parunnkumar@gmail.com.docx

by Parunnkumar@gmail.com.docx Last

Submission date: 01-Jul-2021 07:13AM (UTC+0500)
Submission ID: 1614392970
File name: parunnkumar_gmail.com.docx (793.94K)
Word count: 4098
Character count: 23030

Twitter Sentiment Analysis

Ramachandran [1], Arunnkumar [2], Balachander [3]

[1] *Associate Professor, Department of Information Technology, Sri Manakula Vinayagar EngineeringCollege, Puducherry, India*

[2,3] *Student, Department of Information Technology, Sri Manakula Vinayagar EngineeringCollege, Puducherry, India*

Abstract

Twitter is one of the most used applications by the people to express their opinion and show there sentiments towards different occasions.In the era of accelerating social media users, Twitter has a large number of regular users who post their views in the form of tweets. This paper proposes a method for collecting feelings from tweets as well as a method for categorizing tweets as positive, negative, or neutral. Any business that is listed or tagged in a tweet will benefit from this strategy in a number of ways. Since most tweets are in an unstructured format, they must first be translated into a structured format. Tweets are resolved in this paper using a pre-processing phase, and tweets are accessed using libraries that use the Twitter API. The datasets must be trained using algorithms in such a way that they are capable of checking tweets and extracting the requisite sentiments from the feed. The goal of this research is to give an model to this fascinating problem and to present a framework which will perform sentiment analysis associating XGBOOST Classification. Furthermore, with the recent advancements in machine learning algorithms, the accuracy of our sentiment analysis predictions is able to improve. The main objective of this paper is to perform real time sentimental analysis on the tweets that are extracted from the twitter and provide time based analytics to the user associating XGBOOST and Natural Language Processing Classification Techniques.

***Key Words*:** Sentiment Analysis, Twitter, XGBOOST, opinion mining.

1. INTRODUCTION

This research, which analyses tweet sentiments, falls under the categories of "Pattern Classification" and "Data Mining." Both of these concepts are closely linked and interconnected, and they can be formally identified as the method of identifying "useful" trends in large amounts of data. The project will largely depend on "Natural Language Processing" techniques for extracting significant patterns and features from a big corpus of tweets, as well as "Machine Learning" techniques for precisely identifying different unlabeled data samples (tweets) with whatever pattern model best describes them.

10

There are two types of features that can be used for modelling patterns and classification: formal language-based features and informal blogging-based features. Prior intention polarity of individual terms and phrases, as well as parts of speech tagging of the sentence, are examples of language-based features. Prior sentiment polarity defines the inherent propensity of certain terms and phrases to convey similar and related feelings in general. The word "excellent," for example, has a positive social connotation, while "evil" has a strong negative connotation. When a word with a positive connotation appears in a sentence, the whole sentence is likely to express a positive emotion. On the other hand, Parts of Speech tagging is a syntactical solution to the issue. It means determining which part of speech each individual word in a sentence belongs to, such as noun, pronoun, adverb, adjective, verb, interjection, and so on.

Patterns can be derived by studying the frequency distribution of these parts of speech in a specific class of labelled tweets (either separately or in combination with another parts of speech). Twitter-based features are more casual and relate to how people express themselves on online media networks and compact their feelings into the limited 140-character space provided by Twitter. Twitter hashtags, retweets, term capitalization, word lengthening, question marks, URL presence in tweets, exclamation marks, internet emoticons, and internet shorthand/slangs are only a few examples.

Furthermore, adaptive classification approaches struggle with environmental input. In our context, environmental feedback can take the form of a person asking the classifier whether it did a good or bad job classifying a specific tweet, and the classifier would benefit from this feedback. Passive and responsive adaptive strategies are the other two forms of adaptive techniques

LITERATURE SURVEY

6

Abdullah Alsaeedi, et al.[1] offered a study in which people's reactions to buying a product, public services, and other topics are gathered and analysed. Sentiment analysis (also known as opinion mining) is a typical debate preparation activity that seeks to extract the emotions behind opinions expressed in texts on a variety of topics. Twitter is a popular microblog where customers may express themselves. Opinion analysis of Twitter data is a discipline that has gotten a lot of interest in the recent decade, and it entails examining "tweets" (comments) and thus the substance of those expressions. This study looked at a variety of strategies for analysing Twitter sentiment, including machine learning, ensemble approaches, and dictionary (lexicon) based approaches. Hybrid and ensemble strategies for analysing Twitter sentiment were also investigated. The findings of the study showed that machine learning approaches are effective. The addition of emoticons in Twitter sentiment enhanced categorization accuracy from 79 percent to 85 percent, according to the findings.

The fundamental subtask of opinion summarization is stated in the work proposed by **Gamgarn Somprasertsri, et al**[2]. For mining opinions from online customer evaluations, a dependency and semantic-based approach is proposed, with a focus on extracting relationships between product characteristics and opinions. Because the effectiveness of feature and opinion extraction influences the performance of opinion orientation identification, it is a key task. It's critical to understand the semantic links between product features and customer opinions. This method extracts product features and opinions based on syntactic and semantic information. The approach has been applied to dependencies and ontological knowledge using a probabilistic based model, and the results of the trials reveal that the approach is more flexible and successful, with an accuracy rate of 72.26 percent.

Andrea Esuli [3] introduced a new approach for identifying the orientation of subjective words. The method is based on a quantitative analysis of such terms' glosses, which includes definitions from online dictionaries, and the use of the resulting term representations for semi-supervised term classification applied to term representations obtained by using term glosses from a freely available machine readable dictionary fication. This research yielded an accuracy of about 83.02 percent.

Alec Go et al[4] proposed a study that used distant supervision to classify the sentiment of Twitter posts using machine learning techniques. Tweets with emoticons are utilised as noisy labels in the training dataset. This kind of training data may be obtained using automated methods. When taught on emoticon data, the machine learning algorithms Naive Bayes, Maximum Entropy, and SVM have an accuracy of above 80%. The preprocessing procedures required for high accuracy are described in this work. The use of tweets with emoticons for remote supervised learning is the paper's primary contribution.

ArkaitzZubiaga, et al.[5]proposed that the analyse and experiment with a set of simple language-independent features that rely on the social spread of trends to discriminate among those types of trending topics and provides an efficient method without the need for external data to quickly and accurately categorise trending topics, as well as enabling news organisations to track and discover breaking news in real-time, or immediately identifying viral memes that might enrich marketing decisions. The analysis of social features observed in social trends reveals social patterns that have been associated with each type of trend, such as tweets about ongoing events being shorter because many of the tweets were likely sent from mobile devices, or memes having more retweets than other types of trends because they originated from fewer users. Memes made by a huge number of a celebrity's or group's followers. In this typology, these unique memes may be classed as a subclass of memes, which can then be further subdivided into more specific fandom memes based on the sort of fan.

PimwadeeChaovalit,et al [6] suggested a study in which online content mining is used to assist individuals find useful information in vast amounts of unstructured material on the internet. Movie review mining separates good and negative reviews into two categories. Movie review mining differs from other classifications, according to sentiment-based categorization, and empirical research have been done in this sector. The results demonstrate that the results are equivalent to or even better than earlier findings. This work determines movie review mining utilizing two techniques: machine learning and semantic orientation, and the approaches have been fitted to the movie review domain for comparison**.** The data is always combined with real-life review data, and sarcastic phrases are utilised in composing movie reviews, which makes movie review mining difficult. The use of bag-of-words features in supervised learning approaches is problematic due to the scarcity of words in movie reviews. After changing the dividing baseline, there is a 77 percent categorization accuracy on 100 movie reviews from Movie Vault.

B. Pang et al. [7] established a strategy for identifying subjectivity in sentiment analysis. This is necessary since the audits' irrelevant material may be discarded. This removes the overheads associated with processing huge amounts of data. To produce subjective extracts from the information, the method employs minimal cut. The study has been focused on extracting subjectivity at the sentence level.

J. Wiebe[8] presented the Naive Bayesian classifier. They demonstrated the results of training subjectivity classifiers using unclear material. The approach has been devised training information as a result of this work of learning Subjective and Objective phrases. A rule-based technique is used to achieve this. When a sentence contains two or more subjective guesses, the rule-based subjective classifier labels it as subjective. The principle-based target classifier, on the other hand, looks for intimations that aren't present: It classifies a sentence as a target. There is at least one solid subjective educate the past and next sentence consolidated, and at most two weak subjective educate the present, past, and next sentence consolidated classifiers if there are no solid subjective that enlighten the current sentence. For additional evaluation, they use the Subjective Precision, Subjective Recall, Subjective F measure, Objective Precision, Objective Recall, and Target F measure.

SNO	TITLE	METHODS	PROPOSED TECHNIQUE	ADVANTAGES	DISADVANTAGES	ACCURACY RATE
1	Creating Subjective and Objective sentence Classifiers from Unannotated texts," in Computational Linguistics and Intelligent Text Processing.	Supervised	. Microtext Analysis, . SubjectiveDetection, . SemanticParsing, . Anaphora Resolution	. Very good at retrieving texts,Checking the spelling and counting words . Overall pos and neg can be detected right after their occurrence . Use Microblogging to manage interaction between Human and Robots	. Does not have best performance in evaluation . Capabilities are limited	. This Work yields good Flexibility . Establishes good Accuracy
2	Mining Feature-opinion in online Customer Reviews for Opinion Summerization	Unsupervised	. Dependency Relation forFeatureOpinion, .RelatedOpinionExtraction, . Product Ontology	. Useful for Customers and Manufactures . Written in Natural Language . More Efficient . More Flexible and Effective .Rapid Growth	.Unsatisfactory Performance .Not Practically Implemented .Extraction is Critical .Unstructured Data	Yields a Flexible and Effective result that provides a accuracy rate of 72.26%.

3	A Clusterig based Approach on Sentimental Analysis	Unsupervised	. Machine Learning . Artificial Intelligence	. Effectively utilize data Technologies . Accessible Resources,BenchmarkDatasets,Evaluation Campaigns are Provided	Long term issues Vocabulary of Natural Language makes the opinion Mining Hard	. This Work yields good Flexibility . Establishes good Accuracy
4	Twitter Sentimental Analysi	Supervised	. Naïve bayes . SVM . Lexicon based approach	.Data called noisy lables are abundantly available .Accuracy above 80% .With twitter api easy to extract large amount of data	.Tweets are limited .No previous Research provided .Very difficult to manually collect data	The machine learning algorithms (Naive Bayes, Maximum Entropy, and SVM)have accuracy 80% when trained withemoticon data.
5	Twitter Sentiment Classification using Distant Supervisor	Unsupervised	Nlp(natural language processing)	. Higher accuracy is provided .It lows the feedback to be aggregated without manual intervention	. This process requires lots of Computation power and time . Further message have unique characteristics difficult to classify	. This Work yields good Flexibilty . Establishes good Accuracy
6	Dissatisfaction oriented advertising based on Sentimental Analysis	Supervised	. Sentimental Analysis, . Sentatic Passing	. Improve user experience . Design a prototype for product information	. They cannot be able to classify the tweets Automatically	. This Work yields good Flexibilty . Establishes good Accuracy
7	Determining the semantics orientation of	Supervised	.Opinion mining,	. It is based on quantitative analysis	. This method had outperformed	The accuracy obtained

	terms through Gloss Classification		.Text Classification, .Polarity Detection	. It has increased effectiveness and efficiency of text classifiers	all the published methods . It does not withstand even a small margin	through this work is nearly 83.02%.
8	Markov blankets and Meta Heuristic search	Unsupervised	.Sentimental Analysis, .Meta Heuristic Search	. Able to capture the Dependencies among words . Finds the vocabulary that is efficient for purpose of extracting sentiments	. It has word selection problem . It gives unstructured data	. This Work yields good Flexibility . Establishes good Accuracy

3.1EXISTING SYSTEM

The existing system only works with datasets that are restricted to a single subject. Existing systems often do not assess the magnitude of effect that the calculated outcomes will have on the specific area under consideration, and they do not enable retrieval of data based on a user-initiated query, i.e. they have a limited reach. To put it another way, it works with static rather than dynamic data. Unsupervised algorithms, such as Vector Quantization, are used for data compression, pattern recognition, facial and speech recognition, and other applications and thus cannot be used to evaluate sentiment in twitter data. Since the algorithm struggles with large datasets, it can produce erroneous results.

3.2 DRAWBACKS OF EXISTING SYSTEM:

1. It is not applicable to find continuous time shots present in tweets.
2. It does not give the minimal set of tweets which are relevant to an event.
3. It cannot provide summary of unknown events which cannot be predicted.
4. Noises and background topics cannot be eliminated.
5. Tweets are modeled as binomial mixture where tweets in which most words belong to general topics are considered as general tweets and tweets in which most words belong to specific event as specific tweets. It is totally unreasonable for tweets having short lengths.

3.3 PROPOSED SYSTEM

Our aim is to perform sentiment analysis using data from Twitter. We're going to make a classifier out of a variety of machine learning classifiers. We'll proceed with the steps until our classifier is ready and educated.

Step-1 First we are going to stream tweets using Twitter Streaming API in our build classifier with the help of TextBlob library in python

Step-2 Then we pre-process these tweets, so that they can be fit for mining and feature extraction

Step-3 The preprocessed data is involved in tokenization, normalization. Since, Twitter is our source of data for analysis. We are going to stream the tweets from twitter in our database with NLTK Library. For this we are going to use Twitter Application.

Step-4 Then passed to our trained XGBOOST classifier, which then classify them into positive or negative class based on trained results.

Step-5 The classified data is inferred into various charts and analysis are made. It also allows to generate report of their classified data and files.

ARCHITECTURE DIAGRAM

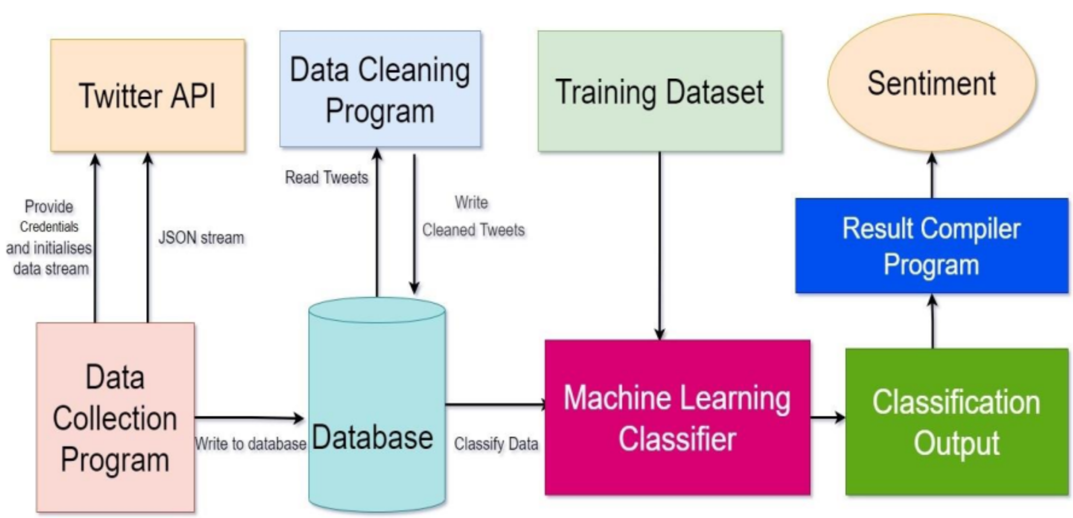


Figure3.3.1

OVERALL ARCHITECTURE

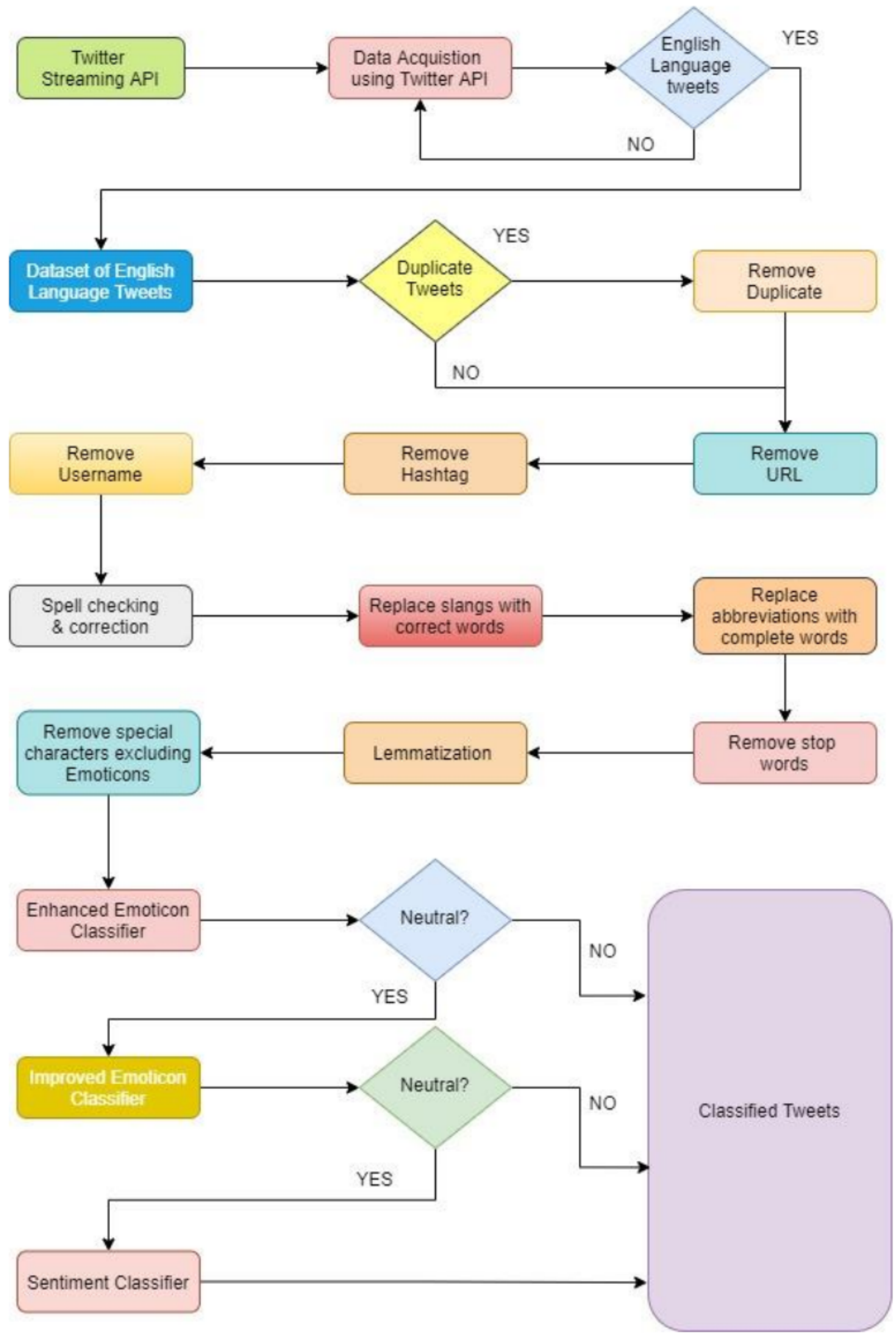


Figure 3.3.2

XGBOOST PSEUDOCOD

Input: Training data $\mathcal{D} = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$
List of n first- and second-order gradients: $\{(g_1, h_1), \ldots, (g_n, h_n)\}$
Regularization parameters λ and γ; Maximum depth of tree, D

Output: Optimal tree structure q_t minimizing the simplified objective $\tilde{\mathcal{L}}^{(t)}$.

$I_1 \leftarrow \{1, \ldots n\}$ ▷ Initialize the tree structure with a single node
$m \leftarrow 1$ ▷ Number of nodes
$d \leftarrow 0$ ▷ Depth level
$(S, S') = (\{1\}, \emptyset)$ ▷ Set of leaf nodes for current/next level
while $d \leq D$ **do**
  for all $j \in S$ **do**
    for feature a in $\mathcal{D}$ **do**
      $(G, H) \leftarrow \sum_{i \in I_j} (g_i, h_i)$
      Let $F_a(\mathbf{x})$ denote the value of $\mathbf{x}$ for feature a.
      $M_a \leftarrow \{i \in I_j : \mathbf{x}_i \text{ has no value for feature } a\}$ ▷ Provision for missing values
      $(G_a^{\text{n/a}}, H_a^{\text{n/a}}) \leftarrow \sum_{i \in M_a} (g_i, h_i)$
      $V \leftarrow \{v : F_a(\mathbf{x}) = v \text{ for some } \mathbf{x} \in \mathcal{D}\}$
      for all $v \in V$ **do** ▷ Consider each threshold candidate
        $L_{a,v} \leftarrow \{i \in I_j : F_a(\mathbf{x}_i) < v\}$ ▷ Split I_j using threshold v
        $R_{a,v} \leftarrow \{i \in I_j : F_a(\mathbf{x}_i) \geq v\}$
        $(G_{a,v}^{\text{left}}, H_{a,v}^{\text{left}}) \leftarrow \sum_{i \in L_{a,v}} (g_i, h_i)$ ▷ Compute gradient sums
        $(G_{a,v}^{\text{right}}, H_{a,v}^{\text{right}}) \leftarrow \sum_{i \in R_{a,v}} (g_i, h_i)$

$$\Delta\tilde{\mathcal{L}}^{(t)}(a, v, T) \leftarrow -\frac{1}{2}\left[\frac{G^2}{H+\lambda} + \frac{\left(G_{a,v}^{\text{left}} + G_a^{\text{n/a}}\right)^2}{H_{a,v}^{\text{left}} + H_a^{\text{n/a}} + \lambda} + \frac{\left(G_{a,v}^{\text{right}}\right)^2}{H_{a,v}^{\text{right}} + \lambda}\right] - \gamma.$$

$$\Delta\tilde{\mathcal{L}}^{(t)}(a, v, F) \leftarrow -\frac{1}{2}\left[\frac{G^2}{H+\lambda} + \frac{\left(G_{a,v}^{\text{left}}\right)^2}{H_{a,v}^{\text{left}} + \lambda} + \frac{\left(G_{a,v}^{\text{right}} + G_a^{\text{n/a}}\right)^2}{H_{a,v}^{\text{right}} + H_a^{\text{n/a}} + \lambda}\right] - \gamma.$$

      end for
    end for
    $(a, v, b) \leftarrow \text{argmax}_{a,v,b}\, \Delta\tilde{\mathcal{L}}^{(t)}(a, v, b)$ ▷ Choose (a, v, b) that minimizes $\tilde{\mathcal{L}}^{(t)}$
    if $b = T$ **then** ▷ Commit the split using optimal (a, v, b)
      $(I_{m+1}, I_{m+2}) \leftarrow (L_{a,v} \cup M_a, R_{a,v})$
    else
      $(I_{m+1}, I_{m+2}) \leftarrow (L_{a,v}, R_{a,v} \cup M_a)$
    end if
    $S' \leftarrow S' \cup \{m+1, m+2\}$ ▷ Mark new nodes as leaves
    $m \leftarrow m + 2$
  end for
  $d \leftarrow d + 1$ ▷ Done with current level
  $S \leftarrow S'$
end while
return tree structure q defined by instance sets $I_1, \ldots, I_m$

4.IMPLEMENTATION

4.1 MODULE-WISE DESCRIPTION

4.1.1 DATA GATHERING

Twitter enables developers with a collection of streaming APIs that provide low-latency access to Twitter data flows. The public streams API was used for data collection; it was discovered that this was the best method of gathering information for data mining purposes because it provided access to a global stream of twitter data that could be filtered as needed. Twitter enables developers with a collection of streaming APIs that provide low-latency access to Twitter data flows. The public streams API was used for data collection; it was discovered that this was the best method of gathering information for data mining purposes because it provided access to a global stream of twitter data that could be filtered as needed. To use this stream, you'll need to install a Python interface package. This library is needed for Python to communicate with Twitter's API v1.1. There were several libraries available for this mission, but python twitter tools v1.14.3 was chosen because it provided the basic filtering and streaming features required for this project.

4.1.2 DATA PRE-PROCESSING

Data obtained from twitter is not fit for extracting features. Mostly tweets consists of message along with usernames, empty spaces, special characters, stop words, emoticons, abbreviations, hash tags, time stamps, URL’s ,etc. Thus to make this data fit for mining we pre-process this data by using various function of NLTK. In pre- processing we first extract our main message from the tweet, then we remove all empty spaces, stop words (like is, a, the, he, them, etc.), hash tags, repeating words, URL’s, etc. We then replace all emoticons and abbreviations with their corresponding meanings like :-), =D, =), LOL, Rolf, etc. are replaced with happy or laugh. Once we are done with it, we are ready with processed tweet which is provided to classifier for required results. A sample processed tweet is shown in Table 4.1.1

Table 4.1.1 Sample Tweet and Processed Tweet

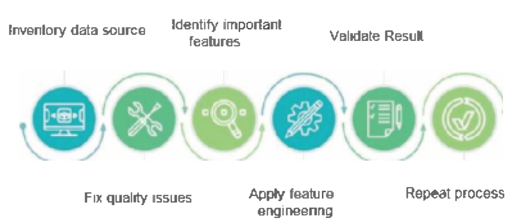
Tweet Type	Result
Original tweet	@xyz I think Kejriwal is a habitual liar, even where he don’t needs to lie he tells a lie >☹#AAP
Processed tweet	think, habit, lie, even, don’t, need, tell, angry

Cleaning of Twitter data is necessary, since tweets contain several syntactic features that may not be useful for analysis. The pre-processing is done in such a way that data represented only in terms of words that can easily classify the class.

We create a code in Python in which we define a function which will be used to obtain processed tweet. This code is used to achieve the following functions:

- ∑ remove quotes - provides the user to remove quotes from the text
- ∑ remove @ - provides choice of removing the @ symbol, removing the @ along with the user name, or replace the @ and the user name with a word 'AT_USER' and add it to stop words
- ∑ remove URL (Uniform resource locator) - provides choices of removing URLs or replacing them with 'URL' word and add it to stop words
- ∑ remove RT (Re-Tweet) - removes the word RT from tweets
- ∑ remove Emoticons - remove emoticons from tweets and replace them with their specific meaning
- ∑ remove duplicates – remove all repeating words from text so that there will be no duplicates
- ∑ remove # - removes the hash tag class
- ∑ remove stop words – remove all stop words like a, he, the, and, etc which provides no meaning for classification

STEPS FOR DATA PRE-PROCESSSING



4.1.3 DATA TOKENISATION

It is the process of breaking a stream of text up into words symbols and other meaningful elements called “tokens”. Tokens can be separated by whitespace characters and/or punctuation characters. It is done so that we can look at tokens as individual components that make up a tweet. Emoticons and abbreviations (e.g., OMG, WTF, BRB) are identified as part of the tokenization process and treated as individual tokens

CONTENT	**ACTION**
Punctuation (! ? , . ” : ;)	Removed
#word	Removed #word
@any_user	Remove @any_user or replaced with “AT_USER” and then added in stop words.
Uppercase characters	Lowercase all content
URLs and web links	Remove URLs or replaced with “URL” and then added in stop words
Number	Removed
Word not starting with alphabets	Removed
All Word	Stemmed all word (Converted into simple form)
Stop words	Removed
Emoticons	Replaced with respective meaning
White spaces	Removed

4.1.4 DATA NORMALISATION

The inclusion of abbreviations inside a tweet is noted for the normalisation process, and then abbreviations are replaced with their actual meaning. We also look for informal intensifiers like all-caps (e.g., I LOVE this show!!!) and character repetitions (e.g., I've got a mortgage!! happyyy) replaced with a single letter. Finally, any unique Twitter tokens are noted, and placeholders indicating the token type are replaced. This normalisation, we expect, will improve the output of the POS tagger, which is the final preprocessing phase.

Raw data	**Clean data**
@jackstenhouse69 I really liked it, in my opinion it def is :)	Really, liked, opinion, def
:(\u201c@EW: How awful. Police: Driver kills 2, injures 23 at #SXSW http://t.co/8GmFiOuZbS\u201d	Sad, awful, police, driver, kills, injures

4.1.4 DATA TRAINING

This dataset is used by a supervised learning algorithm to learn how to map the input examples to their predicted targets. The machine learning algorithm should be able to generalise the training data so that it can accurately map new data that it has never seen before if the training process is performed correctly.

Training data must have a class label; this can be accomplished by manually assigning a class to and tweet, but this is a time-consuming operation, and because Twitter has strict rules about the distribution of its data, finding accurate hand-annotated twitter datasets has proven difficult.

REVIEWS IN TWITTER	**CLASS**
foolish, idiotic and boring it's so lad dish and youngish , only teenagers could find it funny	NEGATIVE
the rock is destined to be the 21st century's new conan and that he's going to make a splash even greater than arnold schwarzenegger	POSITIVE
Barry Sonnenfeld owes frank the pug big time the biggest problem with roger avary's uproar against the map	NEGATIVE
the seaside splendor and shallow , beautiful people are nice to look at while you wait for the story to get going	POSITIVE

5. CONCLUSION

Our solution provides the best technique to predict the sentiment analysis with an efficiency of 95%. By predicting the sentiment, it would be beneficial for manufactures to in assuring the quality of the product. Machine-Learning Algorithm predicts and identifies the exact feedback/review of the user on a specific subject.

This paper presented a rigorous literature review of Sentiment Analysis including Twitter. It also presents a variety of opportunities and challenges for using opinion mining.

.In future, this model focuses to enhance the accuracy of prediction. The various approaches to sentiment analysis, primarily Machine Learning and Cognitive approaches, are discussed in depth in this study. It gives a comprehensive overview of the numerous applications and challenges that Sentiment Analysis can present, making it a difficult task.

REFERENCES:

[1] Andrea Esuli and FabrizioSebastiani, "Determining the semantic orientation of terms through gloss classification", Proceedings of 14th ACM International Conference on Information and Knowledge Management,pp. 617-624, Bremen, Germany, 2015.

[2] Bing xu, tie-jun Zhao, de-quanzheng, shan-yuwang, "Product features mining based on conditional random fields model, "Proceedings of the Ninth International Conference on Machine Learning and Cybernetics, Qingdao, 11-14 July, 2018.

[3] Chaovalit,Lina Zhou, Movie Review Mining: a Comparison between Supervised and Unsupervised Classification Approaches, Proceedings of the 38th Hawaii International Conference on System Sciences – 2015.

[4] Chau, M., &Xu, J. Mining communities and their relationships in blogs: A study of online hate groups. International Journal of Human – Computer Studies, 65(1), 57–70, 2017.

[5] Christopher Scaffidi, Kevin Bierhoff, Eric Chang, MikhaelFelker, Herman Ng and Chun Jin, "Red Opal: product feature scoring from reviews", Proceedings of 8th ACM Conference on Electronic Commerce, pp. 182-191, New York, 2017.

[6] Chunxu Wu, LingfengShen , "A New Method of Using Contextual Information to Infer the Semantic Orientations of Context Dependent Opinions" , 2019 International Conference on Artificial Intelligence and Computational Intelligence.

[7] Gamgarn`Somprasertsri, PattarachaiLalitrojwong , Mining Feature-Opinion in Online Customer Reviews for Opinion Summarization, Journal of Universal Computer Science, vol. 16,

no. 6 (2018), 938-955.

[8] Gang Li, Fei Liu. "A Clustering-based Approach on Sentiment Analysis", 2018, 978-1-4244-6793-8/10 ©2018 IEEE.

5

[9] Hila Becker, Dan Iter, MorNaaman, and Luis Gravano. 2012. Identifying content for planned events across social media sites. In Proceedings of the fifth ACMinternational conference on Web search and data mining. ACM, 533–542.

[10] Yan Chen, HadiAmiri, Zhoujun Li, and Tat-Seng Chua. 2013. Emerging Topic Detection for Organizations from Microblogs. In Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '13). ACM, New York, NY, USA, 43–52.

5

[11] Van Dang, Michael Bendersky, andW. Bruce Croft. 2013. Two-Stage Learning to Rank for Information Retrieval. In Proceedings of the 35th European Conference on Advances in Information Retrieval (ECIR'13). Springer-Verlag, Berlin, Heidelberg, 423–434

5

[12] MalekHajjem and ChirazLatiri. 2017. Combining IR and LDA Topic Modeling for Filtering Microblogs. ProcediaComput. Sci. 112, C (Sept. 2017), 761–770.

5

[13] JeffreyNichols, JalalMahmud, and Clemens Drews. 2012. Summarizing Sporting Events Using Twitter. In Proceedings of the 2012 ACMInternational Conference on Intelligent User Interfaces (IUI '12). ACM, New York, NY, USA, 189–198.

13

[14] Kate Starbird and LeysiaPalen. 2012. (How)Will the Revolution Be Retweeted?: Information Diffusion and the 2011 Egyptian Uprising. In Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work (CSCW '12). ACM, New York, NY, USA, 7–16.

parunnkumar@gmail.com.docx

ORIGINALITY REPORT

25%	29%	18%	25%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Jawaharlal Nehru University (JNU) Student Paper	6%
2	www.ijarcsse.com Internet Source	3%
3	tudr.thapar.edu:8080 Internet Source	3%
4	Submitted to National College of Ireland Student Paper	2%
5	export.arxiv.org Internet Source	1%
6	www.researchgate.net Internet Source	1%
7	Submitted to Higher Education Commission Pakistan Student Paper	1%
8	rcciit.org Internet Source	1%
9	ijcsit.com	

	Internet Source	1%
10	Submitted to Panjab University Student Paper	1%
11	www.ijptonline.com Internet Source	1%
12	esatjournals.net Internet Source	1%
13	www.adityavashistha.com Internet Source	1%
14	www.ijirst.org Internet Source	1%
15	ieeexplore.ieee.org Internet Source	1%
16	Phu Vo Ngoc, Chau Vo Thi Ngoc, Tran Vo Thi Ngoc, Dat Nguyen Duy. "A C4.5 algorithm for english emotional classification", Evolving Systems, 2017 Publication	1%
17	Submitted to University of Bristol Student Paper	1%
18	Submitted to Griffith College Dublin Student Paper	1%
19	Keerthana Pramudi Suresh, Siddhaling Urolagin. "Android App Success Prediction	1%

based on Reviews", 2020 International Conference on Computation, Automation and Knowledge Management (ICCAKM), 2020
Publication

20	Fatima-ezzahra LAGRARI, Youssfi ELKETTANI. "Customized BERT with Convolution Model : A New Heuristic Enabled Encoder for Twitter Sentiment Analysis", International Journal of Advanced Computer Science and Applications, 2020 Publication	<1 %
21	www.pantechsolutions.net Internet Source	<1 %
22	L. Zhou. "Movie Review Mining: a Comparison between Supervised and Unsupervised Classification Approaches", Proceedings of the 38th Annual Hawaii International Conference on System Sciences, 2005 Publication	<1 %
23	Submitted to University of East London Student Paper	<1 %
24	www.scribd.com Internet Source	<1 %
25	Submitted to Guru Gobind Singh Indraprastha University Student Paper	<1 %