DATA ANALYSIS AND RESULT PREDICATION USING TWITTER DATA MINING

Archana Gundavade*1, Dhanshree Dalavi*2, Ranjana Landage*3, Sonali Raijadhav*4, Akshata Sutar*5

*1Professor, Computer Science and Engineering Dr. J. J. Magdum College of Engineering, Jaysingpur, Maharashtra, India.

*2,3,4,5UG Student, Computer Science & Engineering, Dr. J. J. Magdum College of Engineering, Jaysingpur, Maharashtra, India.

ABSTRACT

Social Media houses an immense measure of information which can be used for information mining. It has become an indivisible source that has been impacting the way of life of millions of individuals. The primary goal of this paper is information examination and result expectation utilizing online media mining. In this task exhibit how we can abuse social and primary literary data of Tweets to distinguish the moving points utilizing incessant examples and characterize the notions of Tweets utilizing Natural Language Processing (NLP). Utilizing online media as a stage this work expects to assemble information and to examine it to draw the personal conduct standard of the current youth that assists us with recognizing their inclinations, different preferences.

Keywords: Sentiment Analysis, Research, Data Mining, Text Classification, Tweet, Polarity Score.

I. INTRODUCTION

Social media catches the beat of humankind, it can straightforwardly examine conclusions and conduct of millions of clients to acquire understanding into human conduct, market examination and item suppositions. Web-based media mining is the way toward addressing, examining and removing noteworthy examples and patterns from crude web-based media information. This task intends to plan a framework that can investigate and sum up the information accessible via web-based media. The current work incorporates two primary subjects. Supposition examination. Information mining is the computational interaction of finding designs in enormous informational indexes including techniques at the convergence of man-made reasoning, AI, insights, and data set frameworks. The general objective of information mining measure is to remove data from a dataset and change it into a justifiable construction for additional utilization. The various methods utilized are grouping, bunching, relapse, peculiarity identification, affiliation rules, organized expectation and so forth

The objective is to mine sentiments from social media information at sentence and record level. Text classification: The objective is to bunch comparable instant messages that can make data more sensible. It utilizes a structure that can bunch comparative content have a place with the very theme and a few comparative points that can have a place with a more extensive class. Utilizing text classification as the premise of information mining, we accordingly can anticipate an example from which the conduct of the young via web-based media can be resolved.

II. METHODOLOGY

The Data Analysis and Result Predication Using Twitter Data Mining is separated in 4 stages as:

1. GETTING DATA FROM TWITTER STREAMING API

2. PERUSING AND UNDERSTANDING THE DATA

3. MINING THE TWEETS

4. SENTIMENT ANALYSIS

1. GETTING DATA FROM TWITTER STREAMING API

Programming interface (Application Programming Interface) is an apparatus that is furnished to engineers to communicate with their administrations and to get to information in automatic manner. For this undertaking, we will utilize Twitter Streaming API to download tweets identified with the gave watchwords.

• Step1:Getting Twitter API Keys. Make a twitter account. Go to https://apps.twitter.com/and sign in to your record. Snap "Make New App" Fill out the structure, consent to the terms, and snap "Make your Twitter application" In the following page, click on "Programming interface keys" tab, and duplicate your "Programming interface key" and "Programming interface mysterious". Look down and click "Make my entrance token", and duplicate your "Entrance token" and "Access token mystery".

• Step 2: Connecting to Twitter Streaming API and downloading information. We utilize a Python library called Tweepy to associate with Twitter Streaming API and to download the information. Next make a document called twitter_streaming.py, and compose the code for gathering tweets. Enter your certifications into access token, access_token_secret, consumer_key, and consumer_secret. On the off chance that you run the program from your terminal utilizing the order: python twitter_streaming.py, you will see information streaming. You can stop the program by squeezing Ctrl+C. We need to catch this information into a record by funneling the yield to a document utilizing the accompanying order: python twitter_streaming.py > twitter_data.txt.

2.   PERUSING AND UNDERSTANDING THE DATA

The information that we get is in JSON (JavaScript Object Notation) design. This arrangement makes it simple to people to peruse the information, and for machines to parse it. We utilize 4 Python libraries: json for parsing the information, pandas for information control, matplotlib for making diagrams, adn re for normal articulations. The json and re libraries are introduced as a matter of course Python.

• We will begin first by transferring json and pandas

• Next, we will add the information to an exhibit and call it tweets.

• To print the quantity of tweets, use print order.

Then, at that point, structure the tweets information into a pandas Data Frame to improve on the information control. We will begin by making a vacant Data Frame called tweets.

3. MINING THE TWEETS

Our principal objective through text mining is to analyze the notoriety of watchwords being given. We will add labels to our tweets Data Frame to have the option to control the information without any problem. Target tweets that have required watchwords. Concentrate joins from the important tweets' straightaway, add segments to our tweets Data Frame contingent upon the points.

4. SENTIMENT ANALYSIS

• Call Sentiment Intensity Analyzer () work utilizing Vader supposition investigation instrument. It returns conclusion power scores to the sentences.

• Using the polarity scores () work return a float for assessment strength dependent on the info text. Positive qualities are positive valence, negative worth is negative valence and the excess are nonpartisan valence.

• The polarity scored words are contrasted and a worth having a greatest and least reach.

Then, we will make diagrams plotting the opinion-based outcomes

III.   MODELING AND ANALYSIS

This model is the conceptual model that defines the structure behaviors and more use of system. It consists of four stages: -

1.Data Collection: -In this system, tweets are collected from twitter. These tweets are related to trending topics like organizations, famous personalities, music, sports, politics etc. These tweets are json format.

2.Preprocessing of Data: - The data that we get is in json format. This format makes it easy for humans to read the data and for machines to parse it. That data stored in json format.

3.Data Modeling: -Documents are run through a string to word vector kernel and stored data from the data frame is converted into tokens.

4.Sentiment Analysis & categorize tweets: - Intensity of that token is calculated. According to intensity, polarity get assign to respective tokens. Based on polarity score tweets are classified into positive, negative & Neutral.
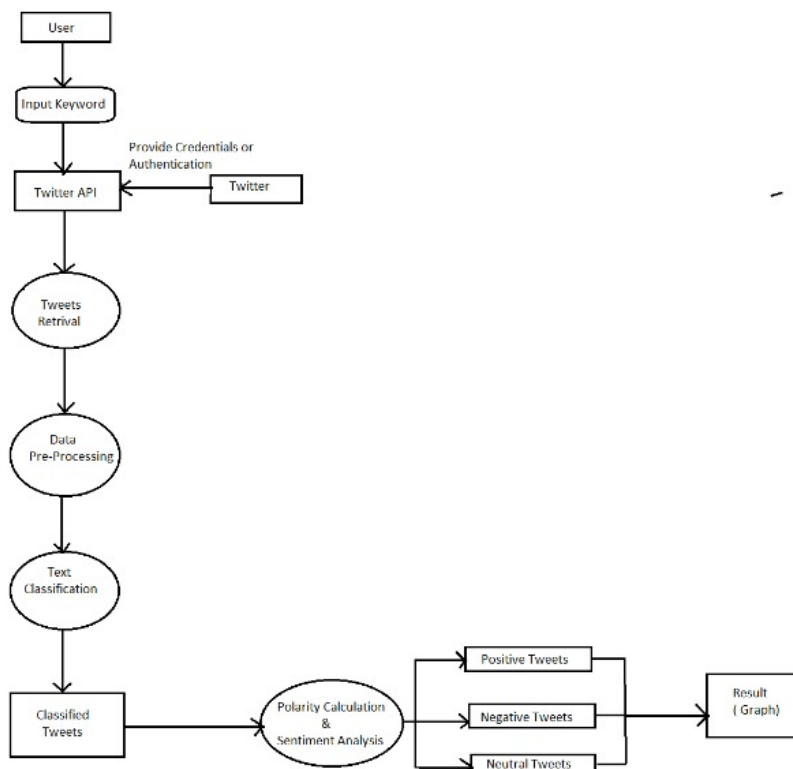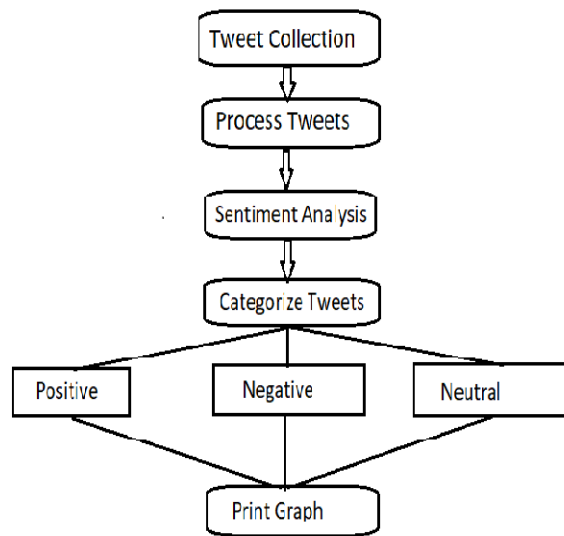
Figure 1: Data Flow Diagram.

IV.   RESULTS AND DISCUSSION

We tested and evaluated the approach with different types and number of Tweets. Sample test results are given below. The keywords passed are (Vaccination)

RESULT: Number of tweets for each keyword passed is displayed below :

General Report: -Weakly Positive

Detailed Report: -15% people thought it was positive.

40% people thought it was weakly positive.

 0% people thought it was strongly positive.

20% people thought it was neutral.

0% people thought it was negative.

25% people thought it was weakly negative.

0% people thought it was strongly negative.

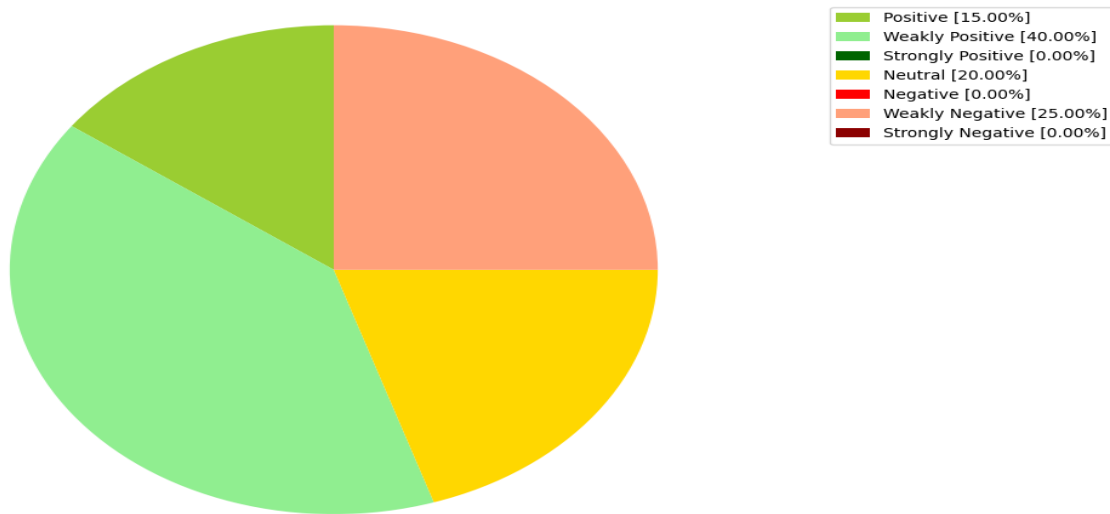How people are reacting on vaccination by analyzing 20 Tweets.

Positive [15.00%]
Weakly Positive [40.00%]
Strongly Positive [0.00%]
Neutral [20.00%]
Negative [0.00%]
Weakly Negative [25.00%]
Strongly Negative [0.00%]

Figure 2: Pie-chart on Reaction of people on Vaccination in Percentage.

V.   CONCLUSION

This examination began concurring with an objective to characterize tweets as per polarity. Distinguishing extremity on tweet in analogs to discovering an extremely elusive little thing. Challenges associated with gathering successful information and preprocessing them so everything text could be utilized. A legitimate technique was utilized to remove tweets was use with the assistance of hag work that sudden spike in demand for consistent schedule. Then, at that point the information was preprocess utilizing numerous strategies to make the information without clamor and anomaly. The spotless information was then further preprocessed to keep away from any tranquility of any data to be taken out. That was not aiding in assessment examination. Investigation and examination were led on genuine datasets to discover the extremity of tweets. The distinctive boundary picked to gauge the extremity were, TextBlob that was a library gotten from Natural Language Processing. Assessment mining on tweets will proceed to be and significant space of exploration with the developing business sector on AI and online media. These theories present a relative investigation on opinion examination in an adaptable and savvy information mining system.

ACKNOWLEDGEMENTS

VI.   REFERENCES

[1]   Zhunchen, L., Miles, O., Ting, W., "An effective approach to tweets opinion retrieval", Springer Journal on World Wide Web, May 2015, vol. 18, no. 3, pp 545–566.

[2]   Sameesksha Shrivastava, Dr. Pramod S. Nair, "Mood Prediction on Tweets Using Classification Algorithm", International Journal of Science and Research, 2015, vol. 4, no. 11, pp. 295-299

[3]   SharadGoel, Daniel C. Goldstein, Predicting Individual Behavior with Social Networks, Marketing Science Journal, 2014, vol..33, no.1, pp 82-93.

[4]   Shruti Wakade, Chandra Shekar, Kathy J. Liszka and Chien-Chung Chan. Text Mining for Sentiment Analysis of Twitter Data,2016, no.1, pp 1-4

[5]   Manisha Rani, Jyoti Arora, Twitter Data Predicting Geolocation Using Data Mining Techniques, June 2016, vol. 4, no. 6, pp. 10446-10453.