IMPLEMENTATION OF MACHINE LEARNING CLASSIFIER TO CLASSIFY THE USER COMMENTS ON SOCIAL MEDIA

[1]Pragati Sahane, [2]Harshada Rodage, [3]Snehal Wakchaure, [4]Sai Pund

[1]UG Scholar, Dept. of Computer Engineering., Amrutvahini College Of Engineering,India
[2]UG Scholar, Dept. of Computer Engineering., Amrutvahini College Of Engineering,India
[3]UG Scholar, Dept. of Computer Engineering., Amrutvahini College Of Engineering,India
[4]UG Scholar, Dept. of Computer Engineering., Amrutvahini College Of Engineering,India
saipund@gmail.com

ABSTRACT:- Now a days the Internet facility is full of information pertaining to opinion of people through applications, for example, web-based media, miniature contributing to a blog destinations, audit locales, individual web journals and so on Nostalgic investigation is a region in text mining where assessment of individuals can be broke down and arranged into positive, negative or unbiased. In this work, the estimations of the tweets or audits distributed in the twitter is recognized via looking for the specific catchphrase in tweets and afterward assess the extremity of the tweets as certain and negative. The estimations of the tweets that are tweeted on a twitter assessed dependent on include choice of each score words. To choose the best highlights Naive Bayes Classifier (NBC) is utilized for preparing and testing the highlights of a words and furthermore assessing the notion extremity of each tweets. Performance evaluation parameters such as accuracy, precision and time is taken into consideration and compared with three machine learning classifiers, namely, Random Forest, Naive Bayes and Support Vector Machine(SVM).

KEYWORDS: Twitter Analytics, Naïve Bayes classifier, Sentiment analytics, Polarity, etc.

I. INTRODUCTION

The development of social network platforms has given people a new way to generate and consume a great deal of information on the web. In the past, people used to get information from portal websites. A large number of websites provide a long list of topics varying from politics to entertainment. These traditional online information sources are useful but less efficient because they often contain redundant information. However, since the arrival of online social network platforms, people tend to get information from these platforms because of their fast and efficient

features. These platforms are available for users to choose the information source they are interested in. And also a large number of social network platforms such as Twitter, Google+, and Facebook provide information for users. Twitter is the most popular microblogging platform in the world. It is also the fastest growing social network platform and has a dominant position in the area of microblogging. More than 500 million registered users post 340 million twitter messages every day, sharing their opinions and daily activities. Compared with regular microblogging platforms, Twitter messages are much shorter. You are only allowedto post 140 characters or less in one Twitter message. This feature makes Twitter easier for people to get the main point from the massive amount of data accessible on the web. Contingent upon the need of the clients, Twitter clients can follow whichever individuals and data source they like. With the entirety of the benefits referenced above, Twitter in this manner has become an incredible stage with numerous sorts of data from overall breaking news to buying items at home.

Over the most recent couple of years, the data streams on Twitter have encountered a unimaginable expansion in the notoriety of this interpersonal organization. The clients arrange a huge measure of data about various angles. In any case, not the entirety of the data is valuable for clients and every client has their own advantages and inclinations. There is earnestness for clients to have customized administrations. These days, an ever increasing number of customized administrations are given to profit the clients. Individuals need this customized administration to make their speedy lives more effective. Consistently, a lot of data is distributed by clients on the Twitter stage. These information identify with clients conduct and many exploration concentrates consequently center around Twitter and this information assortment. One of the examination concentrates in the field of Twitter is client demonstrating. To offer a customized assistance, scientists began to investigate positioning and proposals of web assets referred to from Twitter. A lot of exploration center around demonstrating clients intrigues dependent on clients distributed tweets information. Notwithstanding the tweets substance and Twitters possible use, scientists additionally saw that tweets frequently pass on relevant data about the users emotional states. Emotion analysis on Twitter has thus become an important research issue in the micro blogging area. Most research related to emotion focuses on the sentiment classification on Twitter. A number of features and methods for training classifier for sentiment on Twitter platform have been researched in the past few years with varying results. There are also some other research studies related to emotion analysis on Twitter. One of the studies in this area is about getting feedback about products by extracting the customers emotion on the Twitter platform. Also, investigating public attitudes by extraction of emotions from Twitter messages has been the focus of previous studies.

II. RELATED WORK

In paper [1], authors propose two attributes for visual sentiment analysis model: multiscale perception and different levels of emotional representation. To learn the sentiment correlation degree of different regions for different scale in the image, multi-level context pyramid network composed of multi-scale adaptive context modules is proposed. Proposed MACM module which

has multiscale attribute can combine different scales from different data sources and mine semantic information of images. It is also independent of model structure.

Authors have analyzed the polarity for measuring the amazon product semantical analysis. Thus, if the bar height in the graph towards 1 means the product reviews are positive, and if the bar height looks down, that means the semantics of that product is negative same as well when they analyse for an amazon user review. The user's review is different from another user. they have analyzed the polarity for measuring the amazon user's semantics. Thus, if the bar height in the graph towards 5, that means the product reviews are positive, and if the bar height looks down, that means the semantics of that users are negative.[2]

Visual sentiment analysis studies the impact of images on human emotions, authors propose a novel picture notion arrangement strategy, which applies the relationship between's item semantics and picture assessment to upgrade the visual notion investigation model for more precise picture feeling prediction.Experiments on the well known full of feeling datasets FI and Flickr LDL show that our proposed picture estimation examination technique can accomplish great execution on picture feeling forecast and beat best in class methods.[3]

In this paper[4], creators propose another way to deal with gather passionate states by ordering instant messages of people. To display enthusiastic states, they use the grounded Circumplex model that portrays full of feeling experience along two measurements: valence and excitement. In technique, Hashtags are used as labels which in course trains supervised classifiers, this detects multiple classes of emotions from tweeter databases.: Text messages, in their raw form, do not have labels. However, in order to train a classifier, supervised learning methods require labeled data.

SrinivasuBadugu, MatlaSuhasini: This paper describes a Rule Based approach, which detects the emotion or mood of the tweet and classifies the Twitter message under appropriate emotional category. Earlier authors logic where only able to determine whether the specific sentiment is positive tweet or negative tweet with the proposal of new methods, the tweet is able to give deeper information about the tweet which can be used in various fields such criminology, psychology, Economics etc.When a user is not specifically micro blogging about their personal emotive status, the message can reflect their mood. [5]

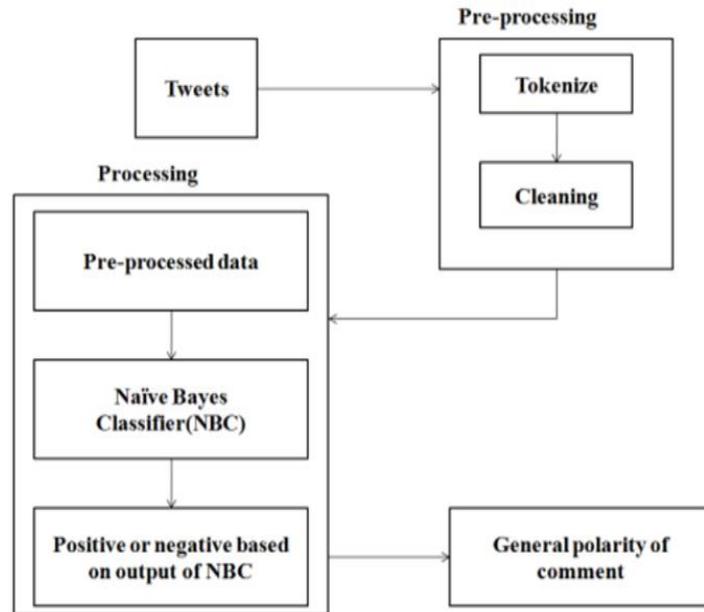## III.PROPOSED METHODOLOGY

A. *System Design :*

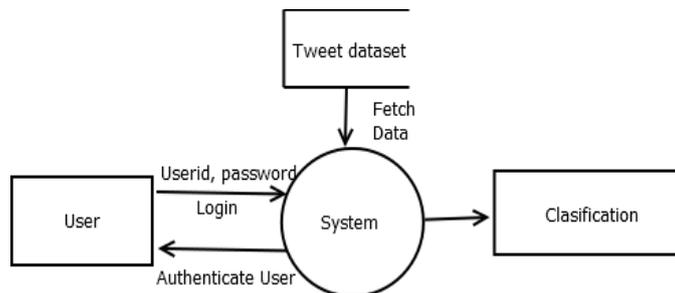Fig 1: System Architecture

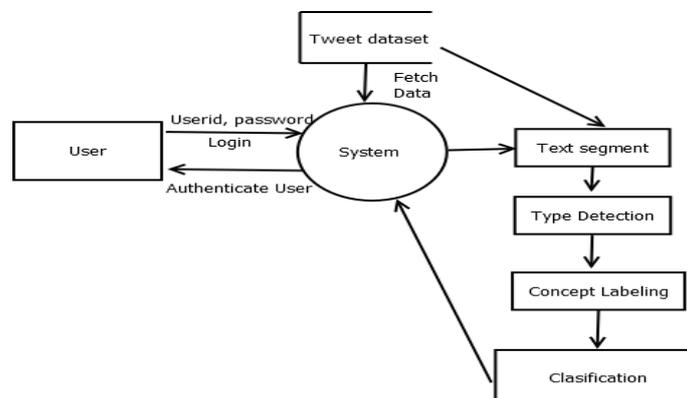B. *Data Flow Diagrams:*

Fig. 2 : Level 0 DFD

Fig 3 : Level 1 DFD

IV. ALGORITHMS

A) Support Vector Machine:

SVM is a machine learning technique to separate data which tries to maximize the gap between the categories. This Algorithm helps to classify the textual feedback and classifies according to emotions.

Input: D Dataset, Semantic of Tokens, Feeds;
Output: Classification of Application
Step1: for each Feed id in D do
Step2: Get on-demand features and stored on vector x   for tweet id
Step3: x.add (Get Features (Feed id));
Step4: end for
Step5: for each Feed in x vector do
Step6: Fetch first feature and stored in b, and other features in w
Step7: h w, b (x) = g (z) here, z = (w T x + b) Step8: if (z≤0)
Step9: assign g(z) = 1;
Step10: else g(z) = -1;
Step11: end if

B) Random Forest

Random forest is a supervised learning algorithm which is used for both classification as well as regression. Be that as it may, notwithstanding, it is for the most part utilized for characterization issues. A backwoods is comprised of trees and more trees implies more hearty woodland. Essentially, arbitrary backwoods calculation makes choice trees on information tests and afterward gets the expectation from every one of them lastly chooses the best arrangement through Twitter. It is an ensemble method which is better than a single decision tree because it reduces the over-fitting by averaging the result.
Working of Random Forest Algorithm
• Step 1: First, start with the selection of random samples from a given Twitter dataset.
• Step 2 : Next, this algorithm will construct a decision tree for every sample. Then it will get the prediction result from every decision tree.
• Step 3 : In this step, Twitter will be performed for every predictedresult.
• Step 4 : At last, select the emotion prediction result as the final prediction result.

C) Naive Bayes

Na¨ıve Bayes Classifier is one of the simple and most effective Classification algorithms which helps in building the fast machine learning models that can make quick predictions. It is a probabilistic classifier, which means it predicts on the basis of the probability of an object .
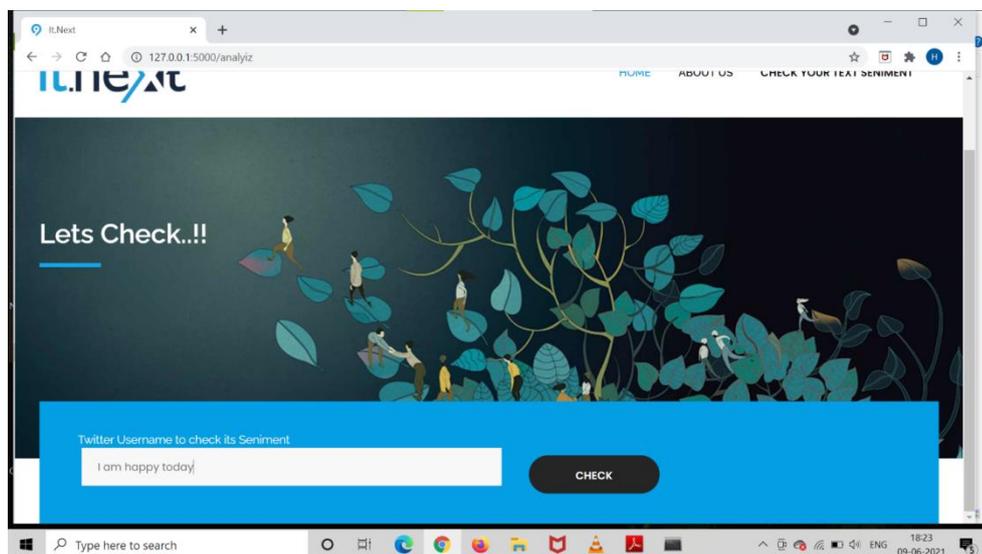
- 1: for Test Class do
- 2: training the classifier naive Bayes
- 3: gauss nb=gaussianNB ().fit (training classes)
- 4: predicting the target variable test classes clf.predict(test class)
- 5: import the models of accuracy score
- 6: accuracy score and matrix of test class prediction given
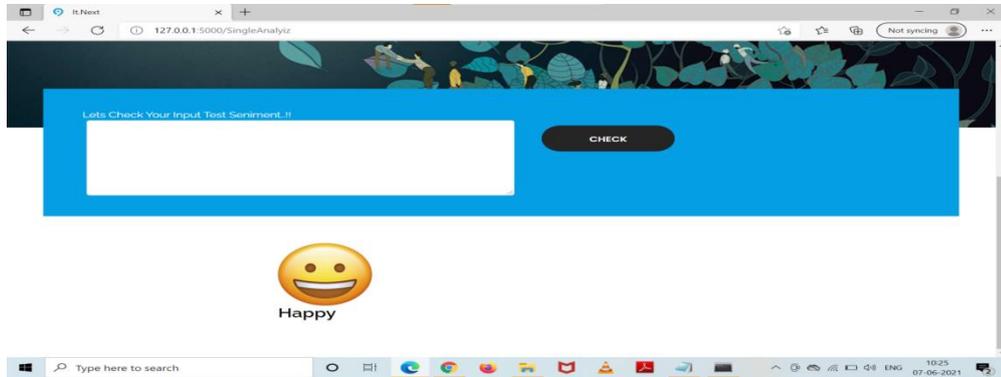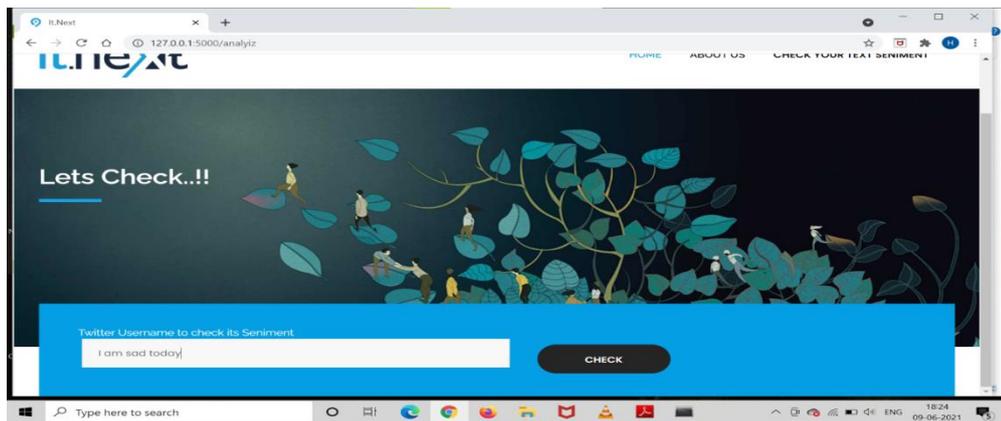- 7: end for

V. APPLICATION RESULTS

A) Web Interface
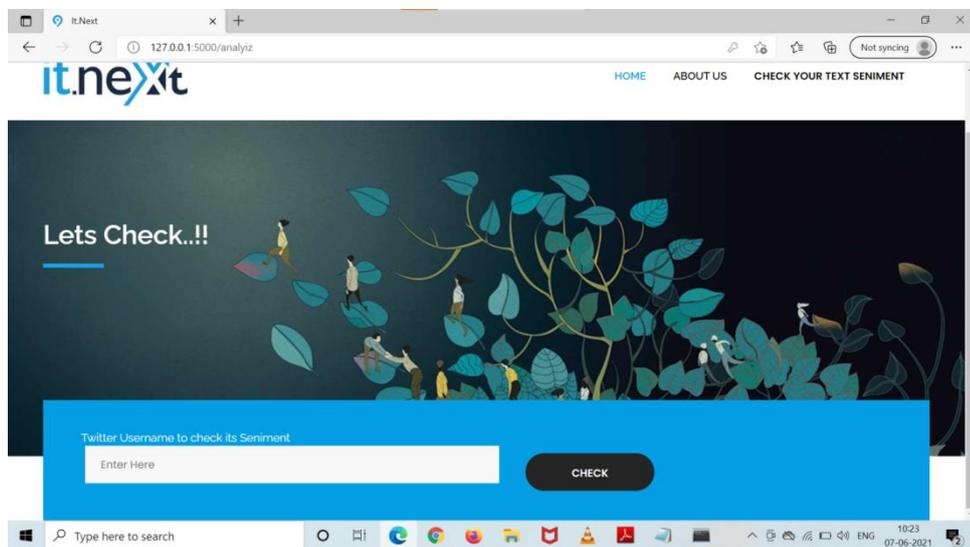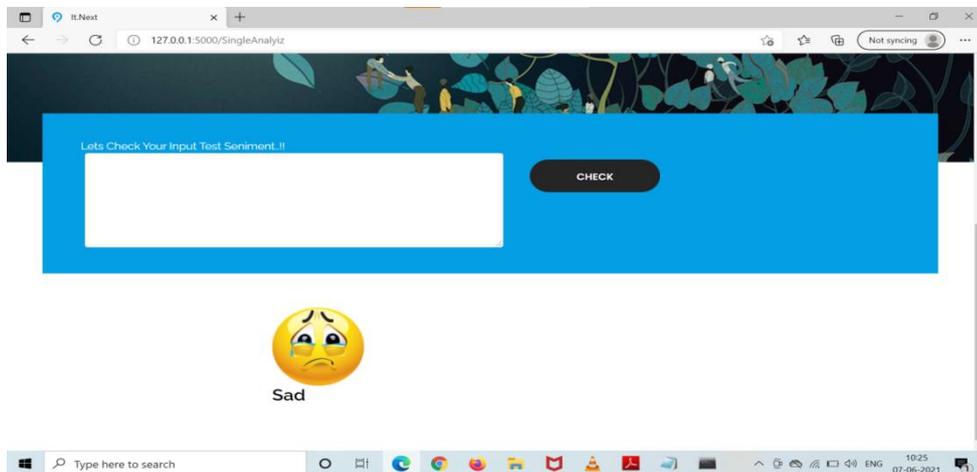

B) **Input tweet for Analysis of Happy Emotion**

C) Output tweet for Analysis of Happy Emotion

D) Input tweet for Analysis of Sad Emotion

E) Output tweet for Analysis of Sad Emotion
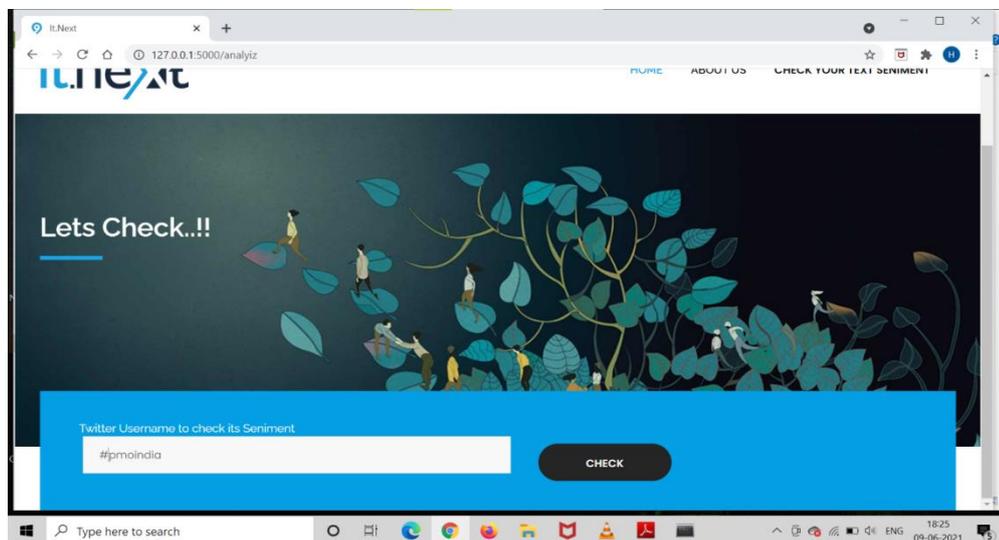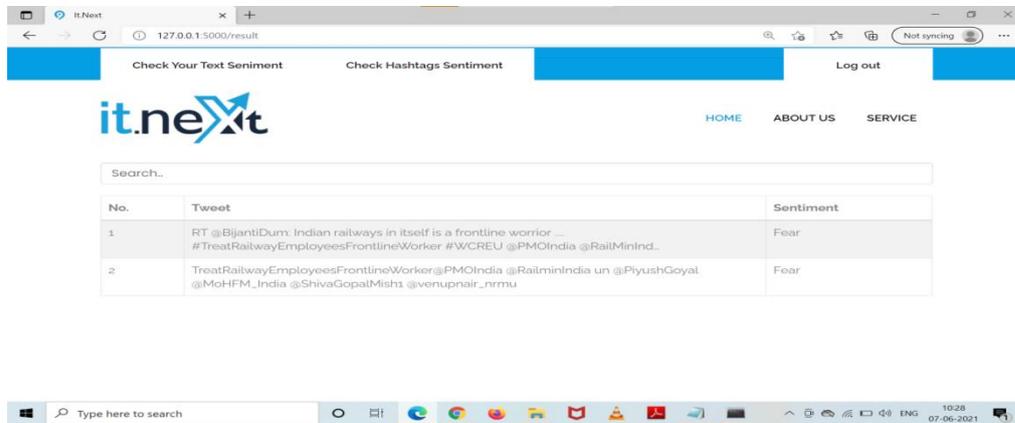
F) Input tweet for Analysis of Angry Emotion
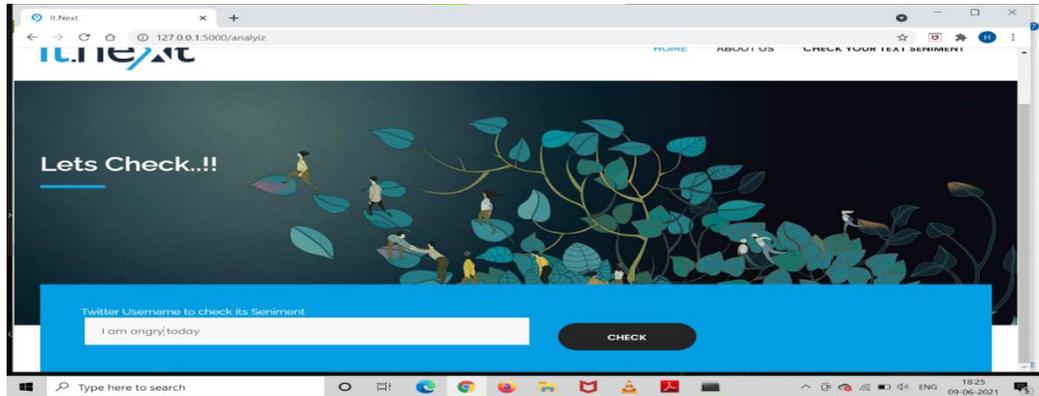
G) Output tweet for Analysis of Angry Emotion
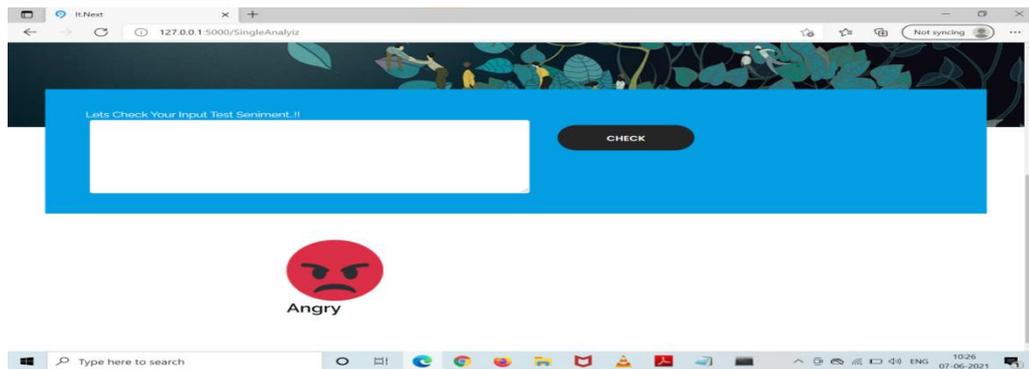
H) Input For Analysis of Hashtag

I) Output for Analysis of Hashtag

VI. CONCLUSION AND FUTURE WORK

Sentiment analysis is used to find the sentiment of author behind his/her comment. In this proposed work, tweets will be removed utilizing a specific string search and these tweets will be exposed to assumption investigation utilizing RF, SVM and NB classifiers to characterize them into positive, nonpartisan and negative. As a piece of investigation, the AI classifiers RF, SVM are thought of and their precision will likewise be estimated considering three features and accordingly increasing the number of tweets. The precision of RF, SVM and NB will be estimated by increasing the number of tweets.

REFERENCES

[1] Ou, H.; Qing, C.; Xu, X.; Jin, J. Multi-Level Context Pyramid Network for Visual Sentiment Analysis. Sensors 2021, 21, 2136. https://doi.org/10.3390/s21062136

[2] Wassan, Sobia & Chen, Xi & Shen, Tian & Waqar, Muhammad & Zaman, Noor. (2021). Amazon Product Sentiment Analysis using Machine Learning Techniques. 30. 695-703. 10.24205/03276716.2020.2065.

[3] Jing Zhang, Mei Chen, Han Sun, Dongdong Li, ZheWang,Object semantics sentiment correlation analysis enhanced image sentiment classification,Knowledge-BasedSystems,Volume 191,2020,105245,ISSN 0950-7051,https://doi.org/10.1016/j.knosys.2019.105245.

[4] Hasan, M., Rundensteiner, E., &Agu, E. (2019). Automatic emotion detection in text streams by analyzing twitter data. International Journal of Data Science and Analytics, 7(1), 35-51.

[5] Suhasini, Matla&Badugu, Srinivasu. (2018). Two Step Approach for Emotion Detection on Twitter Data. International Journal of Computer Applications. 179. 12-19. 10.5120/ijca2018917350.

[6] Niko Colneric and JanezDemsar, "Emotion Recognition on Amazon Product Review: Comparative Study and Training a Unison Model", IEEE TRANSACTIONS ON AFFECTIVE COMPUTING, FEBRUARY 2018.

[7] A. Radford, R. Jozefowicz, and I. Sutskever, "Learning to Generate Reviews and Discovering Sentiment", 2017.

[8] B. Nejat, G. Carenini, and R. Ng, "Exploring Joint Neural Model for Sentence Level Discourse Parsing and Sentiment Analysis",Proc .of the SIGDIAL2017.