

Predicting Best Match Sports person for Product Advertisement

Shah Shahil Shakir Hussain, Sarguroh Junaid Jawed,
Gavandi Abhay, Sayed Areeb
Department of Computer Engineering
AIKTC, New Panvel

Apeksha Gopale
Prof., Department of Computer Engineering
AIKTC, New Panvel

Abstract:- 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.

Keywords:- Sentimental analysis, Machine learning, Prediction, Product advertisement, Sports, Naïve bayes

I. INTRODUCTION

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.

II. LITERATURE REVIEW

Sentiment analysis is a method of data mining to determine the outlook of a speaker or an essayist as for some subject. Many sentiment analysis expertise has been developed in recent years. The simplest classification is to classify text into either a negative or positive sentiment category based on text classification. The fundamental methodology is vocabulary based, which is to look at tweets in view of the words that the content contains. The texts are analyzed and checked if some specific sentimental words are contained. It has been characterized that a few words are positive and some are negative and each of them is given a thought score. The whole text will be persistent based on the score.

However, it is difficult to maintain a dictionary of keywords to calculate the thought score. For this reason, some supervised and unsupervised algorithms are also enlarged and used for text categorization, such as Naive Bayes. For these machine learning algorithms, in order to do the categorization, sufficient labelled data needs to be fed into the classifier to instruct the classifier. Based on the teaching dataset, the classifier will build a probability model that is able to give a prediction of the next input [1].

Public opinions and public point of view have been suggested and generally accepted as possible indicators. There has been an advent of the number of social networks in which a person can share their instinct, feelings, status or locations, and a rush in their popularity. More specifically, Twitter has become one of the most important origins of public sentiment on the various subject matter about companies, products, movies and many others. On Twitter, although each different user post, which is called a tweet, is restricted to just 140 characters, then a large number of tweets could create public opinions and public sentiment at a certain extent [1].

Among all their judgment, the judgment of particular organizations have raised, and it very well may be contended that Twitter posts by a client are an exact representation of a user's opinions and thoughts.

Naive Bayes is one of the most common distribution methods that can be used to perform text categorization. Previous to that, we need to first have a look at what the feature vector is used for. With a specific end goal to perform personality, we have to choose highlights from the information first. For content grouping, the element vector is moreover called the term vector, which is the most essential structure between the preparation and categorization process.

All tweet texts will be transformed to term vectors to be organized by the classifier. Usually, the term vector is generated based on a rare vocabulary, which is generated from the training dataset, and there are no identical twin words in the vocabulary.

III. PROPOSED METHOD

In API we will get JSON which we will convert into SQL format.

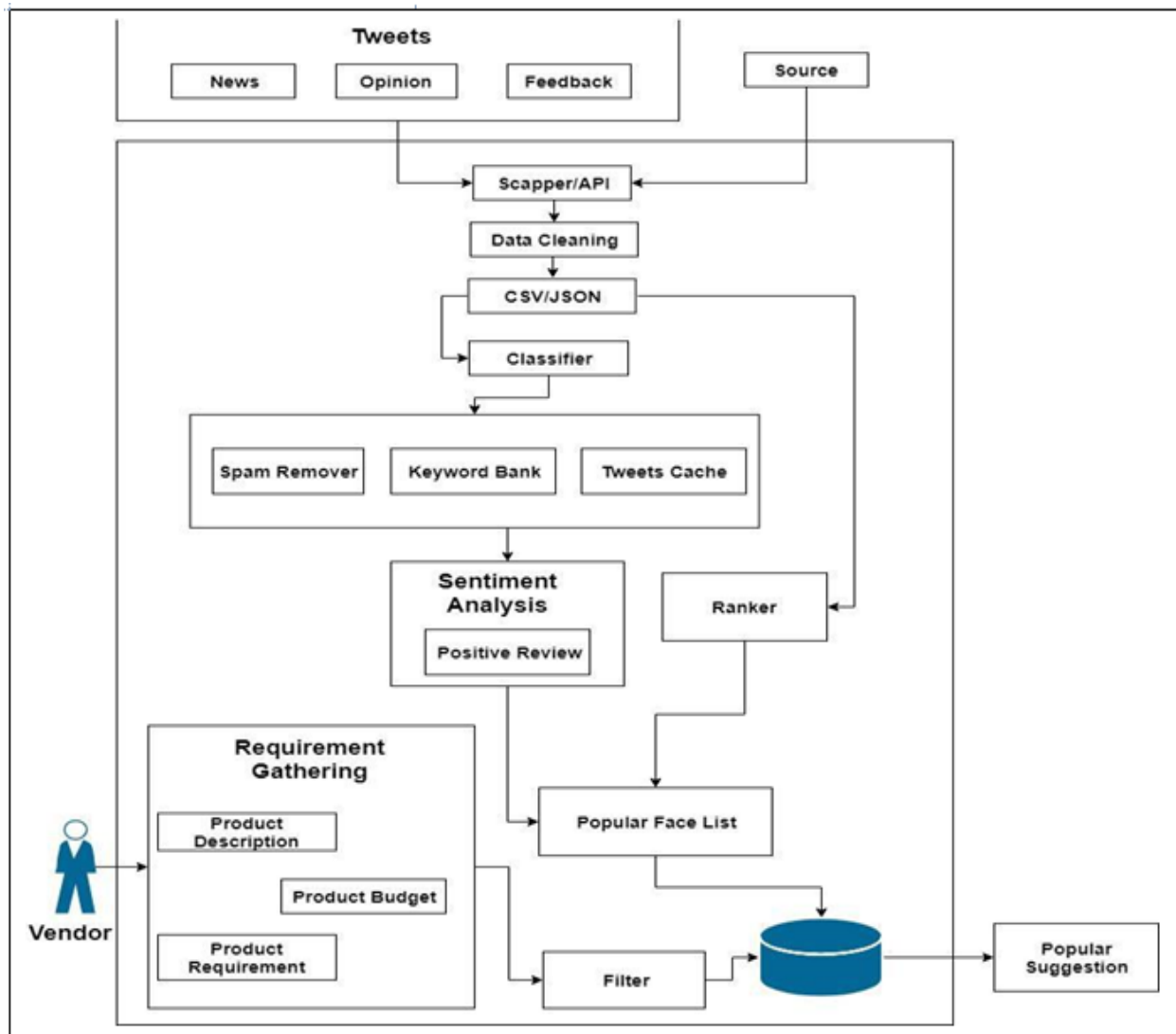


Fig 1:-System Architecture

A. Data Extraction/Scraping

A web scraping programming will consequently load and concentrate information from numerous pages of sites in light of your prerequisite.[2]

As our System is related to predicting sportsmen for an advertisement so we need the best result. If we want the best result we should have better knowledge in our field which needs deep research and analysis skills. As we are human we can't remember every single player information which results in incorrect information. An Individual cannot enter every single player information to the database which also causes incompleteness. So to overcome this problem we will extract / scrap data for certain website/API with help of python.

B. Data Formatting

As we have Extract data so we will not get data in required format i.e unstructured data and also by scrapping it need deep analysis of parent to child tag. First, we will do csv to SQL and

C. Data Cleaning

Data Cleaning is to identify incomplete, incorrect, inaccurate, irrelevant parts of the data and then changing, modifying or remove the dirty or unrequired data.[3]

We will discard the unwanted table or recordset from extracted data and clean incomplete, inaccurate data.

D. Data Analysis

Data Analysis is a process of extracting useful information from a dataset to achieve a certain goal.

We will analysis our dataset and discovered using data to transform into further purification.

E. Data Classification

Naive Bayes Algorithm is based on Bayes Theorem in which we compare it is individual variable with predictors (set of inputs). Naive Bayes is easy to build, unique and works on a large dataset. We take the data which is required and discard the non-required data.[4]

As we have large dataset so our main problem is to classify according to our need which results in better accuracy. We will use Naive Bayes Classifier for filtering, prediction, sentimental analysis etc.

In Sentiment Analysis, as we have large dataset if we use a traditional algorithm to review twitter opinion obviously it will take time which results in a loss of capital. So we will supervise with Naive Bayes algorithm with positive, negative, neutral review keyword which helps to do more fast and efficient way. In Filtering, if we use a database to take certain constraints and show a result but it is a lengthy process to achieve on a large dataset. So by using Naive Bayes Classifier, we will make an independent variable which represents like rank, net value, etc so we can predict the higher probability of result to optimize query result.

F. Popular Facelist

If a Player Rank is average that doesn't mean that he is not popular.

For Example -Ishant Sharma was also popular due to his funny face but he is an average player. Our motive is popularity/trending personality to attract advertisement agency. Now we will use Our Twitter Popularity and Sports Ranking to shortlist candidate using a certain Mathematical Model.

G. Requirement Filter

Requirement Filter works on SQL query in which vendor will get the result from our database.

From Popular Face List, Requirement will be analyzed and send to Filter which will Suggest according to their Budget, Popularity, etc.

H. Popular Suggestion

As Agency Requirement, We will suggest a list of their suitable Requirements

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]. <https://mospace.umsystem.edu/xmlui/bitstream/handle/10355/45675/research.pdf?sequence=1>
- [2]. <https://www.webharvy.com/articles/what-is-web-scraping.html>
- [3]. https://en.wikipedia.org/wiki/Data_cleansing
- [4]. https://www.saedsayad.com/naive_bayesian.html