Scientific Voyage
Vol. 1, No. 1,
Page 45-59 (2020)

SCIENTIFIC VOYAGE
ISSN: 2395-5546

ISSN: 2395-5546

A Study onCustomerSentiment Analysis ofCommuter Airlines using Twitter Data Mining

Arkadip Ray[1] and Samik Datta[2]
[1]*Department of Information Technology,*
Government College of Engineering & Ceramic Technology, Kolkata - 700010, WB, India
Email: arka1dip2ray3@gmail.com
[2]*Department of Computer Science & Engineering,*
Techno Engineering College, Banipur, Habra - 743233, WB, India
Email: samikdatta123@gmail.com

Sentiment mining has mainly been correlated with analysis of text to establish whether an entity is of positive or, negative polarity. Recently, sentiment mining has been broadened to focus on objects such as distinguishing objective from subjective intentions, and determining the cradles and topics of different opinions expressed in textual formats such as tweets, web blogs, message board reviews, and news. Enterprises can leverage the opinion polarity and sentiment topic recognition to achieve a deeper perspective of the drivers and the overall scope of sentiments. These insights can improve customer service, establish better brand image, and enhance competitiveness. This paper solely examines the trend of analysing consumer feedbacks given in Twitter for various US based Airlines. Researchers are applying data mining and machine learning tools to accommodate different business centric evaluations such as Customer Feedback Assessment, Airline Quality Control, and Consumer Loyalty Prediction etc. Various knowledge building tools (lexicon creation, feature extraction, polarity formation) are emphasized in this study.These are applied to process the raw twitter data into characterized sentiment blocks which determine consumer intuition for choosing desired airlines. Also, external metrics such as weather condition, airline punctuality and service, staff oversight are taken into consideration that impact customer conclusions. This paper succinctly looks through sentiment recognition algorithms with their data processing paradigm and finally possible future directions for improvement are discussed.

*Keywords:*Netnography, Airline Quality Rating, Twitter Data, Machine Learning

1. Introduction

The rise of social media and innovative Web 2.0 technologies have marked the beginning decade of the twenty-first century, enabling faster and better online networking. Kaplan et al. [1] define social media as "a group of Internet-based applications that build on the ideological and technological foundations of Web 2.0, and that allow the creation and exchange of user-generated content". Social media comprises online social communities (e.g. Facebook, LinkedIn etc.), blogs, content communities (e.g. YouTube), and wikis (e.g. Wikipedia), social bookmarking, and news sites (e.g. Reddit).

The principal distinction between the "old Web" and social media is accessibility of free and easy-to-use online tools for content editing, publishing, sharing, bookmarking and classification that enable interactive exchanges between individuals and groups. Originally stemmed to facilitate connecting online and distribution of content across the Web, social media quickly became a favoured public platform for sharing experiences with products and services. Consumer commentaries in social media are sought after information on internet which are characterized by their originality and genuineness. Consumers' own reviews of products, services and brands are highly regarded among online audiences. Customers are also willing to voice their disappointment openly with products and services via social media.

Increasing number of blogs, tweets, online polls, and groups offering suggestion on products/services from the first-hand consumer perspective have transformed social media into an enormous depository of information with immeasurable potential for processing and chronicling customer data. Using social media as a source of customer feedback conform the fact that customer commentaries are posted on the internet by users for other enthusiasts, forming an ideal environment for modest and discreetresearch into customer.

This enables the research on customer experiences, and relieves an examiner from the cumbersome process of formal service quality studies. Over the past decade, numerousmethods have been proposed for analysing social media content, from netnography studies to opinion-mining and natural-language processing. In the following section we will focus our discussion on the opinion-mining, also known as sentiment analysis, for analysing consumer insights in social media and to unveil the experiencescustomers encounter when dealing with various types of services and service providers.

2. Sentiment Analysis on Customer Opinion

Sentiment analysis involvedsignificant interest in recent studies, ensuring to deliver effectiveness in analysing subjective and unstructured online content, especially in the social media setting (i.e. products/service blog reviews, video posts, tweet feeds, etc.). Sentiment analysis is an analytical procedure for classifying pieces of text into clusters containing views on specific topics (Jacobson et al. [2]). Every opinion is characterized by its holder, its target and its polarity. Using natural language processing tools, text mining and computational linguistics, sentiment analysis permits researchers to classify subjects of opinions, categorise them according to the standpoints and decide their emotional subtext, thus permitting researchers to associate customer experiences with particular responses.

According to Deshpande et al. [3], the objective of opinion mining is finally to enable decision-making in the view of quality control and performance evaluation. Also, by processing "raw" consumer data shared via social media, the acquired acumen can be used as a handle for further strategic acts. The matters that consumers share in social media is often filled with emotions of happiness, frustration, disappointment, delight and other feelings about experiences with services and service providers. What makes sentiment analysis uniquely fascinating is the ability to transform unstructured online content (i.e. tweets, blog commentaries, etc.) into structured data that can be processed to reveal information about sentiment patterns and trends that affects customer insights.

The analysis involves not only the recognition of themes in the online content (text), but also their emotional polarity, its intensity and degree of prejudice. This gives prospect for researchers to isolate customer opinions about services and also to develop new ideas on how to enrich service qualities and attributes. Sentiment analysis is considered as valuable technique for “customer experience analysis”.Advocates of opinion mining claim that automated text mining is more effectual substitute for opinion analysis than conventional customer surveys (which are often time consumingand costly).Although blogs, review sites and forums are some of the finest sources of subject matters about customer opinions, recent study efforts have indicated to the prospects that Twitter yields for opinion mining research.

3. Twitterand Sentiment Analysis

Micro-blogging site Twitter is among the world’s top ten social media channels in terms of registered users and daily visits. In 2012, more than 100 million users posted 340 million tweets a day, and the service handled an average of 1.6 billion search queries per day. As of 2018, Twitter had more than 321 million monthly active users [4]. Twitter is one of the most popular social media platforms for sharing information and receiving real time updates. Despite its brief format which allows posting messages of only 280 characters at a time and a pretty simplistic goal (i.e. to answer the question “Where are you?”), Twitter sustained as one of the most preferred social mediaapparatus for information sharing. Its popularity attracted many companies to not only get in touch with consumers for sharing publicity information, but also to interact with them in real time and deal with issues arising before, during or, after services.

In sentiment analysis, a relatively new trend concentrates on analysing Twitter feeds to make future projections about customer behaviour, to dissect the effects of individual events and to supervise public attitudes on various topics. When a user reacts to an event by tweeting, he/sheexhibits “information behaviour”, which basically includes some emotions, judgments or,intentions and consequently exposes intuitivesentiments on the tweeted topic.While the scope of Twitter data sentiment analysis is still reticent, there is growing interest in Twitter as a source of wide-ranging consumer information. Researchers are still assessing various techniques to inspect Twitter feeds and to extricate meaningful information from them. The following sectionintroduces the study on sentiment analysis of Twitter datato assess customer experiences in the airline industry.

4. Literature Review onSentiment Analysis

4.1 Marshall et al. [5] have devised a framework that conceptualises micro blogging service of Twitter which has allowed for customers to provide their positive, negative and/or noncommittal sentiment and opinion about air travel. It merges the twitter, natural disaster and aviation communities – briefly 4 U.S.-based severe weather events that included tornadoes, hurricane and massive snowstorm (called a nor’easter), 6 large and medium sized airlines and 3 weather stations. Through their evaluation, it is observed how a suite of keywords correlated to the weather event geographically and the timelines of the notification from the weather service affects airlines’ services. In its proposed Flight Data Analyser framework, primacy of flight-related data on twitter in order to determine a commercial airline’s quality of service to its customers in a significant weather event is optimized as main objective. The components are distributed in 4 modules –

- Designing a completely java-based modular framework for collecting, warehousing and processing flight-related data on twitter.
- Creating a semi-automated process to correlateeach weather event with an airline called the Fetcher Module. Twitter accounts of Airlines and weather services are sourced to identify relevant and accessible twitter data.
- Associating the weather conditions with the gathered tweeted content through Post-Filter and Analyser Modules.
- Testing the performance of Flight Data Analyser framework through the Performance Module, which provides basic statistics about tweets and run time of the Analyser Module.

Twitter Keywords	Targeted U.S. Airlines	Weather Events	Weather Services
airline, flight, delay, delivered, plane, airplane, weather, forecast, airport, schedule, cancelled, reschedule, postponed, postpone, take off, boarding, luggage, bag, bags, on time, October Snowstorm, Hurricane, Irene, disaster.	American, Delta, United, Southwest, JetBlue, Spirit	April 25-28, 2011 Tornado Outbreak, Joplin, Missouri Tornado (May 2011), Hurricane Irene (August 2011), Halloween Nor'easter (October 2011)	The Weather Channel, AccuWeather, National Oceanic and Atmospheric Administration (NOAA)

TABLE I – FLIGHT DATA ANALYSER FRAMEWORK DATASET

This system evaluates quality of service (QOS) measures by airlines through Performance Module.QOSinstruments took the form of number of retweets, relevancy of tweets, number of replies to customer tweets, and the Analyzer Module execution time. The retweet count signifies the efficacy of a particular tweet since the users would retweet only those that are suitable and pertinent. Tweet relevancy is the ratio of post-filtered tweets from the total tweets gathered with the pre-defined keywords. The number of replies to customer's tweets reveals airline customer service and how quickly they are posting replies to customer's concerns. From analysis of QOS, it is observed that American Airlines is very active on twitter in terms of replying to customer tweets but they lack in posting germanetweets in general. United Airlines delivers many useful tweets regarding flights, weather events etc., but performs poorly in terms of replying to tweets from customers. United Airlines is also observed to have a high ratio of relevant tweets. It is also the first airline under consideration to start web presence through twitter. Therefore, it is evident to infer that United Airlines has an edge over the other airlines in terms of twitter presence and customer service.

4.2 Misopoulos et al. [6] present a study that uses Twitter to identify critical elements of customer service in the airline industry. The goal of the study was to disclose customer opinions about services by supervising and analysing public Twitter commentaries; to identify components of customer service that deliver positive experiences to customers; to pinpoint service processes and features that require further improvements.The authors employed the approach of sentiment analysis as part of the netnography study. The authors processed 67,953 publicly shared tweets to identify customer sentiments about services of four airline companies. Sentiment analysis was shepherded using the lexicon approach and vector-space model for assessing the polarity of Twitter posts.The sentiment analysis of the collected Twitter data began with the classification of tweets. Two approaches were used to categorise the tweets according to their content, in order to prepare them for further processing and assessment of sentiment polarity. All gathered tweets were grouped into three pools data, based on whether their subject matter included the keywords mentioned in Table II.

Sentiment categories	Representative words/symbols
Positive sentiment	"good," "awesome," "excellent," "best," "fine," "finest," "greatest," "ideal," "perfect," "top"
Negative sentiment	"fail," "bad," "weight," "delay," "failure," "lost," "sucks," "worst," "charge," "problem"
Neutral sentiment	"take," "fly," "airport, "find"

TABLE II – SENTIMENT MAPPED WITH KEYWORDS

Using the lexicon, the analysis is done by calculating the frequency of positive, negative, and neutral words/symbols in each tweet to define its polarity (i.e. positive, negative, or neutral). Vector Space Model (VSM) of information mining is implemented where each Tweet is represented as a vector and each dimension of the vector tallies with a separate term. The lexicon developed in this study consists of 20 terms (keywords), meaning that each tweet, represented as a vector, has 20 dimensions. In "document indexing," the first of the three stages of VSM, words without any relevant meaning to the content are discarded (i.e. the words "a" or "the"). In the second stage called "term weighting," the frequency of occurrence is calculated for each term (dimension). Using the lexicon, polarity scores are allocated to each word in the lexicon. In the third stage of VSM called "similarity coefficient," the query benchmarks are matched with the terms (keywords) within a tweet, thus indicating similarities.

By analysing Twitter posts for their sentiment polarity the authors are able to identify areas of customer service that caused customer satisfaction, dissatisfaction as well as delight. Positive sentiments are linked mostly to online and mobile check-in services, favourable prices, and flight experiences. Negative sentiments revealed problems with airlines' web sites, flight delays and lost luggage. Similarly, the approach revealed service features and practices that produced positive sentiments among customers (e.g. check-ins via mobile phone applications, good ticket prices, on-board entertainment, use of social media, etc.) and even testimonies of delightful service experiences (e.g. iPad access in lounges). Another interesting insight states that customers of airlines are eager to share both positive and negative aspects of their service experiences, and are often directly addressing companies with requests to address their problems. This proactive stance on customers' behalf to raise awareness about service problems obliges airlines for improved amenities.

4.3 Adeborna et al. [7] propose a sentiment mining approach which detects sentiment polarity and sentiment topic from text. The approach includes a Sentiment Topic Recognition (STR) method that is based on Correlated Topics Model (CTM) with Variational Expectation-Maximization (VEM) algorithm. The efficiency of this model is validated using airline data from Twitter. It also examines the reputation of three major airlines by computing their Airline Quality Rating (AQR) based on the output from the before mentioned approach.

The proposed methodology serves as an intelligent tool to answer questions regarding the drivers and overall scope of sentiments. The STR model is used to examine available data and to determine the reputation of airlines by computing their Airline Quality Rating (AQR). The AQR assessment is based on customer sentiments towards three major airlines (AirTran Airways, Frontier and SkyWest Airlines) accumulated from Twitter. An algorithm is developed to match opinionated tweets to a topic lexicon and an example of how the paradigm works is illustrated using a case study with real world tweets.

The case study includes classifying tweets for three airlines (AirTran, Frontier and SkyWest) as positive, neutral or negative. Then the proposed STR model is used to generate topics for each airline classified under four AQR categories (OT, DB, MB and CC). The tweets used in this experiment contain 452 tweets on AirTran, 499 on Frontier Airlines and 195 on SkyWest Airlines. Naïve Bayes Algorithm is employed which performed better on subjectivity dataset and provided a polarity classification accuracy of 86.4%. There are more positive tweets for AirTran than negative tweets, 57.5% and 27.6% respectively, the remaining being neutral. Frontier has 64.1% positive, 18.0% negative tweets and the rest are neutral. The overall sentiment score for SkyWest airline is highly positive with 82% tweets. SkyWest has approximately 19.4% negative tweets and remaining tweets are neutral (percentages are approximated).

STR model employs the CTM with VEM algorithm to generatelexicons from the airline tweets which are used in building each AQR category. In total, four lexicons are derived – on-time, denied boarding, mishandled baggage and customer complaint lexicons. This model yields better comparative performance to other STR models because the dependency and correlation between sentiment topics are taken into consideration. The STR model helps to rightly sort topic related terms under each AQR criteria for positive and negative polarity.

4.4 Duan et al. [8] built a model for six major U.S. airlines that executes sentiment analysis on customer reviews so that the airlines can have fast and concise feedback; make recommendations on the most important aspect of services they could improve given customers' complaints. The authors performed multi-class classification using Naive Bayes, SVM and Neural Network on US Airline Twitter data set collected from Kaggle.

The data set contains about 15,000 tweets collected in 2015 on various airline reviews. Every review is labelled as either positive, negative or neutral. The sentiment analysis labels are positive (20%), negative (60%), and neutral (20%). The negative reason labels are bad flight (7.45%), cancelled flight (9.62%), customer services issues (39.77%), damaged luggage (0.84%), flight attendant complaints (6.05%), flight booking problems

(6.19%), late Flights (1.99%), long lines (19.97%), and lost luggage (8.23%). In the data set, about 80% of the negative reviews has a negative reason label, yet the rest islabelled as “can’t tell”. By knowing every review’s negative reason, specific suggestions can be given to different airline companies on how to improve their service.Sentiment Analysis and Negative Reason Prediction are the pillars of this model. For both tasks, Naive Bayes, SVM and Neural Network are used on the frequency matrix generated from the twitter dataset.

Naive Bayes with multinomial event model from ‘sklearn’ is used where input is the Laplace smoothed frequency vector. For Support vector machines, linear kernel and RBF kernel are used. SVM uses the same input and implementation package as Naive Bayes. Tensorflow is used in Neural Network implementation where input is the frequency vector that represents a review. The output is a vector with probabilities for different classes and the highest is selected as prediction. Furthermore Recurrent Neural Network which is a Bi-directional Gated Recurrent Unit Network (GRU) is used to capture the structure features of pre-trained word vectors that are trained on twitter data set. Also, it solves the vanishing gradient problem that many recurrent neural network models have.

In sentiment analysis, SVM with linear kernel achieves the best test accuracy. Therefore, SVM is recommended in this section. According to the result from linear SVM, Virgin American performs the best according to its lowest negative review composition in its total reviews. Given by the result of lowest test error, Naive Bayes is used for the prediction of unclassified data. From the results, it is clear that SVM and Naive Bayes perform better than deep learning methods. By associating major negative reasons for each airlines and airlines performance on service issues, negative reasons are classified as shown in the following table.

Airline	1st Negative Reason	2nd negative Reason
Virgin America	Customer Service Issue	Flight Booking Problems
United Airlines	Customer Service Issue	Late Flight
Southwest	Customer Service Issue	Late Flight
Delta	Late Flight	Customer Service Issue
US Airways	Customer Service Issue	Late Flight
American Airline	Customer Service Issue	Late Flight

TABLE III – AIRLINE WISE NEGATIVE SENTIMENT CLASSIFICATION

4.5 Hakh et al. [9] applied SMOTE method to analyse and solve the imbalanced challenge of the twitter datasets collected from US airline companies. In addition, the effect of feature selection and over-sampling techniques are tested on airline datasets. The “Twitter Airline Sentiment” dataset from Kaggle contains tweets covering six U.S. airline companies with 14,640 tweets, each of which is labelled according to sentiment polarity as: positive, negative, and neutral. In addition to the class label, all datasets share seven features namely Airline Sentiment, Airline Sentiment Confidence, Negative reason, Negative Reason Confidence, Airline, Retweet count, Text. Term frequency – Inverse Document Frequency (TF-IDF) is applied to measure weight of significant terms. Further, a tweet-term weight matrix is generated where the terms represent the features and weights are the TF-IDF scores calculated earlier.

In the first level of feature selection, genetic search is used as a filter to select the best subset of elements based on the correlation of features with the class label. In second level, a univariate feature selection process is applied which aimed at selecting the best subset of features based on statistical tests. The result portrays sentiment polarity of all the datasets, except for unevenly distributed data in Virgin America twitter handle. Unless solved, the training process of any classifier considering the imbalanced datasets will result in a bias due to the larger number of instances, sampled for training, and belonging to one class. Synthetic Minority Over-Sampling Technique (SMOTE) is added to increase the number of instances used for the training process from the minority class. Finally, on each of the partial datasets various classification techniques are applied (i.e. AdaBoost, Decision Tree, Linear SVM, Naïve Bayes, Random Forest, K-NN, and Kernel SVM) which aim at predicting the class of each tweet provided the subset of features are obtained after the two levels of feature selection and data balancing.

Classification settings per algorithm are set realistically after performing experiments with different settings. For example, number of trees in the random forest classifier is set to 4 as it showed the best results. On the other hand, the kernel chosen for the Kernel SVM is the Radial Basis Function (RBF). Additionally, for the K-NN classifier, the value of number of neighbours (k) are selected between 1, 3, and 5. Random Forest and Decision Tree have shown a high prediction level, and constancy when applied on all datasets. While K-NN and Linear SVM have shown an acceptable level of performance regarding all the evaluation metrics. On the other hand, Kernel SVM has shown poor results in comparison with other classifiers.

4.6 Joshi et al. [10] aims to provide a decision support for the customers for selecting the best fit US-based Airline, by presenting an Aspect level sentiment analysis on customer opinions available in micro blogging site Twitter and online review site Skytrax. This study supports a modified Knowledge discovery and data mining (KDD) methodology. Several machine learning algorithms (i.e. SVM, Decision tree, Random Forest, Bagging and Boosting, SLDA, Maximum Entropy) are applied in order to find out the best fit algorithm for the system.

The study follows a modified KDD methodology where a few stages are modified like implementation stage modified to Aspect and Sentiment Detection stage. The data for the research is taken from two sources, Twitter and online review site Skytrax. The assembled data was then interpreted as 1-gram, 2-gram and 3-gram manually by inspecting their frequency. Reviews are further categorized into five aspects (Food and Beverages services, Staff services, Luggage, Punctuality and Seat). A bag of words are added for each aspect after analysis of 1-gram and 2-gram. The polarity of each sentence isequated and a score isallocated for negative, neutral and positive as -1, 0 and 1 respectively.

Machine learning algorithms like support vector Machine (SVM), Glment, Maxent, SLDA, boosting, Random forest, neural networks and Decision Tree are applied and tested. The training and test dataset are separated in the ratio of 70:30. The final dataset consists attributes like cleaned sentence, polarity score, food and beverages aspect score, staff services score, luggage score, punctuality score and seat aspects score, airline names and airline code. This dataset is further treated for statistical operations to find out the correlations between the different aspects and polarity score of the airlines. Also, linear regression is performed in order to find the connection between the attributes of the airlines.

The proposed algorithm is measured in terms of Precision, Recall, and F score. It is observed that Random Forest and Decision Tree performed well with the F-score of 0.60 resulting as the best fit algorithm.The first case study is an analysis of relation between three layers - Airlines, Aspects and Sentiments. United Airlines comes out to be 1st rank holder showing high positive response whereas Skywest Airline bags 10th rank with highest negative sentiments. The second case study is an analysis of aspects contribution towards the reviewer sentiments. From the unigram and word cloud implementation, “Seats”, “Staff”, “Punctuality”, “Luggage” and “Food” are found to be the major areas of customer sentiments. The case study projects Seats Aspect as the most important aspect, whereas food is considered to be lowest contributor in the aspects towards customer’s sentiments. Being one of the most popular airlines, United holds first rank in every aspect and there is a competitive edge between flights having mediocre ranking.

4.7 Ahuja et al. [11] focuses on identification of the diverse content typologies being used by Jet Airways on Twitter to implore Customer Engagement. This has been done, using a technique called Netnography. Subsequently, it proceeds to derive customer insights from the twitter page of Jet Airways, using wordclouds. The authors further attempt to generate aLexicon Based algorithm for Sentiment Analysis, using the tweets from the Jet Airways Twitter handle.

Netnography is a qualitative andexplanatoryresearch methodology that uses internet augmentedethnographic research techniques to study virtual networks and communities. Netnography is operatedon Twitter handle of Jet Airways and the analysis helps in distinguishingthe right type of content typology that will drawhigh levels of retweets and likes and successively more engagement from customers. A word cloud is a visual illustrationof text data, normallyused to portraykeyword metadata on websites. It represents the systemsupportedby a particular set of data (in this case twitter data) and is suggestiveof the ideaassociated with that set of textual information (tweets by customers of Jet Airways). The authors have generated word clouds to conduct a textual analysis using the Twitter handle of Jet Airways to extract social and customer insights from the content producedby the organisation and customers.Lexicon-based sentiment classification is possiblybe the most basic technique for computingthe sentimentpolarity of dataset (that is, a corpus). It needs a dictionary of words (a lexicon) which in this case is supplied from the Jet Airways Twitter handle and each word is then associated with Positive, Negative and Neutral polarity score.

Netnographic analysis of the Jet Airways Twitter page helps to extract the following Content Typologies - Organisational, Promotional and Relational Contents. The organisational content is not able to createadequate levels of customer retweets and likes (on an average). The promotional content increases customer interest in the brand. The relational content is able to generate a high level of retweets and likes. The word cloud, producedby at least 1500 tweets, depicts the service excellence that Jet Airways has attempted to achieve across the areas of punctuality, safety, seating comfort, large network, friendly and caring behaviour, professional and efficient staff, quality of food, cleanliness of the aircraft, quick baggage clearance, ease in booking tickets and easy check-ins.

4.8 Khaturia et al. [12] aim to provide a proper mechanism which helps the customers to opt for an airline in accordance to their comfortable experience. The process commenceswith collection of tweets on a large scale, analysing and classifying them as positive, negative and neutral – helps to delivera better rating and review which then assistthe customers to opt for achoiceas per their convenience.

The tweets generated from six major U.S. Airlines, namely Virgin, United, Delta, Southwest, American and US Airways are measuredas research data and the Sentiment Polarity Score for each airline is computedin order to rank them. Naïve Bayes Algorithm and Sentiment Analysis are applied on the classified tweets to create a comparative platform both statistically as well as graphically for the customer recommendation system.

After possessingtweets from Kaggle, they are categorized into three labels – positive, negative and neutral tweets. The subjects of negative tweets are mainly bad flights, cancelled flights, customer services issues, damaged or, lost luggage, flight attendant complaints, flight booking problems, late flights, long boarding lines etc. A feature matrix is then built to convert the textual information into numerical information. In the feature matrix, the number of rows indicates the number of samples, the number of columns indicates the length of the dictionary (i.e. collection of tweets), and each element indicates whether the specific word has appeared in the current review – "1" for existence and "0" for absence. Naïve Bayes classifiers are probabilistic units applied with Bayes' theorem for an applicableindependent assumption between its features (called naive). The classifiers are ascertainedfor every airline into positive, negative and neutral probabilities using the following formulas.

$$Score = \frac{positive + negative + neutral}{Total\,number\,of\,Tweets} \times 100$$

$$p(c \mid x) = \frac{p(x \mid c)\,p(c)}{p(x)}$$

Where p(c|x) is posterior probability, p(x) is predictor prior probability, p(c) is class Prior probability and p(x|c) is Likelihood.

Comparing the projectedscores of each airline, it is observed that Sentiment Analysis and Naïve Bayes Classification give contrasting results for Positive scores; for Neutral scores, both the algorithms provide similar result though the differences are quite high; the algorithms judge Negative scores with similar ratings. Evaluatingthe scores separately, one can predict the efficiency of the airlines based on which one can take the decision to choose an airline.

4.9 Ashi et al. [13] compared two word embedding models for aspect-based sentiment analysis (ABSA) of Arabic tweets. The ABSA instance isframedas a two-step process of aspect detection trailedby sentiment polarity classification of the detected aspects. The compared embedding models include fastText Arabic Wikipedia and AraVec-Web, both available as pre-trained models. Their corpus comprisedof 5K airline service related tweets in

Arabic, manually labelled for ABSA with imbalanced aspect categories. For classification, a support vector machine classifier is used for both aspect detection and sentiment polarity classification.The Arabic tweets dataset has the majority of tweets being in the Saudi dialect since the studied airline being the national carrier airline of Saudi Arabia.A distribution plot of the final dataset for the 13 specified aspects categories related to the airline service is performed manually by a group of native Arabic speakers.Also, for the study of sentiment polarity detection, the collected tweets are distributed regardingsentiment polarity. In this part of ABSA, the tweets are classified into projectedlabels either as positive (1410 tweets) or, as negative (3590 tweets), discarding the neutral tweets.

Aspect	**Tweet Topics**
Schedule	Flight schedule, rescheduling, timing, delays.
Destinations	Airline destination and routes.
Luggage and Cargo	Luggage, air cargo, luggage allowance, luggage delays.
Staff & Crew	All staff such as pilots and flight attendants.
Airplane	Airplane seating, cabin features, maintenance.
Lounges	First-class and frequent flyer service and airport lounges
Entertainment	In-flight entertainment, other media, and wi-fi.
Meals	In-flight meals and in-flight services.
Booking Services	Airline website, mobile app, and self-service machines.
Customer Service	Customer communications and complaint management
Refunds	Refund and compensations.
Pricing	Ticket pricing and seasonal airline offers.
Miscellaneous	It represents all tweets with gratitude and thanking or complaints about the airline in general.

TABLE IV – CATEGORY-WISE DISTRIBUTION OF OPINION ASPECTS

Employingthe two pre-trained word embedding models, simple vector-based features are used to classify the 5k collected airline-related tweets.By assessingthe semantic distance between Arabic words and phrases and byutilizing the Vector Space Mode, two defined sub-tasks are performed, namely Aspect Detection and Aspect-Based Sentiment detection. Two distinct word embedding models are utilized as the lexical resource for the two sub-tasks of Arabic aspect-based sentiment analysis. Such word embedding models utilisethe Vector Space Model (VSM) and cosine similarity between words.

Authors have trained a supervised SVM linear classifier using simple vector-based features on the labelled dataset to exploit it on unseen tweets to predict polarity – positive or, negative. For pre-training, two word embedding models are used, namely AraVec and fastText. Such vector-based features obtainboth semantic and sentiment information and these methods haveenhancedthe sentiment classification task.

This study also compares performance of two word embedding models using various evaluation metrics, namely, accuracy, precision, recall. The adopted pre-trained word embedding model used in both ABSA sub-tasks has performed comparably well in comparison to existing techniques. This vector space approach without any hand-crafted features performs comparably well in comparison to techniques adopting manually engineered features.However, one limitation is thatthe dataset is imbalanced with a significant number of tweets between the 13 aspect classes as well as for the positive and negative polarity classes.

4.10 ---The research focus ofKhan et al. [14]is on analysis of tweets related to airlines based in four regions: Europe, India, Australia and America for consumer loyalty prediction. The tweets are used to calculate and graphically represent the positive, negative mean sentiment scores and a varying mean sentiment score over time for each airline. The terms for complaints and compliments are depicted using visualization methods. A novel method is proposed to measure and predict consumer loyalty using the data gathered from Twitter. Three classifiers are employed, namely, Random Forest, Decision Tree and Logistic Regression.They have collected tweets for 18 airlines based in four selected regions, which are America, India, Europe and Australia.Moreover, they have formed dataset for consumer loyalty analysis by collecting tweets using the search queries “loyal flyer”, “loyal to airline” as well as “left airline” totalling 10,000 tweets.

Airlines			
American	**European**	**Indian**	**Australian**
Delta, United, Spirit, Southwest, Jet Blue	Lufthansa, Air Berlin, Turkish Air, KLM, Easy Jet	Indigo, Air India, Jet Airways, Vistara, Spice Jet	Virgin, Qantas, Tiger Air, Jet Star, Sharp

TABLE V – AIRLINES TAKEN INTO CONSIDERATION

Sentiment Information Visualisation emphasises on user opinions posted on airline Twitter Help Desks or, on official Twitter handles of various airline companies.However, texts, images, links, emoticons and other forms of media that are present in a simple tweet create complications. Hence, the first step of the analysis is to clean the tweets and tokenize. Then sentiment analysis is operated on the tweets denotingmost positive, most negative and neutral. The mean scores of positive and negative sentiments are calculated for each airline. These scores areessentialfor an airline to realiseconsumer opinion about its services. An airline can make sure that its positive sentiment score is greater than negative sentiment score. Airlines may work to build their positive scores greater and negative scores lesser than their competitors. Brand sentiment score for a period of time is valuable for companies, since they indicate the positive attitude towards the brand or, the dissatisfaction with the same.

The positive terms are “awesome”, “excellent”, “delicious”, “perfect”, “superb” and “wonderful” with highest score of 1.0. The tweets with “worst” term expose the areas that need improvement and the positive tweets indicates to areas that are gaining appreciation and thus mandates an airline to keep up the good work.For Airline Passenger Loyalty, normalization is performed by dividing the difference between maximum and minimum loyalty score collected from the tweets using search queries such as “loyal flyer”, “loyal to airline” as well as “left airline”. The various values used to calculate the loyalty measurement are graphically represented in different combinations using the k-means clusters to illustrate the loyal and disloyal passengers.For consumer loyalty prediction, three prediction models are used – Random Forest, Decision Trees and Logistic Regression. The model is fitted with positive and negative sentiment scores, mean of retweets, mean of likes and the number of followers. The models are tested on 10-fold cross validation and the maximum accuracy of 99.05% is observed for Random Forest.

Thus, identification ofcompliments and complaints, variations in sentiment over a period of time and mean sentiment score visualization provide passenger opinion on airline services. The consumer loyalty analysis assistsairlines to retain loyal flyers and bring in new customers.

4.11 Hong et al. [15] present a study that used text mining to develop a unified understanding of keywords in the aviation industry in a data-driven way. Based on the airline review data, it proposes a two-step process forobtainingkeywords using text mining and grouping them for cluster analysis.Particularly, it uses a combination of metrics and clustering algorithms to pre-process and analyse texts related to keyword extraction method, including text from the scientific literature and news articles. These keywords affect corporate marketing performance.

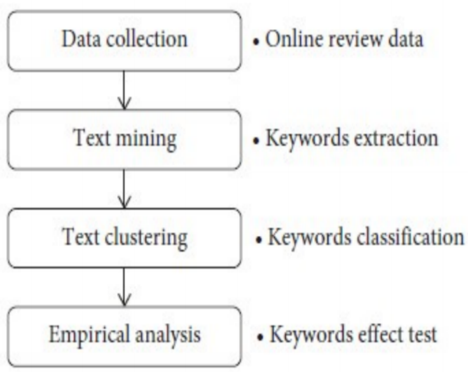

TABLE VI – FRAMEWORK OF DATA ANALYSIS

The authors have used the online review data from customers of two large air carriers in Korea and Japan, provided by global air service evaluation agency Skytrax in United Kingdom. One type of text data focuses on the customer experience with the airline services. The other type is the survey which include the evaluation of facilities, satisfaction, and recommendation after using the airline. Among the assembleddata, text data are used for core keyword extraction andquestionnaire data are used for practicalanalysis to verify the influence of core keywords. Text mining process is applied to extract words and text clustering process is utilised to analyse the collective sentiment of extracted words. Semantic analysis of the extracted 45 keywords is done where classification employed hierarchical clustering and the distance between the keywords are measured using the Euclidean method.

As a result of clustering analysis, 45 keywords aregroupedinto two clusters. Cluster 1 comprises of service content such as “seats”, “cabin”, “class”; “staff” who provides the service; service evaluation related keywords – “good”, “excellent”, “comfortable” etc. On the other hand, Cluster 2 contains more detailed service content such as “meal”, “movie”, “drinks”, “check”, “served” etc. Also, it has service evaluation related keywords such as “great”, “nice”, “well”, “best”, “clean” etc.

Analysis shows“seats”, “staff” and “cabin”are positivelycorrelated withcustomer satisfaction and recommendation. It also shows that the airline customers talk frequently about “class” types (economy, business), but it does not influencethe customer satisfaction. This means that the airline should pay immediate attention to the seating arrangement in different classes along with continuing satisfactory performances in seat quality, staff attitude and cabinatmosphere.

5. Discussion on Recent Works

This study paper imparted an overview on the application of sentiment classification for better understanding of the airline commuter behaviour. The paper focuses on the different algorithms that have already been applied on the sentiment data for analysis. It also focuses on the proposed approach in depth, along with exploration of implementing tweet content in order to further improve the performance.

After going through the series of research work it is very clear that some supervised learning techniques produces good result. It is also seen that SVM or Naïve Bayes systems provide better result as compared to Neural Networks and Aspect-based decision support systems.Naïve Bayes and Support Vector Machines are the most frequently used machine learning algorithms for the purposes of sentiment classification. Many proposed algorithms are compared to these models as reference. But it is too early to judge that Neural Network will not produce better result.

Most of the research works emphasize either on attaining the sentiment polarity of customer experience or aim at deciding the rating or delay time of the airlines. Also, consuming Twitter feeds for three to four months is not sufficient in order to procure appropriate insights into customer behaviour towards airlines; so, there is a need for diversified data sets.

Transforming textual information into sentiment clusters provokes some genuine complications such as negative content processing, simplifying domains, language consideration and dealing with slangs or, profanity. Also, fake or spam reviews should be dealt with to eliminate ambiguity and opinionated results.

The articles presented in this survey mainly observes Twitter data generated in English language by customers of US based Airlines with few exceptions. To get a comprehensive understanding, all major spoken languages from Asian, European, Latin, Middle-East continents should be taken into consideration; the Natural Language Processing tools can be used to administer Sentiment Analysis.

6. Conclusion and Future Work

The objective of this paper is to review or, assemble the idea of sentiment analysis performed by different researcher in the domain of airline service using Twitter data. After analysing before mentioned articles, it is clear that the enhancement of current sentiment analysis algorithms is still a rewarding area of research.

Our future goal is to develop an ensemble of machine learning algorithms (supervised, unsupervised) and lexicon based approaches. By combining them with existing models, we are willing to identify areas of customer service that caused customer satisfaction, dissatisfaction as well as delight. This will be useful to rate major airlines by computing their Airline Service Quality (ASQ) and to build Airline Recommendation System based on customer feedbacks.Repository of data can be enriched by extending the data sources to social media (Facebook), online review sites (TripAdvisor, Skytrax, skyscanner, makemytrip) etc. The scope of customer sentiment modelling can be extended with advancement of data mining techniques into areas like automobile and railways.

REFERENCES

1. A. Kaplan and M. Haenlein, Business Horizons **53**, 59 (2010)
2. L. Jacobson, Public Relations Tactics **16**, 18 (2009)
3. M. Deshpande and A. Sarkar, Business Intelligence Journal **15**, 41 (2010)
4. www.wikipedia.com
5. B. Marshall et al., Journal of Emerging Trends in Computing and Information Sciences **3**, 8 (2012)
6. F. Misopoulos, M. Mitic, A. Kapoulas and C. Karapiperis, Management Decision **52**, 705 (2014)
7. E. Adeborna and K. Siau, Pacific Asia Conference on Information Systems (PACIS) Proceedings, Paper 363.
8. X.T. Duan, T. Ji and W. Qian, CS229, Stanford.edu (2016)
9. H. Hakh, I. Aljarah and B. Al-Shboul, Proceedings of the New Trends in Information Technology (NTIT-2017), The University of Jordan, Amman, Jordan
10. S. Joshi, "Aspect based sentiment analysis for United States of America Airlines", MSc Research Project – Data Analytics, School of Computing, National College of Ireland (2017)
11. V. Ahuja and M. Shakeel, Procedia Computer Science **122**, 17 (2017)
12. D. Khaturia et al., International Journal of Advanced Science and Technology **111**, 107 (2018)
13. M.M. Ashi, M.A. Siddiqui and F. Nadeem, Advances in Intelligent Systems and Computing **845** (2018)
14. R. Khan and S. Urolagin, International Journal of Advanced Computer Science and Applications **9**, 6 (2018)
15. J.-W. Hong and S.-B. Park, Hindawi Mobile Information Systems **8**, 1790429 (2019)