**PES UNIVERSITY**
**(Established under Karnataka Act No. 16 of 2013)**
**100-ft Ring Road, Bengaluru – 560 085, Karnataka, India**

***Report on***

# ' Comparative Study of Sentiment Analysis of Twitter Tweets using Deep Learning Methods'

***Submitted by***


| | |
|---|---|
| **Haritha GB** | **PES1201800325** |
| **Sahana N.B** | **PES1201800443** |
| **Sherine Gifta** | **PES1201801251** |


**August - December 2021**

**under the guidance of**

***Internal Guide***


**Raghavendra M J**
**Associate Professor**
**Dr.BN Krupa**
**Professor**
**Department of ECE**
**PES University**
**Bengaluru -560085**


**FACULTY OF ENGINEERING**

**DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGG**

**PROGRAM B.TECH**

# Table of Contents

# 1. Introduction

Sentiment Analysis is a text mining technique. As a result, it's possible to classify it as a text mining technique for assessing the underlying sentiment of a text message, such as a tweet. Positive, negative, or neutral sentiments or opinions can be conveyed on Twitter.

Companies use Twitter Sentiment Analysis to create their business plans, to assess customers' attitudes toward items or brands, to see how people react to their campaigns or new releases, and to figure out why certain things aren't selling. Sentiment Analysis Dataset in Politics Twitter is used to keep track of political viewpoints and to find consistency and inconsistency between government words and actions. Twitter is being used to analyse election results, according to the Sentiment Analysis Dataset. Twitter Sentiment Analysis is also used to track and analyse social events, foresee potentially dangerous situations, and gauge the mood of the blogosphere.

# 2. Problem Statement

To build a model that will *determine the tone* (neutral, positive, or negative) of the text. The model is trained on the existing data (train.csv). The resulting model will have to determine the class (neutral, positive, or negative) of new texts (test data that were not used to build the model) with *maximum accuracy.*

# 3. Literature Survey

**[1] Go, A., Bhayani, R. and Huang, L., 2009. Twitter sentiment classification using distant supervision. *CS224N project report, Stanford*, *1*(12), p.2009.**

**AIM:** To classify Tweets for a query term into negative or positive sentiment.
This paper provides a novel method for classifying the sentiment of Twitter posts automatically. With respect to a query word, these messages are classed as either positive or negative. This is important for consumers who want to research the sentiment of products before purchasing them, as well as for businesses who want to track public perception of their brands. There has never been any research done on categorising the sentiment of messages on microblogging sites like Twitter. Using distant supervision, this paper describes the results of machine learning systems for classifying the sentiment of Twitter messages.

**METHODOLOGY:** Various approaches are implemented – unigrams, bigrams, and Part-of-Speech and train their classifier on various machine learning algorithms – Naive Bayes, Maximum Entropy, and Scalable Vector Machines and compare it against a baseline classifier by counting the number of positive and negative words from a publicly available corpus. The training data is made out of Twitter messages using emoticons as noisy labels. This type of training data is widely available and can be accessed using automated methods.

**RESULT:** It is observed that Bigrams alone and Part-of-Speech Tagging are not helpful while machine learning algorithms (Naive Bayes, Maximum Entropy, and SVM) have accuracy above 80% when trained with emoticon data. It is observed that Bigrams alone and Part-of-Speech Tagging are not helpful. Naive Bayes Classifier gives the best results. This paper also elucidates the preprocessing steps needed in order to achieve sufficient accuracy. The main contribution of this paper is the idea of using tweets with emoticons for distant supervised learning.

**RESEARCH VOID:** No analysis of negation and its interpretation. Leads to higher amounts of misclassification.

**[2] Saif, H., He, Y. and Alani, H., 2012, November. Semantic sentiment analysis of Twitter. In *International semantic web conference* (pp. 508-524). Springer, Berlin, Heidelberg.**

**AIM**: Semantic based approach to identify the entity being discussed in a tweet, like a person, organisation, etc. Sentiment analysis on Twitter provides businesses with a quick and easy tool to track public perceptions of their brand, company, directors, and other topics. In recent years, a variety of features and approaches for training sentiment classifiers on Twitter datasets have been investigated, with mixed results.

**METHODOLOGY**: In this paper, a unique technique to sentiment analysis that incorporates semantics as additional features into the training set has been introduced. A semantic concept (e.g. "Apple product") has been added as an additional feature to each retrieved item (e.g. iPhone) from tweets, then the connection between the representative concept and negative/positive sentiment has been analysed. For three separate Twitter datasets, this method has been used to predict sentiment. Efficiency is improved by using more basic techniques used in Sentiment Analysis, like stemming, two-step classification and negation detection, and scope of negation. They also demonstrate that removal of stop words is not a necessary step and may have undesirable effects on the classifier. The scope of negation of a cue can be taken from that word to the next following punctuation.

**RESULTS:** Over the baselines of unigrams and part-of-speech features, the results demonstrate an average gain of 6.5 percent and 4.8 percent in F harmonic accuracy score for identifying both negative and positive sentiment. Semantic features have been compared to a sentiment-bearing topic analysis approach, and find that when categorising negative sentiment, semantic features give superior Recall and F score, and better Precision with lower Recall and F score than when classifying positive sentiment.

**RESEARCH VOID:** No emotion classification, only identification of the subject.

**[3] Neethu, M. S., and R. Rajasree. "Sentiment analysis of Twitter using machine learning techniques." In *2013 Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT)*, pp. 1-5. IEEE, 2013.**

**AIM**: Machine Learning techniques use a training set and a test set to build a classification model to classify input feature vectors to corresponding class labels. Sentiment analysis is the process of recognizing and categorizing the views and attitudes represented in a source text. Tweets, status updates, blog entries, and other forms of social media generate a large volume of sentiment-rich data.
Sentiment analysis of this user-generated data is quite beneficial for determining the general consensus. Due to the existence of slang phrases and misspellings, Twitter sentiment analysis is more complex than conventional sentiment analysis. In Twitter, the maximum number of characters allowed is 140.

**METHODOLOGY:** Preprocessing of tweets to replace slang and analyse hashtags. Creating feature vectors using relevant features. Using Naive Bayes, SVM, Maximum Entropy, and Ensemble Classifiers to predict the sentiment. The two methodologies for analysing feelings from the text are the knowledge base approach and the machine learning approach. In this research, a Machine Learning technique has been used to analyse tweets on electronic devices such as smartphones and laptops. The methodology involves Preprocessing of tweets to replace slang and analyse hashtags. Creating feature vectors using relevant features. It is possible to determine the effect of domain information on sentiment classification by doing sentiment analysis in a specific domain. A new feature vector for identifying tweets as good or negative, as well as extracting people's product opinions has also been developed. It also does not require a predefined dataset with all the required emotions.

**DATASET:** Automatically extracted 1200 tweets (1000 train, 200 test) using Twitter API.

**RESULTS**: The accuracy while using the Naive Bayes classifier was 89% whereas, in the SVM as well as the Maximum Entropy Classifiers, the percentage accuracy came up to 90%.

**[4] Kumar, A. and Sebastian, T.M., 2012. Sentiment analysis of Twitter. *International Journal of Computer Science Issues (IJCSI)*, *9*(4), p.372.**

**AIM:** A method has been proposed and tested for extracting sentiment from Twitter.
There has been an explosion of user-generated content since the advent of the social networking age. Because of its unique short and easy method of expression, millions of individuals utilise microblogging sites to share their opinions on a daily basis. This paper proposed and tested a method for extracting sentiment from Twitter, a prominent real-time microblogging service where users submit real-time comments and views on "anything."

**METHODOLOGY**: In this research, a hybrid methodology for determining the semantic orientation of opinion terms in tweets using both corpus-based and dictionary-based methods was presented. The hybrid model allows for more flexibility in model tuning.

**ADVANTAGE:** The hybrid model allows for more flexibility in model tuning. A case study was also carried out to demonstrate how the proposed system can be used and how beneficial it is.

**RESEARCH VOID:** However, this paper did not include the scope of negation analysis.

**[5] Sunarya, P., Refianti, R., Mutiara, A., and Octaviani, W., 2019. Comparison of Accuracy between Convolutional Neural Networks and Naïve Bayes Classifiers in Sentiment Analysis on Twitter.** ***Int. J. Adv. Comput. Sci. Appl*****,** ***10*****, pp.77-86.**

**AIM**: This study used one of the Deep Learning algorithms, Convolutional Neural Networks, to analyse sentiment on English-language tweets on the topic "Turkey Crisis 2018." (CNN). After that, the CNN classifier model has been compared against the Naïve Bayes Classifier model to see whether one can deliver greater sentiment analysis accuracy.
The community's requirements and desires for easy access to information are driving an increase in the usage of social media technologies like Twitter to share, deliver, and search for information. The collection of tweets can be converted into meaningful information using sentiment analysis due to the high number of large tweets shared by Twitter users every second.

**METHODOLOGY:** As a huge number of tweets are required to provide information, a classifier model that can complete the analysis process fast and accurately is required. Deep Learning is a prominent and commonly used approach for constructing classifier models nowadays. Data retrieval, pre-processing, model construction and training, model testing, and visualisation are the research methodologies that were used in this study.

**RESULTS:** In the testing phase of the data test, the CNN classifier model generates an accuracy of 0.88 or 88 percent, whereas the NBC classifier model provides an accuracy of 0.78 or 78 percent, according to the findings of this study. Based on these findings, it can be stated that the Deep Learning classifier model delivers superior sentiment analysis accuracy than the Naïve Bayes classifier model.

**[6] Jianqiang, Z., Xiaolin, G. and Xuejun, Z., 2018. Deep convolution neural networks for Twitter sentiment analysis.** ***IEEE Access*****,** ***6*****, pp.23253-23260.**

**AIM:** In this research, a word embeddings method derived from unsupervised learning on large Twitter corpora, which employs latent contextual semantic links as well as statistical co-occurrence properties between words in tweets has been developed. Twitter sentiment analysis technology allows you to poll public opinion on events or goods that are linked to them. The majority of current research focuses on extracting sentiment features from lexical and syntactic features. Sentiment words, emoticons, exclamation marks, and other symbols are used to express these characteristics.

**METHODOLOGY:** To create a sentiment feature set of tweets, these word embeddings are merged with n-grams and word sentiment polarity score features. The feature set is used to train and predict sentiment classification labels using a deep convolutional neural network. On five Twitter data sets, the following has been compared: the performance of our model to the baseline model, which is a word n-grams model.

**RESULTS:** The findings show that the model developed outperforms the baseline model in terms of accuracy and F1-measure for Twitter sentiment classification.

**[7] Severyn, A. and Moschitti, A., 2015, August. Twitter sentiment analysis with deep convolutional neural networks. In *Proceedings of the 38th international ACM SIGIR conference on research and development in information retrieval* (pp. 959-962).**

**AIM:** The key contribution of this study is a new methodology for initialising the convolutional neural network's parameter weights, which is critical for training an accurate model without having to insert any additional features. The deep learning method for sentiment analysis of tweets is described in this publication.

**METHODOLOGY:** To summarise, an unsupervised neural language model to train initial word embeddings has been used, which are then fine-tuned on a distantly supervised corpus by our deep learning model. The network's pre-trained parameters are used to initialise the model in the end. The supervised training data provided by Semeval-2015's official system assessment campaign on Twitter sentiment analysis is used to train the latter.

**ADVANTAGES:** In both the phrase-level subtask A (among 11 teams) and the message-level subtask B (among 11 teams), a comparison of the outcomes of this approach and the systems participating in the challenge on the official test sets suggests that the model we have tried to implement might be rated in the first two spots (among 40 teams). This is vital evidence of this solution's practical value.

# 4. Methodology

Our methodology uses recurrent neural networks to classify a tweet into positive, negative or neutral sentiments. These tweets come from a dataset of airline reviews. The workflow of this implementation is divided into three parts: data preprocessing, model building, and model evaluation.

## 4.1 Data Preprocessing

### 4.1.1 Dataset

We have used the US Airline Dataset (open source from Kaggle). This is an open source dataset. The dataset that we used has 14641 tweets that are all unprocessed. Unprocessed tweets are defined by tweets that have the Twitter handle (@user), contain URLs, and have redundant or unnecessary words. This dataset contains the following columns: Tweet ID, sentiment, sentiment confidence value, negative reason, negative reason confidence, airline, username, the text of the tweet. The airlines included in the dataset are SpiceJet, IndiGo, GoAir, Vistara, Air Asia, and Air India.

### 4.1.2 Preprocessing

Tweets from the data contain mentions, hashtags, spelling errors, additional and irrelevant words. We only use the data in the text and airline_sentiment column.

The preprocessing has the following steps:

1. Removing additional punctuation
2. Remove web address, Twitter ID, and digit.

## 4.2 Implementation

### 4.2.1 Implementation of RNN using LSTM

When working with sequential data, when previous knowledge is important, Feed Forward NNs will not provide you with decent results. The choice of RNNs can be summed up in one word, that is, memory. RNNs have the ability to store data from prior steps, which works well when dealing with Sequential Data. Because of the vanishing gradient problem, RNN finds it difficult to transmit information from one instance to another when the sequence is long. LSTM (Long Short Term Memory) has been used to solve this problem. The various steps are explained next.

First, Pre-processing of tweets is done by removing punctuation, urls, hashtags, twitter id numbers and stopwords. Then, Encoding of words is done by mapping the words in our vocabulary to integers. Unique words are printed out as well as the result of the first tokenized review. Encoding of labels is completed and the labels are "positive", "negative", or "neutral. To use these labels in the network, we convert the tweet to 0 and 1. Padding Sequences is done to deal with both short and very long reviews, we pad or truncate all our reviews to a specific length. Dataset is split into Training,Validation and Test dataset. Three layers are present which are explained next. An embedding layer that converts our word tokens (integers) into embeddings of a specific size. An LSTM layer defined by a hidden_state size and number of layers. A fully-connected output layer that maps the LSTM layer outputs to a desired output_size. A sigmoid activation layer which turns all outputs into a value 0-1; return only the last sigmoid output as the output of this network.

**Machine Learning Implementations**

The Implementations done are Naive Bayes, Random Forest, Support Vector Machine and logistic regression. Common preprocessing includes removing stopwords, analysis of hashtag usage, removing URLs, and removing the redundant words. The data is then tested with 20% of the airline dataset.

## 4.2.2 Implementation of Naive Bayes Classifier

Naive Bayes classifiers are a collection of classification algorithms based on Bayes' Theorem. The fundamental principle behind using NBC is that every pair of features being classified is independent of each other.

The Feature matrix contains all the vectors(rows) of the dataset in which each vector consists of the value of dependent features. The Response vector contains the value of the class variable(prediction or output) for each row of the feature matrix.

$$\boxed{v_{MAP} = \underset{v_j \in V}{argmax} \frac{P(x_1, x_2, x_3 \cdots, x_n | v_j) P(v_j)}{P(x_1, x_2, x_3 \cdots, x_n)}}$$

## 4.2.3 Implementation of Random Forest Classifier

It is a decision tree based algorithm. Decision tree is a type of supervised learning algorithm (having a predefined target variable) that is mostly used in classification problems. It works for both categorical and continuous input and output variables. In this technique, we split the population or sample into two or more homogeneous sets (or sub-populations) based on the most significant splitter or differentiator in input variables. Random forest builds decision trees on different samples and takes their majority vote for classification and average in case of regression.

## 4.2.4 Implementation of Support Vector Machines

In the SVM algorithm, we plot each data item as a point in n-dimensional space (where n is a number of features you have) with the value of each feature being the value of a particular coordinate. Then, we perform classification by finding the hyper-plane that differentiates the two classes very well. Then, the test point is located on the coordinate axes and evaluated with the separating hyperplane's equation. Based on its value, it is classified as class A or class B.

## 4.2.5 Implementation of Logistic Regression

Logistic regression predicts the output of a categorical dependent variable. Therefore the outcome must be a categorical or discrete value. It is used for classification problems. In Logistic regression, instead of fitting a regression line, we fit an "S" shaped logistic function, which predicts two maximum values (0 or 1).

The sklearn toolkit was used to implement the model and the word cloud's 100 most frequent words as polarity extremes.

# 5. RESULTS

| **Implementation** | **Accuracy** | **F1** |
|---|---|---|
| RNN | 84.687955 | 0.690145 |
| Random Forest | 65.531583 | 0.678558 |
| Logistic Regression (Multiclass) | 64.855732 | 0.672010 |
| SVM | 45.474829 | 0.069594 |
| Naive Bayes | 75.899543 | 0.685469 |

Table 1. Comparison of various implementations

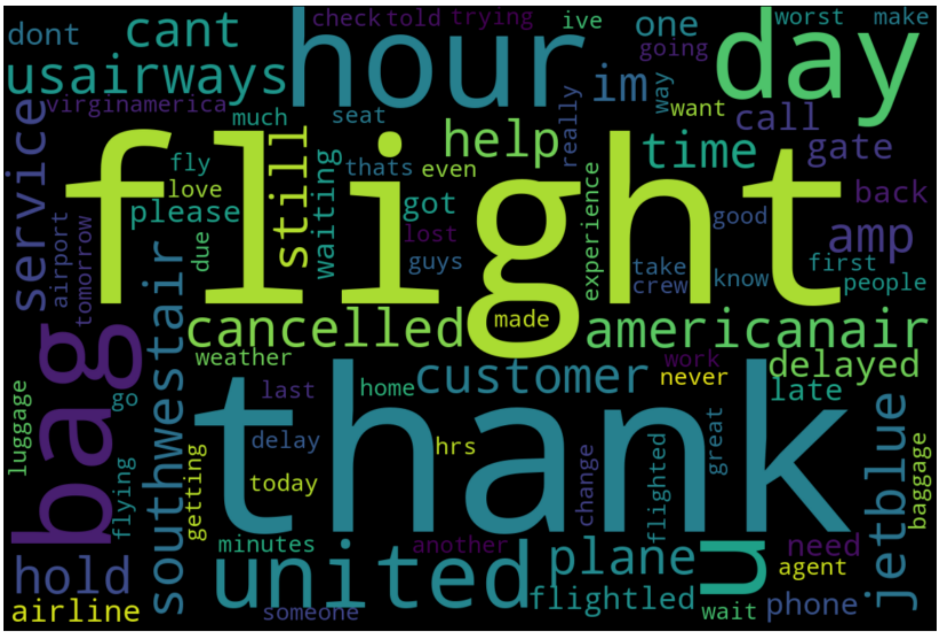

Fig 1. Most frequently used words in the dataset

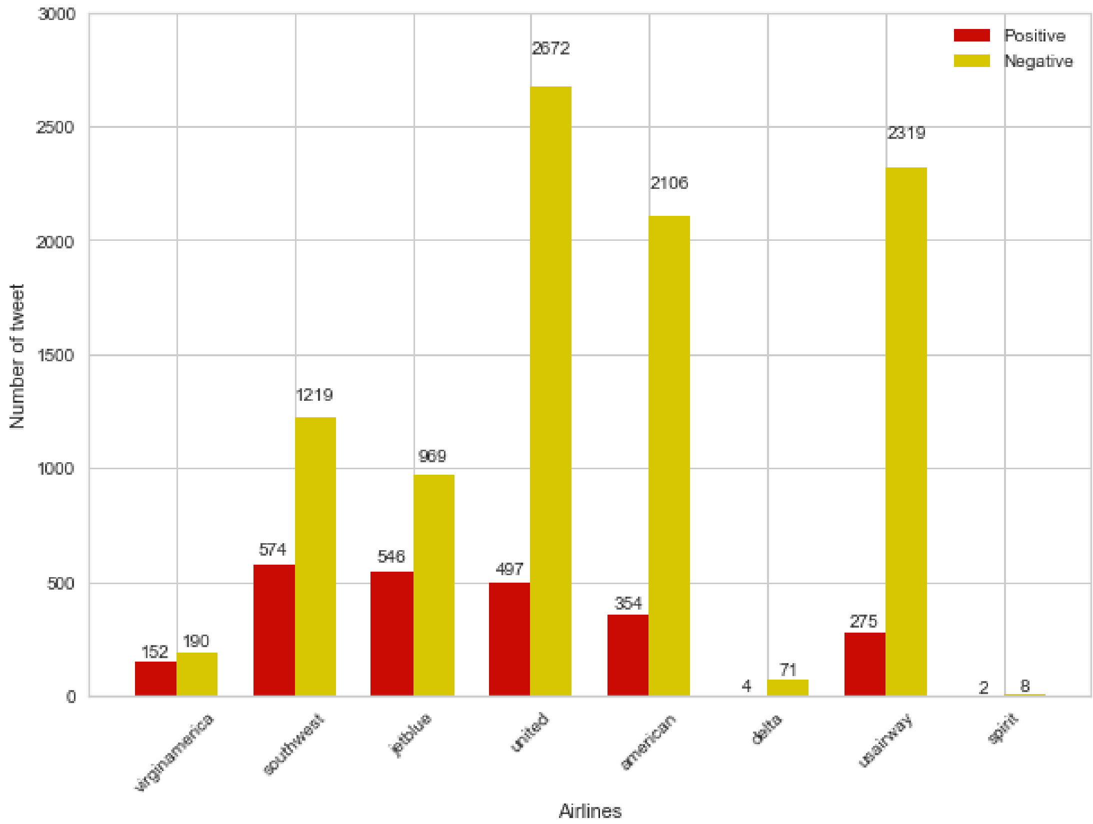


Fig 2. Sentiment Distribution across Airlines

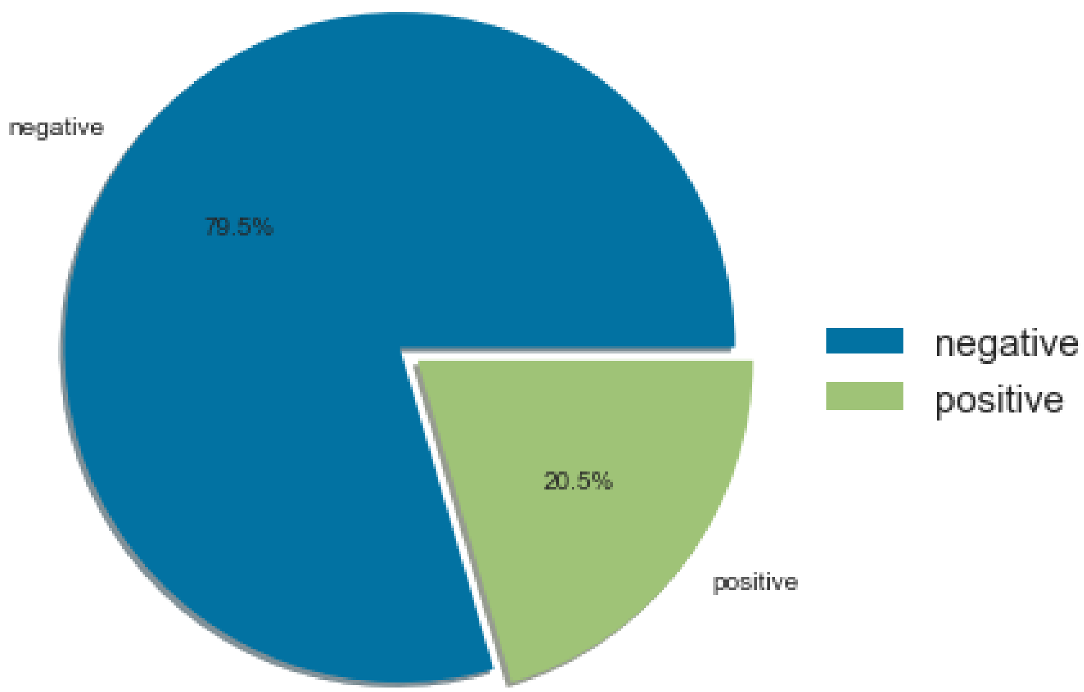


Fig 3. Airlines tweet sentiment distribution

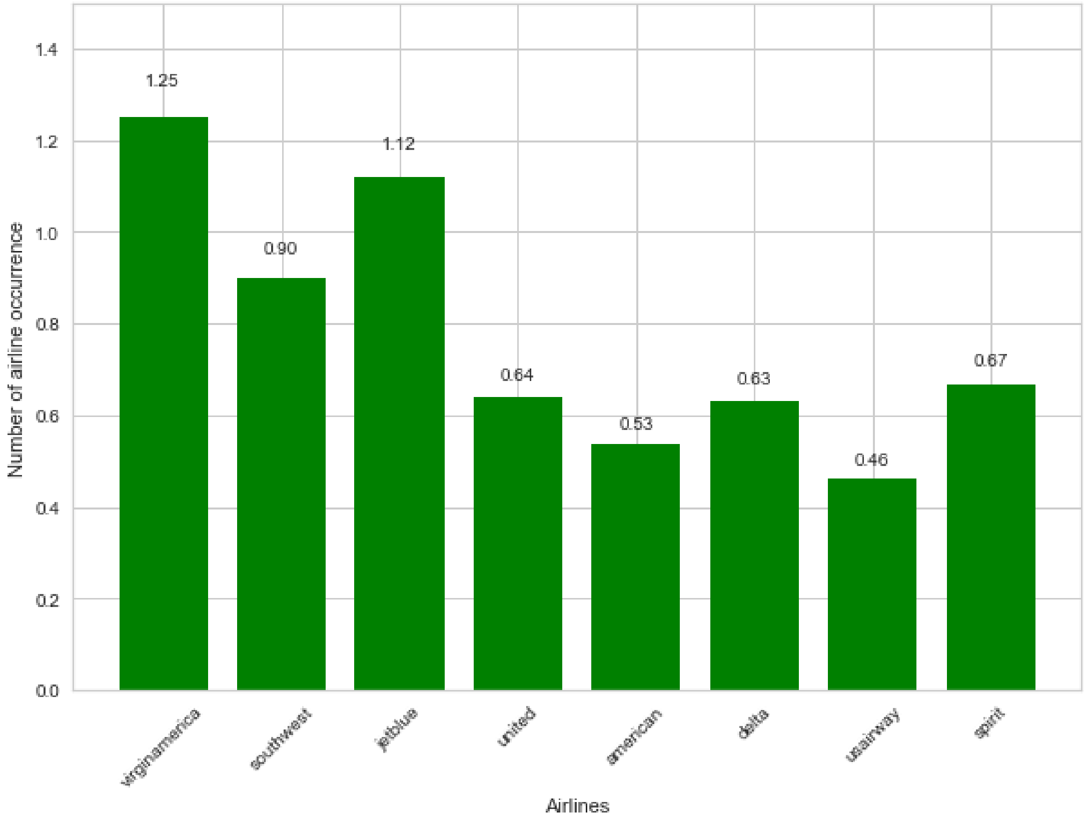


Fig 4. Positive to negative sentiment ratio by airlines

```
In [30]: # call function on negative review
         test_review_neg = "@Indigo you have my money, you change my flight, and don't answer your phones! Any other suggestions so I can
         predict(net, test_review_neg, seq_length)

         Prediction value, pre-rounding: 0.000959
         Negative review detected.

In [31]: # call function on positive review
         test_review_pos = "@JetAirways thank you we got on a different flight to Chicago."
         predict(net, test_review_pos, seq_length)

         Prediction value, pre-rounding: 0.999737
         Non-negative review detected.

In [32]: # call function on neutral review
         test_review_neu = "@SpiceJet i need someone to help me out"
         predict(net, test_review_neu, seq_length)

         Prediction value, pre-rounding: 0.999054
         Non-negative review detected.

In [34]: test_review_neu = "@Vistara This flight provided the best services"
         predict(net, test_review_neu, seq_length)

         Prediction value, pre-rounding: 0.968284
         Non-negative review detected.
```

Fig 5. Output of the RNN  model on testing different tweets

# References

[1] Go, A., Bhayani, R. and Huang, L., 2009. Twitter sentiment classification using distant supervision. *CS224N project report, Stanford*, *1*(12), p.2009.

[2] Saif, H., He, Y. and Alani, H., 2012, November. Semantic sentiment analysis of Twitter. In *International semantic web conference* (pp. 508-524). Springer, Berlin, Heidelberg.

[3] Neethu, M. S., and R. Rajasree. "Sentiment analysis of twitter using machine learning techniques." In *2013 Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT)*, pp. 1-5. IEEE, 2013.

[4] Kumar, A. and Sebastian, T.M., 2012. Sentiment analysis on Twitter. *International Journal of Computer Science Issues (IJCSI)*, *9*(4), p.372.

[5] Sunarya, P., Refianti, R., Mutiara, A., and Octaviani, W., 2019. Comparison of Accuracy between Convolutional Neural Networks and Naïve Bayes Classifiers in Sentiment Analysis on Twitter. *Int. J. Adv. Comput. Sci. Appl*, *10*, pp.77-86.

[6] Jianqiang, Z., Xiaolin, G. and Xuejun, Z., 2018. Deep convolution neural networks for twitter sentiment analysis. *IEEE Access*, *6*, pp.23253-23260.

[7] Severyn, A. and Moschitti, A., 2015, August. Twitter sentiment analysis with deep convolutional neural networks. In *Proceedings of the 38th international ACM SIGIR conference on research and development in information retrieval* (pp. 959-962).