Sentimental Analysis using Dictionary and Machine Learning Approach

Manish Dwivedi
Department of Computer Science and Engineering
GL Bajaj Institute of Technology and Management(GLBITM)
Greater Noida
manishclg640@gmail.com

Ishika Chaudhary
Department of Computer Science and Engineering
GL Bajaj Institute of Technology and Management(GLBITM)
Greater Noida
chaudharyishika67@gmail.com

Sansar Singh Chauhan
Department of Computer Science and Engineering
GL Bajaj Institute of Technology and Management(GLBITM)
Greater Noida
hod.cse@glbitm.ac.in

Satya Prakash Yadav
Department of Computer Science and Engineering
GL Bajaj Institute of Technology and Management(GLBITM)
Greater Noida
Satya.yadav_cse@glbitm.ac.in

Abstract— Nowadays we can see a significant surge in user-generated material on the web as a result of enhanced digitization, which gives people's thoughts on many themes. The computer study of assessing people's sentiments and views regarding an entity is known as sentiment analysis. What do people think? How do you feel about a specific topic? Bringing together computer science researchers, Computing linguistics, data mining, psychology, and even sociology are just a few examples. Sentiment analysis is a text mining approach that automatically analyses text for the writer's sentiment using machine learning and natural language processing (NLP) (positive, negative, neutral, and beyond). Positive, Negative, and Neutral comments may be readily identified using powerful machine learning algorithms. There are various types of sentimental analysis. The main feature of emotional is that it classifies the polarity in text data. The number of smart phones is expanding in lockstep with the growth of the internet. The contemporary Internet allows millions of individuals all over the world to connect with one another and share their ideas and opinions via email, social networking websites like Twitter, Facebook and other means. It is the cheapest and most convenient method of interacting with others. There is a lot of text material on this social networking site. These text data may be utilized to analyse public sentiment on certain issues, as well as the emotion exhibited on any online platform. In this paper we are going to review and compare the Traditional Dictionary Based Approach and Machine Learning with text classifier which is trained with the dataset of U.S. Airline under first propose and second propose.

Keywords: Data Pre-Processing, sentiment analysis, NLTK, matplotlib, Long-short-Term-Memory (LSTM), binary text classifier, Pandas, TensorFlow.

Introduction

One of the industry's most popular initiatives, every customer-facing industry (retail, telecommunications, banking etc.) is interested in determining whether its consumers have favourable or negative feelings about them. Python sentiment analysis is a way for examining a piece of text and determining the hidden sentiment. This is accomplished through the use of a combination of machine learning and natural language processing (NLP). Sentiment analysis is a technique for analysing the emotions represented in a text. The computer analysis of human's views, feelings, emotions, and attitudes about things like products, services, issues, events, themes, and their qualities is known as sentiment analysis or opinion mining (Liu 2015). In view of result, sentiment analysis may be used to track public attitude about a certain entity and generate useful information. This form of information may also be utilized to comprehend, explain, and forecast social processes (Pozzi et al. 2017). Sentiment analysis is critical in the business sphere since it allows companies to develop strategy and acquire insight into client opinion on their goods. Understanding the consumer is becoming increasingly vital in today's customer-oriented company culture.

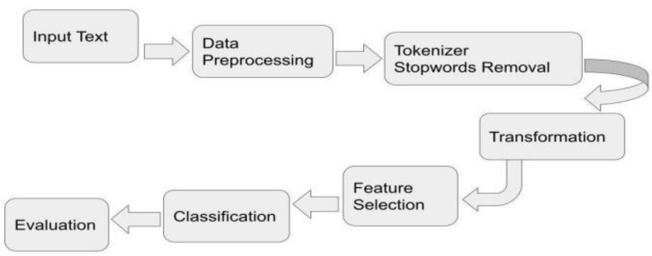


Figure.1. Flow chart of Sentimental Analysis

Fig.1 above demonstrate the flow chart of the steps involve in building the sentimental analysis model using dictionarybased approach ,According to the figure first tweet is taken using scrapper as input text for data processing and sentenceis broken into smallest unit which is word or tokens followed by removal of stop words to transform it into list which contains words which are useful in performing sentimental analysis and finally emotion is classified using the Dictionary which contains words as key and emotion as value of the key.

The Internet has revolutionized the way individuals share their thoughts and ideas nowadays. It is currently mostly accomplished through blog entries, internet forums, product review websites, social media, and other similar mediums. Millions of people use social networking sites like Facebook, Twitter, Google Plus, and others to express their emotions, discuss ideas, and share viewpoints about their daily lives. We receive an interactive media through online communities, where consumers utilize internet forums to educate and influence others. In the form of tweets, status updates, blog posts, comments, and reviews, social media generates a vast amount of sentiment-rich data. Furthermore, social platform allows businesses to engage with their consumers for the purpose of advertising. People make a lot of decisionsbased on user-generated material found on the internet. For example, before making a decision to buy a product or use a service, people will study it online and discuss it on social media platforms. End-user-generated content is simply too huge for a normal user to study. As a result, there is a requirement to automate. Textual information retrieval strategiesare primarily concerned with processing, finding, and interpreting the factual information available. Although facts are objective, there are certain textual components that represent subjective traits. Opinions, feelings, assessments, attitudes, and emotions are the most common contents in Sentiment Analysis (SA). Because of the tremendous proliferation of existing content on the internet via blogs and social media platforms, it presents numerous difficult chances for developing new applications. For example, using SA, it is possible to forecast recommendations of goods provided by a recommendation platform by considering criteria like positive or negative comments about such products. Sentiment analysis includes a variety of tasks such as sentiment extraction, sentiment classification, subjectivity categorization, opinion summarization, and opinion spam detection, to mention a few. Its goal is to examine people's feelings, attitudes, views, and emotions regarding things including products, people, subjects, organizations, and services.

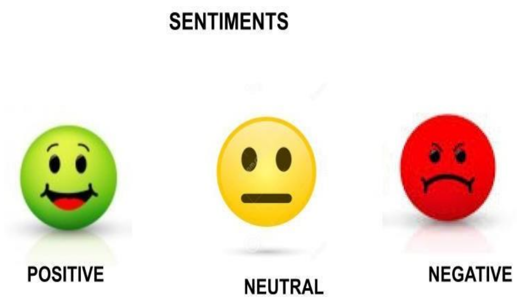


Figure 2. Segments of Sentimental Analysis

Fig.2 is the Pictorial Representation of the types of basic emotion present in the text or words in a conversation which are Positive, Neutral and Negative.

The three levels of sentiment categorization are Document level, Sentence level, and Aspect or feature level. The objective at the document level is to categorize the entire document into a good or bad category. Sentence level sentiment categorization divides sentences into three classes: positive, negative, and neutral. The polarity of each wordin a sentence is decided first, followed by the overall mood of the statement. The sentiment classification at the aspector feature level finds and extracts product attributes from the source data before categorization. Machine learning-based sentiment analysis and dictionary-based sentiment analysis are the two most used techniques to sentiment analysis. To categorize text, a machine learning-based approach uses a classification algorithm such as a support vector machine or aneural network. To identify polarity, a dictionary-based technique employs a sentiment dictionary comprising opinion terms and matches them to the data. hey give opinion words sentiment scores that describe the Positive, Negative, andObjective scores of the words in the dictionary.

I. LITERATURE REVIEW

There has already been a lot of study done on sentiment analysis in the past. The most recent study in this area focuses on doing emotional analysis on any type of text, sentence, paragraph, or someone's voice, with the majority of the data coming from social media platforms such as Facebook, Twitter, and Amazon. Emotional analysis research, in particular, is focused on machine learning algorithms, with the goal of determining whether a given text encourages or opposes recognizing text divisions. In this part, you'll get an in-depth look at one of the most useful research activities: sentiment analysis. The following are some examples of research in sentiment analysis using various techniques:

P. Pang, L. Lee, S. Vaidyanathan, and others: "Thumbs up?" by P. Pang, L. Lee, and S. Vaidyanathan. Proc.ACL-02 conference on Empirical approaches in Natural Language Processing, vol.10, pp. 79-86, 200[10].

They were the pioneers in the area of sentiment analysis. Their major goal was to categorize material based on overall sentiment rather than simply topic, for example, good or negative movie reviews. They use a movie review database to test machine learning algorithms, and the findings show that these algorithms outperform human-made techniques. They employ Nave-Bayes, maximum entropy, and Support vector machines as Machine Learning Algorithms. They also end bylooking at a variety of characteristics that make sentiment categorization difficult. They reveal the root of sentiment analysis is supervised machine learning algorithms.

"NLTK: the Natural Language Toolkit," in Proceedings of the ACL-02 Workshop on Effective Tools and Methodologies forTeaching Natural Language Processing and Computational Linguistics, vol. 1, pp. 63-70, 2002 E. Loper and S. Bird [11]

The Natural Language Toolkit (NLTK) is a collection of software modules, structured files, tutorials, problem sets, statistical functions, machine learning classifiers that are ready to use, computational linguistics courseware, and other resources. One of NLTK's core tasks is natural language processing, which entails assessing human language data. Corpora are provided by NLTK and are used to train classifiers. Developers replace old components with new ones, programs getmore organized, and datasets produce more complex outputs.

The Natural Language Toolkit (NLTK) is a collection of software modules, structured files, tutorials, problem sets, statistical functions, machine learning classifiers that are ready to use, computational linguistics courseware, and other resources. One of NLTK's core tasks is natural language processing, which entails assessing human language data. Corpora are provided by NLTK and are used to train classifiers. Developers replace old components with new ones, programs getmore organized, and datasets produce more complex outputs.

O. Almatrafi, S. Parack, B. Chavan, and others [12]

According to the researcher Sentiment, Sentimental Analysis, is the process of extracting a sentiment from a text unit from a specific location using Natural Language Processing (NLP) and machine learning approaches. They look at a variety of location-based sentiment analysis applications using a data source that permits data to be obtained from numerous locations. easily. A script may easily access a feature of Twitter called tweet location, which allows data (tweets) from a given location to be obtained for the aim of detecting trends and patterns. The following illustration aims at providing an insight into more popular algorithms used in sentiment analysis:

1. Machine Learning Approach

Machine learning is a data analytics technology that trains computers to learn from experience in the same way that people and animals do. Machine learning algorithms employ computer approaches to "learn" information directly from data rather than depending on a model. Machine learning is a data analytics technology that trains computers to learn from experience in the same way that people and animals do. Machine learning algorithms employ computer approaches to "learn" information directly from data rather than depending on a model.

Machine Learning Approach is classified mainly into four parts for Sentimental Analysis:

Supervised Learning: Supervised learning is one of the methods of sentimental analysis that involves training a computersystem on input data that has been labelled for a certain output.

Decision Tree Classifier: In the decision tree classifier, each node in the tree represents a test on an attribute, and eachbranch descending from that node represents one of the property's potential values.

Linear Classification: A linear classifier is a model that uses a linear combination of explanatory factors to categorize a set of data points into a discrete class.

Support Vector Machine (SVM): Support-vector machines are supervised learning models with related learning algorithms for classification and regression analysis in machine learning.

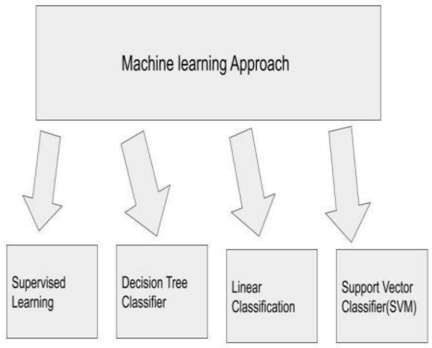


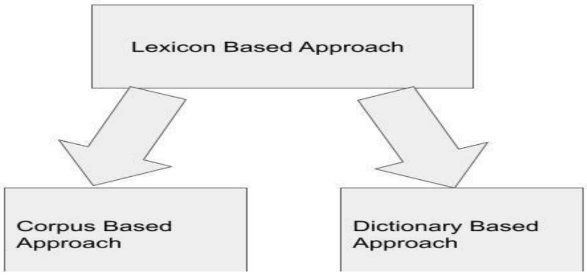
Figure 3. Flow Chart of Machine Learning Approach

Above figure.3 demonstrate the different ways of using Machine Learning Approach for sentimental Analysis

2. Lexicon-based Approach

The Lexicon-based technique assesses a document by aggregating the sentiment ratings of all the terms in the content using a pre-prepared sentiment lexicon. A term and its related sentiment score should be included in the sentiment lexicon.

Fig.4 flow chart of Lexicon Based Approach



Above fig.4 demonstrate the various case used in the lexicon Based Approach for Sentimental Analysis

Corpus Based Approach: The corpus-based approach to language instruction is based on genuine and authentic occurrences of language as it is spoken, written, and used by native speakers in a variety of settings.

Dictionary Based Approach: Dictionary-based sentiment analysis is a computational approach to measuring the feeling that a text conveys to the reader. In the simplest case, sentiment has a binary classification: positive or negative, but it can be extended to multiple dimensions such as fear, sadness, anger, joy, etc.

Table 1: Contrast Study of Various Sentimental Analysis Techniques.

S.no	Title of Paper	Used Methodology	Review Dataset	Accuracy
1)	SVM and Co-reference Resolution in a Feature-Based Approach for Sentiment Analysis (2017)	Resolution based on SVM and co-reference	Product Review Dataset for Training	73.6%
2)	Sentiment Analysis on Twitter using Neural Networks (2015)	Feed-forward neural networks	Data from Twitter	74.15%
3)	Twitter Sentiment Analysis Using Python Machine Learning Algorithms (2015)	Maximum Entropy, Naive Bayes, SVM	Data from Twitter	86.4% 73.5% 88.97%
4)	A New Approach: Sentiment Analysis using Neural Networks (2018)	Neural Network with convolution	Review of Product Data on Twitter	74.15% 64.69%
5)	Twitter Corpus Sentiment analysis in Relation to AI Assistants (2018)	(VADER) Valence Aware Dictionary for Sentiment Reasoning	Electronic product Reviews	87.4%
6)	A framework for sentiment analysis with hotel review opinion mining (2018)	Naive Bayes	Opin Rank hotel reviews.	83.5%
7)	Sentiment Analysis on E-Commerce Data at the Aspect Level (2018)	SVM, Naive Bayes	Review Data of Amazon Customer	90.423% 83.43%

II. METHODOLOGY

1) Dictionary Based Approach:

We reviewed the strategy used in performing the execution of the study and the algorithm utilized in the Sentimental Analysis with using scrapper to extract tweet for text data. System's process flow chart. is given below:

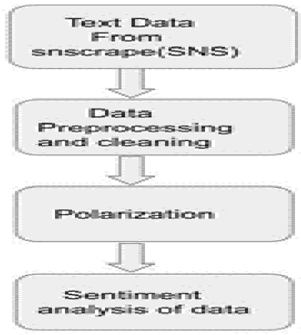


Figure 5. Methodology for Sentiment Analysis using NLP

There are various steps involved in the Methodology of the first propose from the Text data to sentimental analysis of that text data. As we discussed above that Text data is taken as input from snscrape which is scrapping tool and Data isprocessed after breaking big sentence into individual words or tokens and removing the stop-words or unnecessary word from that list, the next step is Polarization which involves the nltk or natural language processing to give the score of the sentiment under class like pos, neg, neu, compound using SentimentIntensityAnalyzer function and the final step is to analyze the sentiment based on the score.

Text data from snscrape [SNS]

Text data is taken from Scraper for social networking services, snscrape (SNS). It scrapes user profiles, hashtags,and tweets and returns the items found, such as related postings.

Data Pre-Processing

The unanalyzed data is handled in preprocessing for feature extraction. It is further broken down into thefollowing steps:

a) **Tokenization:** A phrase is broken down into words by removing white spaces, symbols, and special lcharacters.

b) **Stop words removal:** some words like arti- cle, adjective etc. are removed using NLTK corpus library whichdoes not have any kind of emotion.

c) **Case Normalization:** The entire documents are

 ISSN: 2582-3930

converted into lower-case.

Data Polarization: The orientation of the stated emotion is determined by the element's sentiment polarity, which determines whether the text communicates the user's positive, negative, or neutral feeling toward the entity in question. The main motive of sentimental analysis is to examine a body of text in order to determine the viewpoint communicated. We usually measure this feeling with a positive or negative polarity value. The polarity score's sign is usually used to assess whether an emotion is positive, neutral, or negative.

2) *Machine Learning Based Approach:*

The methodologies utilized in the execution of the study and the Binary text Classifier algorithm that are employed in the class prediction using sentiment analysis are explained in this model. The flow chart of theapproach used in this article is shown below.

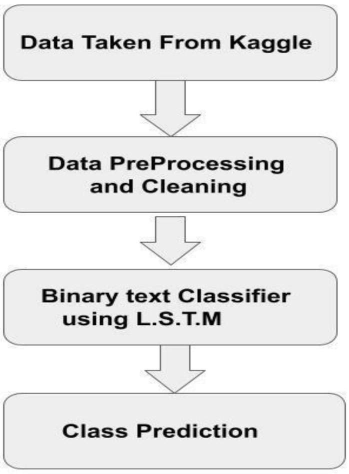


Figure 6. Methodology Of Sentimental Analysis using Binary text classifier.

The steps involve under the propose-4 in the fig.4 are as follows, firstly csv file or dataset is taken from the Kaggle which have more than 14,000 tweets of us-airline sentiment and then huge dataset is converted and cleaned using pandas which make dataset of more than 6 column into 2 column which are text and airline sentiment and binary text classifier algorithm is used with L.S.T.M to predict the class.

	text	airline_sentiment
1	@VirginAmerica plus you've added commercials t...	positive
3	@VirginAmerica it's really aggressive to blast...	negative
4	@VirginAmerica and it's a really big bad thing...	negative
5	@VirginAmerica seriously would pay $30 a fligh...	negative
6	@VirginAmerica yes, nearly every time I fly VX...	positive
8	@virginamerica Well, I didn't…but NOW I DO! :-D	positive
9	@VirginAmerica it was amazing, and arrived an ...	positive
11	@VirginAmerica I <3 pretty graphics. so muc...	positive
12	@VirginAmerica This is such a great deal! Alre...	positive
13	@VirginAmerica @virginmedia I'm flying your #f...	positive

Dataset: Contains:

More than 14000 tweet data samples are included in the collection, which are classified into three groups: Positive, Negative, or Neutral.

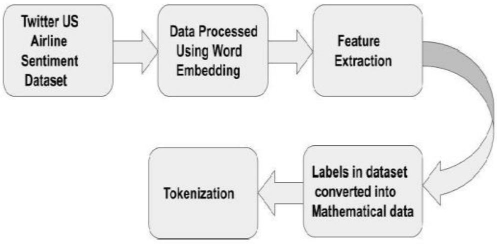


Figure. 7 flow of Data preprocessing and cleaning

Figure.7 utilized the US Airline Twitter Dataset, which comprises over 15,000 tweets, in this work. The dataset initially contains parameters such as twitter id, airline sentiment, text, airline sentiment, confidence, airline, name, which are then reduced to two attributes, airline sentiments and text, using feature extraction. In this paper, we used a binary text classifier that gives 0 and 1 for negative and positive classes for column airline sentiment.

	text	airline_sentiment
0	@VirginAmerica What @dhepburn said.	neutral
1	@VirginAmerica plus you've added commercials t...	positive
2	@VirginAmerica I didn't today... Must mean I n...	neutral
3	@VirginAmerica it's really aggressive to blast...	negative
4	@VirginAmerica and it's a really big bad thing...	negative

Dataset after feature Extraction

Data Pre-Processing and Cleaning:

we have used binary text classifier which only takes two class as we don't need neutral reviews from thedataset therefore it has been removed from the dataset.

VOLUME: 07 ISSUE: 05 | MAY - 2023 IMPACT FACTOR: 8.176 ISSN: 2582-3930

This dataset's labels are categorical. Only numeric data is understood by machines. So, using the factorize () function, transform the category data to numeric values. This gives you an array of numeric numbers as well as a category index.

```
(array([0, 1, 1, ..., 0, 1, 1], dtype=int64),
 Index(['positive', 'negative'], dtype='object'))
```

The 0 signifies good feeling and the 1 represents negative sentiment, as you can see. Now comes the most important part of python sentiment analysis. The input text should be transformed into something that our machine learning model can understand. Therefore, the text has been converted into vector embeddings array. Word embeddings are a lovely method of displaying the relationship between words in a text. To do so, we assign a unique number to each of the unique words, then replace the word with the assignednumber.

Now, before proceeding ahead in python sentiment analysis project let's tokenize all the words in the text with the help of Tokenizer. In tokenization, we break down all the words/sentences of a text into small parts called tokens. The fit-on texts () function establishes a link between the words and the numbers supplied to them. This association is stored in the form of a dictionary in the tokenizer. word index attribute. Using the text to sequence () function, replace the words with their allocated numbers.

The sentences in dataset are not all the same length. To make the sentences equal in length, use padding.

Binary Classifier using LSTM:

In our machine learning model for sentiment analysis, we employ LSTM layers. Our model has three layers: anembedding layer, an LSTM layer, and a Dense layer in the end. We used the Dropout mechanism in-between theLSTM layers to minimize overfitting.

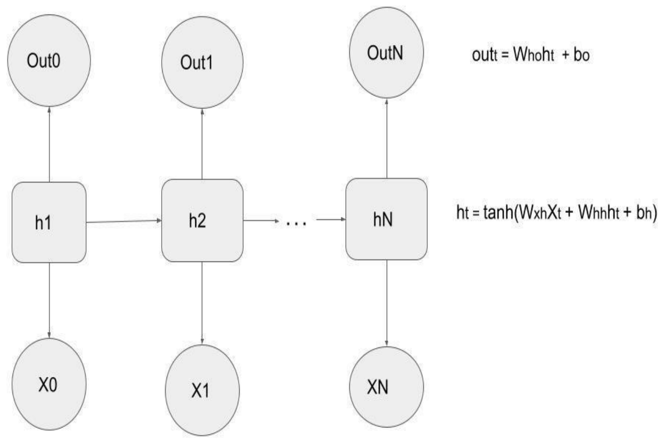


Figure.8 Long- short-term neural network

Figure.8 explains about the Mathematical working of the Long-short-Term Neural Network

Long Short-Term Memory Networks are abbreviated as LSTM. It's a Recurrent Neural Networks variation. Recurrent Neural Networks are typically used to process sequential input like text and audio. The meaning of each word and associated computations (known as hidden states) are usually kept while constructing an embedding matrix. If a word's reference is utilized after 100 words in a text, then RNNs cannot keep all of thesecomputations in their memory. RNNs are unable to learn these long-term dependencies for this reason. One of the regularization techniques is dropout. It is employed to prevent overfitting. We drop some neurons at random in the dropout mechanism. The layer accepts a value between 0 and 1 as an argument, which reflects the likelihood of dropping the neurons. This results in a stable model that avoids overfitting.

```
Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 embedding (Embedding)       (None, 200, 32)           3200000

 spatial_dropout1d (SpatialD  (None, 200, 32)          0
 ropout1D)

 lstm (LSTM)                 (None, 50)                16600

 dropout (Dropout)           (None, 50)                0

 dense (Dense)               (None, 1)                 51

=================================================================
Total params: 3,216,651
Trainable params: 3,216,651
Non-trainable params: 0
_________________________________________________________________
None
```

Class Prediction:

With a batch size of 32 and a validation split of 20%, train the sentiment analysis model for 5 epochs on the whole dataset.

```
In [32]:  1 history = model.fit(padded_sequence,sentiment_label[0],validation_split=0.2, epochs=5, batch_size=32)

Epoch 1/5
289/289 [==============================] - 52s 166ms/step - loss: 0.3942 - accuracy: 0.8358 - val_loss: 0.2109 - val_accuracy:
0.9168
Epoch 2/5
289/289 [==============================] - 53s 183ms/step - loss: 0.2160 - accuracy: 0.9183 - val_loss: 0.1786 - val_accuracy:
0.9415
Epoch 3/5
289/289 [==============================] - 55s 191ms/step - loss: 0.1762 - accuracy: 0.9344 - val_loss: 0.1788 - val_accuracy:
0.9355
Epoch 4/5
289/289 [==============================] - 55s 189ms/step - loss: 0.1336 - accuracy: 0.9501 - val_loss: 0.1783 - val_accuracy:
0.9394
Epoch 5/5
289/289 [==============================] - 58s 200ms/step - loss: 0.1097 - accuracy: 0.9581 - val_loss: 0.1978 - val_accuracy:
0.9411
```

We have defined a function to predict sentiment that takes the input as a sentence and classifies it into one of the two-class that is Positive and Negative. The accuracy of the model is approximately 94.0 percent. as we see that the test_sentence1 which seems to be negative is also predicted negatively by our text classifier model and text_sentence2 which seems to be positive is also predicted correctly by our text classifier model.

III. RESULT AND DISCUSION

Below table 2 illustrates the comparison and similarities between two proposed approaches in this paper using strength and challenges given below:

	Approach	Strength	Challenges
Dictionary based Approach	To classify the individual word/sentence/dataset, this approach requires reference dictionary.	1.This approach is mainly used for Unsupervised dataset. 2. No prior training of Dataset is needed this approach. 3. This approach works efficiently on small dimension dataset. 4. Faster in Execution and has higher accuracy as this approach works smaller datasets. 5. This approach has Lesser Computation as no prior training and modeling is required. 6. This approach has lesser risk as it is already tested and widely used.	1. This approach has no learning capability. 2. This is a rule based approach and requires an Expert to create rules at first. 3. This approach face challenges on high dimension dataset. 4. This approach is not sensitive to the type of document being coded hence, not suitable for every type of document.
Machine learning Based Approach	For Classification of dataset this requires Supervised and Un-Supervised Machine Learning Approaches.	1. This approach works for both supervised and Un-supervised dataset. 2. There is no need of Dictionary for this approach. 3. This approach demonstrates highest accuracy of classification independent of size of dataset. 4. Efficiency of this approach does not depend on any specific rule like dictionary based approach. 5. This approach works efficiently on any dimensional dataset. 6. This approach has learning capability.	1. Classifier trained in one domain does not work on the texts of other domains. 2. This approach is affected by imbalances caused by classes and other linguistic variations. 3. This approach may produce contradictory result for the same input in some rare cases.

We successfully constructed a sentiment analysis model in Python. We created a binary text classifier in this machine learning project that divides tweet sentiment into positive and negative categories. On validation, we got more than 94 percent accuracy. Let's use matplotlib to plot these metrics. Matplotlib is a visualization tool that uses a low-level graphplotting toolkit written in Python.

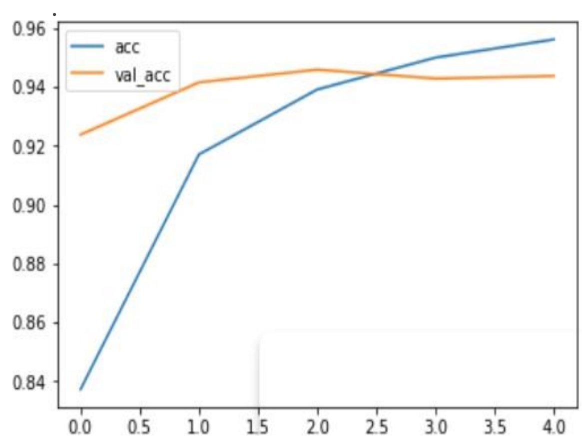


Figure .9 Accuracy Function of the Model

```
test_sentence1 = 'tdayshae we apologize for the disruption dm us your bag tag number lets take a closer look'
predict_sentiment(test_sentence1)

Predicted label:  negative
```

In Figure.9 the orange line in the graph shows the training accuracy is approx. to 96 percent and testing accuracyapprox. to 94 percent respectively.

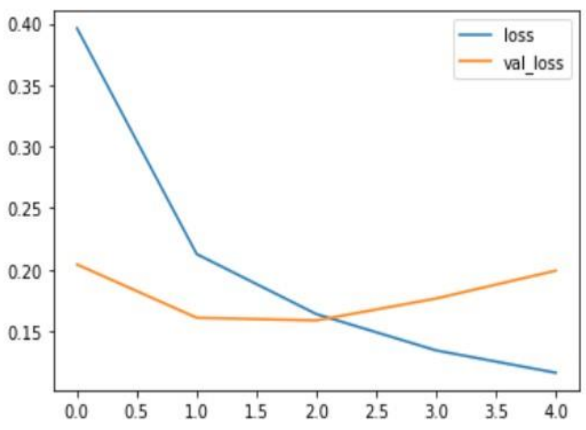


Figure.10 loss function of the Model

Now, we have compared the both Dictionary based approach and text-classifier approach for the given tweetand the following results are obtained:

```
C:\Users\91639\PycharmProjects\pythonProject3\venv\Scripts\python.exe C:/Users/91639/PycharmProjects/pythonProject3/main.py
tdayshae we apologize for the disruption dm us your bag tag number lets take a closer look
['tdayshae', 'we', 'apologize', 'for', 'the', 'disruption', 'dm', 'us', 'your', 'bag', 'tag', 'number', 'lets', 'take', 'a', 'closer', 'look']
Counter()
class : Positive
```

Figure 10: output of first propose

Figure 10 shows the result of a sentiment analysis for the following tweet using Dictionary based Approach as apositive class.

```
test_sentence1 = 'tdayshae we apologize for the disruption dm us your bag tag number lets take a closer look
predict_sentiment(test_sentence1)
test_sentence2 ='I had a nice fly experience from air america'
predict_sentiment(test_sentence2)

Predicted label:  negative
Predicted label:  positive
```

Figure 11:output of second propose

Whereas, according to the fig.11 When the identical tweet is fed into the text classifier model, it produces anegative classification.

IV. RESULT AND DISCUSION

This work is divided into two sections under the first propose and second propose. The first part of the project focuses on sentimental analysis using a dictionary-based technique, while the second portion focuses on sentimental analysis using a machine-learning approach. After comparing both, it was determined that the second propose advised the correct sentiment for a particular tweet with a testing accuracy of 94%. our second proposal, a machine learning technique, has certain limitations. Therefore, if the provided tweet contains terms that are not included in the dataset, the class prediction may be incorrect for few tweets as input.

V. References

1. G. Vinodhini and RM. Chandrasekaran, "Sentiment Analysis and Opinion Mining: A Survey", Volume 2, Issue6, June 2012 ISSN: 2277 128X International Journal of Advanced Research in Computer Science and Software Engineering

2. A. Pak and P. Paroubek. „Twitter as a Corpus for Sentiment Analysis and Opinion Mining". In Proceedings ofthe Seventh Conference on International Language Resources and Evaluation, 2010, pp.1320-1326

3. Madhoushi, Z. Hamdan, A. R. and Zainuddin, Sentiment Analysis Techniques in Recent Works," Science andInformation Conference (SAI), 2015. IEEE

4. Singh and Vivek Kumar, —A clustering and opinion mining approach to socio-political analysis of theblogosphere", Computational Intelligence and Computing Research (ICCIC), 2010 IEEE International Conference

5. Madhoushi, Z. Hamdan, A. R. and Zainuddin, Sentiment Analysis Techniques in Recent Works," Science andInformation Conference (SAI), 2015. IEEE

6. Jha, I. Manjunath, N., Shenoy, Venugopal, K. R. and Patnaik, L., "HOMS: Hindi Opinion Mining System," IEEE2nd International Conference on

Recent Trends in Information Systems (RTIS) M. 2015.

7. E. Loper and S. Bird, "NLTK: the Natural Language Toolkit", Proc. ACL-02 Workshop on Effective tools andmethodologies for teaching natural language processing and computational linguistics,vol. 1,pp. 63-70, 2002