

Product Reviews based on N-Gram Model and Random Forest Classifier

Dhruvi D. Gosai Department of Computer Engineering, Chandubhai S. Patel Institute of Technology (CSPIT), CHARUSAT University, Anand, Gujarat, India <u>dhruvidgosai4297@gmail.com</u> Amit Ganatra Faculty of Technology & Engineering (FTE), CHARUSAT University, Anand, Gujarat, India amitganatra.ce@charusat.ac.in

Article Info Volume 83 Page Number: 8200 - 8210 Publication Issue: May - June 2020

Article History Article Received: 19 November 2019 Revised: 27 January 2020 Accepted: 24 February 2020 Publication: 18 May 2020 Abstract:

The gigantic growth of online informal communities (OSN) like Twitter, Facebook, and other interpersonal interaction gateway has made a need to decide an individual's assessment and states of mind. Posting client input on items has gotten increasingly common for individuals to state their sentiments toward items and administrations. A model is available to divide Twitter information about current issues and concentrate data and make forecasts about their conduct on detailed occasions. The organizations feel that there is a chance to recover the market of an item[21]. In this investigation, there is the utilization of AI systems to discover about various current issues. In the first procedure, tweets are assembling by pulling information from witter posts with the topographical area. By applying the tokenization process on tweets and apply emoticons and synonyms features. After that utilization n-gram model with an arrangement of unigram, bigram and trigram and distinctive word reference for score count. At last, characterization is finished with various AI approaches, for example, SVM, ANN, RF and discover hitter among it. The framework is diminishing the expenders on the study by utilizing on the web client tweets.

Keywords: Sentiment Analysis; Pre-Process; Feature Extraction; Grams; Classification.

I. INTRODUCTION

Sentiment is one kind sort of thought, view or attitude of mind dependent on feeling. Sentiment Analysis is additionally called as opinion mining examination. Opinion Analysis has been a trendy appearance in the world of Deep Learning and Machine Learning [7].

Sentiment Analysis is a recent field of research that is closely related to emotion detection and recognization from the text. There are two types of analysis available to detect emotion. I) Sentiment Analysis: In Sentiment Analysis, we can detect positive, negative or neutral feelings from the content. II) Emotion Analysis: In Emotion Analysis, we can detect types of feelings such as Happy, Sad, Anger, Disgust, Fear and Surprise from the content [11]. Face and Voice expression will give better

Published by: The Mattingley Publishing Co., Inc.

result compare to the Textual expression. But Textbased emotion detection system is also very attractive for the research side because nowadays most people write their expression on their person's blogs, journals, product reviews, and comments, etc [11].

Client review is the most important factor. Ecommerce sites like Amazon, Flipkart, etc give a stage for clients to share their incident[8]. Sentiment Analysis is broadly used to follow purchaser reviews. Using this one can progress the product quality, and also raise the earnings margin of the business. In this way, it is also helpful in Business Analysis. Based on customer reviews, we can analyze positive response (feedback) or negative response (feedback) of the particular selected product. One cannot purchase a product without meaningful the view of the other users. E.g. To 8200



purchase a phone, we will ask our friends, relatives or neighbors. Then evaluate the pros and cons of that particular product. Afterward, we will decide whether to buy or not [6].

There are many applications of sentiment analysis. Here in this research, work on an efficient product review based on location. Sentiment analysis also used inefficient current affairs based on location. Current affairs include product or movie rating, upcoming events like sports or political election, etc. It also generates a Geo-graphical vision (view) of the particular event. In this way, analyze location-based event results e.g. analyze political election results, movie reviews based on location. Sentiment analysis also generates a graphical representation of the positive, negative and neutral emotions. By using sentiment analysis, identify which type of people buy which type of the product. By adding location module in sentiment analysis, analyze product reviews based on location. Tweets are progressively easygoing and restricted to 140 characters of content. Twitter is a smaller scale blogging administration that was propelled officially on July 13, 2006 [12].

Sentiment Analysis is one kind of task that can be identifying positive, negative or neutral emotion from the give input data. E.g. "I am so glad today, good morning to everyone", is a general positive text & the text is: "Sholay is such a superior movie highly recommended by 9/10", state positive sentiment towards the movie [14].

Sentiment Analysis is one kind of research field, particularly in light of the measure of open and available existent data that could be removed. In this exploration, we are getting down to business on the live information for example effective item survey dependent on Location. We are going to test include extraction technique with three unique classifiers, for example, Innocent Bayes classifier, Support vector machine, ANN (Back proliferation) with Afinn-111-word reference [14].

In this exploration, we are going to utilize equivalent words just as emoji. Equivalent words are utilized to check all related given words that aren't accessible in the lexicon. Emoji is utilized to change the overall sorts of emoji which is as content in the tweets [14].

There are many applications of Sentiment Analysis like, Product and Service Reviews: Numerous sites that give computerized synopses of surveys about items and their particular viewpoints. A striking case of that is "Google Product Search" [14]. Reputation Monitoring: Twitter and Facebook are a point of convergence of numerous supposition examination applications. most The widely recognized application is observing the notoriety of a particular brand on Twitter or potentially Facebook Decision Another [14]. making: significant application that assumption is examination can be utilized as a significant factor in helping the basic leadership frameworks.

II. LITERATURE SURVEY

In [1] Bouazizi, Mondher, and Tomoaki Ohtsuki describe "A pattern-based approach for multi-class sentiment analysis in Twitter." This Proposed framework is constrained up to seven distinctive assumption classes and this proposed methodology is adaptable and rushed to characterize writings into more classes. In this proposed framework, use SENTA instrument, it is a simple to-utilize graphical UIs and runs multi-class grouping on it. This Proposed methodology demonstrates to be exceptionally precise in twofold grouping and ternary characterization[1].

In [2] Ahuja, Shreya, and Gaurav Dubey describe "Clustering and sentiment analysis on Twitter data." This paper proposes grouping ideas like partition "Clustering Algorithm, Fuzzy C Clustering Algorithm, and K-Mean Means Clustering. This paper directed an investigation and discovered that the utilization of grouping can rapidly and effectively recognize tweets based on their assumption score and can discover week after week and emphatically positive or negative tweets when bunched with after-effects of various word references[2].

In [3] Algur, Siddu P., and Rashmi H. Patil describe "Sentiment analysis by identifying the speaker's polarity in Twitter data." This paper proposes the devices for opinion examination of the information by giving a portion of the tweet information as information and getting the particular scores as yield. This paper likewise characterizes the Acronym Dictionary and Emoji Dictionary in detail. This paper clarifies two order strategies in detail that is, two-way characterization and the three-way arrangement techniques. The two-way arrangement depicts the positive versus negative method for



grouping the tweet's information, the three-way order portrays positive versus negative versus impartial characterization of the given tweets information. The bit tree was planned and organized to demonstrate that it is superior to the example of the unigram model[3].

In [4] Bhan, Namrata, and Mitchell D'silva describe Sarcasmometer using sentiment analysis and topic modelling. This paper proposes a framework that will give a different entry to verify the attain of any sentence/content entered by the customer and decide its score utilizing the most exact calculation. The framework likewise ascertain parameters like exactness, accuracy, review, and Fmeasure for the various calculations like Linear Support Vector Classifier (SVM) under SVM, Regression Logistic under Linear Model. BernolliNB under Naïve Bayes. Since the F-score produced for Linear SVC and Logistic Regression are the same, yet the review for Linear SVC is more than Logistic Regression, consequently utilizing Linear SVC shows signs of improvement precision[4].

In [5] Kumari, Upma, Arvind K. Sharma, and Dinesh Soni describe Sentiment analysis of smart phone product review using SVM classification technique. The point of this paper is to apply Support Vector Machine (SVM) characterization. The trial work incorporates three execution highlights, for example, Precision, Recall, and Fmeasure. Based on these highlights, the precision of the various items has been figured. The presentation coming about models tried to acquire the estimation of exactness, accuracy, review, and f-proportion of Support Vector Machine (SVM). At long last, the Support Vector Machine (SVM) calculation has been accomplished higher precision for example 90.99% and found that the SVM is a hearty and better one[5].

In [6] Lavanya K., and C. Deisy describe Twitter sentiment analysis using multi-class SVM. The calculation Proposed is practical to fixed information of various areas. As tweets are spilling, it very well may be useful to active tweets for a given course of events. The proposed work utilizes 3 class names, for example, positive, negative and unbiased. To be increasingly explicit, three class names can be reached out to five class names. They are sure, positive, impartial, negative, and negative[6].

In [7] Mahmud, Quazi Ishtiaque, Asif Mohaimen, and Md Saiful Islam describe A Support Vector Machine mixed with statistical reasoning approach to predict movie success by analyzing public sentiments. This paper investigates the intensity of open assumptions on foreseeing the achievement of motion pictures. This framework utilized the non-straight RBF portion for our assessment classifier which accomplished preferable precision over the classifiers that utilization direct bits in the acclaimed IMDB Movie Review Dataset (89.51% exactness) and the Pang and Lee Movie Review Dataset (86.86% precision). This framework can anticipate whether a film will be fruitful or not with a precision of 90.3%. It performs very well on motion pictures that have either a high or low appraising yet the exhibition diminishes when considering the hazy when areas meaning considering films that have a normal rating[7].

In [8] Singla, Zeenia, Sukhchandan Randhawa, and Sushma Jain describe "Sentiment analysis of customer product reviews using machine learning." In this proposed work, more than 4,000,000 surveys have been arranged into positive and negative supposition utilizing Sentiment Analysis. Out of the different grouping strategies, Naïve Bayes, Support Vector Machine (SVM) and Decision Tree have been utilized for characterization of audits. The assessment of models is finished utilizing 10 Fold Cross Validation. This examination is expecting to accomplish by directing supposition investigation of cell phone audits and grouping the surveys into positive and negative feelings. In the wake of offsetting the information with the practically equivalent proportion of positive and negative surveys, three grouping models have been utilized to characterize audits. Out of three classifiers, i.e., Naïve Bayes, SVM, and Decision Tree, the prescient precision of SVM is seen as the best. The precision results have been cross approved and the most noteworthy estimation of exactness accomplished was 81.75% for SVM among the three models[8].

In [9] Haque, Tanjim Ul, Nudrat Nawal Saber, and Faisal Muhammad Shah describe "Sentiment analysis on large scale Amazon product reviews. The purpose of this paper is to arrange the positive and negative criticisms of the clients over various items. The proposed model is an administered learning technique. Additionally depicted the fundamental hypothesis behind the model, and contrasted outcome and a portion of the comparative works for item audit. In this paper, likewise attempted distinctive recreation utilizing cross approval, preparing testing proportion, and diverse element extraction process for contrasting differing measures of information with accomplishing promising outcomes. In the greater part of the cases, 10 overlaps gave a superior precision while Support Vector Machine (SVM) if best-arranging outcomes[9].

In [10] Jianqiang, Zhao, and Gui Xiaolin describe Comparison research on text preprocessing methods on twitter sentiment analysis." This paper examined the impacts of content prepreparing strategy on notion grouping execution. The examinations show that the precision and F1-Twitter assumption arrangement quantify or classifier is superior when utilizing the prepreparing strategies for extending abbreviations and supplanting nullification, yet hardly changes while evacuating URLs, expelling numbers or stop words. The Naïve Bayes and Random Forest classifiers are more delicate than Logistic Regression and Support Vector Machine (SVM) classifiers when different pre-handling techniques were applied[10].

III. BACKGROUND METHODOLOGY

The sentiment analysis process, contains some basic modules like Input Data, Pre-Processing Part, Feature Extraction, Classification, etc.

A) Input Data: Sentiment analysis contains two types of input data.

i) Online Data: This can be done by calling API. If you want to retrieve Twitter online data then simply call Twitter API. After collecting data successfully generate Token. Online data contains one disadvantage, i.e. it works only for less than 100 Tweets.

- ii) Offline Data: In Offline data, extract data from the past. Offline data contain one advantage i.e. it works for more than 100 Tweets.
- **B) Pre-Processing Stage:** Before applying the algorithms on the input, pre-processing on the text is done. These transform the raw input into another format which is easy and effective for

processing. There are various methods for preprocessing data such as Cleaning in which it deals with punctuation, stop words, repeated letters, capitalization, etc. After collecting the input data, the pre-processing stage comes into the picture. The Pre-processing stage is used to remove noisy data from the input text. In Preprocessing stage, eliminate deafening data from the input text, follows some basic rules like,

- i) Lower case conversion: By using this rule, convert uppercase input text to lowercase input text. i.e. "CONTACT" "contact"
- ii) Remove of @ username and # value: Twitter users use @ character to indicate the exact user. Twitter users use the # character to raise tweet's visibility. In the Pre-processing stage, eliminate @ and # symbols from the input text. i.e. #happy happy
- **iii)URLs are removed:** In the Pre-processing stage, eliminate the URL part from the input text, because the URL will not contain many informational data. E.g. remove www., http://, and https:// content.
- iv)Tweets/ Documents are split into smaller text (like Tokenization method): In Tokenization method,

e.g. "We are going on vacation. I'm very excited." Each text is a list of sentences.

e.g. [{'We are going on vacation.'}, {'I'm very excited.'}] Each sentence is a list of Tokens [11].

e.g. {'We', 'are', 'going', 'on', 'vacation', '.'}, {'I', ''m', 'very', 'excited'} each token is a tuple of three elements: -

A word form: The exact word that appeared in the text.

A word lemma: A generalized version of the word.

A list of associated tags: There are many types of tags available like[11]

CC-Conjunction-and, or, but, neither, or, etc.

NN - Noun-apple, bat, ball, etc. PRP - Pronoun-he, she, it, etc. JJ - Adjective-great, best, beautiful, etc. WP - Wh-pronoun-who, when, what, which, where, etc. e.g. Tuple of these attributes, ('excited',

e.g. Tuple of these attributes, ('excited', 'excited', 'VB')

- v) Removing of common English words: Nowadays people on social media do not strictly follow grammar. They use different spells and shortcuts to represent their emotion as using words like ohhhh, wowww, cooool, etc. They will write things such as "I likeee it" to word "like". emphasize the However. computers don't understand that "likeee" is one of the variants of "like" so they must be told. This method removes these annoying repeating characters to end up with an appropriate meaningful word in the English Dictionary [11].
- vi) Removing punctuation symbol from the text: We want interesting keywords from the given input on which processing can be done. The punctuations are uninteresting tokens in our input structure which has to be removed. One way would be to split the input into words by white space, then use string translation to replace all punctuation with nothing [11]. Remove punctuation symbol from the text, because punctuation symbol doesn't contain any emotion. e.g. ?, :, ', ", !, ; etc.
- vii) Stop Word: The main goal of preprocessing is to eliminate unwanted word which does not have any importance in an application such as search queries in the search engine. Stop words are these unwanted words that just occupy unnecessary space in the database and increases processing time. These stop words vary from system to system. Following are some of the stop words in the English language: 'some', 'against', 'at', 'can', 'these', 'ourselves', 'because', 'from', "wasn't", 'theirs', 'is', 'very', 'just', etc.

We will remove these stop words to save valuable space and time. They can be easily *Published by: The Mattingley Publishing Co., Inc.* removed by storing these stop words then ignoring such stop words when encountered [11].

viii) Removing all non-English words: Remove all non- English words from the text. E.g. Removing Hindi, Gujarati words from the text and generate proper English text.

C) Feature Extraction: Feature Extraction Contains Two Steps:

i) Score Calculation using Dictionary: After getting meaningful words, assign particular score value to the words.



ii) No of Positive/Negative Dictionary: After assigning positive, a negative score to each word, calculate total no of positive and negative values. i.e., the summation of each positive/negative score.

Table 1: Comparative analysis of feature extraction[21]

Name	Method	Advantage	Limitation
Feature	Unigram Method [1]	-Unigram model extracts the adjective and segregates it. Example: "Driving Happy" is equivalent to "Happy"[21]	-Not supported particular Characters. Example: "-):" is Equivalent to Happy[21]
[21]	Negation Method [2, 4, 5]	-The form of negative words may modify the attitude direction in sentence. Example: "not happy" is equal to "sad"[21]	-Only Work For Negative Words[21].

D) Classification: There are two types of Classifications available:

i) **Classification:** In Training Classification, Check "whether the tweet is positive or negative." In Training Classification, Train the Model.



ii)Testing Classification: After Training, Testing can be conducted. In Testing Classification, Test the Model.

E) Dictionary

i) AFINN Dictionary:

AFINN is a listing of English words that has valence with an integer between -5 (negative) and +5 (positive).

There are two versions: **AFINN-111:** This is a latest edition with 2477 words and phrases [13]. **AFINN-96:** This dictionary contains 1468 unique words and phrases on 1480 lines. Note that it contains 1480 lines and some words are listed twice [13].

ii) Lexicon Dictionary:

Lexicon: the statement "lexicon" infers starting with those Greek (lexicon). Lexicon Dictionary contains 3382 words in the dictionary. Its score is between -5 and +5 as an affine dictionary [13]. The aggregated sum of these scores determines the sentiment behind the text [13].

IV. PROPOSED SYSTEM



Fig 1: Proposed Block Diagram

In our framework, we have proposed a procedure that is separated into various stages as appeared in Figure 1. The five phases are as per the following:

- 1) Collection of tweets
- 2) Pre-preparing
- 3) Feature Extraction
- 4) Classification

5) Predictive Analysis dependent on Location"

Step 1:- Retrieving Tweets using twitter API.

Step 2:- Apply Preprocessing Remove Special Characters () Hash tags () URL () Common English Words () Repeated Words () Step 3:- Replace emoticons into word :) – happy, etc.

"Step 4:- Replace synonyms words with its dictionary words

Glad – Happy, etc. Step 5:- Apply N-gram Model Sentence to Word break () for i=1:all pair If one_pair()=Ø or value else If two_pair ()=Ø or value else If three_pair= Ø or value else score=0 or max(score) end; End Step 6:- Classify using Multi Level Classification["]"

Flow of the Proposed System

As demonstrated in the figure, the methodology to extract the sentiment contains the several steps that are described below:

1. Collection of tweets

The Twitter Search API is a committed API for organization requirements beside the continuous record generally tweets. Every appeal will return up to in more than 100 tweets, for a singular question. For our structure, we collected our dataset by directing the Twitter API and making usage of word spotting in perspective on the occasion of the word we are scrutinizing the late tweets [15]."

2. Pre-Processing

Twitter information is formless information. It should be handled before it very well may be utilized. This makes information simpler to process in the later stages.

The method for pre-preparing comprises of the accompanying advances:

1) Remove all non-English Tweets.

2) Convert all of the tweets assembled to the lower case.

3) Remove the URLs – killed all string that depicts relations or hyperlinks current in the tweets."

4) Replace any usernames acquaint in the tweets with @username – removed the username and considering the way that these are not considered for estimations.

5) Convert the hash marks to common words since hash names can give some steady information, so it is significant to restore them with the same word without the hash. For example #Happy displaced with Happy.

6) Removing any unnecessary characters, extra spaces, etc. [15]"

3. Models: Utilization of Negation Method

Published by: The Mattingley Publishing Co., Inc.



The nearness of negative words may change the end presentation like not happy is corresponding to hopelessness [15]."

Use of Unigram Model

The Unigram model concentrates the descriptive word and isolates it. For the above model, for example, "Driving Happy" throughout the unigram model, just Happy is extricated from the sentence [15].

4. Proposed System Focus on Two Main Features:

i) Emoticons Features: Tweets messages containing emojis are recovered by utilizing the Twitter APIs ("Data Collection stage") and afterward they are gathered into two sets: positive set and negative set, containing positive © and negative © emojis separately ("Positive/Negative Emoticon Class Creation stage")."

Table 2. Emoticons icons and meaning	Table 2:	Emoticons	icons	and	meaning
--------------------------------------	----------	-----------	-------	-----	---------

ICON	MEANING
:-) =) :) 8) :] =] => 8-) :-> :-] :'') :')	smiley
:3 :> :^) :-3 => :-> :-V =v :-1	happy face
$^{^{^{^{^{^{^{^{^{^{^{^{^{^{^{^{^{^{^{$	happiness
(*)_*	kiss, couple kissing
:-) :) :-] :] :-> :> %-}	wink, smirk
<3	heart
:-D :D =D :-P =3 xD	laughing
:P =P	tongue sticking out, playful
0.0 0.0	surprised
IV IV	gape
B) B-) B 8	feel cool
('-)	tears of happiness
:1	exclamation
-X	Sealed lips, wearing braces
=* :-* :*	kiss

ii) Synonym Features: It takes all equivalent words of word and searches out with the lexicon and afterward gives the outcomes. Model: Good = handy, plunge, pleasant and so forth. Supplanting at least two rehashing letters in a tweet by two letters of the equivalent in tweets, some of the time clients rehash letters to pressure the feeling or sentiments. Words with reiterated letters, for example, "coooool", are ordinary in tweets and people will in general use in this manner to state their inclinations. For example, "coooool" is superseded by "cool". e.g. hunggrryyy, huuuuuungry for "hungry". e.g. Happpyy, Happyyyyyy for "Happy".

5. Classification

We have used two coordinated classifiers: Support Vector Machines, Naive Bayes and Random Forest that model the probability data being in a particular class by the grouping of things. (Positive and negative) [15]."

Classification Techniques: 1) Support Vector Machine (SVM) [13]

SVM is the greatest separator for the particular classes.

+ representing data points of type 1, and

- representing data points of type –1

SVM has been used effectively in many real-world problems

- Text categorization
- Image classification
- Bioinformatics (Protein classification, Cancer classification)
- Hand-written character recognition"

Text data are perfectly suitable for SVM classification because of the spare nature of the text, in which few features are unrelated, but they tend to be connected with one another and generally prepared into linearly separable categories.



Fig 2: support vector machine (SVM)

2) Naive Bayes (NB) [12]

There are two first request probabilistic models for Naïve Bayes characterization are Bernoulli model and the Multinomial model. The Bernoulli model is a Bayesian Network with no word conditions and parallel word highlights, or nonattendance; along these lines how the Bernoulli model likewise considers words that don't show up in the report [12]. The Multinomial model is a unigram language model with whole number word checks. Credulous Bayes Classification Model is anything but difficult to translate and give proficient calculation yet Assumptions of properties being free, which may not be fundamentally substantial.

Likelihood Class Prior Probability



Posterior Probability Predictor Prior Probability

 $P(c | X) = P(x_1 | c) * P(x_2 | c) *....* P(x_n | c) * P(c)$

(2)

Above,

"P(c|x) is the posterior probability of class (c, target) given predictor (x, attributes)."

"P(c) is the prior probability of class."

"P(x|c) is the likelihood which is the probability of predictor given class."

"P(x) is the prior probability of predictor."

3) Random Forest (RF)

It is a supervised classification Algorithm. This algorithm creates a forest with numerous Trees.

In the Random Forest Classier, the superior the numeral of trees in the forest gives the high accuracy results.

The similar random forest algorithm or the random forest classifier can use for both classification and the regression task. Random Forest Classifier will handle the absent values.

When we have more trees in the forest, a random forest classifier won't over fit the form.



N-Gram Model Example

Table 3.	Example	of N-Gram	Model
	Example	01 IN-Ofain	MOUCI

Grams	"This is not very good feature"	Scor
		e
Unigram	"This", "is", "not", "very", "good",	2
S	"feature"	
[1]	0 0 -1 0 3	
	0	
Bigrams	"This is", "is not", "not very", "very	4
[8]	good",	
	0 0 0 4	
	"good feature"	

Published by: The Mattingley Publishing Co., Inc.

	0	
Trigrams	"This is not", "is not very", "not very	-4
[10]	good",	-
	0 0 -4	
	"very good feature"	
	0	
N-gram	-Check in sequence Unigram, Bigrams	Trigra
(Propose	and Trigrams score.	m
d)	-Condition 1: if unigram score is not	-4
	null and trigram is null consider	
	unigram score.	
	-Condition 2: if unigram, bigram score	
	is not null and trigram is null consider	
	bigram score.	
	-Condition 3: if unigram, bigram and	
	trigram score is not null consider	
	trigram score.	





Fig 4: SVM Classification

As shown in the figure 4 Support vector machine classification use to classify airlines dataset. The dataset include positive words so our result is good.



Fig 5: NB Classification

As shown in the figure 5 Naiver Bayer's classification use to classify airlines dataset. The



dataset include positive words so our result is good but accuracy is low as compare to SVM.



Fig 6: RF Classification

Table 5: Table of Result Analysis

	SVM			NB			RF		
Datasets	Acc	Р	R	Acc	Р	R	Acc	Р	R
Dataset 1	75	79.33	75	70	76.66	70	95	96	95
Dataset 2	80	81.33	80	70	73.33	70	95	96	95
Dataset 3	80	84	80	70	73.33	70	95	96	95
Average	78.33333	81.55333	78.33333	70	74.44	70	95	96	95

For the table 5 it can be said that the Accuracy, Precision and Recall for Random Forest (RF) classifier is batter that all the classifier.



rating, upcoming events like sports or political election, etc. It also generates a Geo-graphical vision (view) of the particular event. In this way, analyze location-based event results e.g. analyze political election results movie reviews based on location. The work is completed by making use of Twitter Reviews. The work deals with real-time and location. The system uses the emoticon feature to calculate the actual rating or popularity of any product. It also gives the perfect or actual rating of any product or thing which will be very helpful to the customers or to the businessman to do the changes in their products.

extend it into an efficient current affair based on

location. Current affairs include product or movie

REFERENCES

[1] Bouazizi, Mondher, and Tomoaki Ohtsuki. "A pattern-based approach for multi-class sentiment

VI. FUTURE WORK AND CONCLUSION

In this research paper mainly focus on an efficient product review based on location. In Future,

As shown in the figure 6 Random Forest classification use to classify airlines dataset. The dataset include positive words so our result is good but accuracy is high as compare to all.



analysis in Twitter." *IEEE Access* 5 (2017): 20617-20639.

- [2] Ahuja, Shreya, and Gaurav Dubey. "Clustering and sentiment analysis on Twitter data." In 2017 2nd International Conference on Telecommunication and Networks (TEL-NET), pp. 1-5. IEEE, 2017.
- [3] Algur, Siddu P., and Rashmi H. Patil. "Sentiment analysis by identifying the speaker's polarity in Twitter data." In 2017 International Conference on Electrical, Electronics, Communication, Computer, and Optimization Techniques (ICEECCOT), pp. 1-5. IEEE, 2017.
- [4] Bhan, Namrata, and Mitchell D'silva. "Sarcasmometer using sentiment analysis and topic modeling." In 2017 International Conference on Advances in Computing, Communication and Control (ICAC3), pp. 1-7. IEEE, 2017.
- [5] Kumari, Upma, Arvind K. Sharma, and Dinesh Soni. "Sentiment analysis of smart phone product review using SVM classification technique." In 2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS), pp. 1469-1474. IEEE, 2017.
- [6] Lavanya, K., and C. Deisy. "Twitter sentiment analysis using multi-class SVM." In 2017 International Conference on Intelligent Computing and Control (I2C2), pp. 1-6. IEEE, 2017.
- [7] Mahmud, Quazi Ishtiaque, Asif Mohaimen, and Md Saiful Islam. "A Support Vector Machine mixed with statistical reasoning approach to predict movie success by analyzing public sentiments." In 2017 20th International Conference of Computer and Information Technology (ICCIT), pp. 1-6. IEEE, 2017.
- [8] Singla, Zeenia, Sukhchandan Randhawa, and Sushma Jain. "Sentiment analysis of customer product reviews using machine learning." In 2017 International Conference on Intelligent Computing and Control (I2C2), pp. 1-5. IEEE, 2017.
- [9] Haque, Tanjim Ul, Nudrat Nawal Saber, and Faisal Muhammad Shah. "Sentiment analysis on large scale Amazon product reviews." In 2018 IEEE International Conference on Innovative Research and Development (ICIRD), pp. 1-6. IEEE, 2018.
- [10] Jianqiang, Zhao, and Gui Xiaolin. "Comparison research on text pre-processing methods on twitter sentiment analysis." *IEEE Access* 5 (2017): 2870-2879.

- [11]Gosai, Dhruvi D, "A Review on a Emotion Detection and Recognization from Text Using Natural Language Processing." (2018).
- [12]Dharmesh Ramani and Hazari Prasun, "Sentiment Features based Analysis of Online Reviews" IOSR Journal of Computer Engineering (IOSR-JCE) e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 17, Issue 3, Ver. V (May – Jun. 2015), PP 82-86
- [13]Ms. Rekha K Dimke; , Mr. Aniruddh Fataniya.
 SOCIAL MEDIA E-CUSTOMERS' BEHAVIOUR MINING. International Education and Research Journal, [S.l.], v. 3, n. 5 , jun. 2017.
 ISSN 2454-9916. Available at: <http://ierj.in/journal/index.php/ierj/article/view/10 72/1032>. Date accessed: 13 Apr. 2020.
- [14]"Twitter sentiment analysis of current affairs".
 [Online]. Available: https://www.slideshare.net/editorijritcc1/twittersentiment-analysis-of-currentaffairs?from_action=save. [Accessed: 13- Feb-2020].
- [15] Ms. Sujal Shah and Ms. Reshma Lakhani, "A Survey Paper on Product Review based on Geographic Location using SVM approach in Twitter", IJSRD - International Journal for Scientific Research & Development| Vol. 4, Issue 10, 2016 | ISSN (online): 2321-0613
- [16] Zala, Dhruvi K., and Ankita Gandhi. "A Twitter Based Opinion Mining to Perform Analysis Geographically." In 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI), pp. 59-63. IEEE, 2019.
- [16] "Twitter", Twitter.com, 2020. [Online]. Available: https://twitter.com/. [Accessed: 13- Apr- 2020].
- [17] Developer.twitter.com, 2020. [Online]. Available: https://developer.twitter.com/en/apps/16808251.[Accessed: 13- Apr- 2020].
- [18]"Analyzing Twitter with MATLAB", Loren on the Art of MATLAB, 2020. [Online]. Available: https://blogs.mathworks.com/loren/2014/06/04/ana lyzing-twitter-with-matlab/. [Accessed: 13- Apr-2020].
- [19] "Geocode Finder Lookup Latitude and Longitude | Map Developers", Mapdevelopers.com, 2020. [Online]. Available: https://www.mapdevelopers.com/geocode_tool.php. [Accessed: 13- Apr- 2020].

Published by: The Mattingley Publishing Co., Inc.



[20]"Smiley Face and Emoji Meanings - Webopedia Guide", Webopedia.com, 2020. [Online]. Available: https://www.webopedia.com/quick_ref/textmessag

eabbreviations_02.asp. [Accessed: 13- Apr- 2020].

[21] Varma, Kajal S., Arpana Mahajan, and Sheshang D. Degadwala. "Product Reviews based on Location using N-gram model." In 2018 3rd International Conference on Inventive Computation Technologies (ICICT), pp. 100-104. IEEE, 2018.