**TWITTER DATA ANALYSIS**

**(ENGLISH PREMIER LEAGUE)**

**Project Report – II**

Submitted in partial fulfillment of the requirements

for the degree of

**Bachelor of Engineering (Computer Engineering)**

|  |  |
| --- | --- |
|  | by: |
| Nadeem Yoosuf | TU3F1213008 |
| Shyam Nair | TU3F1213041 |
| Umair Akhtar | TU3F1213004 |
|  |  |

Under the Guidance of

Prof. D.K. Chitre



Department of Computer Engineering

TERNA ENGINEERING COLLEGE

Nerul (W), Navi Mumbai 400706

(University of Mumbai)

(2015-2016)

**Internal Approval Sheet**



**TERNA ENGINEERING COLLEGE, NERUL**

**Department of Computer Engineering**

Academic Year 2015-16

**CERTIFICATE**

This is to certify that the project entitled **“TWITTER DATA ANALYSIS (ENGLISH PREMIER LEAGUE)”** is a bonafide work of

|  |  |
| --- | --- |
| Nadeem Yoosuf | TU3F1213008 |
| Shyam Nair | TU3F1213041 |
| Umair Akhtar | TU3F1213004 |
|  |  |

submitted to the University of Mumbai in partial fulfilment of the requirement for the award of the Bachelor of Engineering (Computer Engineering).

**Guide** **Project Convener**  **Head of Department** **Principal**

ii

**Approval Sheet**

Project Report Approval

This Project Report – II entitled “***Twitter Data Analysis (English Premier League)***”by following students is approved for the degree of ***B.E. in "Computer Engineering"***.

***Submitted by:***

|  |  |
| --- | --- |
| Nadeem Yoosuf | TU3F1213008 |
| Shyam Nair | TU3F1213041 |
| Umair Akhtar | TU3F1213004 |
|  |  |

Examiners Name & Signature:

1.---------------------------------------------------------

2.----------------------------------------------------------

Date:

Place:

iii

Declaration

We declare that this written submission represents our ideas in our own words and where others' ideas or words have been included, we have adequately cited and referenced the original sources. We also declare that we have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in our submission. We understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

|  |  |
| --- | --- |
| Nadeem Yoosuf | TU3F1213008--------------------------- |
| Shyam Nair | TU3F1213041--------------------------- |
| Umair Akhtar | TU3F1213004--------------------------- |
|  |  |

Date:

Place:

iv

**Acknowledgement**

We would like to express our sincere gratitude towards my guide **Prof D.K. Chitre** and Project Convener **Prof. V. B Gaikwad**, for the help, guidance and encouragement, they provided during the Progress seminar. This work would have not been possible without their valuable time, patience and motivation. We thank them for making my stint thoroughly pleasant and enriching. It was great learning and an honour being their student.

We are deeply thankful to **Dr. Lata Ragha (H.O.D Computer Department)**, and entire team in the Computer Department. They supported us with scientific guidance, advice and encouragement, they were always helpful and enthusiastic and this inspired us in our work.

We take the privilege to express our sincere thanks to **Dr. L. K. Ragha,** Vice Principal, and **Dr. S. M. Jagade**, our Principal for providing the encouragement and much support throughout our work.

.

|  |  |
| --- | --- |
| Nadeem Yoosuf | TU3F1213008 |
| Shyam Nair | TU3F1213041 |
| Umair Akhtar | TU3F1313004 |
|  |  |

v

|  |  |  |  |
| --- | --- | --- | --- |
| SR NO | | CONTENTS | PAGE NO. |
| 1 | Abstract | | viii |
| 2 | List of Figures | | ix |
| 3 | CHAPTER 1 – INTRODUCTION   * 1. Aim   2. Review Of Literature   3. Scope   4. Motivation   5. Problem Statement | | 1  1  4  10  11  12 |
| 4 | CHAPTER 2 – REPORT ON THE PRESENT INVESTIGATION  2.1 Software Requirements  2.2 Non Functional Requirements  2.3 Timeline Chart  2.4 Process Model | | 13  13  19  19  20 |
| 5 | CHAPTER 3 – DESIGN  3.1 Data Flow Diagram  3.2 E-R Diagram  3.3 System Flowchart  3.4 Feasibilty Analyis  3.5 Technical Analysis  3.6 Algorithm | | 22  22  23  24  24  25  26 |
| 6 | CHAPTER 4 – IMPLEMENTATION  4.1 Features  4.2 Datasets  4.3 Screenshots | | 28  28  31  39 |
| 7 | CHAPTER 5 – RESULTS AND DISCUSSION  5.1 Performance Measure  5.2 Analysis  vi | | 49  49  51 |
| 8 | CHAPTER 6 – CONCLUSION | | 53 |
| 9 | CHAPTER 7 - REFERENCES | | 54 |

vii

**Abstract**

Twitter has proven to be a notable source of analytic data of various domains such as the stock market, natural disasters, consumer preferences, political attitudes, different forms of governance, and other new applications which are emerging every day for big data, and even sports outcomes. However, such a study has not been conducted to determine the popularity of different soccer teams. Hence, the purpose of this project was to study whether data mined from twitter can be used for this purpose and then displaying the findings onto a website for everyone to access. We have chosen to examine the English Premier League (EPL) since it’s the most popular league in the world. The project aims to procure the findings by considering different metrics such as the number of tweets, most popular tweets, number of followers, trending tweets, and trending topics (related to EPL). Some of these attributes described above can potentially aid professional data analysts of statisticians or even football teams themselves in making an informed decision about the right choice related to their work. Our findings will be displayed on a website, so we aim to make the platform visually appealing which would make the browsing experience better for users, and could also spur other humans to visit our site. Fellow users can provide their own knowledge about the English Premier League and can even spark off an argument in the discussion forum provided in the website. Nothing is more entertaining than an interactive website!

All of this data will be uploaded onto our website, <http://www.footweets.com>. I am sure the content on the website will fancy football fans all over the world. We aim to make [www.footweets.com](http://www.footweets.com) a must-visit website for football fans.

viii

**LIST OF FIGURES**

ix

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Sr.** |  |  | **Name** |  |  | **Page** |  |  |
|  |  |  |  |  |  |  |
|  | **No.** |  |  |  |  |  | **No.** |  |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  | |  |  |  |  |  |  |  |  |
|  | Fig. 1.1 |  |  | Chelsea vs. Norwich (2016) |  |  | 2 |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  | |  |  |  |  |  | |  |  |
| Fig. 1.2 | |  |  | Location of Premier League Clubs |  | 3 | |  |  |
|  | |  |  |  |  |  | |  |  |
|  | Fig. 1.3 |  |  | Location of Premier League Clubs in London itself |  |  | 4 |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  | |  |  |  |  |  | |  |  |
| Fig. 2.1 | |  |  | OAuth Workflow |  | 14 | |  |  |
|  | |  |  |  |  |  | |  |  |
|  | Fig. 2.2 |  |  | R Programming in Statistics |  |  | 15 |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  | |  |  |  |  |  | |  |  |
| Fig. 2.3 | |  |  | Match Popularity Graph created by RStudio |  | 16 | |  |  |
|  | |  |  |  |  |  | |  |  |
|  | Fig. 2.4 |  |  | Storage space occupied by a traditional SQL Database compared to Non-SQL Database, i.e. MongoDB |  |  | 18 |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  | |  |  |  |  |  | |  |  |
| Fig. 2.5 | |  |  | MongoDB Architecture (far below) |  | 18 | |  |  |
|  | |  |  |  |  |  | |  |  |
|  | Fig. 2.6 |  |  | Timeline Chart |  |  | 19 |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  | |  |  |  |  |  | |  |  |
| Fig. 2.7 | |  |  | Incremental Model Procedure |  | 20 | |  |  |
|  | |  |  |  |  |  | |  |  |
|  | Fig. 3.1 |  |  | Data Flow Diagram |  |  | 22 |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  | |  |  |  |  |  | |  |  |
| Fig. 3.2 | |  |  | E R Diagram |  | 23 | |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  | |  |  |  |  |  | |  |  |
|  | Fig. 3.3 |  |  | System Flowchart |  | 2 | 23 |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
| Fig. 3.4 | |  |  | Organization of MongoDB Data |  | 24 | |  |  |
|  | |  |  |  |  |  | |  |  |
|  | Fig. 4.1 |  |  | Manchester United Tweet count |  |  | 29 |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  | |  |  |  |  |  | |  |  |
| Fig. 4.2 | |  |  | Display of tweets on the website |  | 30 | |  |  |
|  |  |  |  |  |  |  |  |  |  |

|  |  |  |
| --- | --- | --- |
| Fig. 4.3 | MongoDB: Showing Collections in which tweets are stored | 32 |
|  |  |  |
|  |  |  |
| Fig. 4.4 | IntelliJ: Showing Java code synced with MongoDB database. Contains code to implement both Rest and Streaming | 33 |
|  |  |  |
| Fig. 4.5 | IntelliJ: Contains Twitter API Codes | 34 |
|  |  |  |
|  |  |  |
| Fig. 4.6 | IntelliJ: Tweets are extracted and shown | 35 |
|  |  |  |
| Fig. 4.7 | Dashboard of the Website | 36 |
|  |  |  |
|  |  |  |
| Fig. 4.8 | Main Page of the Website | 37 |
|  |  |  |
| Fig. 4.9 | The Premier League Fixtures Grid | 38 |
|  |  |  |
|  |  |  |
|  |  |  |

|  |  |  |
| --- | --- | --- |
| Fig. 4.10 | RStudio, OAuth Procedure | 39 |
|  |  |  |
|  |  |  |
| Fig. 4.11 | RStudio, Tweet Extraction | 40 |
|  |  |  |
| Fig. 4.12 | RStudio, Creating Donut Chart | 41 |
|  |  |  |
|  |  |  |
| Fig. 4.13 | Donut Chart | 42 |
|  |  |  |
| Fig. 4.14 | Wordcloud of #epl, Created by RStudio | 46 |
|  |  |  |
|  |  |  |
| Fig. 4.15 | Rstudio Creating Followers Map | 47 |
|  |  |  |

x

|  |  |  |
| --- | --- | --- |
| Fig. 4.16 | Followers Map of Newcastle United | 48 |
|  |  |  |
|  |  |  |
| Fig. 5.1 | The Premier League Table | 50 |
|  |  |  |
| Fig. 5.2 | Players popularity graph | 50 |
|  |  |  |

xi