Twitter-based Opinion Mining for Flight Service utilizing Machine Learning

Prayag Tiwari
Department of Information Engineering, University of Padova, Italy
E-mail: prayagforms@gmail.com

Hari Mohan Pandey
Department of Computer Science, Edge Hill University, Ormskirk, UK
E-mail: pandeyh@edgehill.ac.uk

Aditya Khamparia
School of Computer Science and Engineering, Lovely Professional University, Phagwara, India
E-mail: aditya.khamparia88@gmail.com

Sachin Kumar
Department of System Programming, South Ural State University, Chelyabinsk, Russia
E-mail: sachinagnihotri16@gmail.com

Abstract. Twitter is one of the most prominent social networking platforms far and wide. Thousands of users utilize Twitter to share their thoughts and views on various topics of interest every day, producing a huge yet increasing amount of data. The rich information in those data, when properly extracted and processed by utilizing machine learning methods, can give rise to effective recommender frameworks for individuals to manage their lives in a much more convenient way. In this paper, we propose the use of two Machine Learning methods that classify the passenger's tweets regarding the airplane services in an effort to understand their opinions. We adopt Random Forest and Logistic Regression to classify each tweet into positive, negative and neutral sentiment. The evaluation of the collected real data demonstrates that these two methods perform better than compared to a predefined benchmark of around 80% by accuracy.

Keywords: sentiment analysis, Random Forest, Logistic Regression.

1 Introduction

These days, big companies are investing in far more time and energy than before on enhancing consumer loyalty. This can create more opportunities to interact

with customers and get their feedback, with the very belief that it will bring about an increase in the income and efficiency, and accelerate the development of the company.

A crucial but challenging step is to automatically analyze the customer feedback by extracting useful information from it. Based on the analysis of customer feedback, many user-centric tasks can be addressed, and sentiment classification is a very important one among them. By using the extracted patterns companies are seeking to recognize the polarity of the opinion, i.e., whether the opinion is positive or negative, or the customer emotion, i.e., happy, excited or sad. Companies use these polarity values to achieve an overall understanding of customers' sentiments, and in turn, improve customer services.

Sentiment analysis or opinion mining helps to figure out such a question like what others are saying, what they mean, how they are saying, etc. Sentiment analysis is the domain to understand the emotions by utilizing several software's. In today scenario, we utilize natural language processing, text analysis for feature extraction and to figure out the sentiment of text into a negative, positive or neutral class.

The sentiment is another name for the view and assessment that is held or communicated. The sentiment may depict euphoria, bliss, bitterness or outrage. Travelers communicate their sentiments on Twitter. Each trip on carriers can bring either delight or uneasiness amid movement for any traveler. If a passenger is not happy with the services, his tweet will show distress. On the other hand, if he is very much happy with the services, he will demonstrate a sentiment of joy in his tweet. Figure 1 depicts a furious tweet by a passenger on British Airways. The British aviation routes considered it as an exceptionally important twitter and settled the issues of the particular passenger. Another negative tweet about the Indigo carriers is shown in Figure 2. It can be seen from the figure 2 about sarcasm that passenger baggage is sent to Hyderabad and passenger is flying to Calcutta. Some tweets are difficult to classify as it can be seen from this tweets that it means negative from human view but difficult to classify in the case of the machine.

One solution is to use these tweets to understand the problems of the travelers during the journey and improve it by the time. However, around more than millions of people are traveling on the flights every day and expressing their opinions on Twitter, and sometimes their tweets are very general in nature. As a result, it is extremely hard to tweet about a particular flight and time in a short period of time. Therefore, the idea is to analyze all the tweets about any airline and try to understand the sentiment of travelers.

Another challenge is that the size of the dataset and the number of tweets are often large. Therefore, we need a technique which is efficient enough to deal with large datasets. Machine learning is such a technique, which is efficient in handling

large dimensional data. It can be viewed as a set of methods that can be used to extract hidden and meaningful information from large datasets.

Don't fly with @British_Airways. They can't keep track of your luggage.

↗ Promoted by
9/2/13, 8:32 PM

Figure 1: A tweet representing discomfort during travel [1]

@IndiGo6E
Thank you for sending my baggage to Hyd and flying me to Calcutta at the same time. Brilliant service. #DieIndigo

Figure 2: A tweet representing discomfort during travel [2]

Machine learning is playing an important role in transportation [11, 12, 13], bio-informatics [9], computer vision [8], social media [5-10] and healthcare analytics [10]. Therefore, the power of machine learning is known to researchers and practitioners. In this study, we apply machine learning to analyze Twitter data. The main motivation of this work is to provide a better model for predicting user sentiment and help airlines to improve customer service and avoid passengers from facing such problems in the near future. This study can be beneficial for airlines services in order to improve their customer service. The rest of the paper is organized as follows: in section 2, related work is demonstrated in a literature survey. Section 3 describes the technical background. Section 4 provides a description of the dataset. Section 5 presents and discusses the obtained results and Section 6 gives the conclusion.

2 PREVIOUS WORKS

The author Kusen et al. [3] have analyzed and extracted twitter dataset consisting of 343645 tweets about 2016 Austrian presidential election. This analysis amalgamated approaches from sentiment analysis, network science, and bot detection. It was shown that the immediate relationship between's the winner of the 2016 Austrian presidential races was more famous and had a high impact on Twitter than other rivals.

Ahmed et al. [18] have demonstrated how the first time twitter utilized as a campaign tool in the Indian election 2014 by different parties. They demonstrated computer-aided and multi-level manual analysis of 98363 tweet messages by 11 parties during the campaign. It had a high impact on twitter of winning party than other parties.

Stigleitz et al. [20] examined whether opinion persisting in online networking content is related to a client's data sharing coordination. They conducted an examination with regards to political correspondence on Twitter. On the basis of two dataset collections of about 165,000 tweets altogether, they found out that candidly charged Twitter messages had a tendency to be retweeted all the more regularly and more immediately contrasted with biased ones. As a general suggestion, organizations should give careful consideration to the examination of opinion identified with their brands and items in social networking correspondence, in addition to planning promoting content that triggers emotions.

Priyanga et al. [21] investigated the objection resolution experience of passengers of U.S. aircraft, by utilizing an interesting data collection amalgamating both customers– brand cooperation's on Twitter and how clients felt toward the end of these associations. They located that objection Customer who is more dominant in online networking communities will probably be fulfilled. Customers who have beforehand objection to the brand via social networking media and customers who grumble about process-related instead of result related issues are less inclined to feel better at last. To the best of our insight, this examination is the first to recognize the key factors that shape client sentiments toward their brand– client communications via social networking media. Their outcomes give useful direction to effectively settling clients' objection using social networking field that expects exponential development in the coming decade.

Park et al. [23] showed social networking examination utilizing Twitter data alluding to cruise travel. This examination likewise incorporated an inside and out an investigation on tweets by three kinds of group users: private, commercial and blogs. The outcomes demonstrated that not exclusively were words identified with industry, travel, emotions, and destination most often utilized as a part of organizing tweets, but also proficient bloggers, cruise lines, celebrities and travel

organizations really drove significant subgroups on cruise themes on Twitter. On the basis of such outcomes, this examination gives attainable marketing approach.

3 PROPOSED WORK

In this section, our proposed model consists of several steps like preprocessing, feature extraction, etc in order to train the model and use the test dataset to check the evaluation metric on the test dataset. Precision, F1-measure, and Recall are used as an evaluation metric.

A. System Architecture

Proposed architecture can be seen in the figure no. 3 that how flow started of our model from the dataset, text pre-processing, feature extraction, a division of dataset into training and testing set, the trained model then tested on the test dataset.
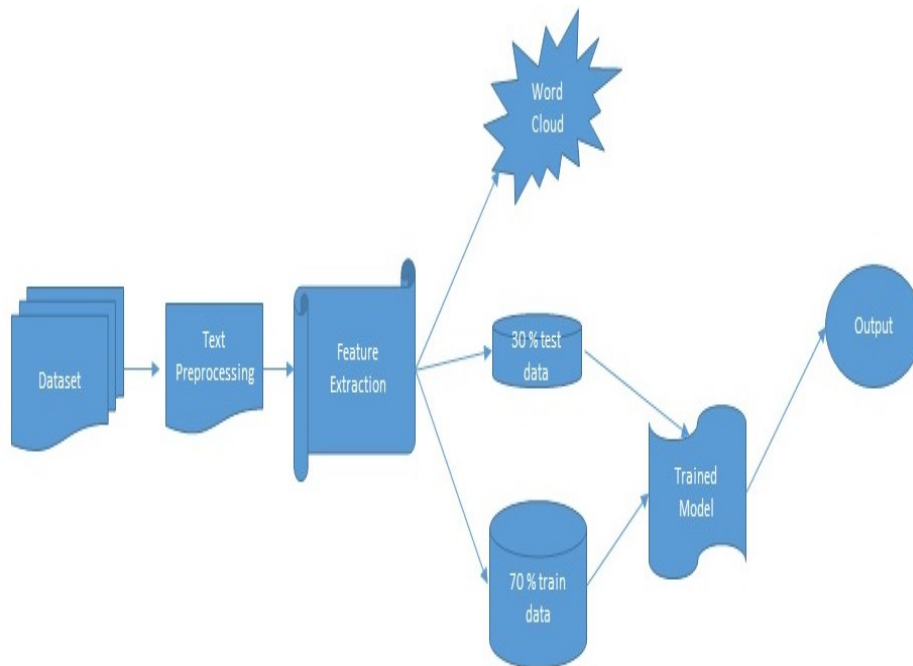

Figure 3: Architecture of Proposed Sentiment Analysis Model

B. Text Preprocessing

As a pre-processing step, we do a basic statistical analysis on the collected data. The statistics include the number of words (denoted as word_counts), the number

of hashtags (denoted as hashtag_counts), and counts for other punctuation marks. The distribution of those textual variables over the three sentiment classes is depicted in Figure 9.

We then remove the hashtags, mentions, URLs etc= to make text data more clean for further analysis. We also removed punctuations, stop words and digits. Finally, we stem words and convert them to lowercase. This is the standard procedure for pre-processing textual data.

The examples of tweets after pre-processing can be seen in Figure 7.

C. *Random Forest*

Decision Trees are the most widely used machine learning methods. Random Forest provides an effective way of averaging several decision trees, trained in different segments of the same training dataset with the aim to deteriorate the variance and provide a stable and accurate prediction. Random forest could be an ensemble learning procedure for regression, classification, and elective undertakings, which is achieved by building a large group of decision trees at training phase and provoking the classes which are the model for the mean prediction (regression) or classifications (classes) of the distinctive trees. In a distinct computation, classification is implemented recursively until every leaf is pure. The aim is to dynamically predict the best decision tree until it catches up the adaptability, precision, and balance. There are three measures in the decision tree which are described here:

$$Entropy = \sum_j P_j \log_2 P_j \qquad (1)$$

$$Gini = 1 - \sum_j P_j^2 \qquad (2)$$

$$Classification\ Error = 1 - maxP_j \qquad (3)$$

Where P_j is the probability of class j.

The algorithms starts as follows: we pick a bootstrap observation from the S in which $S^{(i)}$ represents the ith bootstraps for every tree in the given forest. Then train the decision tree utilizing a revised decision tree algorithm. The revised decision tree algorithms as follows: in contrast of analyzing all feasible feature split, some random features $f \subseteq F$, at every node of the tree where F is the feature sets. The given node split on the top features in f comparably than selecting F. In this, f is much more compact and smaller than F. The most challenging task is to choosing on which feature to split in the decision tree learning that is why making narrow the feature set makes faster learning. The following pseudocode can be seen in figure 4.

```
Algorithm    Random Forest
```
Precondition: A training set $S := (x_1, y_1), \ldots, (x_n, y_n)$, features F, and number of trees in forest B.

```
1  function RANDOMFOREST(S, F)
2      H ← ∅
3      for i ∈ 1,...,B do
4          S⁽ⁱ⁾ ← A bootstrap sample from S
5          hᵢ ← RANDOMIZEDTREELEARN(S⁽ⁱ⁾, F)
6          H ← H ∪ {hᵢ}
7      end for
8      return H
9  end function
10 function RANDOMIZEDTREELEARN(S, F)
11     At each node:
12         f ← very small subset of F
13         Split on best feature in f
14     return The learned tree
15 end function
```

Figure 4 Random Forest Pseudocode

D. Logistic Regression

Logistic Regression is a statistical method for investigating a dataset in which there are at least one or more than one independent variables that decide a result. The result is estimated with a dichotomous variable (in which there are just two conceivable results).

The objective of logistic regression is to locate the best fitting model to depict the connection between the dichotomous feature and the set of independent factors.

Our Hypothesis function can be written like as given below,

Our Hypothesis function can be written like as given below,

$$Y = W^T X \qquad (4)$$

A sigmoid function is implemented across the notable hypothesis function to keep into the range of (0, 1). The sigmoid function can be described as,

$$sg(y) = 1/(1 + e^{-y}) \qquad (5)$$

So our new hypothesis is

$$sg(y) = sg(W^T X) = 1/(1 + e^{-WTX}) \qquad (6)$$

Boundary Estimation:

Our new hypothesis function provides us the values in between 0 and 1so it can be clarified probability of y would be 1 for given X and this can be written in this form,

$$sg(y) = P(y = 1/x, W) \tag{7}$$

Cost Function:

Taking a square error function does not work from the transformed hypothesis function so we make a new form of cost function which is as follows,

$$E(sg(W, x), y) = -log(1-sg(W, x)) \ if \ y = 0$$
$$E(sg(W, x), y) = -log(sg(W, x)) \ if \ y = 1$$

Therefore, the mean of cost function will be as follows,

$$H(W) = \frac{1}{m}\sum_{i=1}^{m} E(sg(W, x_i), y_i) \tag{8}$$

Parameter Estimation:

We utilize an iterative approach known as Gradient Descent to enhance the parameters across every step and reduce the cost function to the most feasible value. Gradient Descent requires a convex cost function to avoid getting stuck in a local minimum at the optimization stage. We begin with irregular parameter values and update their values at every stage to reduce the cost function to some extent until we reach the lowest point or equivalently there are not any changes to the value of the target function. The gradient descent step is as follows,

$$\beta_{(i+1)} = \beta_i - p\frac{\delta H(W)}{\delta \beta_i} \tag{9}$$

For every i=1, 2, 3 ...., n and p is the learning rate controlling the speed that it moves across the slope on the curve to reduce the cost function.

Above process cab be shown in the pseudocode can be seen in figure 5 with L1 regularization. The procedure starts with providing input dataset D with corresponding labels and iteration numbers. In this, w_h is the temporary variable. Our algorithm start working as mentioned in the pseudocode.

Algorithm Logistic Regression with L1 regularization

1: **procedure** STOCHASTICGRADIENTDESCENT(D, *Labels, Iter*)
Input: *Dataset D, Labels of Dataset, Iteration num*
Output: optimal weight of logistic regression
2:   $w \leftarrow [1, 1, ..., 1]$
3:   Initialize q_i with zero for all i
4:   **for** $k = 1 \rightarrow Iter$ **do**
5:     $chooseData = D$
6:     **for** $i = 1 \rightarrow m$ **do**
7:       $\gamma \leftarrow Learning\ Rate$
8:       $\lambda \leftarrow Regularization\ Lambda$
9:       $u = u + \gamma\lambda$
10:      Select a index of chooseData idx randomly
11:      $x \leftarrow chooseData[idx]$
12:      $del\ chooseData[idx]$
13:      **for** $i \in featuresinsamplex$ **do**
14:        $w_i = w_i - \gamma\frac{\partial loss(w,x_i)}{\partial w}$
15:        $wh \leftarrow w$
16:        **if** $w_i > 0$ **then**
17:          $w_i \leftarrow \max(0, w_i - (u + q_i))$
18:        **else if** $w_i < 0$ **then**
19:          $w_i \leftarrow \min(0, w_i + (u - q_i))$
20:        **end if**
21:        $q_i \leftarrow q_i + (w_i - wh)$
22:      **end for**
23:    **end for**
24: **end for**

Figure 5 Logistic Regression Pseudocode

E. *Evaluation Metric*

In order to measure the accuracy of classification [4], we used different parameters such as Recall, Precision, and F-measure. Recall can be regarded as the measure of completeness whereas Precision can be seen as a measure of exactness. Formally, precision can be defined as the ratio of correctly classified instances of one class and a total number of instances classified in the same class, whereas recall is the ratio of correctly classified instances of one class and overall instances of the same class. Both precision and Recall can be calculated using the confusion matrix. Confusion matrix represents the number of correctly classified and incorrectly classified instances of all classes. Using the confusion matrix, all performance evaluation measures can be calculated. For a twitter dataset with a binary classification problem, if the total 600 tweets are classified to one class, among which 500 of them are correctly classified, and the total number of tweets in this

class are 700. Then, the precision of the classifier is 500/600= 83.3%, and the recall of the classifier is 500/700=71.4%. The Recall and Precision are integrated to develop a new measure known as F-measure or F-score. The formula to calculate F-measure is given in Equation 12.

$$Precision = \frac{TP}{TP+FP} \qquad (10)$$

$$Recall = \frac{TP}{TP+FN} \qquad (11)$$

$$F\text{-}measure = \quad 2\left(\frac{Precision*Recall}{Precision+Recall}\right) \qquad (12)$$

Where TP is True Positive, TN is True Negative, FN is False Negative and FP is False Positive.

4 EXPERIMENTS

A. *Experiment Introduction*
In this study, we performed classification based on negative, positive and neutral categories. For classification purpose, logistic regression and random forest classifiers have been used. Our proposed model can classify the customer comments with higher accuracy than previously proposed models.

B. *Experimental Data*
In this study, we experiment on the US Airlines 2016, which contains 14500 passenger tweets. Since the number of original features is too large, we manually select the textual based features, because are easily accessed by passengers. As can be seen from Figure 6, the class labels are highly unbalanced. The dataset can be found in [22].
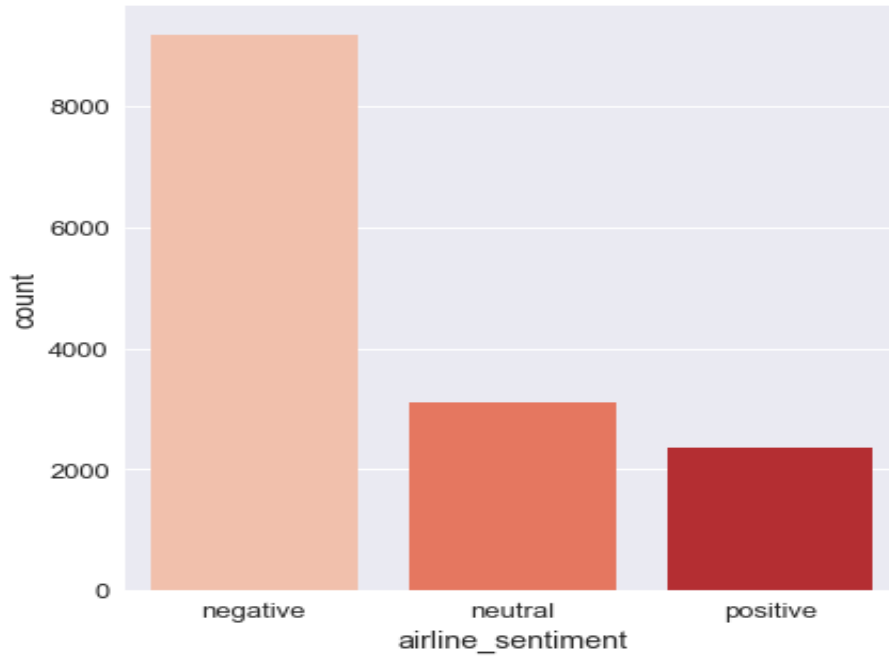
Figure 6: Distribution of Class Labels

```
13892    pilot told us would release bags well offer hotel vouchers neither happened
8817     always happy help
6270     wanted southwest know think great used anymore nothing look
420      helping step game tindertips tinderchamp
7170     no come
7529     landing usual great flight wiyh great crew hello sunny west palm beach jetbluerocks
10131    please thank mellie cae tammy baggage claim clt excellent customer service day complaint
609      worstunitedflightsever ua iad las mechanical problems switched aircraft delayed hours
12166    sounds like date
4232     next flights miami another airline
Name: text, dtype: object
```

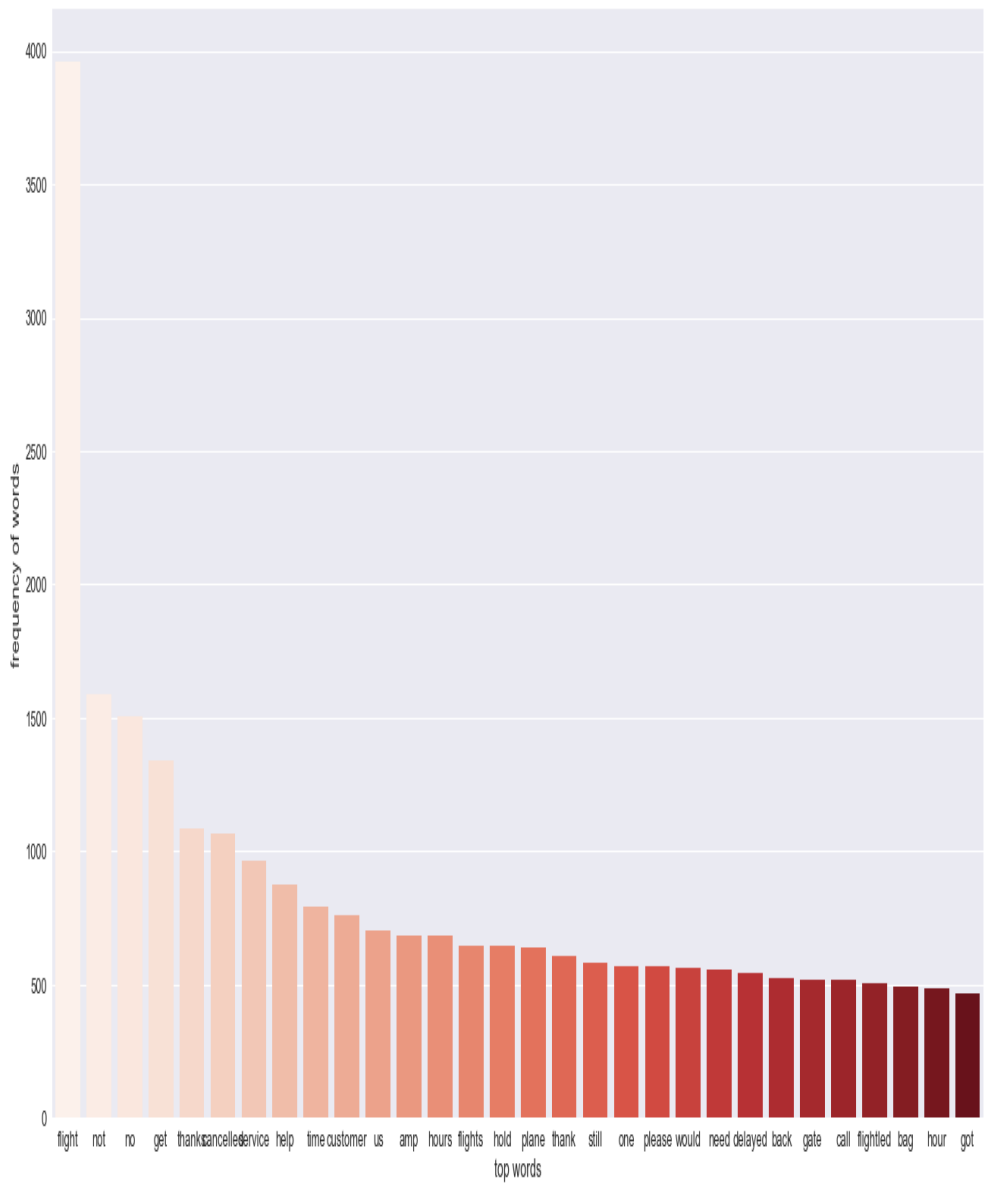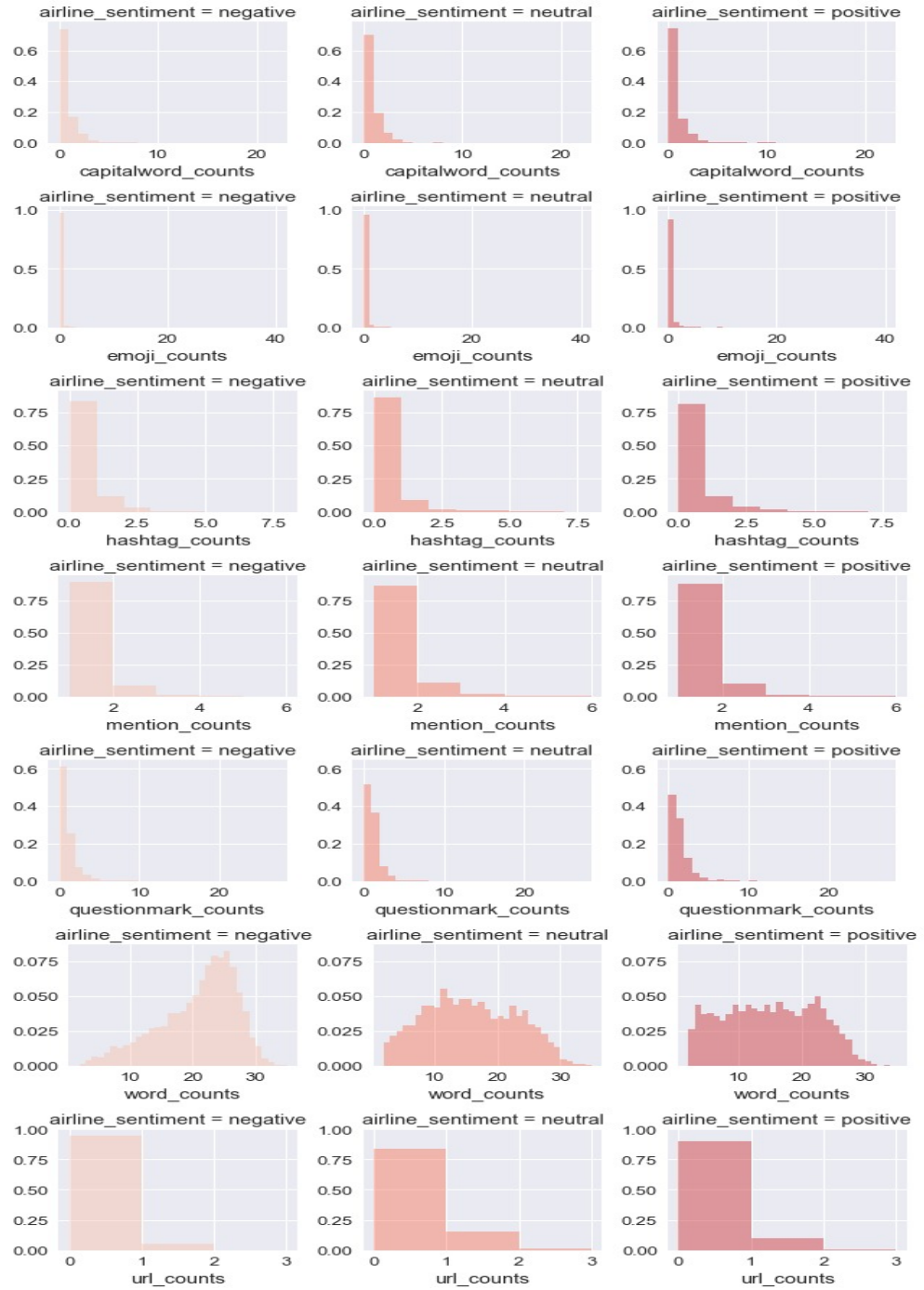Figure 7: Top 10 Sample of cleaned Data

Figure 8 Top 30 frequent words

Figure 9 Distribution of Text variables

After the preprocessing step, we identified the top 30 frequent words in the dataset, which is shown in Figure 8.

C. Experimental Process

For further evaluation, it is necessary to have test data that could be helpful to evaluate several measures of our model. Data was divided into 70 percent train and 30 percent test set Text count variable has been combined with cleaned data to create a data frame.

For opting better parameters, it is needed to assess on a different validation from training. By utilizing just a single validation set one might not deliver reliable validation result. To get a more precise estimation, cross-validation is performed.

In this study, we conduct k-fold validation on the data at hand and utilize GridSearchCV to search for the best-performed parameter combination. We select precision as the metric for optimization for both logistic regression and Random Forest classifiers. In order for bag-of-word features to be properly fed into classifiers, we use CountVectorizer to transform words into vectors.

The word cloud in Figure 10 gives a decent visual depiction of the word recurrence for each kind of opinion, in which the left one corresponds to the positive opinion and the right one the negative. The span of the word relates to its recurrence across all tweets. This figure gives us a rough idea of what passengers are discussing. For instance, for negative opinion, passengers appear to gripe about delayed of flight, cancellation of flights, the low-quality of the flight service, the hours holding up and etc. Be that as it may, for positive opinion, passengers are thankful and they discuss extraordinary administration/flight. A cloud of the word has been mentioned in Figure 10 to visualize those positive and negative tweets more properly.

Several other approaches have been used but Logistic Regression and Random Forest gave better result on train and test dataset. The main advantage of using Random Forest for text classification is that it ensemble multiple and different kinds of decision trees and utilize an assortment of the different trees to improve the result of the model.

D. Experimental Results

Our proposed model provided this result on the test dataset. As it can be seen that in the case of positive, negative or neutral categories, our proposed model can classify with high precision, recall and f-measures. After applying logistic regression and random forest on the dataset, the performance values are recorded in table 1 and table 2.

Table 1: Evaluation Metric of Logistic Regression

	Precision	Recall	F1-score	Support
Positive	0.80	0.74	0.77	233
Neutral	0.73	0.53	0.62	320
Negative	0.83	0.93	0.88	911
Avg/Total	0.81	0.81	0.80	1464

Figure 10: Word Cloud of Positive and Negative Opinion

Table 2: Evaluation Metric of Random Forest

	Precision	Recall	F1-score	Support
Positive	0.82	0.74	0.78	233
Neutral	0.75	0.60	0.65	320
Negative	0.84	0.95	0.90	911
Avg/Total	0.82	0.82	0.81	1464

As from above tables, it can be seen that both classifiers performed very well, but Random Forest works better as compared to logistic regression, with a consistent higher value in Precision, Recall, and F-score than logistic regression. The 82 % accuracy value on the test data is superior to our prefined target, which is to the maximum value we can achieve by setting the prediction labels for all samples to be the dominant class. The precision is also high for all the three classes and the recall rate is relatively low for the neutral classes.

For better illustrating the effectiveness of our proposed models, we also present examples of some negative and positive tweets classified by our proposed approaches.

Table 3: Example of our test data classified into positive and negative tweets

Negative Tweets	Positive Tweets
" @united It's a shame choosing #United may be the difference between reuniting with aging friends and never seeing them again #PoorService",	"@united Big thanks to Ms. Winston for assisting me over the phone with a baggage claim issue today. She really went the extra mile!"
"@united flight attendant doesn‚Äôt understand not understanding English doesn‚Äôt mean they are deaf. Stop yelling English slowly at them"	"@United THANK U! Secured room for the night Thx to VERY helpful customer service rep N. Dorns. I thanked her. Can u 2? #goodenoughmother"

Model-predicted accurately like Negative, Negative in the first column and Positive, Positive for the second column based on the test set.

5   Conclusion and Future Works

This study tackles the sentiment classification problem by utilizing two machine learning models. On the collected data, we achieve an accuracy of 82%. This study has impacts on the aviation industry in that it provides an effective and efficient way to monitor the passengers' sentiments for aviation companies to improve their service. For future work, we would like to conduct a deeper analysis of the data and extract more useful information for providing recommendations for several airplane organization and passengers. It would be also used to use a bigger dataset than the used dataset because a larger dataset may provide some

better result than used one. The author would like to use also deep learning models and especially focus on how to identify the sarcasm because there are several sentences seems positive but their meaning is negative so this is a really big issue to sort out and currently, mo9dels are not efficient to sort out.

References

1. http://business.time.com/2013/09/03/man-spends-more-than-1000-to-call-out-british-airways-on-twitter/ accessed on 16.12.2017 at 1115 hours.
2. http://www.thegenxtimes.com/trends/man-tweets-indigo-about-misplaced-luggage-glad-to-hear-that-replies-indigo/ accessed on 16.12.2017 at 1124 hours.
3. Kušen, Ema, and Mark Strembeck. "An Analysis of the Twitter Discussion on the 2016 Austrian Presidential Elections." arXiv preprint arXiv:1707.09939 (2017).
4. Gräbner, D., Zanker, M., Fliedl, G., & Fuchs, M. (2012). Classification of customer reviews based on sentiment analysis (pp. 460-470). na.
5. Zhang, L., Sun, Y., & Luo, T. (2016, December). A framework for evaluating customer satisfaction. In Software, Knowledge, Information Management & Applications (SKIMA), 2016 10th International Conference on (pp. 448-453). IEEE.
6. Han, J. and Kamber, M.P., Data Mining: Concepts and Techniques, Morgan Kaufmann Publishers Inc. San Francisco, CA, USA, 2005 ISBN:1558609016.
7. Misopoulos, Fotis, et al. "Uncovering customer service experiences with Twitter: the case of the airline industry." Management Decision 52.4 (2014): 705-723.
8. Bollen, J., Mao, H., & Pepe, A. (2010, October). Determining the Public Mood State by Analysis of Microblogging Posts. In ALIFE (pp. 667-668).
9. Tiwari, P., Mishra, B. K., Kumar, S., & Kumar, V. (2017). Implementation of n-gram methodology for rotten tomatoes review dataset sentiment analysis. International Journal of Knowledge Discovery in Bioinformatics (IJKDB), 7(1), 30-41..
10. Efron, Miles. "Information search and retrieval in microblogs." Journal of the Association for Information Science and Technology 62.6 (2011): 996-1008.
11. Tiwari, P., Kumar, S., & Kalitin, D. (2017, March). Road-User Specific Analysis of Traffic Accident Using Data Mining Techniques. In International Conference on Computational Intelligence, Communications, and Business Analytics (pp. 398-410). Springer, Singapore.
12. Tiwari, P., Dao, H., & Nguyen, G. N. (2017). Performance evaluation of lazy, decision tree classifier and multilayer perceptron on traffic accident analysis. Informatica, 41(1), 39.

13. Kumar, Sachin, Prayag Tiwari, and Kalitin Vladimirovich Denis. "Augmenting Classifiers Performance through Clustering: A Comparative Study on Road Accident Data." International Journal of Information Retrieval Research (IJIRR) 8.1 (2018): 57-68.

14. Yee Liau, Bee, and Pei Pei Tan. "Gaining customer knowledge in low cost airlines through text mining." Industrial Management & Data Systems 114.9 (2014): 1344-1359.

15. Fotis Misopoulos, Miljana Mitic, Alexandros Kapoulas, Christos Karapiperis, (2014) "Uncovering customer service experiences with Twitter: the case of airline industry", Management Decision, Vol. 52 Issue: 4, pp.705-723, https://doi.org/10.1108/MD-03-2012-0235.

16. Park, M., Cha, C., & Cha, M. (2012, August). Depressive moods of users portrayed in Twitter. In Proceedings of the ACM SIGKDD Workshop on healthcare informatics (HI-KDD) (Vol. 2012, pp. 1-8). New York, NY: ACM..

17. Breen, Jeffrey Oliver. "Mining twitter for airline consumer sentiment." Practical text mining and statistical analysis for non-structured text data applications 133 (2012).

18. Ahmed, Saifuddin, Kokil Jaidka, and Jaeho Cho. "The 2014 Indian elections on Twitter: a comparison of campaign strategies of political parties." Telematics and Informatics 33.4 (2016): 1071-1087.

19. Alessandro Bessi and Emilio Ferrara. Social bots distort the 2016 U.S. Presidential election online discussion. First Monday, 21(11), November 2016.

20. Stieglitz, Stefan, and Linh Dang-Xuan. "Emotions and information diffusion in social media—sentiment of microblogs and sharing behavior." Journal of management information systems 29.4 (2013): 217-248.

21. Gunarathne, Priyanga, Huaxia Rui, and Abraham Seidmann. "Whose and what social media complaints have happier resolutions? Evidence from Twitter." Journal of Management Information Systems 34.2 (2017): 314-340.

22. https://data.world/crowdflower/airline-twitter-sentiment

23. Seunghyun "Brian" Park, Chihyung "Michael" Ok, Bongsug "Kevin" Chae. (2016) Using Twitter Data for Cruise Tourism Marketing and Research. Journal of Travel & Tourism Marketing 33:6, pages 885-898.