1. EXECUTIVE SUMMARY

Sentiment Analysis is the process of 'computationally' determining whether a piece of writing is positive, negative or neutral. It's also known as opinion mining, deriving the opinion or attitude of a speaker. Data analytics is useful for developing a model for sentimental analysis, and the source of data here are 'tweets' from twitter. Although many techniques can be applied, the one in this paper is "Random Forest". And in order to define the metadata, we have used "Word Cloud" This paper provides an insight into the existing algorithm and it gives an overall summary of the existing work.

2. RESEARCH OBJECTIVE

This work aims at developing a system which tells us about the sentiment prevailing for the particular industries, product etc. In this paper, I have used the random forest as the data mining technique and for visualization, I have applied "Word cloud". The dataset deals with the sentiments of people for US-based airlines industry.

3. Technique Used

Random Forest: Random forests or random decision forests are an ensemble learning method for classification, regression and other tasks that operates by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes or mean prediction of the individual trees.

Word Cloud:

A tag cloud is a novelty visual representation of text data, typically used to depict keyword metadata on websites, or to visualize free form text. Tags are usually single words, and the importance of each tag is shown with font size or color

4. RESEARCH PLAN AND METHODOLOGY

In this study, an efficient machine learning algorithm was chosen from some available algorithms in, 'python is an interpreted, high-level, general-purpose programming language. It was created by Guido van Rossum and first released in 1991, Python's design philosophy emphasizes code readability with its notable use of significant whitespace. The step by step workflow of the complete system has been mentioned below:

- Collecting and selecting the necessary datasets.
- Cleaning the dataset to be used, train various machine learning algorithm.

- Comparison of the data mining algorithm's accuracy and performance in predicting negative, positive neutral sentiment.

.

5. ANALYSIS OF DATA AND FINDINGS

The dataset taken for data mining application includes 1t5 kinds of input which are as follows

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 14640 entries, 0 to 14639
Data columns (total 15 columns):
tweet_id                      14640 non-null int64
airline_sentiment             14640 non-null object
airline_sentiment_confidence  14640 non-null float64
negativereason                9178 non-null object
negativereason_confidence     10522 non-null float64
airline                       14640 non-null object
airline_sentiment_gold        40 non-null object
name                          14640 non-null object
negativereason_gold           32 non-null object
retweet_count                 14640 non-null int64
text                          14640 non-null object
tweet_coord                   1019 non-null object
tweet_created                 14640 non-null object
tweet_location                9907 non-null object
user_timezone                 9820 non-null object
dtypes: float64(2), int64(2), object(11)
memory usage: 1.7+ MB
```

There are various kinds of algorithms are available, which can be applied to the dataset. But in this report I have only applied one machine learning algorithm and for the performance measure I have used the following criteria:

a) Sensitivity
b) Recall
c) F1
d) Support

e) Accuracy

```
[[1715  109   46]
 [ 328  239   47]
 [ 134   64  246]]
              precision    recall  f1-score   support

    negative       0.79      0.92      0.85      1870
     neutral       0.58      0.39      0.47       614
    positive       0.73      0.55      0.63       444

   micro avg       0.75      0.75      0.75      2928
   macro avg       0.70      0.62      0.65      2928
weighted avg       0.73      0.75      0.73      2928

0.7513661202185792
```

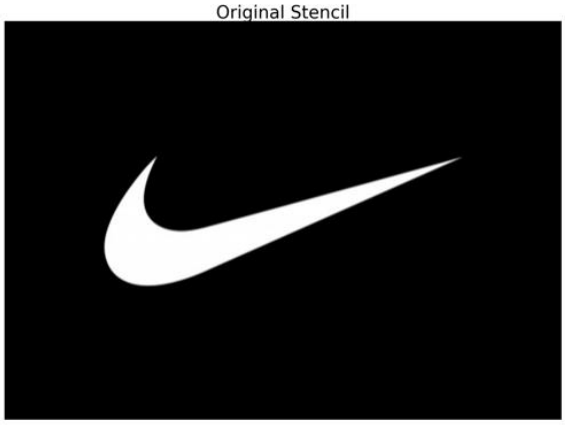
6. CONCLUSION & RECOMMENDATIONS

Thus, we can conclude that the negative sentiment about the particular industry is too high. Thus I would like to recommend the CEO (New comer) to make prior arrangement to tackle these unfavorable situation before entering into this market. And, in case the CEO is an existing player then he should try improve his grievance redressal system, so that he can take advantage existing situation and make profit in long run through taking the advantage of uniqueness

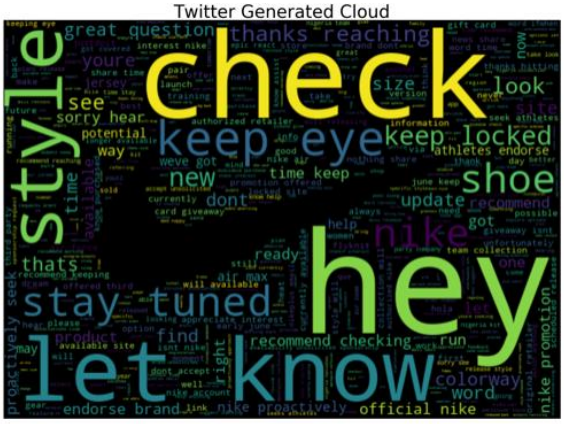
7. Appendix

Word Cloud

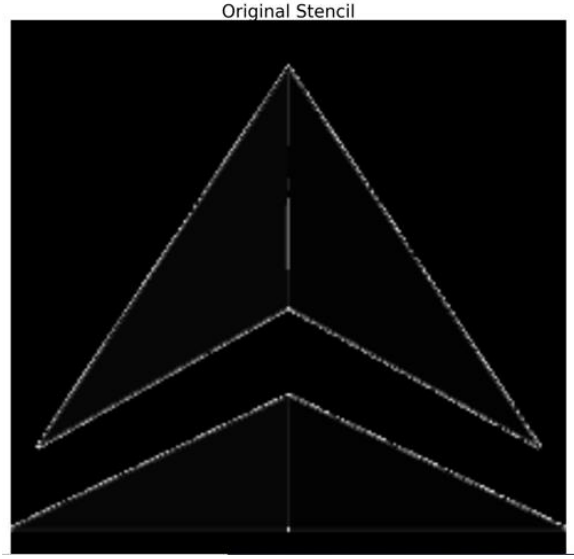
Original Stencil



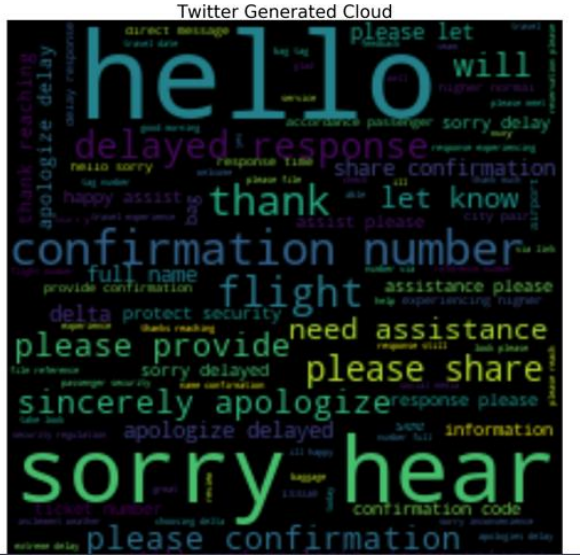
Twitter Generated Cloud



Original Stencil



Twitter Generated Cloud



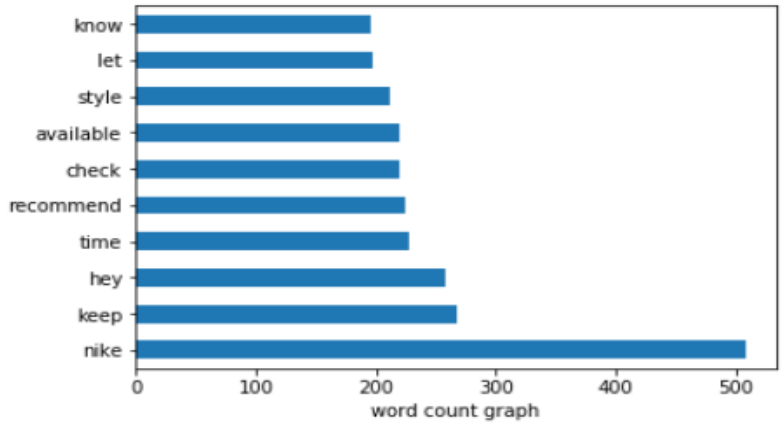
Above word cloud is for Delta airlines and since the log has multiple colors in its logo it is hard to give the word cloud the exact shape like in the case of Nike.

Top 10 most occurring world

Top ten words count

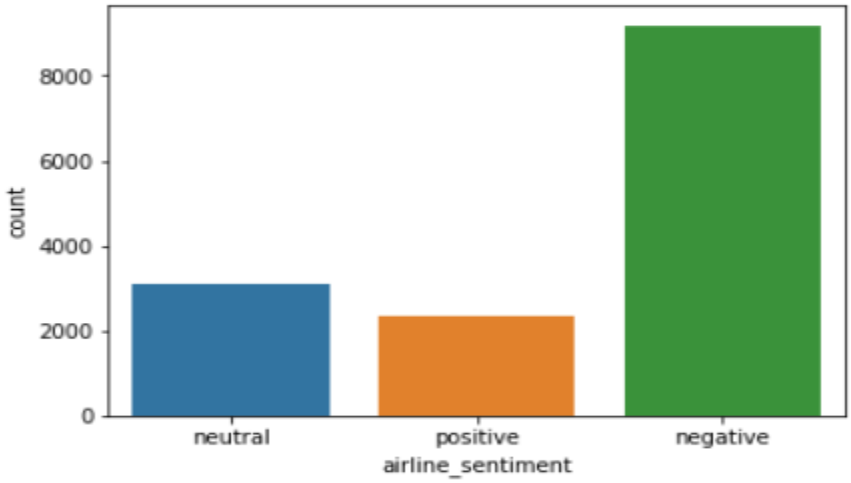
```
top_words= twords.wordc.value_counts()
top_words[:10].plot(kind='barh')
plt.xlabel('word count graph')
```

```
Text(0.5, 0, 'word count graph')
```



Graph showing the sentiment of the US air line industry

```
<matplotlib.axes._subplots.AxesSubplot at 0x232af6e6fd0>
```



Firm wise sentiment graph

```
<matplotlib.axes._subplots.AxesSubplot at 0x232af861f98>
```