International Journal of Aquatic Science
ISSN: 2008-8019
Vol 12, Issue 02, 2021

A Survey Paper On Airlines Sentiment Analysis

Miss. Pragati O. Kadu[1], Miss. Prarthana R. Bhoyar[2], Miss. Sayali A. Katolkar[3], Miss. Shivani S. Jaiswal[4], Mr. Swarup G. Thamke[5], Mr. Vyankatesh K. Barlawar[6], Prof. A.V. Mahalle[7]

[1,2,3,4,5,6]*Students B.E. Computer Engineering (JCET), Jagadambha College of Engineering and Technology, Yavatmal, India*
[7]*Department of Computer Engineering, Jagadambha College of Engineering and Technology, Yavatmal*

Abstract* – *Social medias today are an integral part of everyone as a result it is abundant in user's opinions. Airlines Sentiment Analysis is a sentiment analysis technic about the opinions or problems of each major U.S airline. To understand the customer's voice, Twitter data was scrapped from various years and contributors asked to first clarify positive, negative and neutral tweets followed by categorizing negative (such as "late flight" or "rude service"). The data were provided on a slightly reformatted version of original source. It includes both CSV and SQLite database files. It all contains sentiment tweets in set that was positive, negative, or neutral for six US airlines. People around world are more actively by using social medias such as Facebook, Twitter, etc. They share information, opinions and ideas using social media. The business communities have become more aware of these social medias so see which information have in their favor. This kind of sentiment analysis makes different airlines to understand customer feedback in a very constructive manner. Airlines companies can improve customer services and also it will make market strategies a better. Based on this kind of analysis companies will know what their customer thinks to take a competitive advantage.

1. INTRODUCTION

The main purpose behind the Airlines sentiment analysis system is to mine or study customer's opinion for major airlines. As a result, it is abundant in user's opinions. It analyse the brand's specific opinions that can inform companies level of satisfaction within consumers. It focuses on the analysis of the tweets related to airlines. It makes airlines to understand customer's feedback in a constructive manner. The sentiment analysis uses Naïve Bais and VADER algorithm that uses a collection of words associated with the sentiments. The sentiment in the result will be Positive, Negative and Neutral. The tweets are stored in .csv file. Using sentiment analysis on tweets the airlines services can whether have either a good or bad opinions about their services. The tweets for a specific airline can be analysed over a period of time which result the score of that airline.

Proposed Work

Since customer express their thoughts and feelings are more openly than before, sentiment analysis is becoming essential tool to monitor and understand that sentiment. So far, we have trained our model with tokenized data and the model gave a performance of 78%. To improve

the performance of Naïve Bayes classifier, we are going to apply other text processing methods and reduce the size of dictionary.
This model will be useful for both passengers and airline companies as they can see what user thinks about the particular airline. It will provide the mined data from social media so that no need of checking different review sites or any other sources as everything will be here in consistent and sorted manner. It will contribute the experience of users towards their good and bad flying experience so that other can decide which airline will be suitable for the user.

Features

- System controlled by admin.
- Data can be added manually by admin.
- Data will be summarized in real time.
- Graphical Representation.
- User personalization. This feature refers to collecting user Automated System.
- Sorted Data.
- Easy to analyze.

Literature Review

Yasmin Yashodha (2012) examined the extensive strategic analysis of Air AsiaBerhad that has enabled it to sustain its competitive advantage as Asia's leading low-cost carrier (LCC). The study demonstrates the diverse business-level, corporate level and competitive strategies of Air AsiaBerhad, played crucial roles in the LCC to successfully penetrate the under-served market segment of the airline industry within the ASEAN region. An in-depth analysis using a wide array of academic resources, relevant financial, legal and management resources and authorized websites, including face-to-face interviews were used to provide a more consequential comprehension on the varied business and international strategies that were implemented by AirAsiaBerhad. This research exhibits critical analysis pertaining to the current macro environment of the aviation industry which includes the PESTEL framework and Porter's Industry Analysis. The competitive environment analysis for Air AsiaBerhad is thoroughly scrutinized to examine the driving determinants that attributed to the organization's competitive advantage in the industry.

Yun Wan and QigangGao (2015), In airline service industry, it is difficult to collect data about customers' feedback by questionnaires, but Twitter provides a sound data source for them to do customer sentiment analysis. However, little research has been done in the domain of Twitter sentiment classification about airline services. In this paper, an ensemble sentiment classification strategy was applied based on Majority Vote principle of multiple classification methods, including Naive Bayes, SVM, Bayesian Network, C4.5 Decision Tree and Random Forest algorithms. In our experiments, six individual classification approaches, and the proposed ensemble approach were all trained and tested using the same dataset of 12864 tweets, in which 10-fold evaluation is used to validate the classifiers. The results show that the proposed ensemble approach outperforms these individual classifiers in this airline service Twitter dataset.

Mustafa Altinkök (2016) conducted for the purpose of analyzing the effect of the movement education program through a 12-week-coordination on the development of basic motor movements of pre-school children. A total of 78 students of pre-school period, 38 of whom

were in the experimental group and 40 of whom were in the control group, were incorporated into the study in line with their own consent after their families had also been informed.

Bee Yee Liau and Pei Pei Tan (2016), studied the consumer opinion towards the lowcost airlines or low-cost carriers (LCCs) (these two terms are used interchangeably) industry in Malaysia to better understand consumers' needs and to provide better services. Sentiment analysis is undertaken in revealing current customers' satisfaction level towards low-cost airlines. About 10,895 tweets (data collected for two and a half months) are analyzed. Text mining techniques are used during data pre-processing and a mixture of statistical techniques are used to segment the customers' opinion. The results with two different sentiment algorithms show that there is more positive than negative polarity across the different algorithms. Clustering results show that both K- Means and spherical K-Means algorithms delivered similar results and the four main topics that are discussed by the consumers on Twitter are customer service, LCCs tickets promotions, flight cancellations and delays and post-booking management.

Guoning Hu et al. (2017) Analyzed the opinion of19M Twitter users towards 62 popular industries, encompassing 12,898 enterprise and consumer brands, as well as associated subject matter topics, via sentiment analysis of 330M tweets over a period spanning a month. We find that users tend to be most positive towards manufacturing and most negative towards service industries. In addition, they tend to be more positive or negative when interacting with brands than generally on Twitter. We also find that sentiment towards brands within an industry varies greatly and we demonstrate this using two industries as use cases. In addition, we discover that there is no strong correlation between topic sentiments of different industries, demonstrating that topic sentiments are highly dependent on the context of the industry that they are mentioned in. We demonstrate the value of such an analysis in order to assess the impact of brands on social media. We hope that this initial study will prove valuable for both researchers and companies in understanding users' perception of industries, brands and associated topics and encourage more research in this field.

T. Hemakala and S. Santhoshkumar (2018), designed a framework for sentiment analysis with opinion mining for the case of airlines service feedback. Most available datasets of hotel reviews are not labelled which presents many works for researchers as far as text data pre-processing task is concerned. Twitter is a SNS that has a huge data with user posting, with this significant amount of data, it has the potential of research related to text mining and could be subjected to sentiment analysis. The airline industry is a very competitive market, which has grown rapidly in the past 2 decades. Airline companies resort to traditional customer feedback forms which in turn are very tedious and time consuming. In this work, worked on a dataset comprising of tweets for 6 major Indian Airlines and performed a multi-class sentiment analysis. This approach starts with pre-processing techniques used to clean the tweets and then representing these tweets as vectors using a deep learning concept to do a phrase-level analysis. The analysis was carried out using 7 different classification strategies: Decision Tree, Random Forest, SVM, K-Nearest Neighbors, Logistic Regression, Gaussian Naïve Bayes and AdaBoost. The outcome of the test set is the tweet sentiment.

2. METHOLOGY

The data for this research was collected from different tweets related to major airline reviews. The data contains a huge amount review among which some are positive while some are negative. In addition, dataset also contains emojis classifying positive and negative. We created emoticon lexicon which contains different types of emojis that are vastly used on social medias. The total analysis has been taken into two steps. In the first step the data were tokenized and other words such as URL or numbers has been removed, but punctuation and emojis has not been removed as it represents some sentiments. The pre-processed information was used in finding sentiments. In the second step, punctuations and emojis were removed and sentiments were calculated without emoticon and using same feature except emoticons.
The proposed algorithm starts with pre-processing the airlines tweet data by tokenizing. The data are collected from Twitter and saving the data into .csv file format. After that the system will work into two phases, Training phase in which system will be trained as per data and another one is Testing Phase where system will work in real time to give the expected result in the form of bar graph, pie chart and plain text.

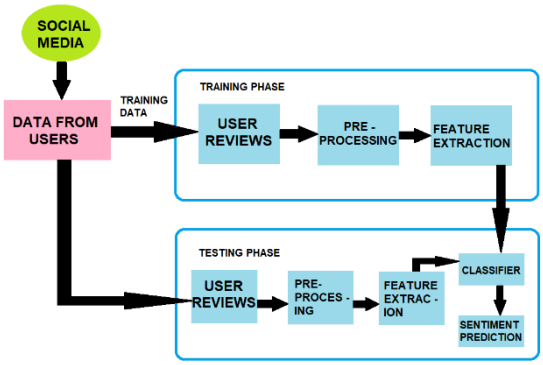

Fig. Working of the System.

3. CONCLUSION AND FUTURE WORK

Social media now a days plays an important role for marketing and presenting opinion, reviews, etc. And it is used in various applications such as academy, travel, medical, etc. With the advancement in the technology, there is a massive amount of data on social media in the increasing terms of volume, subjectivity and become a challenging process manually. This Sentiment Analysis system present for airline tweets customer satisfaction by using various machine learning technique which will be developed and implemented and after reviewing the literature about this. So in the future we need to design a classifier which satisfy the customer and consumer problem more efficiently and accurately.

4. REFERENCES

[1] Arockia Xaiver Annie's applied to airline feedback article 2015.
[2] IOP Science (https://iopscience.iop.org/article/10.1088/1757-899X/263/4/042067)
[3] IEEE Explore (https://ieeexplore.ieee.org/document/9003903/metrics#metrics)

[4] Big Data Journal (https://journalofbigdata.springeropen.com/articles/10.1186/s40537-015-0015-2)
[5] Sentiment Analysis using AdaBoost Approach article (https://ijert.org)