Project I

Predicting Best Match Sportsperson for Product Advertisement

Submitted in partial fulfillment of the requirements for the degree of

Bachelor of Engineering

by

Shah Shahil Shakir Hussain Hafizunnisa	14CO45
Sarguroh Junaid Jawed Sadaf	14CO43
Gavandi Abhay Audumbar Varsha	14CO22
Syed Areeb Iqbal Ahmad Rehana	13CO61

Supervisor

Prof. Apeksha Gopale

Department of Computer Engineering,

School of Engineering and Technology
Anjuman-I-Islam's Kalsekar Technical Campus
Plot No. 2 3, Sector -16, Near Thana Naka, Khanda Gaon,
New Panvel, Navi Mumbai. 410206

Academic Year : 2018-2019

CERTIFICATE

Department of Computer Engineering,
School of Engineering and Technology,
Anjuman-I-Islam's Kalsekar Technical Campus
Khanda Gaon,New Panvel, Navi Mumbai. 410206

This is to certify that the project entitled *Predicting Best Match Sportsperson for Product Advertisement* is a bonafide work of **Syed Areeb Iqbal Ahmad Rehana**(13CO61) submitted to the University of Mumbai in partial fulfillment of the requirement for the award of the degree of **Bachelor of Engineering** in **Department of Computer Engineering**.

Prof. Apeksha Gopale

Supervisor

Prof. Kalpana Bodke

Project Co-Ordinator

Prof. Tabrez Khan

Head of Department

Dr. Abdul Razak Honnutagi

Director

Project I Approval for Bachelor of Engineering

This project entitled *Predicting Best Match Sportsperson for Product Advertisement* by *Syed Areeb Iqbal Ahmad Rehana(13CO61)* is approved for the degree of *Bachelor of Engineering in Department of Computer Engineering.*

Examiners

1. ....

2. ....

Supervisors

1. ....

2. ....

Chairman

....

Declaration

I declare that this written submission represents my ideas in my own words and where others ideas or words have been included, I have adequately cited and referenced the original sources. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission. I 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.

Syed Areeb Iqbal Ahmad Rehana
14CO61
B.E. (Computer Engineering)
University of Mumbai.

Abstract

Tittle: Predicting Best Match Sportsperson for Product Advertisement

Sports are one of the popular forms of entertainment in today's world. People do like to express their views on social sites regarding sports, players etc. As we all know that people do watch television, advertisements and show interest in the products endorsed by their favourite sports person. The proposed system is considering the performance or ranking of a sports person and their popularity on social site to decide on the best suitable candidate for particular product endorsement in order to increase the sale of the product.

Syed Areeb Iqbal Ahmad Rehana

13CO61

B.E. (Computer Engineering)

University of Mumbai.

Contents

List of Figures

Keywords And Glossary

Keywords :
Sentimental analysis, Machine Learning, Product Advertisement, Sports, Naive bayes, Prediction.

Glossary :
A:
API An application programming interface (API) is a set of subroutine definitions, protocols, and tools for building software and applications.
Analysis: Detailed examination of the elements or structure of something.
Artficial Intelligence: he theory and development of computer systems able to perform tasks normally requiring human intelligence, such as visual perception, speech recognition, decision-making, and translation between languages.
C:
CSS Cascading Style Sheets (CSS) is a style sheet language used for describing the presentation of a document written in a markup language.
H:
HTML HyperText Markup Language (HTML) is the standard markup language for creating web pages and web applications.
J:
JSON JSON (JavaScript Object Notation) is a lightweight data-interchange format. It is easy for humans to read and write. It is easy for machines to parse and generate.
N:
NLP Natural language processing is a field of computer science, artificial intelligence, and computational linguistics concerned with the interactions between computers and human(natural) languages.
S:
System Architecture A system architecture or systems architecture is the conceptual model that defines the structure, behavior, and more views of a system.
Social Networking A social networking service (also social networking site, SNS or social media) is an online platform that is used by people to build social networks or social relations with other people who share similar personal or career interests, activities, backgrounds or real-life connections.
W:
Web Browser A web browser is a software application for retrieving, presenting, and traversing information resources on the World Wide Web.
X:
XAMPP A XAMPP is an all in one webserver which consist apache, mysql,tomcat

Chapter 1

Introduction

1.1 Statement of Project

This software will give the list of suitable candidate for a specific product. To predict the best match, the system will consider the players ranking and popularity amongst the people. Such a system can be used by advertising agencies for the promotion of the product. To promote a product one needs a popular personality in order to generate good revenue. Product companies or vendors can select player according to their needs, popularity and criteria which can act as a great business deal for the company.

1.1.1 System Architecture

Figure 1.1: System Architecture
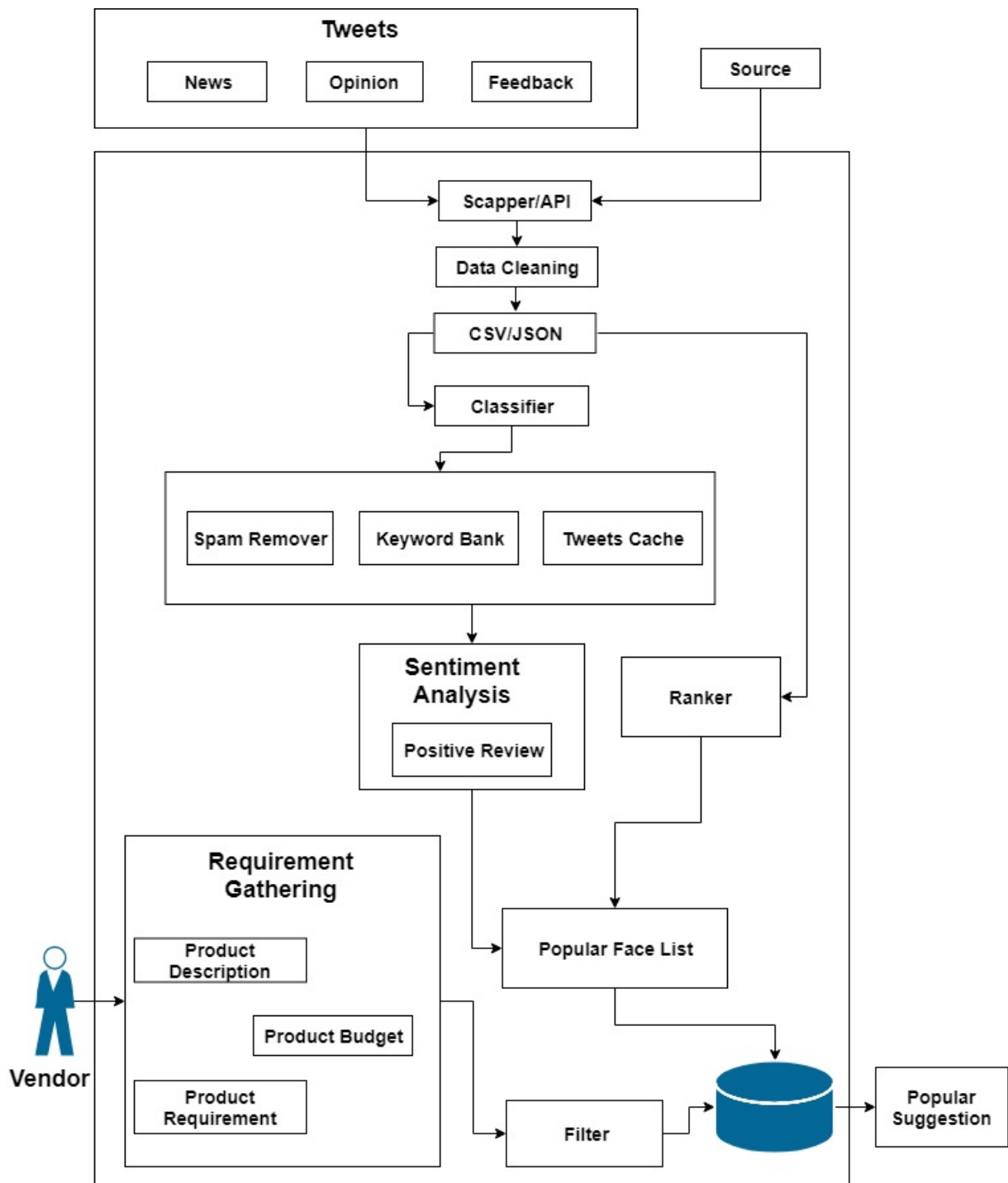
Our system architecture description is a formal description and representation of our system , organized in such a way that supports reasoning about the structures and the behaviours of the system.Our system architecture is based on the plan to predict the best sportsperson for a product advertisement which can lead to revolutionary changes in the way of booking players for an advertisement .

1.1.2 Motivation

This topic is choosen because advertising and marketing is really important in today's business infrastructure. Advertising is important for maintaining profit and improving standings in the product world.

1.2 Objective and Scope

1. **Best sportsperson for the product:-**Through our software we are going to determine suitable player for the advertising agency according to their requirements, budget and scale in profit.

2. **Rise in product sales:-** Selecting a suitable sportsperson who is trending can lead the companies or advertising agency with a boost in their product sales which will lead to a successful campaign.

3. **Profit for both:-** This leads to the profit of both the sportsperson as well as the advertising agency, as the sportsperson gains prominence as well as money and for the advertising agency it leads to a profitable quarter.

Chapter 2

Literature Review

2.1 Quantifying the performance of players in Football Match

2.1.1 Description

It's hard to evaluate the impact of a player's performance on the team. The tradition method and rating systems involve looking at a few metrics which include goal scored, assists, key passes, tackles, intercept, etc. Often this methodology makes the goal scorer and the assist provider the most important players of the team, which might not always be the case. No wonder Ballon d'Or winners are forwards and not defenders. These numbers make sense when comparing similar metrics. However, when comparing a forward, whose primary job is to score and assist goals, with a defender, whose primary task is to clear the ball, tackle, it's difficult[1]. Football is a team sport; there is a complex interaction between the players. A winning goal might be a result of a threading pass by the midfielder and the winger making a diagonal run to take one of the central defenders with him creating space for the striker to score the goal. In such complicated scenario, it is tough to allocate the contribution of the goal. The paper discussed a statistical method, using regression and optimization, to qualitatively allocate the points contributed to the team by a particular player during a season. Thus, even thou the player who scored the winning goal might have secured 3 points for his team; his contribution to the team is not 3 points. The paper provides a methodology for distributing those 3 points to their rightful contributors[1].

2.1.2 Advantage

Predicting the impact of player: The software is having the capability to use divination to predict the impact of a player in the upcoming football match.

2.1.3   Weakness

Inaccurate results: The main problem is that it can sometimes give inaccurate results due to the large sum of datasets present as the mathematical formula can vary due to the huge amount of data.

2.1.4   How to Overcome

Naive Bayes Algorithm: Naive Bayes Algorithm is a family of simple probabilistic classifiers which can accept large datasets and provide accurate results in return.

2.2   Evaluating Player Performance via Statistical Network Modeling

2.2.1   Desription

The major diculty in evaluating individual player performance in basketball is adjusting for interaction eects by teammates. With the advent of play-by-play data, the plus-minus statistic was created to address this issue. While variations on this statistic do correct for some existing confounders, they struggle to gauge two aspects: the importance of a player's contribution to his units or squads, and whether that contribution came as unexpected (i.e. overor under-performed) as dened by a statistical model[2]. We quantify both in this paper by adapting a network-based algorithm to estimate centrality scores and their corresponding statistical signicances. Using four seasons of data, we construct a single network where the nodes are players and an edge exists between two players if they played in the same ve-man unit. These edges are assigned weights that correspond to an aggregate sum of the two players' performance during the time they played together. We determine the statistical contribution of a player in this network by the frequency with which that player is visited in a random walk on the network, and we implement bootstrap techniques on these original weights to produce reference distributions for testing signicance[2].

2.2.2   Advantage

Estimate individual performance: The software is having the capability of estimating and predicting the individual performance of a player in the league games.

2.2.3 Weakness

Neural network: The neural networks can sometimes become very complicated due to a large number of algorithms and data required for it to function and come to a satisfactory result.

2.2.4 How to Overcome

Combination: To overcome this problem of neural networks, we can combine neural networks with regression for better results in estimating the performance of the player.

2.3 Twitter Sentiment Analysis Using Hybrid Cuckoo Search Method

2.3.1 Description

Sentiment analysis is one of the prominent fields of data mining that deals with the identification and analysis of sentimental contents generally available at social media. Twitter is one of such social medias used by many users about some topics in the form of tweets[3]. These tweets can be analyzed to find the viewpoints and sentiments of the users by using clustering-based methods. However, due to the subjective nature of the Twitter datasets, metaheuristic-based clustering methods outperforms the traditional methods for sentiment analysis. Therefore, this paper proposes a novel metaheuristic method (CSK) which is based on K-means and cuckoo search. The proposed method has been used to find the optimum cluster-heads from the sentimental contents of Twitter dataset. The efficacy of proposed method has been tested on different Twitter datasets and compared with particle swarm optimization, differential evolution, cuckoo search, improved cuckoo search, gauss-based cuckoo search, and two n-grams methods. Experimental results and statistical analysis validate that the proposed method outperforms the existing methods. The proposed method has theoretical implications for the future research to analyze the data generated through social networks/medias. This method has also very generalized practical implications for designing a system that can provide conclusive reviews on any social issues[3].

2.3.2 Advantage

Accurate: The results of the sentiments of Twitter-basedareomments is highly accurate using Hybrid Cuckoo search method.

2.3.3   Weakness

Unstructured and grammatical mistakes: The tweets can be highly unstructured and with the presence of grammatical mistakes or typos, it can lead to complications for the software to analyze the tweets.

2.3.4   How to Overcome

Optimum Cuckoo search: The solution for this is to use the optimum Cuckoo method to solve the highly unstructured tweets and grammatical mistakes.

2.4   Technical Analysis Of Player's Performance

2.4.1   Description

We will go on to identify the best performing footballers in comparison to their teammates. This type of analysis is particularly useful in unveiling the potential of footballers who do not yet play for the most competitive clubs. It also allows us to measure the clubs' dependency on their key players[4]. In the conclusion, we will underline some of the numerous advantages that a well-grounded approach to measure players' technical performance as presented in this report can bring to forward-thinking teams. The CIES Football Observatory research team is at the disposal of professional clubs to help enhance their effectiveness in this field.

2.4.2   Advantage

Potential of players: It is used to determine the potential of players who will be playing in the upcoming match.

2.4.3   Weakness

Complicated: The analysis of the players' performance can be complex at times due to the many factors involved in the process.

2.4.4 How to Overcome

KPI: Key Performance Indicator (KPI) can be used to indicate the performances of players and can lead to an easy analysis of the players.

Chapter 3

Technical Details

3.1 Methodology

For Predicting sportsmen, we will extract statistics from a particular website and also data cleaning is done then the finest part is to store in CSV/JSON which become our dataset. Further we use twitter popularity to calculate Popular Face list where data is stored in Main DB. Now we acquire agency requirement where our filter the requirement from main DB to display popular suggestion. Predicting sportsmen can be used in Advertisement, Statistics Analysis, Popularity Index etc. In Today's World, Advertisement are quick and have a large network to reach its audience.
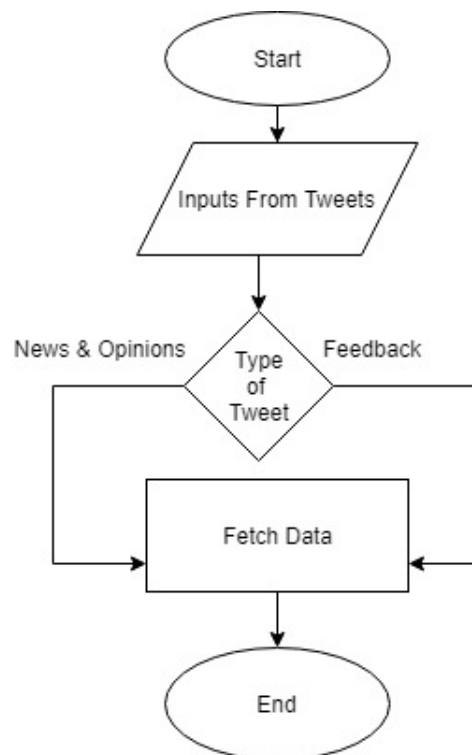

Figure 3.1: Module 1
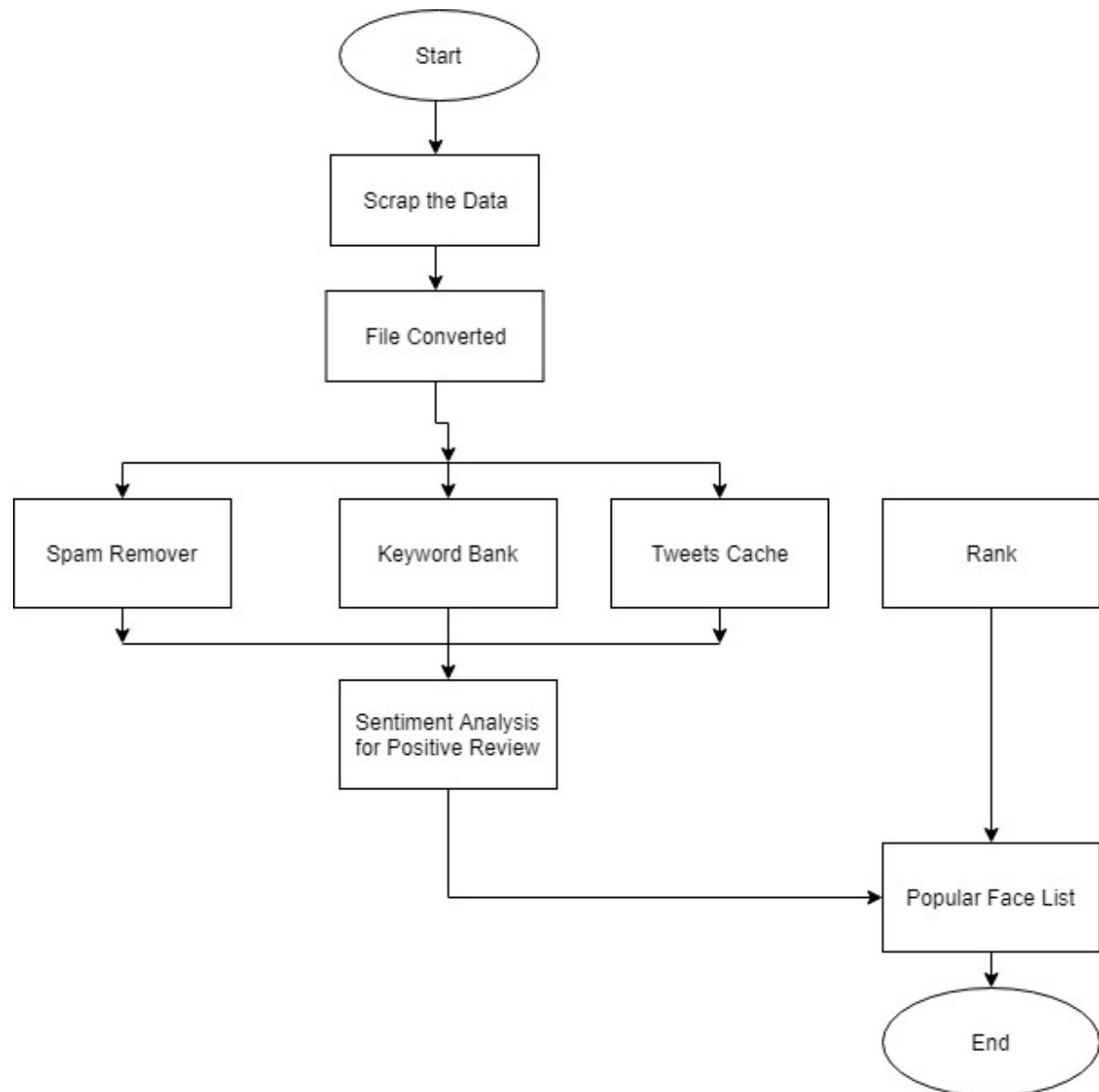
3.1.1 Dataset

Figure 3.2: Module 2

1. **Extraction:-** As we want data to analyze we will extract from the various website where we use a script to extract data and sometimes we get readymade data structure i.e API.

2. **Data Cleaning:-** We will use this data to make dataset for that it should be perfect, no dirty course, organized etc. We will check our dataset variable which will be useful and the remaining variable are been discarded to attain consistency.

3. **Export to CSV/JSON:-** Our data variable is now fixed so we can write this data to set

of a column which we call dataset. Most Dataset Format is CSV but when we use API then it returns JSON which is not that trouble still we want CSV so use JSON to CSV converter.
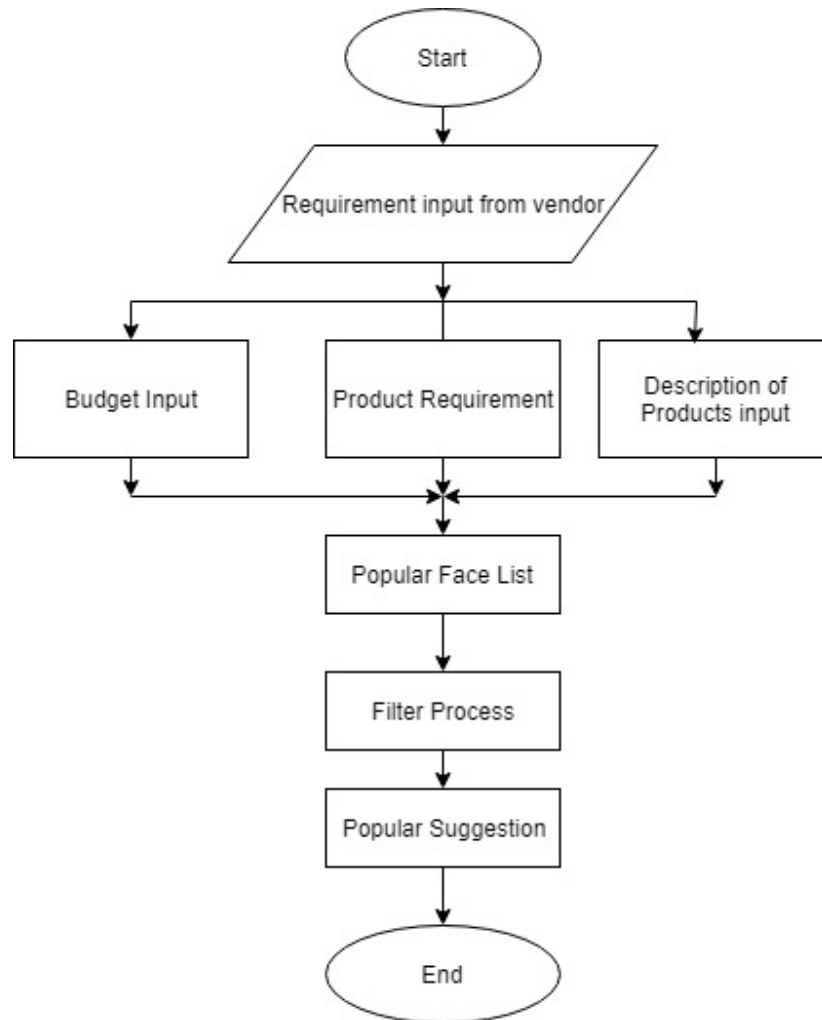
3.1.2 Classifier

Figure 3.3: Module 3

Our classifier is based on Naive Bayesian. This classifier is supervised.

1. **Spam Remover:-** They are various tweets where tweets tag is about sports and the main tweets are about cars then it comes under spam which a classifier should recognize which words are forbidden so we must drop that tweet from the further process.

2. **Keyword Bank:-** As we are using supervised environment we will feed some keyword to analyze tweets which tweets is of sportsmen, popularity, emotion etc.

3. **Tweets Cache:-** Tweets Cache is a small memory but its too fast when it comes to access

time. If Tweets is re-tweets from the same source then it stores in tweet cache where it loads from the cache without requesting server.

3.1.3 Semantic Analysis

1. **Semantic Analysis:-** Semantic Analysis is a classification of emotion, expression from given input where system give appropriate decision.

2. **Positive Review:-** As we want individual sportsmen popularity from twitter but in the social site, there are three types of review i.e Positive, Negative, Neutral. So to get the popularity of an individual we should take positive review into consideration.

3.1.4 Popular Face List

Popular Face List is a mathematical model where we need Sportsmen Rank and also Twitter Popularity.

3.1.5 Requirement gathering

Requirement gathering has sub module i.e Product Description, Product Budget, Product Requirement. Agency will feed their details of Requirement in our Website so we can suggest the best sportsmen.

3.1.6 Filter

Filter is just an query selector or we can say suggestion to fit our requirement also it use classifier to sort optimal options.

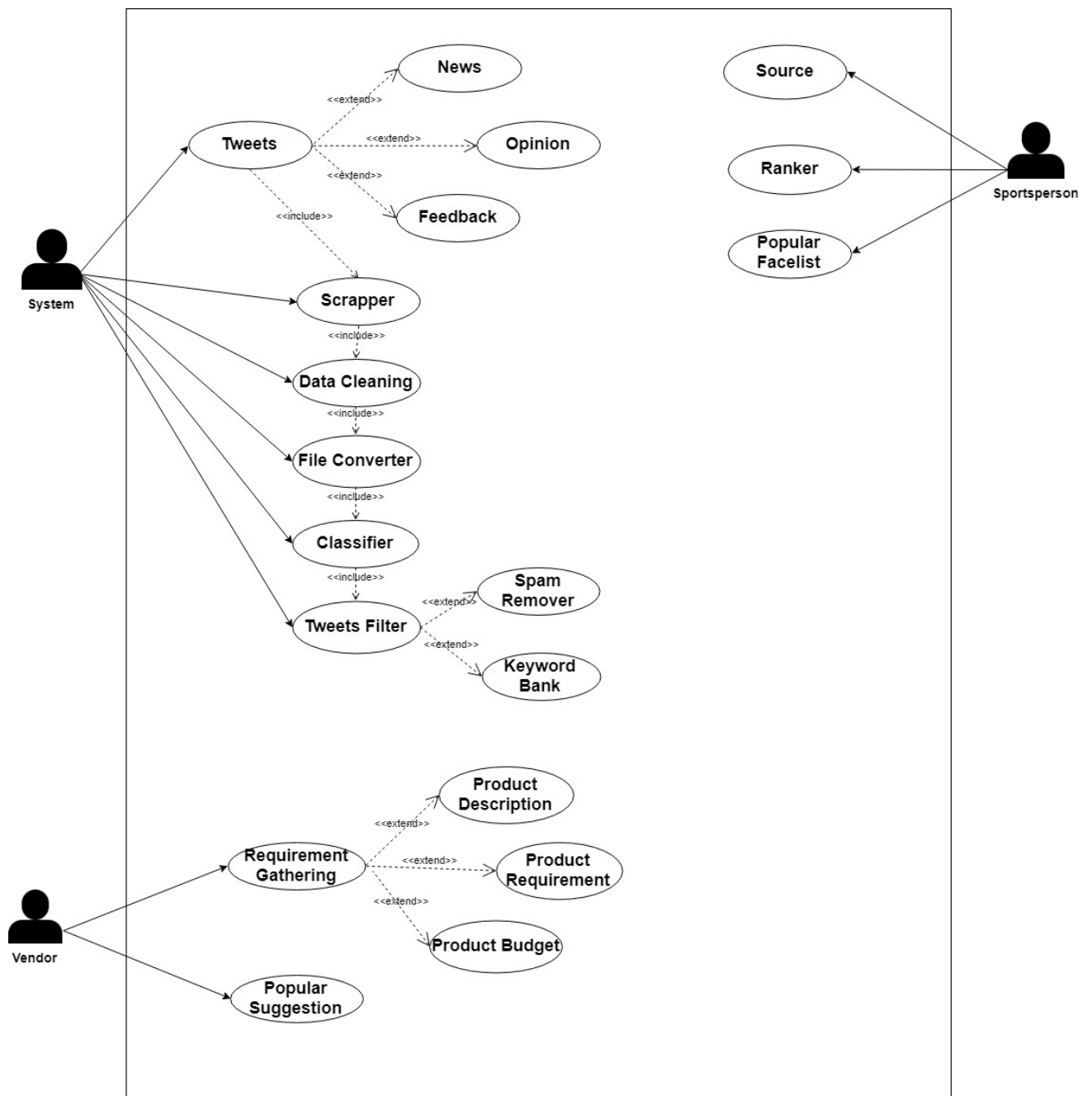## 3.2 Software Architectural Design

3.2.1 Use Case

Figure 3.4: Use Case

3.3 Project Requirements

3.3.1 Software Requirements

I **Software Requirement for Implementation**

(a) Python

Python is one of the unique languages which recognize as simple and powerful. It has a solution to a complex question that's why developer prefers it for Machine

learning. It has a powerful library which make data analysis simple and lots more.

(b) MySql

Mysql is the most popular open source relational SQL database mangement system. It is easy to use and command interface to interact with it. Using PhpMyAdmin, we can also use Mysql but it has a web interface which provides great management to handle crud operations.

(c) PHP

PHP(recursive acronym for PHP: Hypertext Preprocessor) is a wide open source scripting language that is used extensively for web development and application and can also be embedded in HTML.

(d) XAMPP

XAMPP is free and open source cross-platform web server solution web stack. It is only used for development tool to grant web developer to test the code in their own computer.

(e) HTML and CSS

HTML is a markup language to design a structure of web layout.CSS is to design and style the layout.

II **Software Requirement for Deployment**

(a) XAMPP

XAMPP is free and open source cross-platform web server solution web stack. It is only used for development tool to grant web developer to test the code in their own computer.

(b) Browser A Web Browser, or commonly a "Browser" is an application used to access and view the website

3.3.2 Hardware Requirements

I **Hardware Requirement for Implementation**

- Recommended Minimum: Pentium 1 GHz or higher with 1GB RAM or more

- Space Required: 500MB

- Display:800 x 600, 256 colors (Minimum); 1024 x 768 high color, 32bit

II Hardware Requirement for Deployment

- Memory: 1 Gigabyte Minimum

- Processor: 1.0 GHz

- Internet: 200kbps minimum

Chapter 4

Market Potential

4.1 Market size

Since social media and internet is inexpensive in sharing our own personal views on things, this can be used as a tool to scout out what the people actually think about a certain sportsperson whether he is popular in a positive way or a negative way. The advertising agencies need sports person with a good image for their products so to identify this system is pretty useful.

4.2 Market growth rate

According to the survey by Internet World Stats, active internet users over Internet has been tripled in last decades i.e in 2008 it was 1.3 billion and in 2018 it is almost touching 4 billion, and same in the case of social media as well, since sports has a huge market all over the world, hence there is huge growth for advertising agencies as well.

4.3 Profitability

Sportsperson can gain more fame, popularity and money through this software and can attract more offers from other advertising companies as well as companies can use it to their advantage of getting a rise in profit over their product sales.

4.4   Competitive Advantages of Project

Since there is no software like this in the market today, the company who has it will have a huge advantage over the competitive market because of the advanced analysis to book a player before anyone else can do it. And also there is no need to go through the whole process.

Chapter 5

Conclusion and Future Scope

5.1 Conclusion

The project will be build keeping into consideration that , it will be helpful for the advertising agencies to get the report of the players according to the sports that the company wants the certain player to promote a product , which can result into huge profit for the company as well as the player.

5.2 Future Scope

The emergence of artificial intelligence means we're better able to understand customer needs. In fact of spending advertising money to remind consumers that brands exist, the brands can view customer needs in real time and focus their messaging on the things that matter to those customers. We can suggest looking to automation and other emerging technologies to deliver products and operational efficiencies that can create the cost savings businesses are looking for, rather than stripping out product features. With the advancement in digital media, it is easy to interact with the sportsperson and also it can help in the advertising industry. By the use of classifier, we can automatically classify the model well suited for the industry. For example, we have an advertisement based on smartphones we can choose a female model who is fair and beautiful. And along with that, we can choose a male model on the basis of budget.

References

[1] Abhinav Yashkar *Quantifying the performance of players in a football match, IJETT - Volume 45 Number 1- March* 2017

[2] Ramya Nagarajan, Yihang Zhao, Lin Li *Effective NBA Player Signing Strategies Based on Salary Cap and Statistics Analysis, IEEE 3rd International Conference on Big Data Analysis*

[3] Avinash Chandra Pandey, Dharamveer Singh Rajpoot, Mukesh Saraswat *Twitter sentiment analysis using hybrid cuckoo search method " ," Information Processing and Management 53* 2017

[4] Drs Raffaele Poli , Loic Ravenel , Roger Besson *Technical analysis of player performance " , " CIES Football Observatory Monthly Report Issue 15* May 2016

ACKNOWLEDGMENT

I would like to take the opportunity to express my sincere thanks to my guide **Prof. Apeksha Gopale**, Assistant Professor, Department of Computer Engineering, AIKTC, School of Engineering, Panvel for his invaluable support and guidance throughout my project research work. Without his kind guidance & support this was not possible.

I am grateful to him/her for his timely feedback which helped me track and schedule the process effectively. His/her time, ideas and encouragement that he gave is help me to complete my project efficiently.

I would also like to thank **Dr. Abdul Razak Honnutagi**, AIKTC, Panvel, for his encouragement and for providing an outstanding academic environment, also for providing the adequate facilities.

I am thankful to **Prof. Tabrez Khan**, HOD, Department of Computer Engineering, AIKTC, School of Engineering, Panvel and all my B.E. teachers for providing advice and valuable guidance.

I also extend my sincere thanks to all the faculty members and the non-teaching staff and friends for their cooperation.

Last but not the least, I am thankful to all my family members whose constant support and encouragement in every aspect helped me to complete my project.

Syed Areeb Iqbal Ahmad Rehana
13CO61
(Department of Computer Engineering)
University of Mumbai.