceded, Bowling Average by every bowler**)
- Averages at different grounds (Spin and Seam)
- Wickets Comparison between spin and seam bowling (**Deciding whether above average or below average**)
- Runs Scoring comparison on spinners and seamers (**Deciding whether it is above average below average**)
- Player Comparison data tables (In New Zealand Matches during the last 5 years)
- Average Prediction based on Wickets, Runs Conceded, (Based on different circumstances)
- Player selection based on Insert parameters (Input Average and Input Number of matches)
- Player stats Insertion in to the existing collection (Career Performances Collection)
- New Series Player Records Insertion in to the existing collection (Series Player Collection)
- Calculations of Average , Economy, Strike Rate and LBP (Lemmer's Bowling Performance)

9.2 Non-Functional Requirement

The main focus of creating an interactive user interface is to establish an interesting attraction between the user with the simplification of the tasks in the application by implementing the functionalities of the application in a methodical way and improving the usability factor.

The following articles helps to understand how the user interface should be developed in a such a manner to display the important information in the best possible way

“The visual data profiling capabilities should be integrated with a table view interface that lets the user manipulate columns and rows directly”(von Zernichow and Roman, 2017). Furthermore, author states that necessary data cleansing and functionality processes are required, to develop the application which addresses the achievements that needs to be reached by the user.

Another aspect of visualizing projects are to develop the application with a proper color combination to suppress the uneasiness when the user interacts with the application.

The following article explains how certain colors can be used in different parts of the application

“It states that blue and orange can be used to visualize the comparison between two datasets”. These colors can reduce the complexity and will be ideal to be represented in graphs.

The color principal indicates that two main principles govern the design colors, namely; Contrast and Analogy. Both provide different aspects in which enhance the efficacious nature of usability factors (von Zernichow and Roman, 2017).Labelling terms and identities are also important in certain occasions since the system takes in input values and necessary instructions need to be provided to the user. Validation is another prominent attribute that is required to be in tact with the functional features to verify the user input data, whether it is up to the standard form of values that is essential to proceed forward with the task. Since otherwise there will be outrageous abnormal behaviors from the models implemented and embedded in the system due to mis matched data that is processed to produce an output.

Currently the smoothness of the application is up to standard, performing the implemented features without causing sudden failures. Implications that be drawn from this is, even though the system is performing necessary action uninterrupted, a relevant backup methodology is expected to handle the exceptions that can be raised in the future. Currently the Flask integration is handling the models with utmost cautiousness to prevent the system from crashing due to lagging by improving the latency of HTTP request functions (Dash Core principles) which helps to maintain the performance of the application.

<https://dash.plotly.com/performance>

<https://dash.plotly.com/loading-states>

End Project Report

Project Idea – Bowling Performance Analysis of the Opponent Team

Project Vision – Statistical and Predictive Analysis Web Application

Target Audience – Analysts, Coaching Staff, Team Captains

Summary 1 : Overall Project

- ❖ Based on the identified requirements which were gathered through the available resources, the project development was conducted to satisfy the above-mentioned statements. A suitable technology was selected in order to embed **machine learning and mathematical algorithms** in the software. As a part of the learning curve, thorough **research paper** analysis was conducted in order to create the structure of the project and to identify tools and sources of help to establish the idea.
- ❖ **Dash** Framework was used for the development process since it has been built using several other technologies to ease the workload of the developer.
- ❖ Cloud technology was integrated with the use of **MongoDB** as the database to store and retrieve the collection of data.
- ❖ Machine Learning technique will predict an important **performance measurement** (Average) attribute of a cricket player and depict it through a visualization technique.
- ❖ Several **calculations** will be made by the application as a part of performance measurement (Initial Average, Economy, Strike Rate, LBP-Lemmer's Bowling Performance value).
- ❖ Statistical graphs and tables will indicate past records and user entered new player records as well relevant series data.
- ❖ Lastly as the administration feature, delete option was given to alter the records in the database.

Project Objectives

- To keep a track of important player performance details of other teams
- Significant ground records relating to venues in New Zealand
- Visualization of statistical information in form of charts
- Performing important calculations to ease the process (Average, Economy and Strike Rate)
- Predicting the ‘Average’ of active players based on the increment in number of matches
- Calculating the ‘Bowling Performance’ of a bowler to evaluate among other bowlers
- Depiction of player attributes
- Updating current databases with new player/performance details
- Indicating records from previous series (Match Summaries)

- Comparison selection against benchmark attributes
- Ability to alter previous data (Add or Remove player profile details)

Summary 2 : Project Objectives

- Retaining Player Performance Records
- Player performance measurement calculations (Average, Economy, Strike Rate and LBP)
- Statistical Visualization of Past Records (2018-2022) (Graphs and Charts)
- To develop strategies to counterattack the strong bowling unit of New Zealand Team

Changes Made (Additions/Removals)

- ❖ Initial intension was only to analyze the prescriptive data of the bowling statistics and visualize the data in manner that is more understandable to the user and differentiate the values in more significant way.
- ❖ Later the machine learning techniques were added to improve the quality of the product and to ensure the durability of the software. With the application having the ability to make predictions by analyzing several factors, the productivity and the importance of the application will be increased.
- ❖ Performance measurement is equally important as performance analysis therefore integrating methods and equations were conducted to evaluate and consider the past records.
- ❖ Player-player comparison segment have been removed and replaced with Player Selection feature.
- ❖ The technology used (Dash) have turned out to be surprisingly effective than expected since initially the method of approach was to separately transfer the charts and graphs from the Jupyter notebook through API links (Flask API's) to the frontend development using ReactJS.
- ❖ Toggle feature with other teams having the same feature was taken off the context since it was causing complications and expanding the scope of the project.
- ❖ Comparison Visualization was converted in to representing individual player details.
- ❖ Inclusion of animation and dynamic charts was carried out.
- ❖ Even though data visualization was the initial key concept, a mix of formal way of representing (Tabular) and Visualizing techniques were used, since statistical implementation was brought to context at a later period.
- ❖ Hiding information was not a part of the plan, initially creation of administrative privileges was not mentioned in the scope, but later alterations was a concern raised, therefore delete options was enabled in the system. Reason is statistical data is publicly available and these data can be used for analyzes at any time, so requirements of specific user authorization is not mandatory unless more sophisticated and sensitive calculations with more insight is provided regarding a player which will be useful to the opponent team as well.

Project Post - Mortem

- ❖ Technically a variety of functions can be implemented under the specific project domain as discussed in the literature review. Out of which performance measurements is only one segment where the story have been developed around the initial idea.
- ❖ Clinically evaluating past information can bring out more understanding of the back story as to why the data is situated the way it is. Therefore, prescriptive analysis is a must that need to be established in the system in order to use ‘Big Data Concepts’ to capture and convert the available sources in to more categorized manner by breaking down values and terminology in to more atomic meanings.
- ❖ Dash being a low-code open-source framework allowing developers to easily adopt to the nature of the language since a high-level language is used as the primary programming starting point for the application (Python). Newer Versions will bring out new features and modernize the code in simpler manner for designing in the layout and databases etc.
- ❖ In the field of Data Science, a majority of data is used to analyze it from a statistical point of view to highlight the significant nature of the dataset. Generally, sports data is at a high data volume and keeps accumulating day by day in the data warehouses. In order to make use of this data research teams have come up with different methods and techniques to make predictions using the existing information. Currently in today’s world these techniques have managed to bring success encouraging more and more systems to be built in order to stimulate and predict even before the game have taken place.
- ❖ Therefore, specifically the sport of cricket has different types of data generated after each and every match. Systems need to upgrade the performance level (Efficiency of the system) to store and analyze these data for future use. Deriving sub projects under this domain will not result in failure. Continuity is the key to maintain as the system need to upgrade its pool of data to further to re-evaluate and compare new data with the existing data to predict something.
- ❖ More objectives can be situated with different intensions indicating different other predictions based on several other attributes than which are discussed in this entire project. There is no particular barrier narrowing the ideology.
- ❖ The application will be consisting with main priorities pertaining to the performance analysis conducted through and predictive and descriptive scrutiny, summarizing the evaluation in forms of graphical representation.
- ❖ More intense depth can be reached with the statistical data than what is already established in the current system, where the bowling segment can be split in to sub section because Spin and Seam Bowlers can be either Left Arm Bowlers or Right Arm Bowlers. Therefore, more precise separate analysis can be conducted to discover acute details. Furthermore, variety of bowling styles are present within the sub sections where spinners are categorized in four main types (leg spinners, off spinners, leg break spinner and orthodox), so more unique evaluations and predictions can be separately made without limiting to the current standard. Same with the seam bowlers, these bowlers can be categorized in to as medium pacers or fast.
- ❖ Minor faults in the dataset led to incomplete data retrieval even though it was not disastrous.

Conclusion

Once after the problem statement was defined, necessary requirements were identified and after a thorough analysis and cross diagnosing existing systems through research articles. Necessary techniques and methods were found during the next step, which was taken into consideration, was to define the scope of the project and plan the outline. After that with the correct use of technology the desired output was achieved. Proper justifications with evidence were provided to validate the point of view and built software to align with the mentioned objectives.

The overall idea of the project is described through a diagram below.

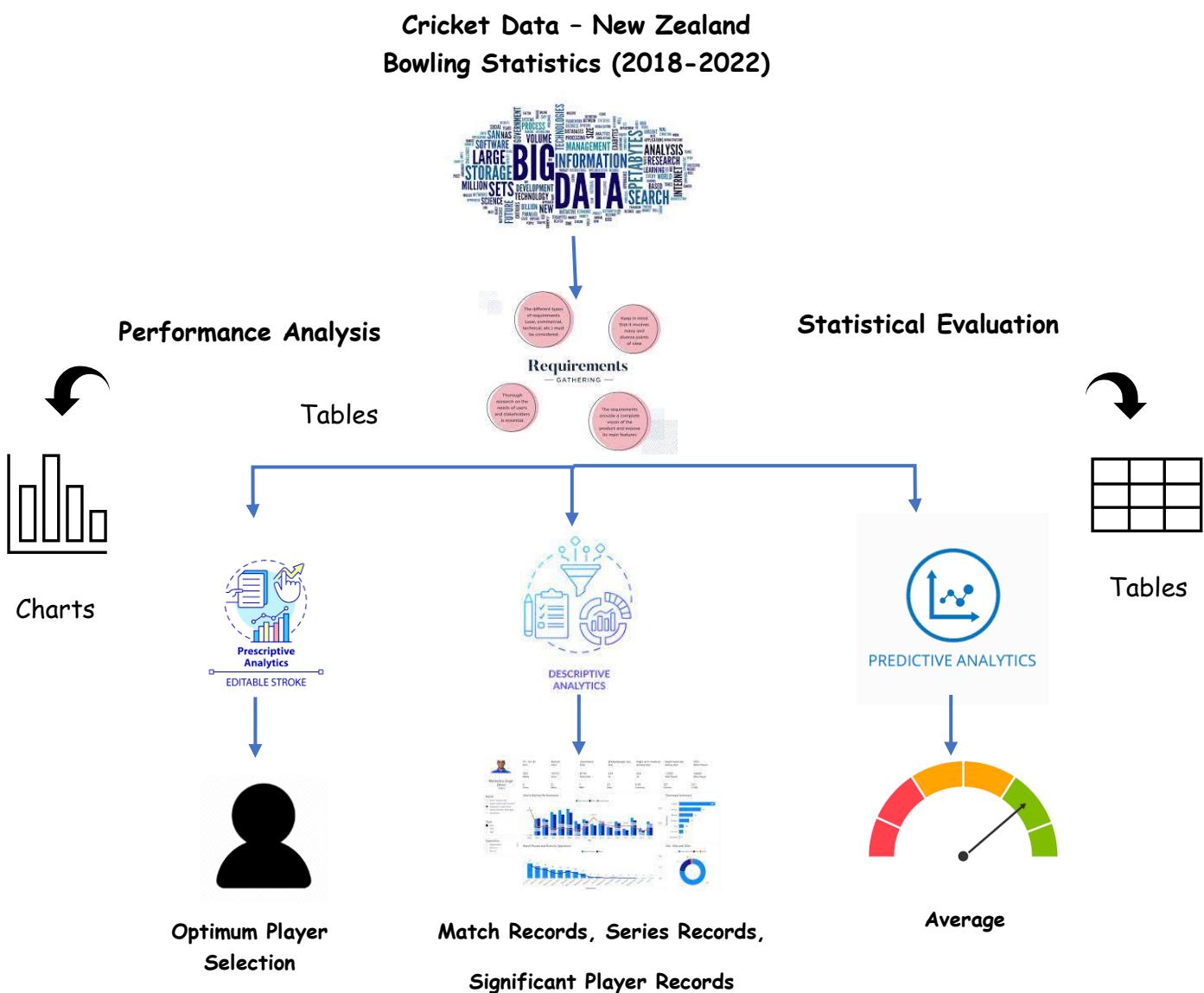


Figure 28 : Overall Idea of the Project

Reference List

- Attygalle, D. and Sunethra, A.A. (2014) *A Study on Selecting the Best Batsmen for the Next One-Day International Cricket Match: In Sri Lankan Context*. Available at: <https://www.researchgate.net/publication/311717277>.
- Awan, M.J. *et al.* (2021) ‘electronics Article’. Available at: <https://doi.org/10.3390/electronics>.
- Bhattacharjee, D., Pahinkar, D. and Pahinkar, D.G. (2012) *Analysis of Performance of Bowlers Using Combined Bowling Rate T20 Cricket View project Adsorption Based Gas Separation View project Analysis of Performance of Bowlers using Combined Bowling Rate*, *International Journal of Sports Science and Engineering*. Available at: <https://www.researchgate.net/publication/327388396>.
- Dananjaya Weerasinghe, T. (no date) ‘Downfall of Sri Lankan Cricket: Local Fans’ Perspective Enhancing Livelihood of Low Income and Economic Marginal Communities through Technology Mediated Education and Training View project Downfall of Sri Lankan Cricket: Local Fans’ Perspective View project’. Available at: <https://doi.org/10.13140/RG.2.2.12800.56329>.
- Kanungo, V. and Tulasi, B. (2019) ‘Data visualization and toss related analysis of IPL teams and batsmen performances’, *International Journal of Electrical and Computer Engineering*, 9(5), pp. 4423–4432. Available at: <https://doi.org/10.11591/ijece.v9i5.pp4423-4432>.
- Krishnamohan, T. (2021) ‘Computing the Runs that Should Be Scored Every over When Chasing a Target in Limited-overs Cricket Using the A* Algorithm’, *Applied Artificial Intelligence*, 35(15), pp. 2087–2101. Available at: <https://doi.org/10.1080/08839514.2021.1999011>.
- Kudale, H.S. *et al.* (2093) *A REVIEW OF DATA ANALYSIS AND VISUALIZATION OF OLYMPICS USING PYSPARK AND DASH-PLOTLY*, *International Research Journal of Modernization in Engineering Technology and Science* [@International Research Journal of Modernization in Engineering](http://www.irjmets.com). Available at: www.irjmets.com.
- Malhotra, A. and Krishna, S. (no date) *A Statistical Analysis of Bowling Performance in Cricket A Statistical Analysis of Bowling Performance in Cricket View project A Statistical Analysis of Bowling Performance in Cricket*. Available at: <https://www.researchgate.net/publication/312492379>.
- Mckinney, W. (no date) *pandas: a Foundational Python Library for Data Analysis and Statistics*. Available at: <https://www.researchgate.net/publication/265194455>.
- Mukherjee, S. (2014) ‘Quantifying individual performance in Cricket - A network analysis of batsmen and bowlers’, *Physica A: Statistical Mechanics and its Applications*, 393, pp. 624–637. Available at: <https://doi.org/10.1016/j.physa.2013.09.027>.
- Pramanik, M.A. *et al.* (2022) ‘Performance Analysis of Classification Algorithms for Outcome Prediction of T20 Cricket Tournament Matches’, in *2022 International Conference on Computer*

Communication and Informatics, ICCCI 2022. Institute of Electrical and Electronics Engineers Inc. Available at: <https://doi.org/10.1109/ICCCI54379.2022.9740867>.

Saikia, H., Bhattacharjee, D. and Lemmer, H.H. (2012) ‘Predicting the performance of bowlers in IPL: An application of artificial neural network’, *International Journal of Performance Analysis in Sport*, 12(1), pp. 75–89. Available at: <https://doi.org/10.1080/24748668.2012.11868584>.

Sayeed, M.A. (2020) ‘Tracking the Spread of COVID-19 Cases in India using Data Visualizing and Forecasting Techniques’, *International Journal for Research in Applied Science and Engineering Technology*, 8(5), pp. 474–492. Available at: <https://doi.org/10.22214/ijraset.2020.5076>.

von Zernichow, B.M. and Roman, D. (2017) ‘Usability of visual data profiling in data cleaning and transformation’, in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Springer Verlag, pp. 480–496. Available at: https://doi.org/10.1007/978-3-319-69459-7_32.

Appendices

User Guide

History Page (Functions)

The diagram illustrates the user flow through the History Page. It starts with a section titled "MATCH SUMMARIES (2018-2022)" containing buttons for various matches. A callout box points to these buttons with the text: "Can check past match scores through a button click". Below this is another section titled "BOWLERS PLAYER LIST BASED ON THE ABOVE SERIES OF MATCHES (2018-2022)" with buttons for different NZ lists. A callout box points to these buttons with the text: "When the buttons are clicked, past player list is". This leads to a third section titled "PERFORMANCE LEVEL IN EACH GROUND (2018-Present)". Inside this section is a button labeled "Show Ground Statistics". A callout box points to this button with the text: "This reveals the ground statistics in form of bar charts within the span of 5 years".

Entries Page (Functions)

The diagram shows the "Entries Page (Functions)". On the left, a callout box contains the text: "When these fields are filled and submitted The below show properties are shown with accurate calculated values". To the right is a form with several input fields: "ENTER THE NUMBER OF MATCHES PLAYED" (Matches Played), "ENTER THE NAME OF THE OPPONENT TEAM" (Country Name), "ENTER THE PLAYER NAME" (Player Name...), "ENTER THE BOWLING STYLE OF THE BOWLER" (Eg Spin or Seam), "ENTER THE NUMBER OF RUNS CONCEDED BY THE BOWLER" (Runs Conceded by the Bowler), "ENTER THE NUMBER OF OVERS BOWLED BY THE BOWLER" (Overs Bowled), and "ENTER THE NUMBER WICKETS TAKEN BY THE BOWLER" (Wickets Taken). At the bottom right of the form is a button labeled "UPDATE THE COLLECTION" and a note "Please fill out this field".

Year	Matches	Opponent	Player	Bowling Style	Runs Conceded	Overs	Wickets	Average	Economy	Strike Rate	Overall Bowling Performance
------	---------	----------	--------	---------------	---------------	-------	---------	---------	---------	-------------	-----------------------------

Records Page (Functions)

These buttons will depict the visualized transformation of series tabular data. The rest of the section below this set of buttons will indicate the same for different series held

PAK VS NZ 2018

OVERS WICKETS AVERAGE ECONOMY STRIKE RATE DOT BALLS RUNS GIVEN

Preference Page (Functions)

When the previous data entry page is updated, the user can enter a search query to find a particular player with current performance within the pool (Average, Economy, Matches, Wickets)

PLAYER PERFORMANCE (2022-PRESENT)

Enter player name (Ex: Search for names like AF MILNE, TG SOUTHEE ...)

SEARCH

ENTER YOUR PLAYER'S AVERAGE VALUE: 0

ENTER THE NUMBER OF MATCHES PLAYED BY THE PLAYER: 0

RUN SCAN

When certain player values are given and the button is clicked, the user can obtain possible closest player compared data

Prediction Page (Functions)

SELECT A PARTICULAR PLAYER FROM BELOW
Cde Grandhomme

SELECT A DEFAULT VALUE FOR NUMBER OF MATCHES
Increase matches by 3

DEFINE CUSTOM VALUES FOR NUMBER OF MATCHES
Inter number of matches: 5

When a player is selected from the dropdown and custom inputs values are given for number of matches, a visualized predicted average value will be showcased

Modernize Page (Functions)

INSERT STATS IN TO THE COLLECTION (CAREER PERFORMANCE STATISTICS)

Player Name	Player
Matches Played	Matches
Wickets Taken by the Player	Wickets
Runs Conceded by the Player	Runs Given
Bowling Type (i.e Spin or Seam)	Bowling Style
Role in the Team (i.e Part-Timer or Main Bowler)	Role

Update the Collection Clear

View Collection (Player Stats in New Zealand from 2018-Present)

Player	Matches	Wickets	Runs Given	Average	Bowling Type	Role
--------	---------	---------	------------	---------	--------------	------



The Prediction page details will be updated once these fields are filled and submitted

Administration Page (Functions)

User has the privilege to delete any records which was entered earlier from the collections



DELETE PLAYER RECORD FROM CAREER PERFORMANCE STATISTICS

Enter player name (Ex: Type Player Names such as Cde Grandhomme, Lockie Ferguson)

DELETE

DELETE NEW ENTRIES FROM NEW RECORDS

Select the specific Player:

Select a player...

Select the specific Year:

Select a year...

DELETE RECORD



Introduction

Problem Statement

Cricket is a well-known sport which is played in many of the European countries as well as in the Asian Continent. Sri Lankan Cricket Team has recently been underperforming in countries such as South Africa, New Zealand, Australia and England (which can be referred to as Sena Nations), while the lowest encountered nation would be with New Zealand. Over the past decade, the number of times both teams have played matches in New Zealand are considerably lower.

New Zealand in Sri Lanka ODI Series	2013/14	drawn	(3)	1-1
Sri Lanka in New Zealand ODI Series	2014/15	New Zealand	(7)	4-2
ICC Cricket World Cup (in Australia/New Zealand)	2014/15	Australia		
Sri Lanka in New Zealand ODI Series	2015/16	New Zealand	(5)	3-1
ICC Champions Trophy (in England)	2017	Pakistan		
Sri Lanka in New Zealand ODI Series	2018/19	New Zealand	(3)	3-0
ICC Cricket World Cup (in England)	2019	England		
ICC Men's Cricket World Cup Super League	2020-2022/23	-		

Due to this reason, the **following factors** would lead SLC Team towards losing matches in New Zealand as well as in other major ICC events.

- Unawareness of Opponent Player Statistics
- Unable to Identify the Performance level of Players in certain Grounds
- Unable to Prep the Relevant player for the Relevant Conditions and Strategies
- Unable to Spot and Adjust with the Pitch Conditions

Stating the Solution

Previous match data is vital to retain, as analyzing these data would help to understand the nature of the game and foresee a game plan and make suitable decisions when preparing players for the upcoming games. Out of the above stated factors, statistical views can be given to express and give a full picture of the players seen over the years with the statistics records and the performance level of each and every player played at grounds in New Zealand (International ODI Cricket) as an IT

solution. The idea is to visualize data in a way in which the computer analyst can grasp the content and interpret the information.

Reasoning

Proposing Solution – Web Application

Concept – Data Visualization

Embedding Components – Statistical Values, Illustrations, Dynamic Charts

- A Web Application can be considered as the ideal outcome to illustrate and indicate the records in an eye-catching manner (Important data and relevant functionalities will be built using a lightweight framework)
- Data Visualizing is the process of expressing the meaningful statistical values in a diagrammatic view to ensure convenience and easy reference
- Numerical values are important to showcase, since performance level of a player is measured from different aspects with the acquired stat records from a match (Ex: Runs, Wickets, and Overs etc.)

Match records are hard to obtain, and ICC (International Cricket Council) is the only place where all the data are being governed. Player records and Player performance levels are indicated during the match period as seen on the TV screens but there is no specific place to dive into the data and do any analysis regarding the previous records thoroughly at a single instance.

Teams tend to read or search for the previous match records along with the player stats as the ordinary way of obtaining information before preparing for a game.

Development of such an application would set steppingstones towards the establishment of networks full of information relating to each and every international cricket team and find similarities and differences in order to make decisions beforehand.

Business Perspective

Business Need

Players tend to play sports as a profession, and due to globalization different groups and societies engage in the activities and take up sponsorships and legality in hosting tournaments. There are 3 different formats to play a cricket match, there is a considerable number of players from different parts of the world specializing to a particular format and plays international level cricket as well as domestic level matches throughout the year. Due to the evolving of franchise cricket (where players are chosen according to the performance level during the international matches and are hired as foreign players to play among teams belonging to a particular nation i.e., IPL, BBL, CPL,

PSL etc), many quality players are spotted and tend to engage and play under these conditions, since it can be seen as a way of income for the players.

Franchise cricket selectors select the players who are performing well in the current international level matches (especially in T20I and ODI cricket) to boost up the team attributes to win the title of a tournament. Such selections are made by analyzing the rankings as well as the player stat records issued by the ICC board.

Business Objectives

- Visualizing information will help the competitors and authorities to choose relevant talented players for their teams based on the performance level
- To make comparison between player stats
- To obtain accurate and precise details regarding the player's capabilities
- To construct strategies based on the information to defeat the opponents (ex: selecting off spinners against left arm batsman considering the opponent team player prior to the match)

Project Objectives

Deliverable

Website

An Exploratory Web Application with 'Dynamic Charts' and Illustrations with Actual Values pertaining to Significant Records of a Cricket Match

Technologies

Dash is the main element handling all the information and charts in the application. The three main components integrated in dash are;

- Flask
- React.JS
- Plotly

Where Flask is the **Server-Side** technology handling the Python Codes interconnected with the **Frontend Framework** developed using React.JS. while, with the help of Plotly, 'Dynamic Charts' too can be integrated to the Frontend Framework. The data pertaining to the different section of the application is pulled from the **NOSQL Database**, where all the data is stored in the cloud and through API links the relevant data will be retrieved to the necessary Frontend Exploratory View. The database with the specific mentioned functions is;

- MongoDB

Method Of Approach

Defining the Project Scope

Setting targets and defining an outline is important to ensure the boundary on the project topic, therefore the development process can be taken in to account smoothly since a proper plan is set from the start

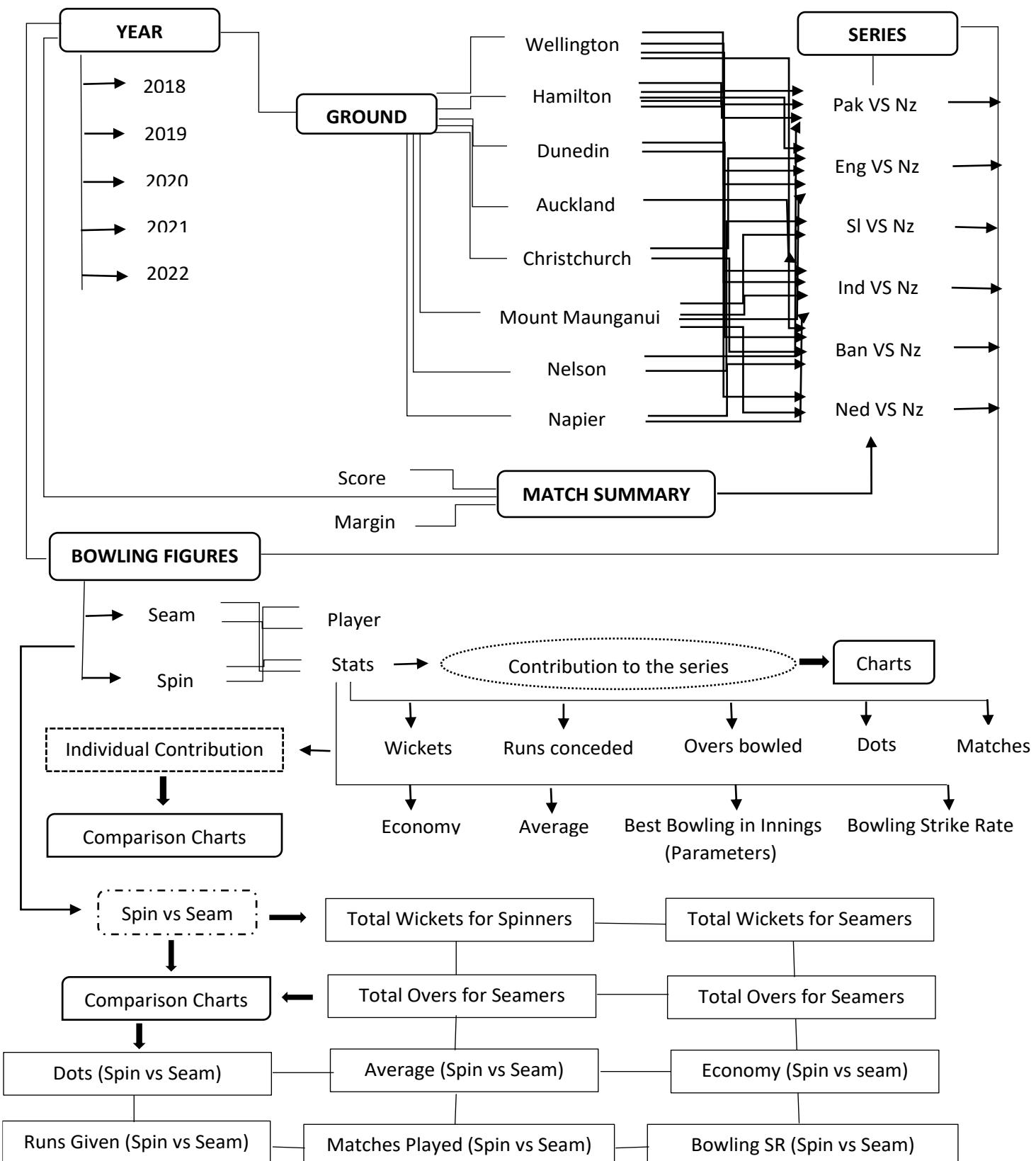
1st Step : Documenting the idea; mentioning the desired outcome required from the project

Main Objectives/ Visualizing Data (Possible Questions from the Analyst Point of View) - Exploratory Data Visualization

- Main Bowlers seen throughout the years (Stats)
Ex: **Matt Henry's Avg, Eco, Wickets at the Wellington Ground etc.** (Visualizing based on the number of matches played at the venue and performances)
- Comparison between grounds
Ex: **Wickets at Nelson vs Napier etc.** (categorizing based on spin vs seam)
Ex: **Runs Given at Nelson vs Napier etc.** (categorizing based on spin vs seam)
Ex: **Bowling Average at Nelson vs Napier etc.** (categorizing based on spin vs seam)
Ex: **Economy at Nelson vs Napier etc.** (categorizing based on spin vs seam)
- Comparison between years
Ex: **Number of wickets taken by spinners/ Number wickets taken by seamers**
Ex: **Number of runs scored for spinners and seamers**
- Player comparison (seam/spin)
Ex: **Trent Boult stats vs Matt Henry stats** (All Ground Records-Number of matches played, Wickets taken, Average, Economy, Bowling Strike Rate, Best Bowling Figures)
- Ground Win Percentage over Matches played (Datasets need to be made)
Ex: Seamer's wickets more than 20(In 5 matches) - win % at ground Wellington

2nd Step : Illustrating a Data-Flow diagram to understand the components of the Web Application

NEW ZEALAND BOWLING SIDE ANALYSIS (IN NEW ZEALAND DURING LAST 5 YRS)



3rd Step : Gathering the necessary datasets required to visualize the identified components

Ex; Downloading the necessary data to visualize the wickets, overs, runs etc in the form of charts with the relevant parameters of a match and saving in the format of (.csv)

Bowler	O	R	W	Econ	WPA
TG Southee	10	47	0	4.7	0.026
MJ Santner	10	54	2	5.4	0.01
C Munro	6	31	1	5.17	0.009
TA Boult	10	64	2	6.4	-0.054
C de Grandhomme	4	20	0	5	-0.066
IS Sodhi	10	63	2	6.3	-0.111

4th Step : Data Preparation accordingly to accomplish the identified objectives depicting through dynamic charts

Ex: Categorizing number of wickets, runs, dot balls, economy, average separately in different csv files considering the year and the ground.

The following information will be indicated in the Web Application based on outline created.

5th Step : Prototyping

6th Step : Development Phase

Pending Criteria's

- Prototyping**

A user interface design along with the wireframing will be required before the development process. Since prototyping will act as a guide representing all the necessary components easing the development process by saving time (Ex: Pre determining the colors, navigations).

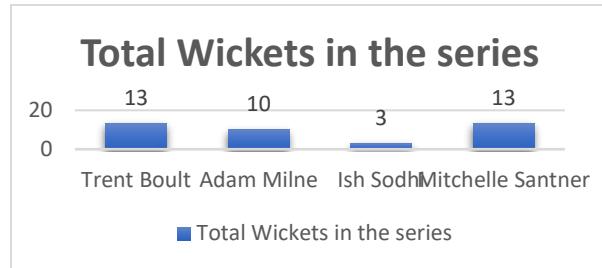
- Development Phase**

Technical research is needed to be carried out to understand the compatibility nature of the mentioned technologies, since by the time certain libraries can be outdated even though it is being mentioned in certain articles

Related with the Step 4

CONTRIBUTION TO THE SERIES (STATS PROPERTIES/ ILLUSTRATION TECHNIQUE)

- Wickets – Total wickets by each bowler (seam/spin)



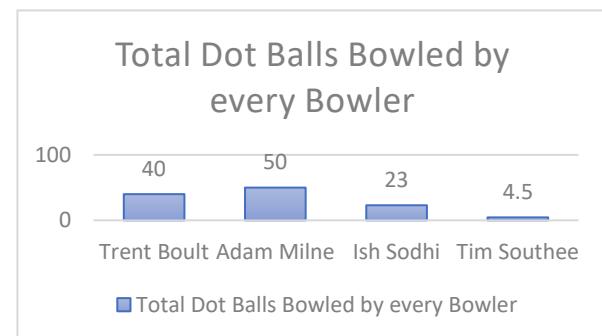
- Runs Conceded – Total runs conceded by every bowler



- Overs Bowled – Total Overs bowled by each bowler in the Series



- Dots – No of Balls Bowled by Bowlers without Conceding Runs

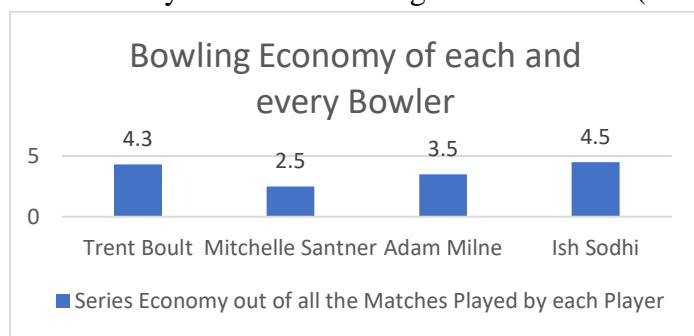


Number of Matches Played out of the 5 Games in the Series

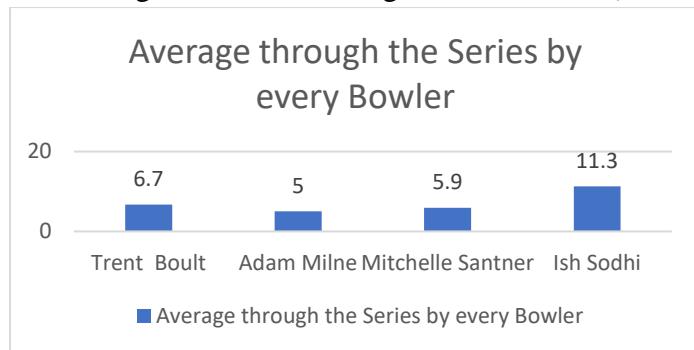


- Trent Boult
- Ish Sodhi
- Tim Southee
- Mitchelle Santner

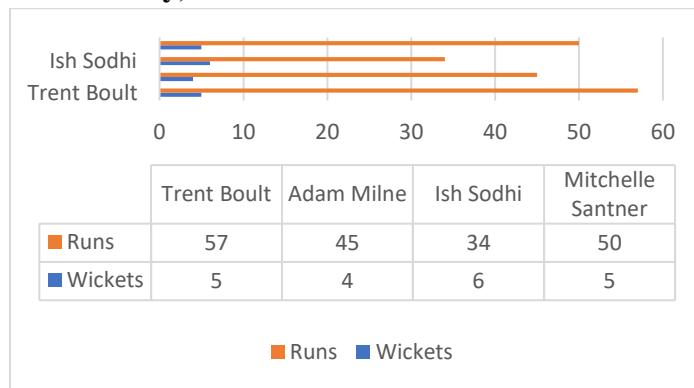
- Economy Variations Among all the Bowlers (**Runs Conceded/Overs Bowled=Eco**)



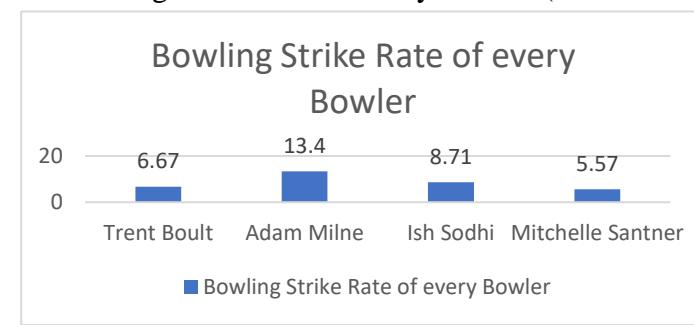
- Average Variation Among all the Bowlers (**Runs Conceded/Wickets Taken**)



- Best Bowling Figures in an Innings (**Highest Wickets for least runs conceded in a match or through the series irrespective to the games played by a player and average and economy**)

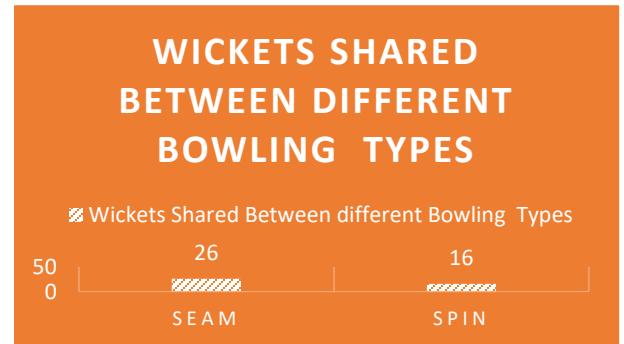
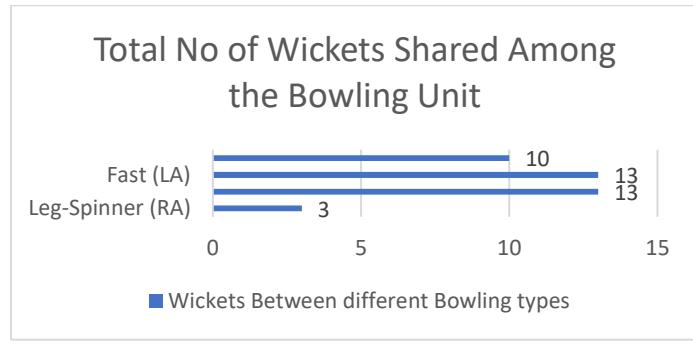


- Bowling Strike Rate of every Bowler (**Balls Bowled/Wickets Taken**)

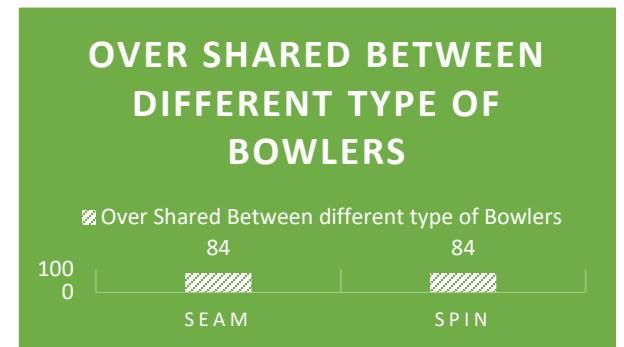
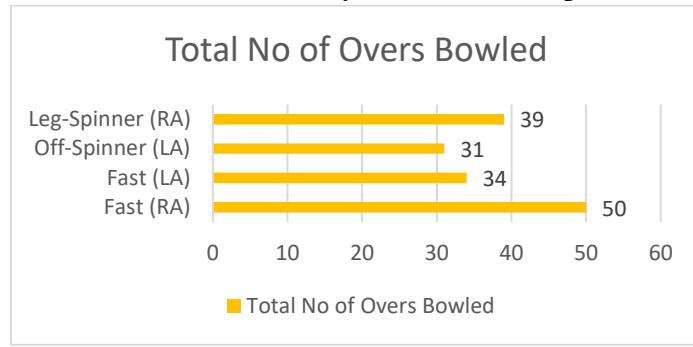


COMPARISON CHARTS (SEAM VS SPIN)

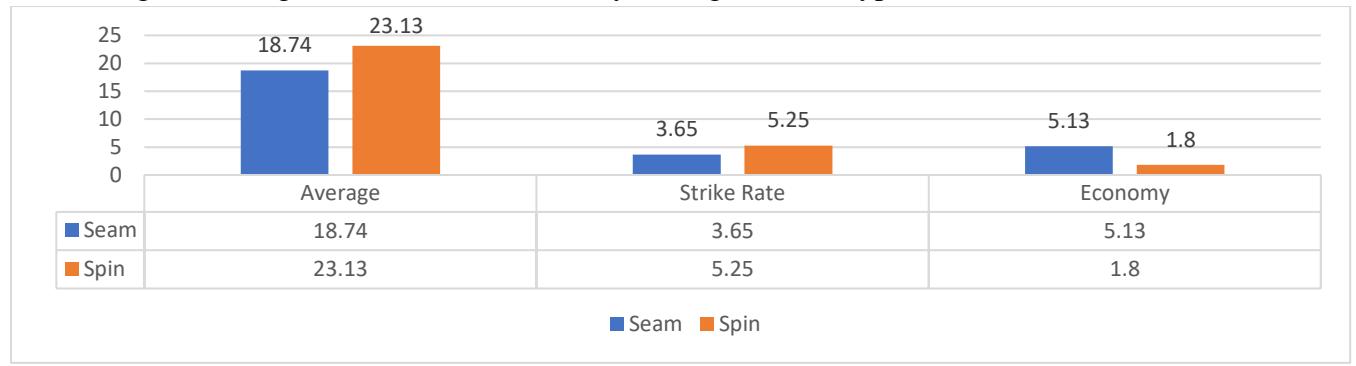
- Total Wickets for the Spin and Seam Bowlers



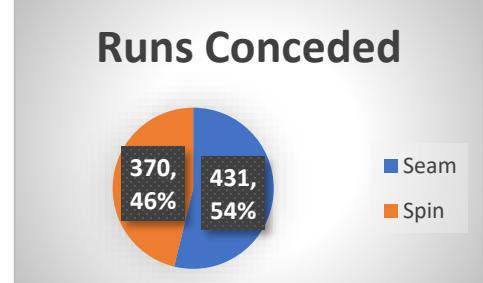
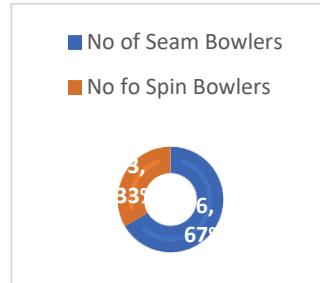
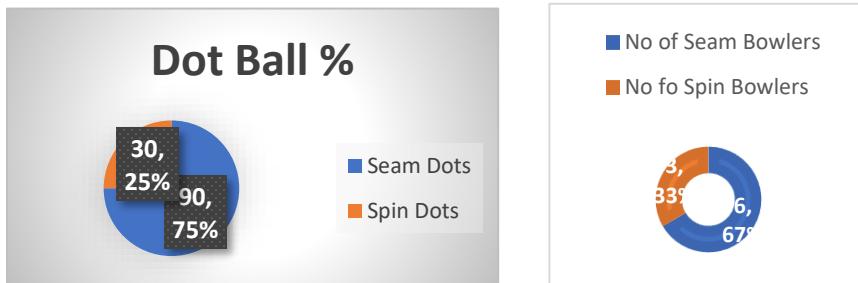
- Total Overs Bowled by the Seam and Spin Bowlers



- Average, Bowling Strike Rate and Economy among different type of Bowlers

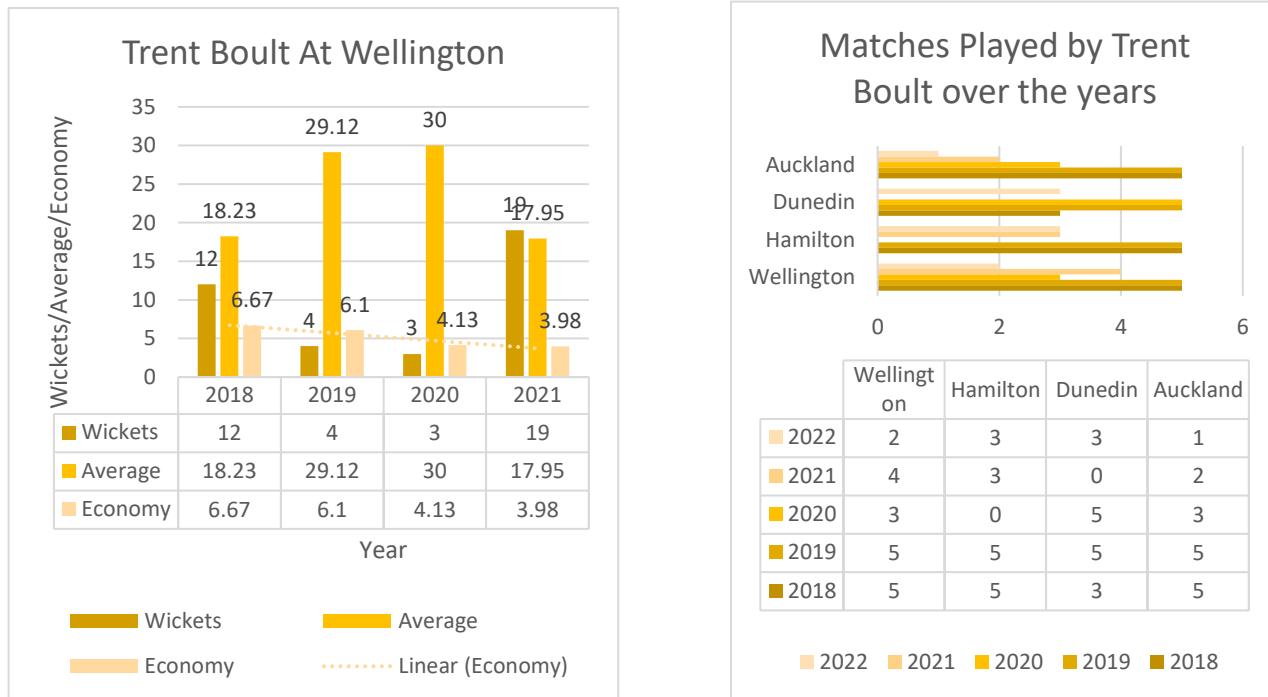


- Dots balls, Runs Given, Matches Played by Seam and Spin

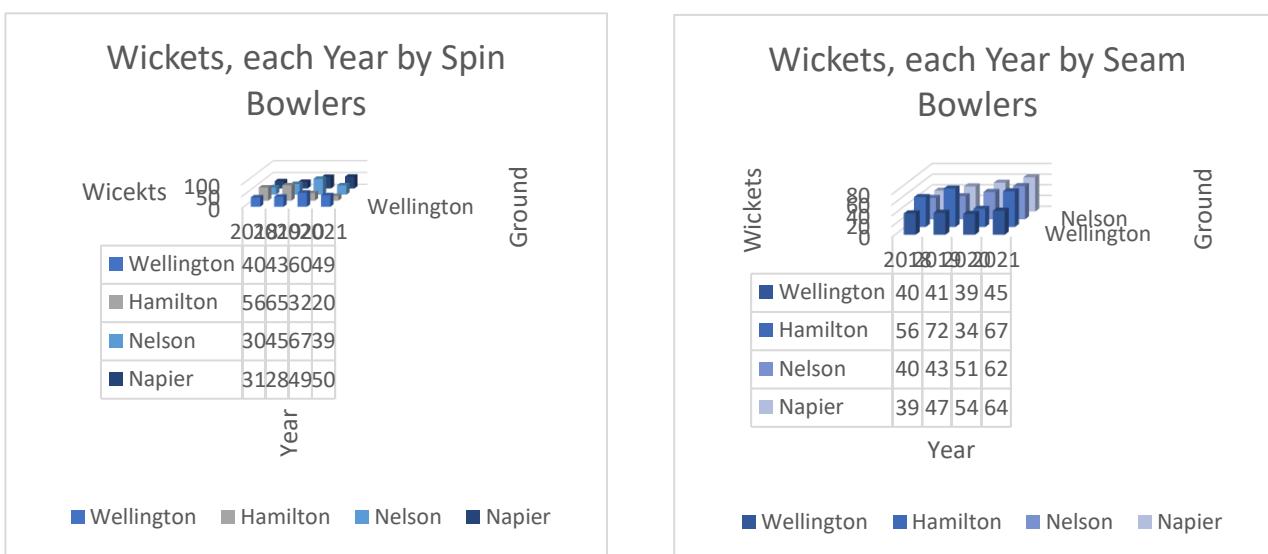


DEPICTING THE MAIN OBJECTIVES USING SOME SAMPLE ILLUSTRATION

- Bowler Stats over the years at a particular Ground (**Total Wickets, Totals Runs Conceded, Total Overs Bowled and No of Matches Played at that Venue are considered**)

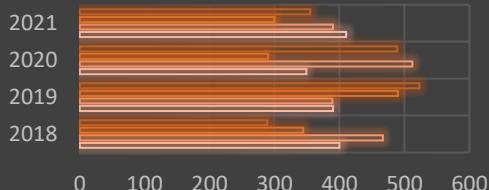


- Seam Vs Spin Ground Stats (**Total wickets, Runs Conceded, Bowling Average and Economy by every bowler**)



Runs Given by Fast Bowlers

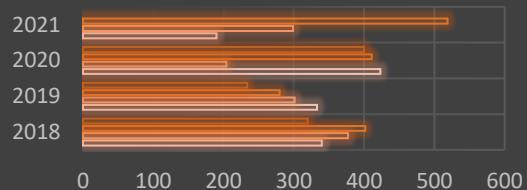
Napier Nelson Hamilton Wellington



	2018	2019	2020	2021
Napier	289	523	489	355
Nelson	344	490	290	300
Hamilton	467	389	512	390
Wellington	400	390	349	410

Runs Given by Spin Bowlers

Napier Nelson Hamilton Wellington



	2018	2019	2020	2021
Napier	320	234	400	
Nelson	402	280	411	519
Hamilton	377	301	204	299
Wellington	340	333	423	190

- Average and Economy at different ground (Spin vs Seam)

Average of Spinners over the years in different Grounds

Wellington Hamilton Nelson Napier



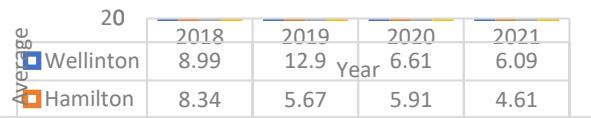
Average of Seamers over the years in different Grounds

Wellington Hamilton Nelson Napier



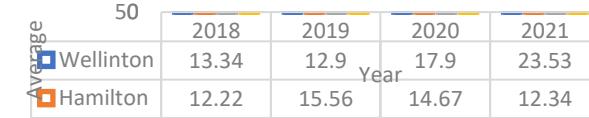
Economy of Spinners over the years in different Grounds

Wellington Hamilton Nelson Napier

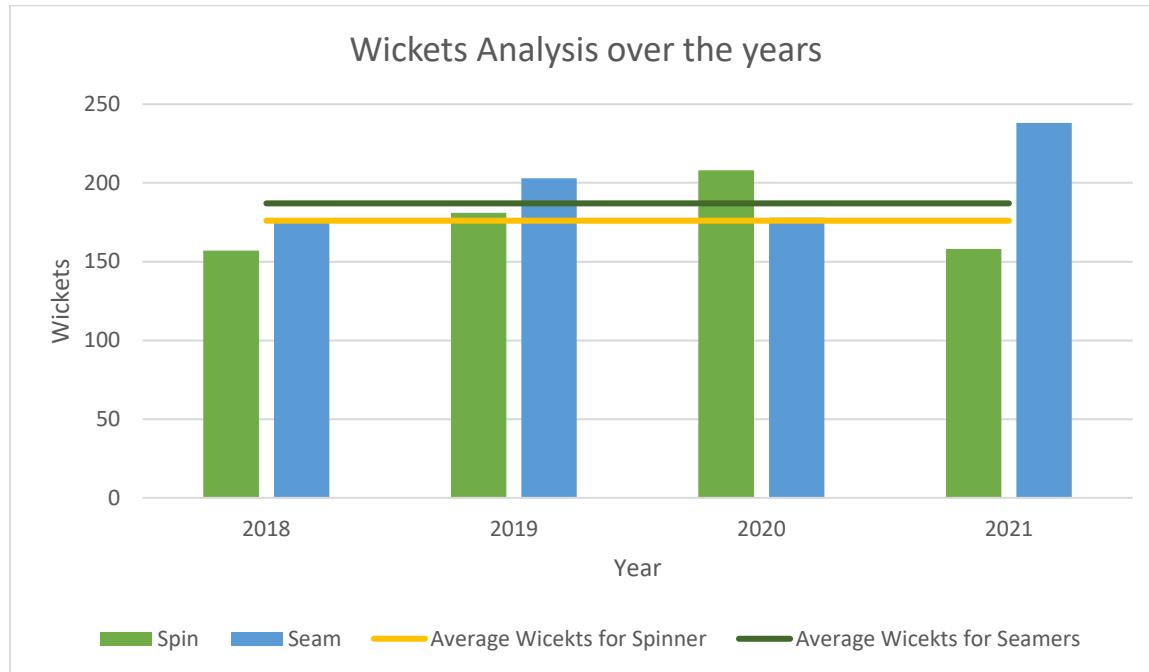


Economy of Seamers over the years in different Grounds

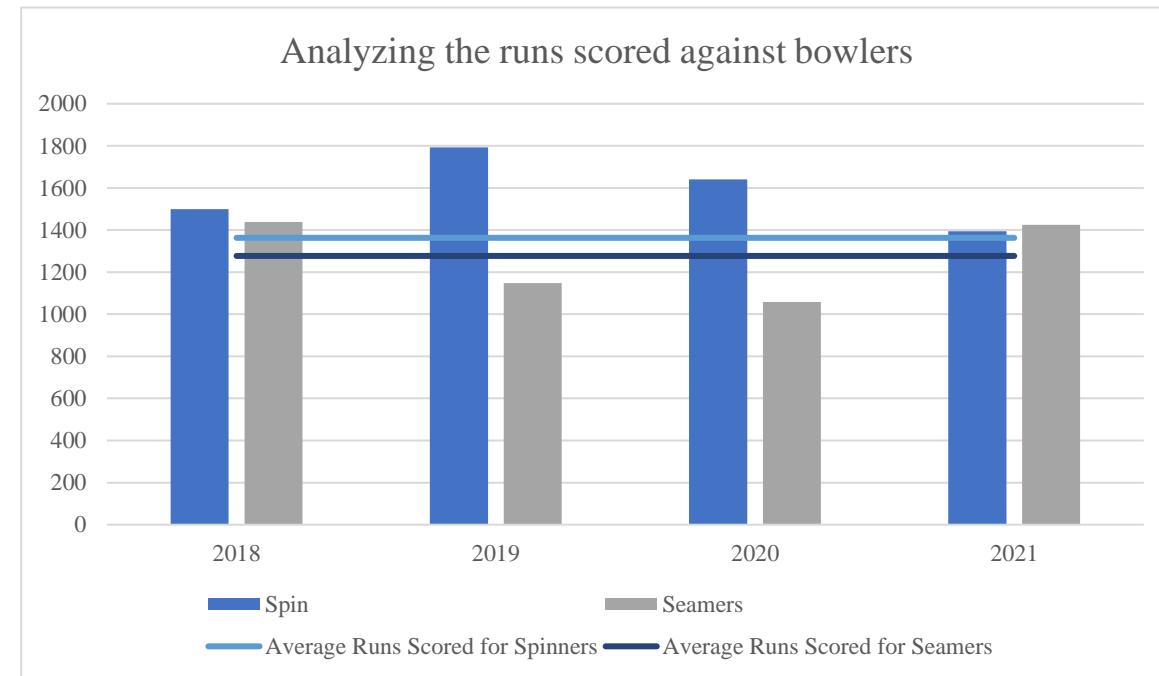
Wellington Hamilton Nelson Napier



- Wickets Comparison between spin and seam bowling (**Deciding whether above average or below average**)



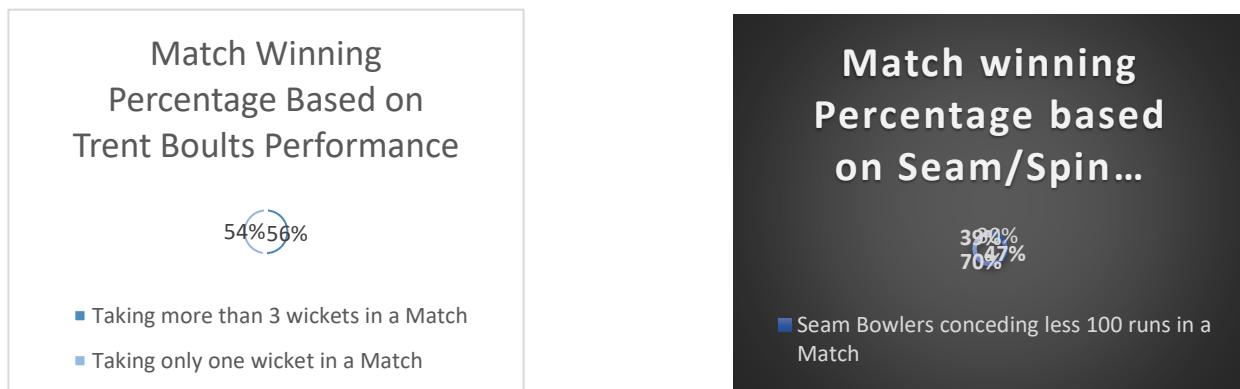
- Runs Scored comparison to spinners and seamers (**Deciding whether it is above average below average**)



- Player Comparison (In New Zealand Matches during the last 5 years)



- Prediction based on percentages (Based on different circumstances)

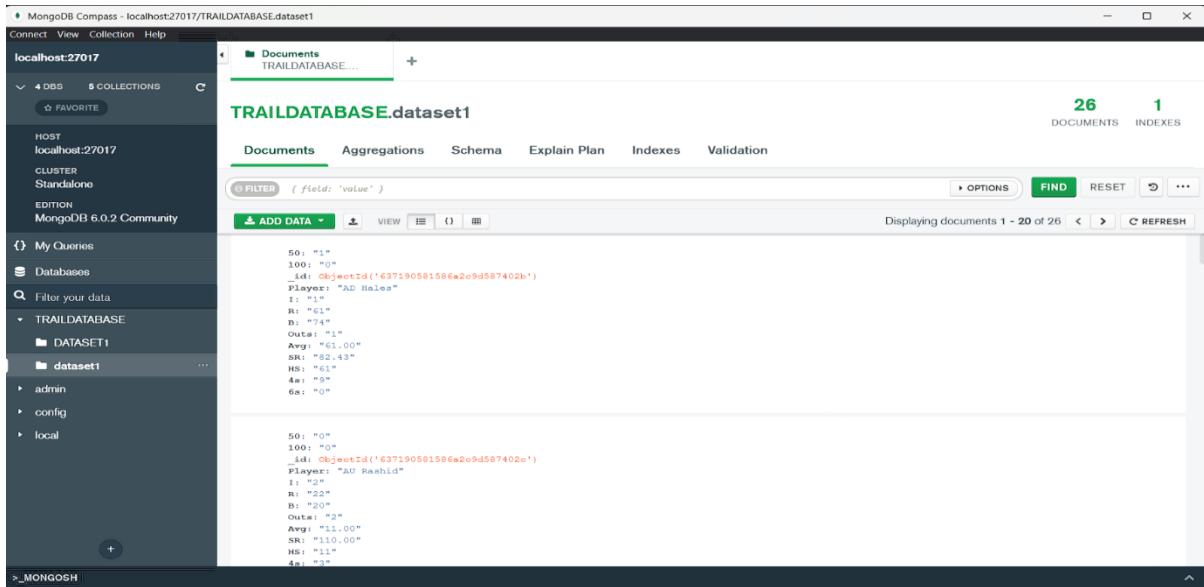


Initial Project Plan

Development Methodology (Example)

1. Connection to the Database

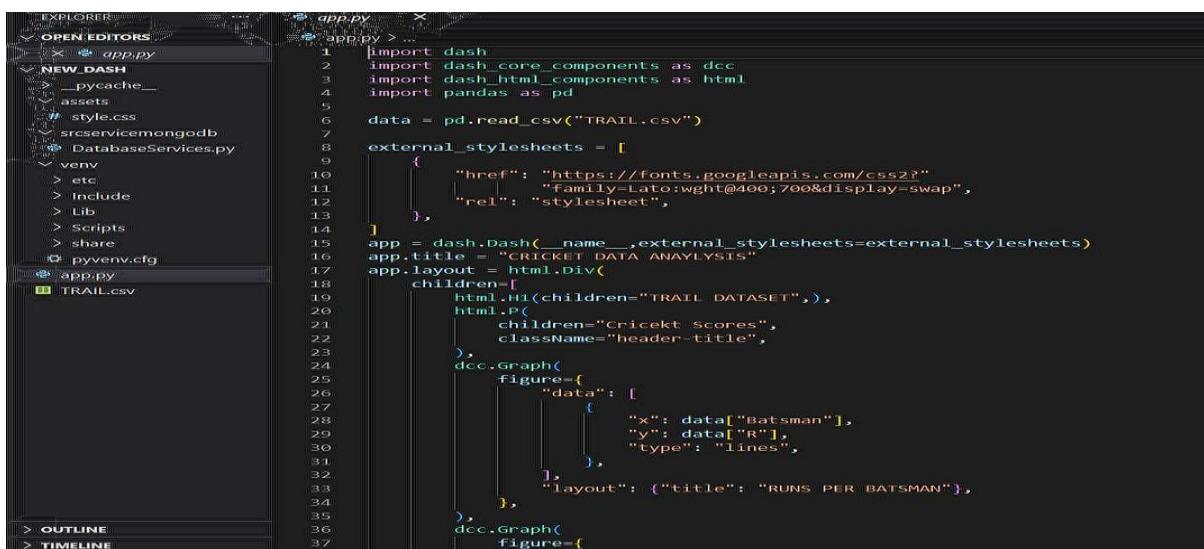
After completing the data preparation process, relevant csv files will be imported to the Mongo Database using the MongoDB Compass (mongo import libraries are required)



The screenshot shows the MongoDB Compass interface. On the left, the sidebar displays the connection to 'localhost:27017' and the database 'TRAILDATABASE'. Under 'COLLECTIONS', there is a single collection named 'dataset1'. The main pane shows the 'Documents' tab with 26 documents listed. One document is expanded to show its fields: Player: "AB Hales", I: "1", R: "61", B: "74", Out: "1", Avg: "61.00", SR: "82.43", HS: "61", 4s: "9", 6s: "0". Another document for 'AU Rashid' is partially visible below it.

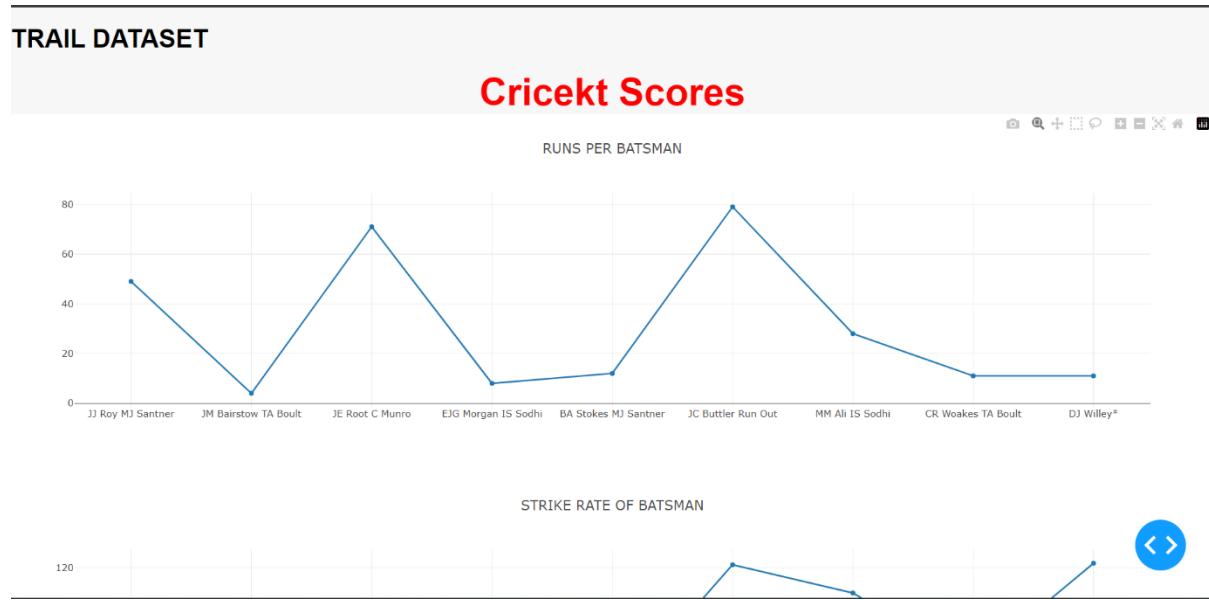
2. Downloading and Installing the required Dependencies and Libraries to code in dash

Inside the virtual environment python pandas and dash components will imported to create dynamic charts



The screenshot shows a code editor with an open file named 'app.py'. The code imports 'dash', 'dash_core_components' (dcc), 'dash_html_components' (html), and 'pandas' (pd). It reads data from 'TRAIL.csv' using pd.read_csv. The code then defines an external stylesheet with a link to Google's CSS library. It creates a Dash app with the title 'CRICKET DATA ANALYSIS' and a layout consisting of a header with the title 'TRAIL DATASET' and 'cricket scores', and a main area with a dcc.Graph component. The graph has a 'data' field containing a line chart with 'x' and 'y' values from the CSV, and a 'layout' field with a title 'RUNS PER BATSMAN'. The code editor also shows imports for 'app.py' and 'TRAIL.csv' at the bottom.

3. Running the localhost to get the output while attempting to style using CSS and HTML from DASH components



Duration

Initiation of Development Process

- Data Preparation – 3 Weeks
- Prototyping – 2 Weeks
- Developing Charts – 4 Weeks
- Developing the Frontend Application – 3 Weeks
- Integration of technologies – 1 Weeks
- Reconsidering and re-evaluating integration of technologies – 3 Weeks

Total Number of Weeks = 15

Total Number of Months = 4

Risk Analysis

Potential Risk Factors

1. Errors while importing data from the cloud into the dash application

MongoDB provides the service of importing .xlsx and .csv files in to the cloud and in order to retrieve the imported data from the database and load in to the dash application the ‘Pymongo’ library is required to be installed. The correct link is essential to connect the application with the cloud using the mongoclient server.

2. The size of the application if a database is not in use

There is a possibility of retaining the prepared data in the root directly and deploying the application along with it, but there would be a significant increase in weight of the application which tends to reduce the performance of the application.

3. Errors while setting up Development Environment

The development phase needs to be carried out inside of a virtual environment. Organizing file folders and installing necessary dependencies are tricky. CLI commands are necessary downloads and installation. Therefore, it is necessary to make sure new versions of ‘pip and compatibility nature of the packages along with it.

Mitigating the Circumstances

✓ Different technologies can be used to obtain the same output

Python scripts can be written in Jupyter Notebook to derive the chart using the prepared data and using Flask the charts can be interconnected to the frontend built using React.JS.

Another a way that development can be carried out is by creating charts using the MongoDB charts clusters and integrating them with the frontend React.JS.

Also, there is the option of connecting D3 charts with the react frontend in order to achieve the same desired output.

❖ These can be considered as the other alternative options to rely on if technical issues arises during the development phase.

Overview

Introduction

Data is far more important nowadays with the evolution of Artificial Intelligence and Internet of Things. Majority of free data is analyzed and customized to target to a particular audience and increase the business productivity. Data analytics is a vast concept used in almost any of the industries to gather data and transform to a manner which benefits a certain group of people.

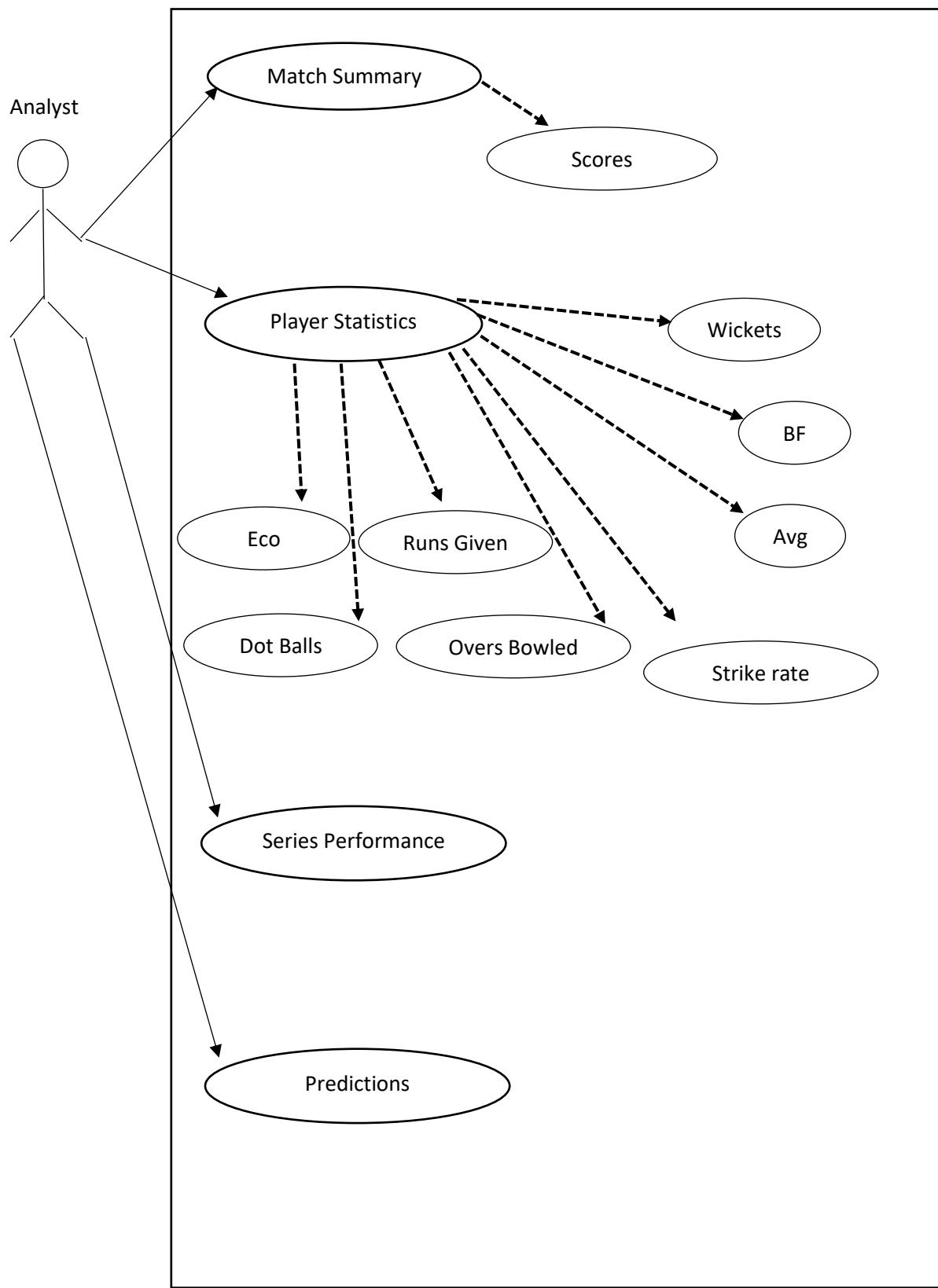
In the field of sports, player statistics and match summaries are vital to predict assumptions and depict chances of winnings while estimating scores based on analytical methods (using algorithms)

System Analysis

Facts Gathering Techniques

In the recent past, a nation which was recognized as top of its game was getting negative thoughts due to loss of matches consecutively across bi-lateral and in ICC events. A major factor that has affected to losing of matches was changing key players from time to time and players regularly getting injured. Many technical mistakes might have happened when choosing a proper team prior to the game. Also, in an article it has mentioned a proper team balancing to the opponent team was not selected and due to that main reason the outcome has been disastrous. Newspapers and blogs were populating comments on how the team should have been selected according to how the game went and criticisms was given when necessary players were left out of the squad.

Use Case Diagram



Drawbacks of the Existing System

Data Visualization is the technique of representing information in a particular manner where certain values are highlighted as important points in the processed data while explaining the nature of the information. The systems that already exist are written algorithms to calculate run rates, predicted scores, winning chances and so on. The systems that already exist are biased with the intention of predicting rather than evaluating a certain amount of information. The scoring rate algorithm will predict the relevant score based on the number of balls conceded, so basically calculations and assumptions are taken place instead for an actual analysis. Therefore, data visualization is bringing the insight and the actual meaning of the dataset in a simpler manner. The primary objective of prescriptive analytics is to understand and utilize the best cause of action based on identifying insights.

Requirement Specification

Hardware/Software Requirement

VSCode

The main development process is carried out in this IDE where necessary extensions are installed namely; Jupyter Notebook, MongoDB, Flask and Plotly.

Anaconda Environment

Since the project is based on python, the necessary framework needs to be installed. Once after the environment is created the newest version of python is installed while upgrading the ‘pip’ version.

Dash

Dash is a python integrated framework containing components such as flask, ReactJS and Plotly where necessary visualizing graphs can be hard coded through using necessary functions.

MongoDB

The online storage will be containing the prepared data (csv files) which will be imported to the frontend application via API links.

Feasibility Study

Operational Feasibility

How well the proposed system supports the cause

The main objective of creating a visualizing project is to understand the true nature of a situation. Visualizing concept is used in almost every industry to depict the current status of the business, event or an occurrence. In the sports sector visualizing projects can bring out the prescriptive nature

of the game and give an idea of the past records to make comparisons during similar situations and analyze to predict a conclusion.

Cricket is considered as a sport where players need to maintain consistency and mental toughness to battle the pressure in the field. Performance is very important, and performance can be measured by the effort the player is dedicating through batting, bowling and fielding. The way the performance measure is different from player to player since some players are considered as all-rounders while some only prioritize either bowling or batting also there will be a keeper supporting with the external role of batting.

The section where the visualizing project is focusing on is bowling, Players which play as all-rounders and with the main priority of bowling tend to get injured most of the time since a lot of effort needs to be put to minimize the run scoring by the opponent team. Identifying the right player for the right conditions is extremely important since analyzing the opponent team would bring the logic of counter attack through utilizing the bowling unit.

For the above stated reason there should be a proper picture which clearly depicts the strength and weaknesses of the players in order to select the relevant player for the current situation detected. Visualizing project can bring out the slightest details of every player who commits the role of bowling where past records can give an understanding whether the player is suitable to take part in the future games or not. Players with different capacities and potential are taking part in the game while cricket is containing three different formats where certain players play all the three formats while some stick to a particular format. Identifying the right player for the key moment decides the winning factor for a team.

Technical Feasibility

The project will be carried out in a python integrated framework, due to high scalability and high demand the development process can be conducted smoothly. Data Visualization project developments are vastly used in the business sectors to illustrate financial status, by environmentalists to locate and navigate wildlife affairs etc. Therefore, in any sector the development process can be initiated based on the data that is available to understand the hidden key points in the dataset.

Dataset required for analysis

<http://cricmetric.com> provides the necessary data tables for analysis containing bowling statistics for 5 years from 2018 to 2022 (where the analysis will be carried out)

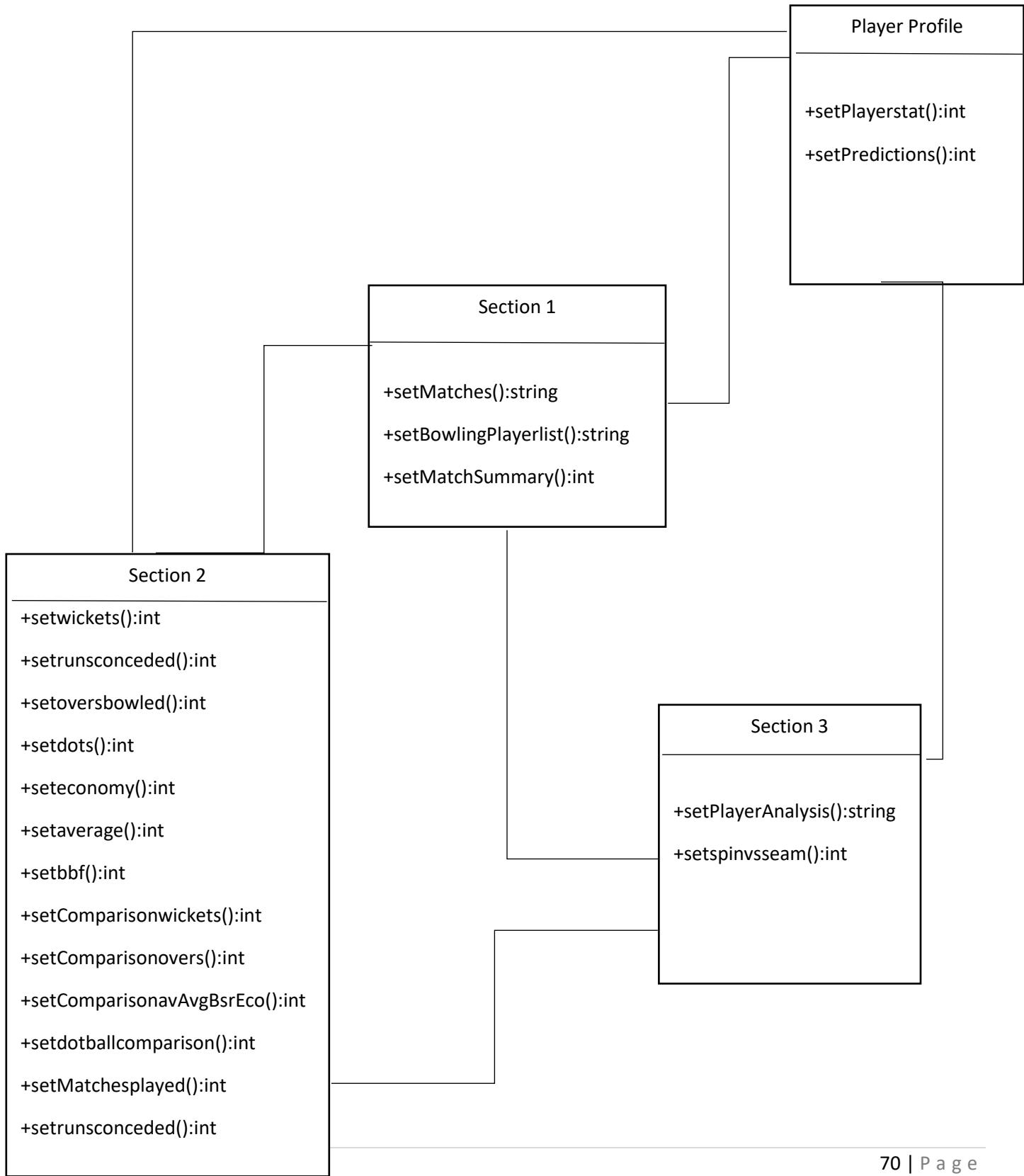
Data transformation

According to identified requirements for the application necessary data cleansing needs to be taken place out therefore, use of excel csv is required in order to categorize and divide the dataset in to separate parts.

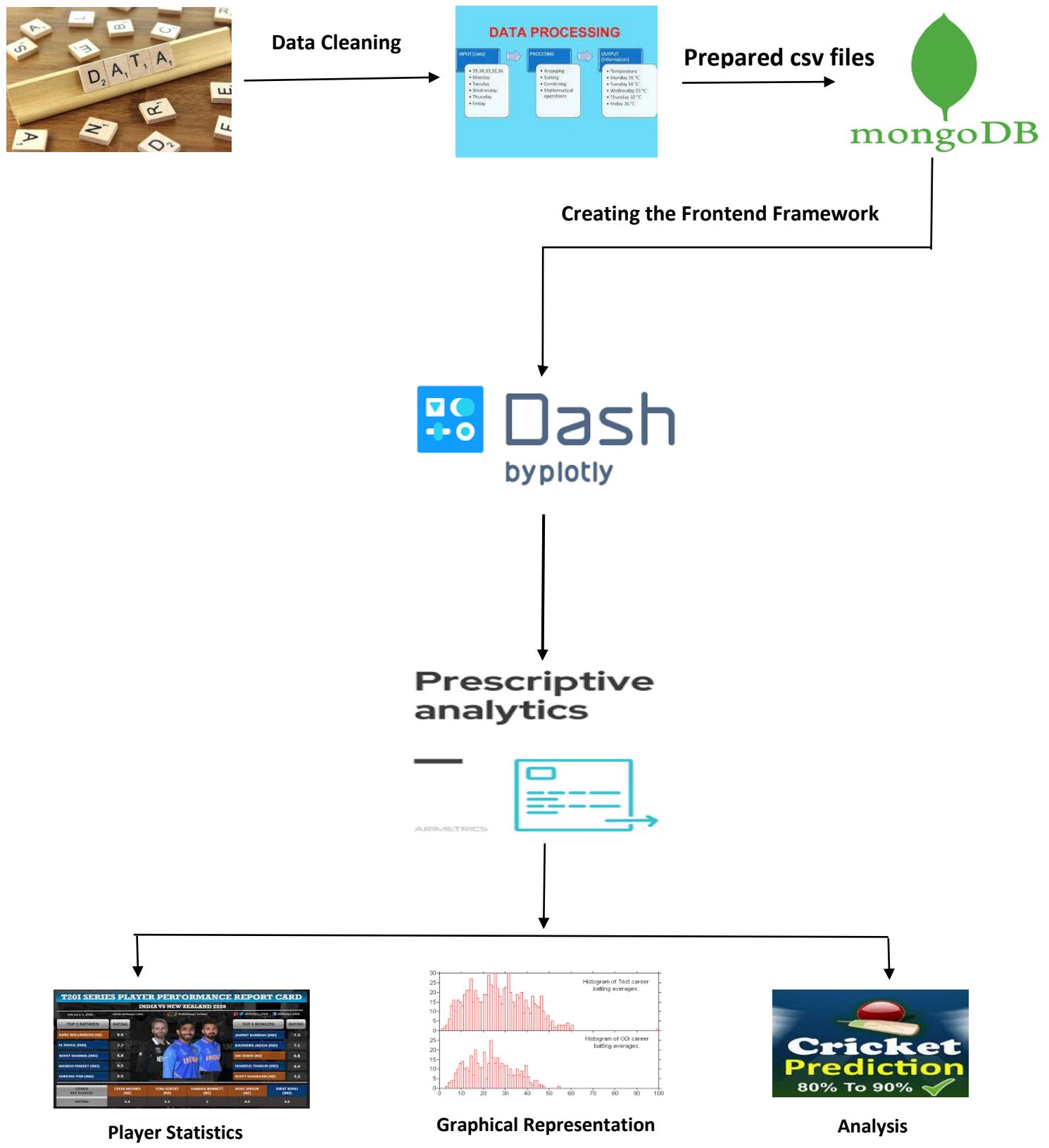
Another factor that helps the development process is the documentation of python and dash where necessary steps are mentioned to help developers understand the concept of developing such applications.

System Architecture

Class Diagram of the Proposed System



High-level Architecture Diagram



Discussion

Overview of the Interim Report

Firstly, an introduction to the problem identified is stated (lack of team selection strategies). Furthermore, the solution is elaborated by stating the concept which will be developed in order to solve the problem (Visualizing data in an attractive manner)

Summary of the Report

The development methodology for the identified cause is stated while reasoning the validity and stability of the development process (python-based languages are used for the development process)

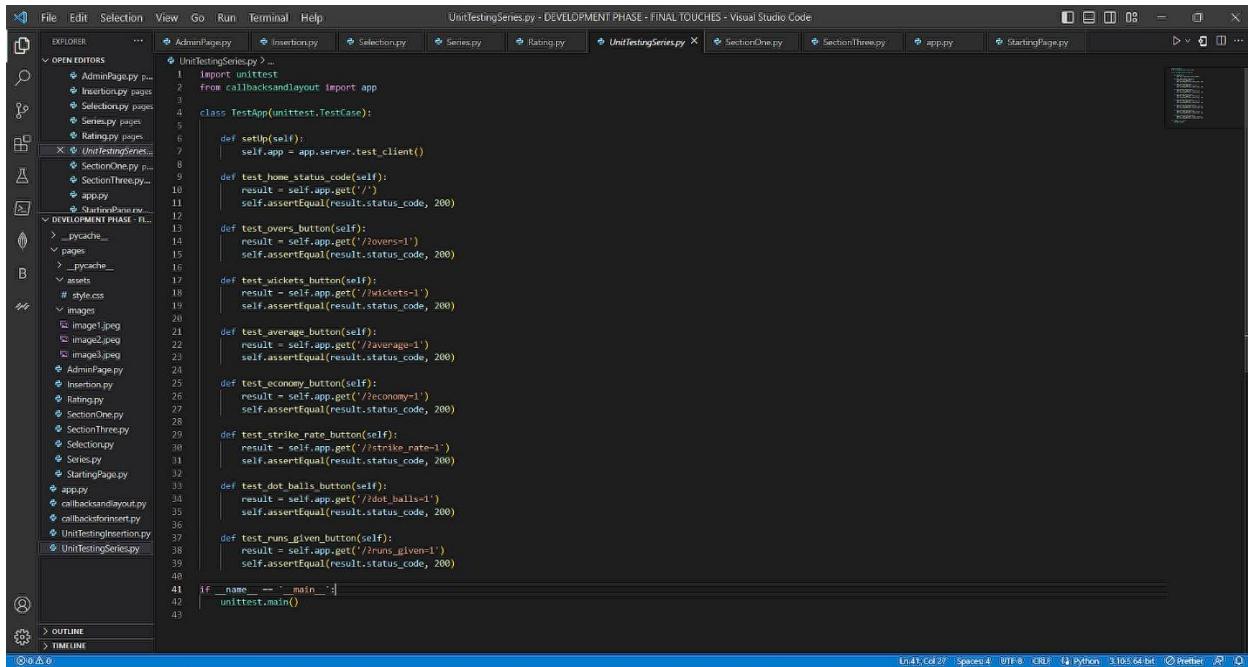
Challenges Faced

- Relevant dataset for the identified issues were difficult to be found.
- Selecting the ideal graphs and illustrations for the available data was difficult

Future Plans

The above-mentioned development is only carried out for the bowling unit; therefore, the same methodology can be carried out for the batting unit as well focusing on other teams as well.

Unit Test Cases



```
# Import unittest
from callbacksandlayout import app

class TestApp(unittest.TestCase):
    def setup(self):
        self.app = app.server.test_client()

    def test_home_status_code(self):
        result = self.app.get('/')
        self.assertEqual(result.status_code, 200)

    def test_loves_button(self):
        result = self.app.get('/?loves=1')
        self.assertEqual(result.status_code, 200)

    def test_wickets_button(self):
        result = self.app.get('/?wickets=1')
        self.assertEqual(result.status_code, 200)

    def test_average_button(self):
        result = self.app.get('/?average=1')
        self.assertEqual(result.status_code, 200)

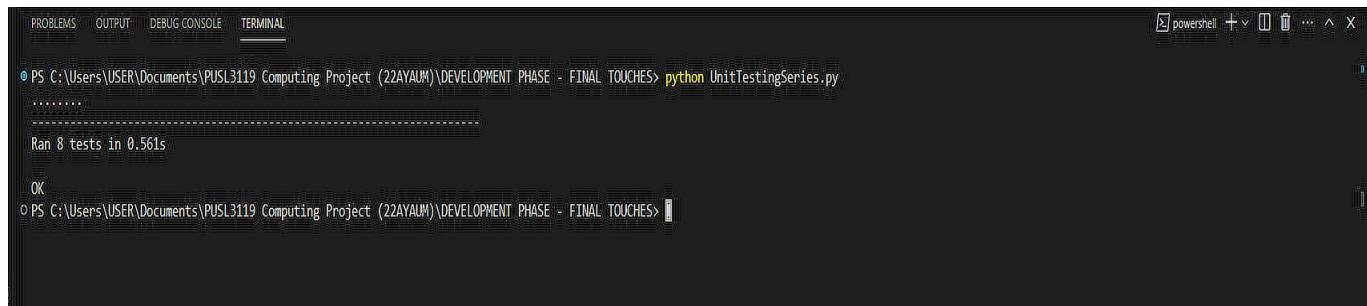
    def test_economy_button(self):
        result = self.app.get('/?economy=1')
        self.assertEqual(result.status_code, 200)

    def test_strike_rate_button(self):
        result = self.app.get('/?strike_rate=1')
        self.assertEqual(result.status_code, 200)

    def test_dot_balls_button(self):
        result = self.app.get('/?dot_balls=1')
        self.assertEqual(result.status_code, 200)

    def test_runs_given_button(self):
        result = self.app.get('/?runs_given=1')
        self.assertEqual(result.status_code, 200)

if __name__ == '__main__':
    unittest.main()
```



```
PS C:\Users\USER\Documents\PUSL3119 Computing Project (22AYAUM)\DEVELOPMENT PHASE - FINAL TOUCHES> python UnitTestingSeries.py
.....
-----
Ran 8 tests in 0.561s

OK
PS C:\Users\USER\Documents\PUSL3119 Computing Project (22AYAUM)\DEVELOPMENT PHASE - FINAL TOUCHES>
```

The same procedure can be carried out for the rest of the button list under the series list since, the same techniques are used to structure the rest (Other series)

```

File: UnitTestingInsertion.py - DEVELOPMENT PHASE - FINAL TOUCHES - Visual Studio Code
OPEN EDITORS
  Explorer AdminPages.py Insertion.py Selection.py Series.py Rating.py UnitTestingInsertion.py SectionOne.py SectionThree.py app.py StartingPage.py
  DevTools
  pyright
  assets style.css
  images image1.jpg image2.jpg image3.jpg
  AdminImage.py
  Insertion.py
  Rating.py
  SectionOne.py
  SelectionThree.py
  Series.py
  StartingPage.py
  app.py
  callbacksAndLayout.py
  callbacksForInsert.py
  UnitTestingInsertion.py
  UnitTestingSeries.py

  # importing necessary modules
  import unittest
  from unittest.mock import MagicMock
  from callbacksForInsert import app, update_data

  # define test case class
  class TestApp(unittest.TestCase):

    # defining the setup method to set up the test environment
    def setUp(self):
      self.app = app.server.test_client()

    # defining the tearDown method to tear down the test environment
    def tearDown(self):
      pass

    # defining test update_data method to test the update_data function
    def test_update_data(self):
      # define dummy input values
      player = "THISARA PERERA"
      matches = "10"
      wickets = "5"
      runs = "13"
      type = "Seam"
      role = "Main Bowler"
      table_data = []

      # calling the update_data function with the dummy input values
      result = update_data(player, matches, wickets, runs, type, role, table_data)

      # asserting that the function returns the expected output
      expected_output = None
      self.assertEqual(result, expected_output)

    # running the unit tests
    if __name__ == '__main__':
      unittest.main()

```

```

PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL
powershell + ×

PS C:\Users\USER\Documents\PUSL3119 Computing Project (22AYAUM)\DEVELOPMENT PHASE - FINAL TOUCHES> python UnitTestingInsertion.py
C:\Users\USER\Documents\PUSL3119 Computing Project (22AYAUM)\DEVELOPMENT PHASE - FINAL TOUCHES\callbacksforinsert.py:10: UserWarning:
The dash_html_components package is deprecated. Please replace
  'import dash_html_components as html' with 'from dash import html'
  import dash_html_components as html
:
-----
Ran 1 test in 0.010s
OK
PS C:\Users\USER\Documents\PUSL3119 Computing Project (22AYAUM)\DEVELOPMENT PHASE - FINAL TOUCHES>

```

_id: ObjectId('64625e6aae12ca726438466d')
Player: "THISARA PERERA"
Matches: "10"
Wickets: "5"
Runs Given: "13"
Average: 2.6
Bowling Type: "Seam"
Role: "Main Bowler"

New Values are added to the existing collection in the database

Project Management

The screenshot shows a Trello board titled "INDIVIDUAL PROJECT". The board is organized into three main columns: "Done", "Done", and "To Do".

- Done Column 1:** Contains cards for "Requirement Gathering" (Oct 16, 2022 - Dec 3, 2022), "Identifying Relevant Technologies" (Dec 4, 2022 - Jan 11), "Research on Existing Systems" (Jan 12 - Jan 19), "Machine Learning Model Implementation" (Feb 28 - Mar 11), "Backend Processes" (Mar 12 - Mar 31), and "Performance Analysis" (Apr 6 - Apr 15).
- Done Column 2:** Contains cards for "WireFraming" (Feb 1 - Feb 15), "Prototyping" (Feb 8 - Feb 21), "Other Component Establishment" (Apr 1 - Apr 5), and "Frontend Development" (Apr 16 - Apr 30).
- To Do Column:** Contains cards for "Machine Learning Model Implementation", "Backend Processes", "Other Component Establishment", "Performance Analysis", "Frontend Development", "Prototyping", "Wireframing", "Requirement Gathering", "Identifying Relevant Technologies", and "Testing".

A sidebar on the left provides access to workspace settings and other boards.

Trello Timeline : <https://trello.com/b/Uee01ERp/individual-project>

Supervisor Meeting Records

 PUSL3119 Computer Individual Project Student Progression Report							
01. Student Name	Mahendu Bondorenayake						
02. Plymouth Index Number	10749841						
03. Degree Program	BSC (Hons) Computer Science						
04. Supervisor Name	Mr Premudya Thiplakoratne						
05. Project Title	E-ANALYTICA (DATA ANALYSIS WEBSITE) Developed using DATA VISUALIZATION Concepts						
Meeting Number	Meeting 01	Meeting 02	Meeting 03	Meeting 04	Meeting 05	Meeting 06	Meeting 07
Date	11/11/22	22/02/23	21/01/23	19/02/23	30/01/23	24/02/23	24/04/23
Student Signature							
Supervisor Signature							
Meeting Number	08	Meeting 09	Meeting 10	Meeting 11	Meeting 12	Meeting 13	Meeting 14
Date							
Student Signature							
Supervisor Signature							

Final Year Project – Supervisory meeting minutes

Meeting No: 01

Date : 11/11/2022

Project Title : Data Visualizing website

Name of the Student : Mahendu Bandaranayake

Students ID : 10749841

Name of the Supervisor : Mr. Pramudya Thukkaratne

Items discussed:

DataVis2 charts (different types)

Items to be completed before the next supervisory meeting:

Navigation of charts (prototyping issues)

 11/11/2022

Supervisor (Signature & Date)

Instructions to the supervisor: Do not sign if the above boxes are blank.

Final Year Project – Supervisory meeting minutes

Meeting No: 02

Date : 22/02/23

Project Title : Data VP2 / Statistical website

Name of the Student : Mahindu Bondara Nayake

Students ID : 10749841

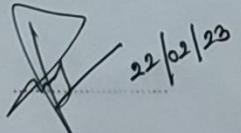
Name of the Supervisor : Mr. Pramudyo Thakkaratne

Items discussed:

Data Gathering
Data collection sources

Items to be completed before the next supervisory meeting:

- Prediction model types



22/02/23

Supervisor (Signature & Date)

Instructions to the supervisor: Do not sign if the above boxes are blank.

Final Year Project – Supervisory meeting minutes

Meeting No: 03

Date : 26/02/2023

Project Title : Data VP2/stotpstpcol create

Name of the Student : Mahindu Bandaranayake

Students ID : 10749841

Name of the Supervisor : Mr. Premudyo Thelokarathne

Items discussed:

- Modal evaluation
- Components in predicting average

Items to be completed before the next supervisory meeting:

- Database design and Implementation

Supervisor (Signature & Date)

Instructions to the supervisor: Do not sign if the above boxes are blank.

Final Year Project – Supervisory meeting minutes

Meeting No: 04

Date : 19/03/2023

Project Title : Data Visualization website

Name of the Student : Mahindu Bandaranayake

Students ID : 10749841

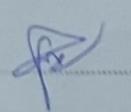
Name of the Supervisor: Mr. Pramudyo Thilakaratne

Items discussed:

- Implementation error discussion
(Login errors)

Items to be completed before the next supervisory meeting:

- Descriptive Data implementation



Supervisor (Signature & Date)

Instructions to the supervisor: Do not sign if the above boxes are blank.

Final Year Project – Supervisory meeting minutes

Meeting No: 05

Date : 30/03/2023

Project Title : Data VP2/ Statistical website

Name of the Student : Mahindu Bandaranayake

Students ID : 10749841

Name of the Supervisor : Mr. Premudya Thilakarathne

Items discussed:

- Data insertion types and methods

Items to be completed before the next supervisory meeting:

- User input functions implementation

Supervisor (Signature & Date)

Instructions to the supervisor: Do not sign if the above boxes are blank.

Final Year Project – Supervisory meeting minutes

Meeting No: 06

Date

: 02/04/2023

Project Title

: Data VP2 / Global School Website

Name of the Student

: Mahindu Bandaranayake

Students ID

: 10749841

Name of the Supervisor:

: Mr. Premudya Thelakkatne

Items discussed:

- Admin portal issue discussion

Items to be completed before the next supervisory meeting:

- Implementation of necessary privileges for users
Ex: Delete feature of entered records

Supervisor (Signature & Date)

Instructions to the supervisor: Do not sign if the above boxes are blank.



Final Year Project – Supervisory meeting minutes

Meeting No: 07

Date

22/04/23

Project Title

Project V21 Globalstar website

Name of the Student

Mohmdy Bandarangyako

Students ID

10749841

Name of the Supervisor:

Nr. Pramudyas Thakoratne

Items discussed:

- Structuring the report
What to include

Items to be completed before the next supervisory meeting:

- Testing (unit, usability)
- Finalizing content for the report

Supervisor (Signature & Date)

Instructions to the supervisor: Do not sign if the above boxes are blank