# Unveiling the Ambivalence from the Airline Reviews: the Airline Recommendation System Using CNN

Muhammad Usman Javeed, Waheed Ramay,
Shafqat Maria Aslam and Aqib Rehman Pir Zada



August 3, 2022

# Unveiling the Ambivalence from the Airline Reviews: The Airline Recommendation System using CNN

Muhammad Usman Javeed
Department of Computer Science
Comsats Univeristy
Sahiwal, Pakistan
usmanjavveed@gmail.com

Dr. Waheed Ramay
Department of Computer Science
Comsats Univeristy
Sahiwal, Pakistan
waheedramay@cuisahiwal.edu.pk

Shafqat Maria Aslam
Department of Computer Science
Comsats Univeristy
Sahiwal, Pakistan
shafqatmaria34@gmail.com

Aqib Rehman Pir Zada
Department of Computer Science
Comsats Univeristy
Sahiwal, Pakistan
aqibrehmanpirzada75@gmail.com

***Abstract*—There are several advantages to traveling when it comes to crossing international boundaries. People have the right to write reviews on websites or online platforms to express their opinions. Reviews have a direct impact on consumer relationships. These opinions could be expressed within a single review (positive or negative) or across reviews (conflicting). Conflicting online reviews, a little-studied topic, have exploded in prominence in recent years. These reviews not only offer assistance to customers in selecting a suitable airline, but also assist airline firms in identifying and correcting flaws in their service. We address this gap by proposing a research paradigm that conceptualizes the characteristics of conflicting airline reviews, which identify the traveler perceptions that cause their attitudinal ambivalence or uncertainty in developing their actions. We examined how conflicting attributes of airline reviews trigger travelers' attitudinal ambivalence, which leads to indecisiveness. In this work, we suggested different ways to remove this ambivalence and provide recommendations to customers and airlines. The contribution of this work is twofold: first, we used NLP (Natural Language Processing) techniques to preprocess traveler reviews in the recommender system. Second, the Convolutional Neural Network model was implemented, which proposes data collection on different social networks. This machine learning approach recommends an appropriate airline to travelers. This unique strategy to utilize online social networks, promote low-cost airlines to travelers, and machine learning model increases the CNN model's recommendation accuracy.***

***Keywords*—Scraping, Natural Language Processing, EDA, CNN Model.**

## I. Introduction

In terms of maintaining passenger contentment and produce future income, the air transport industry's severe competition obligation effective online and offline client relationship management [1]. In recent years, the airlines have made a substantial contribution to the global economy's growth and development. In recent years, the travel and tourism sector has grown significantly, with airlines serving as the industry's primary sponsor. Air travel is the fastest mode of transportation and competes with buses, trains, and other modes of transportation around the world [2]. Flying enables planes to avoid natural and man-made obstacles including seas, rivers, and mountains. Flying was regarded as a costly luxury until the latter half of the twentieth century. Thanks to larger passenger aircraft and technological developments, it's now a rather inexpensive mode of transportation, and prices are continually reducing in some locations. Travelers are also provided with safety, Wi-Fi, food beverage, seat comfort) economic rewards, and employment opportunities, and a comfortable environment is maintained throughout airplane travel.

Air travel is preferred by travellers because it is speedier and less prone to accidents than other modes of transportation [3]. Air travel is the quickest means of transportation, saving you hours or even days on a trip. The speed of air travel also allows for the delivery of critical commodities like organs and drugs to patients in need. The aviation industry employs mechanics, pilots, stewards, customer service agents, and designers.

Travelers use blogs, tweets, emotive analysis, and reviews to communicate their ideas and facts. It benefits the industry, company, and organization, as well as those who want to shop their brand, travel their firm, and utilize their application, among other things. They have the ability to boost retailer trustworthiness and help shoppers compare shopping results. End user Benefits (Increased trust in the product and the firm, Gives consumer a voice). Ecommerce Advantages (valuable market research, profit gain).

Different types of airlines are Qatar airways, Turkish, PIA, Emirates and so on. All airline is working good compared to each other. Traveller is confused to select airline. Currently, social media play an important role to provide reviews or comments. According to prior study on an average day, just 15% of flights (out of a maximum of 34%) produced contrail [4]. The bulk of contrails were made in the southeast and along the Pacific coast, and between June and September, 63 percent of total Contrail Along Track Distance was generated.

When a traveller wishes to secure a seat, they confront numerous challenges because not everyone understands how to utilize websites, apps, or social media. The field of social media is vast. Companies provide access to their websites or webpages to their users, customers, and travellers. Customers are constantly turning to social media to leave company reviews and discuss their thoughts on specific services [4]. Consumer feedback is a critical component of any business owner's economic success in competitive markets. These online pages often have a shoer blog or comment box

[2], and sometimes reviews are written forms or present in the feeling’s forms. These reviews help other travellers who desire to travel with airlines.

A review is a tool for us to communicate our thoughts, opinions, and inspiration to the company or consumer. These are not simple reviews. Different types of reviews exist, but we'll focus on two of them. People are commenting as a result of rapidly expanding online resources, online discussion groups, forums, and blogs, which generate a massive quantity of new data in the form of consumer evaluations, comments, and thoughts about a product, events, and entities. Reviews of any institution, banking, hotels, airlines, and online shopping products including books, image sensors, mobile phones, and notebooks benefit both customers and decision-makers [5]. The first is internal reviews, which include both good and negative feedback. The second is across reviews, where there are different opinions. These reviews are essentially a mash-up of reviews. For example, "This chocolate is amazing, but the sentiments I get after eating it are not good."

Domestic and international airline flights are estimated to reach 22.2 million in 2021. There will be approximately 16 million foreign flights in 2020The Federal Aviation Administration formerly said that its Air Traffic Organization (ATO) managed 45,000 localized flights per day, or more than 16,405,000 per year, prior to COVID-19. With growing awareness of transportation competition, airlines must recognize that their most valuable competitive asset is their client base, and they must do everything possible to suit their needs if they are to survive and profit [6]. Airlines offer the most relevant material, such as products, services, advice, notices, feature introductions, and information sharing, to improve passengers' travel experience, contentment, and loyalty.

After covid social media has more popular and rating is higher because of social distance everyone prefers online boking, shopping, entertainment etc. The globe is not interconnected at the same moment. eWOM includes things like views, recommendations, ratings, video testimonials, tweets, photographs, and blog posts [7]. Today Filieri et al with the extensive use of internet Individuals rely significantly on electronic word-of-mouth to spread information (eWOM) while planning their trips [8],[9]. According to a recent report of Forbes90percent of consumers view eWOM information before making a purchase choice. The eWOM is a key source of electronic information that refers peer generated evaluations spread existing, future, and previous customers on third-party websites about a particular firm, product, or service [10]. Therefore, eWOM is a type of online review., which is one the important influential form of eWOM. Different type of benefits is provided but with the passage of time online working are more than manual so is providing a lot benefits to the world. Online reservation is also providing the eWOM.

However, increasingly online reviews contain within reviews (polar Reviews), and across reviews contradictory, conflicting word-of-mouth about product or service attributes that influences consumer attitude [11]. Ruiz at al conducted study, a total of 14000 consumers reviews on TripAdvisor deceive, complaining, and praising comments. All the review comment and recommendation area present in the blogs and sites. Wang et al. Customer are use online platform to express their feelings these feelings are positive or negative and sometimes are both mixed/conflicting reviews. Customers can contribute information, ideas, and information on products, services, and businesses through Online platforms include Twitter, Instagram, and e-commerce sites [12].

From the perspective of research, the Reviews that are bivalent (e.g., positive and negative), inconsistent, or conflicted practices by airlines firms and consumers, there’s a higher chance that customers will process information incorrectly. which develop feelings of uncertainty and ambivalence in their attitude [13]. Therefore, we argue that paradoxes coexisting at the same time in object (e.g., online reviews) create attitudinal ambivalence in customers by instilling in them an ambiguous and contradictory attitude regarding the object, and the consumers not take any decision and confused.

Since it has been well established that few studies on online airline reviews have examined information processing of univalent message-based qualities (e.g., positively or negatively) and their influences on consumers attitude [14]. However, few studies in examining of airline reviews have bivalent (e.g., good and bad) and contradicting qualities.

Previous literature mentioned the reasons, readers of airline reviews, it is hypothesized are prone to attitudinal ambivalence as a result of contrasting aspects of internet airline evaluations.

Therefore, scholars should conduct additional research. the questions of what regarding the causes of ambiguous attitudes in online airline reviews recommendation recipients, what can be done to reduce this attitude ambivalence, and how to persuade reluctant customers to read contradicting airline ratings online. Furthermore, hospitality businesses are eager to comprehend the reasons why consumers are feeling attitudinal ambivalence when they are using conflicting airlines reviews. Hence, present study findings will assist airline firms and they can create effective techniques to reduce ambivalence in attitudes and improve overall performance. more chance of recommending an airline based on contradicting internet airline reviews [15].

It's worth noting that the current study uses a unique method and examine airline consumers’ perception for conflicting online airline reviews from the perspectives of informativeness, persuasiveness, font diagnostic and understanding of the language. Thus, our study's main goal is to address a vacuum in the literature by proposing a research paradigm for examining attitude ambivalence in the face of contradictory information processing and related outcomes for conflicting online airline reviews recommendation. In present research model, we aim to investigate some textual or non- textual data, prepossessing, sentimental analysis etc. In several ways, our research adds to the body of knowledge. The reviews are helpful for other to select airline in an easy manner. Read the reviews then see and goes to rating direction three or four stars are appearing. The airline sector operates in a highly competitive environment where enterprises must overcome a variety of obstacles in order to flourish. This research examines and infers recommendations reviewers' comments in internet reviews of tourism services, notably airline services, from several points of view. We are collected data/reviews from top 10 airlines websites. Applying in dataset pre-processing method then find out textual and non-textual data. Then we applying it pre-processing. We construct a single base classifier by using machine learning algorithm Convolutional neural network (CNN). The most typical application is for analysing visual imagery. As a result, in this paper, we look at how well the Our new suggested model for airline systems that rely on suggestion uses a convolution neural network model [4].

## II.LITRATURE REVIEW

### A. *Online reviews*

Online platforms are defined in a variety of ways [4]. Online commerce websites, Forums, blogs, social media sites, and online review sites are all examples of online communities have all received online reviews. In most cases, online customer ratings and reviews (OCRs) are provided by consumers who have had first-hand (and usually recent) experience with a product or service (e.g., opinion in text format) [16]. Customer reviews on the internet have also been discovered to have a significant influence on customer-cantered product development [17].

According to Zhang et al.,[18] there are substantial links between changes in smart phone functionality and online reviews. Researchers have noticed the relevance of online consumer evaluations in product creation, and more studies are focusing on product-related concerns. Qiao and colleagues [19] developed a sophisticated LDA model to discover and collect critical information about product problems from internet reviews. Most individuals in the United States have read (82%) and written (61 percent) internet evaluations [20].

A flaw, according to Merriam-Webster, is a lack of anything that is required for completeness, sufficiency, perfection. As a result, recognizing a need is not the same as producing innovative ideas or creative thinking.

Another line of study tries to create various approaches for understanding consumer demands from customer reviews on the internet, such as perceived advantages and requested product features.

Timoshenko and Hauser [21] used Amazon reviews to extract client demands. Consumers' needs are a statement explaining the advantages that customers desire, which includes both complaining about present features and advocating new improvements. On the other side, innovation ideas are new ideas or proposals for meeting client demands. The most important of the various aspects associated to customers' online product reviews is review helpfulness (Malik and Hussain [22]. Because online reviews with more review votes have higher sales correlations, this is most likely the case. Online assessments are divided into numerous pieces such as linguistic traits, views, semantic factors, and other portions as a source of information. [23].

In this study Dooms purposed a study [24] Content information about persons or objects does not have to be machine identifiable in collaborative filtering systems. Pure Collaborative Filtering approaches rely solely on ratings and don't require any other user or item information, because they consider the experiences of others and base their recommendations on that knowledge. The user modelling method is at the heart of content-based filtering (Content Based Filtering). We are defined are as following table.

The interests of the target user are derived from those with whom other users have interacted. One of the most popular content-based filtering techniques is Inverse Document Frequency-Term Frequency (TF-IDF). In comparison to the other ways, Content Based Filtering is advantageous since it allows for a user-based recommendation approach that may create accurate recommendations based on specific characteristics.

The pseudo-code illustrates the operation of the proposed system by Saurabh Bahulikar [25]. We had to create our own training dataset with the necessary criteria for filtering and, as a result, generating suggestions to finish the project, no currently accessible datasets were able to supply all the needed tuples inside the dataset, there are 480 flights in the dataset, each with its own set of content-based indicators. Google provides coordinates validation as part of the ‘google API. The findings are determined using precision, recall, or the F measure, and they have a high level of accuracy.”

In this study Wagh [26] purposed a theory is the lack of a proper recommendation system for the user based on previous customer experience is a big issue in the airline sector. We want to develop a system that allows users to choose an airline based on their preferred level of comfort. When it comes to analysing data in a textual format, emotional analysis is the most effective method. Collecting vast numbers of tweets, evaluating them, and categorizing them as positive, negative, or neutral results in better ratings and reviews, allowing customers to choose the best option for them. Here, Nave Bayes presents a comparative platform to the sentimental analysis research in order to determine how accurate the recommendations made on airlines utilizing sentimental analysis are. We use tweets from six major US airlines and divide them into good, negative, and neutral categories to create a score for each airline in order to evaluate them against one another.

The tweets for the six airlines mentioned above are taken from www.Kaggle.com. Customer Tweets are statements made by consumers who have previously travelled with certain airlines. Based on the phrases used in the tweets, they are evaluated as positive, bad, or neutral, and then passed to the next phase. Naïve Bayes and emotional analysis were used. Following the categorization of tweets into positive, negative, and neutral categories, the total number of positive, negative, and neutral tweets is calculated using two algorithms: Naïve Bayes and sentimental analysis. Scores and method comparisons.

From different viewpoints, this study examines and infers recommendations asserted by critics of travel services, particularly airline services, in internet reviews. This explanatory study is especially valuable for service companies in determining which components of their services influence consumer eWOM promotion via online reviews. The research builds on the previous study's conclusions. to see if the contents conveyed in online reviews can predict reviewers' service recommendations. In the realm of obtaining customer recommendations from online evaluations, there is a research gap. We pay particular attention to the word lists for positive (pos) and negative (neg) terms, which are used to identify the polarity of emotion portrayed in the various evaluations. NB (Nave Bayes) is a relatively simple learning algorithm, while more sophisticated learning techniques include NN (Neural Network) and SVM (Support Vector Machine) [27].

Another line of the study focuses on price and marketing techniques that businesses may use in a multi-period game to build favourable online evaluations in first phase in the second quarter, capitalize on the positive influence of those reviews on their pricing, sales, and profitability [28].

In this study Dr Swagato [29] focuses on pricing and marketing methods that businesses may employ in a multi-period game, they generate good online reviews in the first period and then capitalize on the beneficial influence of those reviews on their pricing, sales, and profits in the second period. (Pos neg) / (pos + neg) was used to compute overall sentiment, with pos representing the frequency of positive terms and neg reflecting the frequency of negative keywords in the text. Consumer satisfaction is a powerful determinant of both consumer outcomes and service settings. The proportional relevance of good and negative emotions has also been discovered.

Detrimental emotions including disgust, fear, and rage have been shown to have a negative impact on customer outcomes.

In this study [4] we used Bidirectional Encoder Representations from Transformers to develop a Convolutional Neural Network model that can suggest airline tickets based on the sentiment of online reviews (BERT). Using reviews online from six social media sites, several sentiment categorization systems are examined. Based on a range of characteristics such as airline consumer happiness, as well as travel destinations, hotel details, restaurants, and tourist trap the new model may categorize airline tickets as inexpensive or not affordable. Data is gathered from several airlines. The model's first component is the classification of Tweet sentiment. In the first part of this investigation, BERT categorizes tweets into positive and negative categories. 50,000 movie reviews were used to teach BERT at first. The BERT model is SST2 [12] is based on the Stanford Sentiment Treebank. SST2 is a human-annotated movie review-based single-sentence categorization job. The data set was divided into two sections: Training accounts for 70% of the budget, while testing accounts for 30% CNN model accuracy performance is good.

In this study Amine Dadoun [30] purposed a study as in which During the data collection and preparation stages, the received material is examined for recognizing opinions. It is very dependent on the unique technique taken as well as the words chosen to express emotions. This is where superfluous opinions are removed by pre-processing. Positive, negative, and neutral orders are the most common classifications. There are three basic machine learning models to consider. SVM, KNN, and decision tree are three types of support vector machines (DT). These models' performance is compared using accuracy, precision, recall, and measure.

Praphula Kumar Jain purposed a study [31]. Our data comes from internet Skytrax is a website that provides airline evaluations and ratings, which can be found at www.airlinequality.com. To enhance the efficiency of the study, some factors that provide less influence are deleted during the pre-processing step. The reviews' text is subdivided, and each one is characterized by a vector of standard word weightings that indicate the value of each word in the review. Tf-idf is the most common weighting used here. All these characteristics are used to predict consumer recommendations. Logistic regression is demonstrated via SVM, Decision Tree, and Random Forest. For prediction based on algorithms, we used four machine learning techniques: are all examples of logistic regression. The models are tested and validated using Stratified K- fold cross-validation. All four models are produced and evaluated in each of the feature kinds. The performance of all the models is shown. The highest accuracy (91%) is achieved by Logistic Regression, while the lowest accuracy is achieved by decision tree (87%).

Praphula Kumar Jain [32] was purposed a study viewer material in internet reviews provides a rich supply of customer feelings that may be used to evaluate services. For our experimental research, we collect online review data from the Skytrax website. 1 Skytrax is a travel consultancy company in the United Kingdom. The website offers airline passengers a free and open forum to share their experiences with various services. Customer polarity may be determined using sentiment analysis algorithms feedback, i.e., measuring whether consumers felt happy, negative, or neutral about an airline's overall service. To address this, the Valence Aware Dictionary for Sentiment Reasoning was used. The Cuckoo Search (CS) algorithm, created by scholars, a meta-heuristic algorithm is a computer program that uses a set of rules to solve a problem. The technique for parameter optimization is stated as an optimization problem, and the optimal parameters are generated using the CS algorithm. Our suggested CS method is used in this study to SVM, DT, LGB, and XGB are four classification techniques that can be used [22]. we prudently defined the hyper- parameter range space for the four separate classifications based on the expertise of specialists in the subject. The data was split into two groups: training (80%) and testing (20%). (20 percent). On the training set, -fold cross-validation with CS-based hyper- parameter adjustment is used for validation. The accuracy of results was much better and above 90%.

In another research by Liau and Tan [33] mentioned 10,895 tweets from Malaysian low- cost airline customers. In order to examine the thoughts of airline passengers, two algorithms for subject recognition and two algorithms for sentiment analysis were evaluated. According to the data, client service, ticket sales, aircraft cancellation and delays, and post-booking management were among the subjects cited by customers on Twitter. Regarding sentiment analysis, clients' attitudes toward the four key issues were more favourable than negative. Customers were more likely to tweet favourable things about customer service, booking managing, and ticket promotions than they were to post bad things about airline cancellations and delays.

Filipe R. Lucini was purposed a study [34] We used data from the Air Travel Review (ATR) website, which was developed and is currently administered by Skytrax and can be accessed at airlinequality.com. ATR is a non-profit consumer forum that has evolved to become the most popular website for user reviews of airlines, airports, and air travel. Individual OCRs are recorded in ATR, allowing us to evaluate and aggregate individual evaluations for analysis. On August 11, 2016, we created a web crawler for this purpose, which gathered all ATR OCRs. The Python-based web crawler visited the webpage and grabbed the HTML code. The pre-processing processes used in this study are similar to those used in other studies. To offer -fold cross-validation for models trained to a range of 2 to 100 topics, the dataset was split into train and test sets. The Nave Bayes Classifier is based on the supervised Bayes theorem. If the adjective has been used in a review, the emotion intensity of an adjective is determined by the likelihood of a class occurring (e.g., positive or negative sensation). For the five perplexity values obtained, we estimated the mean, standard deviation, and 95 % confidence interval.

Aditi Saxena [35] The Internet plays a crucial part in this study, as the dataset was acquired from www.kaggle.com, which comprises tweets for the six U.S. airlines listed above. Tweets from Consumers refers to the remarks made by numerous customers who have travelled with these airlines. The tweets are classified as favourable, negative, or neutral based on the terms they include, and then the following step is taken. Using Naive Bayes and emotional analysis, after tweets are categorized as positive, negative, or neutral, two algorithms, Naive Bayes and sentimental analysis, are employed to determine the overall quantity of positive, negative, and neutral tweets. Utilizing Sentiment Analysis, social media items such as tweets and status updates are analysed for their sentiment. Sentimental analysis uses the following procedure: Select a sequence of characters and evaluate them as favourable, bad, or neutral. Nave Bayes classifiers are basic probabilistic classifiers that function when Bayes' theorem is applied with effective independence requirements between their features (called naive). Since the early 20th century, Naive Bayes has been the subject of much research. Here, the x-axis indicates the name of each airline and the y-axis reflects the sentiment analysis and Nave Bayes-calculated ratings. The examination of sentiment is represented by diamonds, whereas Nave Bayes is represented by squares.

Nikkkolao Korfiats [36] the dataset was consisted of publicly accessible reviews from the prominent travel website Booking.com1 that were obtained with the help of a web crawler. We validate the homogeneity of the data before proceeding with data processing for suggestion provision by learning individual user profiles. The point cluster indices are returned by the k-means algorithm as an i-by-1 vector. K-means uses squared Euclidean distances by default. The method of iterative clustering described in Section 3.2 was used. The first clustering, which is the outcome of 14 rounds, serves as an example of this procedure.

### *B. eWOM*

Lobel trong Thuy Tran [37] have sought to create advanced automation data- driven prediction algorithms that identify critical patterns and insights from vast online consumer assessments, eventually assisting travellers and marketers in making successful judgments. Although the aims and techniques of these research varied, they were all focused on using internet reviews to gather, analyze, and understand user experiences A. Navitha Sulthana [10] purposed a study they contrasted the significance eWOM marketing to conventional marketing. eWOM improves the amount of people who join online social networking sites [38]. Current members exchange referrals and recommendations using social media in addition to traditional marketing. Consequently, new members are linked to social networking sites and exchange eWOM information. The greater the reach of eWOM recommendations, the greater the acquisition of new customers. With the computerized tracking of WOM referrals, the acquisition of new customers in social network sites may be monitored. With the aid of eWOM posted on social media websites, prior to making a purchase, customers gather information about a product or service.

P. Yu. Michell [38] The author examined I customer reviews, (ii) reviewer personality, (iii) review website characteristics, (iv) product review characteristics, (v) environmental influence, and (vi) interpersonal to determine which component had the most impact on buy intent. Three hundred and thirty-seven university students were polled for information. The author discovered that six eWOM characteristics had a favourable effect on purchasing intent. Customer reviews are the most influential aspect on a consumer's buying intent. According to Francesca Di Virgilio and Gilda Antonelli [39] eWOM and Trust serve as a moderating variable for online purchasing intent. On social networking websites, consumers are becoming more familiar with products and product-related information. Trust and electronic WOM communication have an effect on buying intent. Users may produce and distribute user-generated content using Web 2.0 technologies. The sharing of product and service material in multiple online venues is known as online word of mouth/eWOM.

Francesca Di Virgilio and Gilda Antonelli [10] Web 2.0 technology has been identified that enable users to produce and exchange user-generated content. The spreading of information about a product or service across many online contexts is known as online word of mouth (eWOM). eWOM and Trust serve as a moderating variable for the intention of online purchase behaviour of users. On social networking sites, consumers are becoming more familiar with products and product-related information. Social media, electronic WOM communication, and trust all impact purchase intent.

### C. *Polar Reviews on Airlines*

Positive sentiment exposure can also predict higher negative sentiment expression [40]. Customers' positive and negative reviews of a product or service, made public to a huge number of individuals and organizations through the Internet, are an effective marketing tool [41]. Consumers' perceptions of utility and ease of use are positively influenced by the quality, quantity, and timeliness of unfavourable internet reviews [42]. Positive (firms grow and invest in resources and capabilities, resulting in benefits for consumers, workers, regions, and other firms) or negative (firms shrink and invest in resources and capabilities, resulting in benefits for consumers, workers, regions, and other firms) or negative (firms shrink and invest in resources and capabilities, resulting in benefits for consumers, workers, regions, and other firms) or negative (firms shrink and invest in resources and capabilities, resulting in benefits for consumers, workers, regions, and other firms) (i.e., the negative effects of these investments on these subjects).

Rajesh Tolety [43] Using import.io, we were able to retrieve the dataset from three distinct websites, namely Trip Advisor, Consumer Affair, and Airline quality. We attempted to collect data from Twitter but discovered that it only supplies the last weeks' worth of information. The Text Parsing node analyses a collection of documents in order to quantify information about the words. The node is used to parse text data using several grammatical categories and noun groupings. Fewer words are issued by examining the significance of the words. Text Filter node is utilized to further decrease the number of parsed phrases to be evaluated. The objective is to reduce unnecessary information so that only the most useful and pertinent data are examined. The text rule builder node is used to produce a collection of rules that predict a target variable using subsets of phrases. In this situation, the goal variable is binary, indicating whether the feedback is favourable or negative. While collecting the data, based on the value of the customer rating, we classified it as positive or negative.

### D. *Conflicting reviews*

According to Wang et al. [44] product or service-related information that contains both favourable and unfavourable information at the same time, or just positive information, will have boring effects on attitudes. Mixed information, such as contradicting evaluations, has been demonstrated to contribute to favourable, negative, or inconclusive sentiments in the past [45]. According to Lee [46]. Online reviews are subjective and may contain contradictory information Previous study has found that mixed information, such as contradicting appraisals, leads to positive, negative, or ambiguous sentiments. As a result of these reviews, a consumer develops a mixed and bivalent subjective state known as attitude ambivalence. According to Jacoby [47], subconscious processing occurs as a result of exposure to stimuli (conflicting internet reviews). Working memory's limited capacity drives humans to select stimuli in a competitive manner, according to attention researchers [48].

### E. *Attitude Ambivalence*

The existing work ignores the impact of online and offline elements on ambiguous attitudes that affect customer behaviour, and on contradictory internet reviews Unlike attitude, which records attitude ambivalence, univalent (i.e. positive or negative) assessments records conflicting or conflicted sentiments about a stimulus. [49]. In another significant study, Boukamcha [54] alleged that attitude ambivalence causes contradiction between attitude and behaviour requiring additional cognitive effort and motivation to repair the research gap. The antecedents of attitude ambivalence and its influence on consumer behaviour have been widely addressed in the information processing literature. Individuals' perceptions of the legitimacy of content are influenced by the star ratings of contradicting evaluations from several reviewers, resulting in attitude ambivalence [47]. When consumers are exposed to contradictory online information (i.e., good and negative information at the same time), they are likely to become confused and rate the information's dependability as lower.) [50]. Ambivalence has been used in significant study issues like recall and cognition, as well as green marketing. Uncomfortable tension will result from cognitive dissonance [51] discovered that French customers who hold both good and negative opinions of the United States had a more ambivalent attitude toward American brands. When important persons hold strongly opposing viewpoints to the focus person's, the latter is likely to have conflicted feelings about the topic at hand. Furthermore, Roster and Richins [52] discovered that a shopper's personal view, as well as the opinions of crucial persons, contribute to ambiguity in product replacement decisions as opposed to cognitive dissonance, ambivalence is also regarded an intra-attitudinal discrepancy [53]. A customer develops a confused and bivalent subjective state known as attitude ambivalence as a result of these reviews [50]. In another significant research, Boukamcha [54] claimed that attitude ambivalence generates contradiction between attitude and conduct. To narrow the research gap, greater cognitive effort and motivation are required. In a decision-making scenario, Ambivalence is defined as the presence of both positive and negative evaluations of an attitude object. Which can lead to an uncomfortable emotional state [55, 56].

## III. MATERIALS AND METHODS

We are constructing a recommendation system for tourists' convenience based on their internet reviews in this study. We are developing a methodology for this aim, which is detailed below.

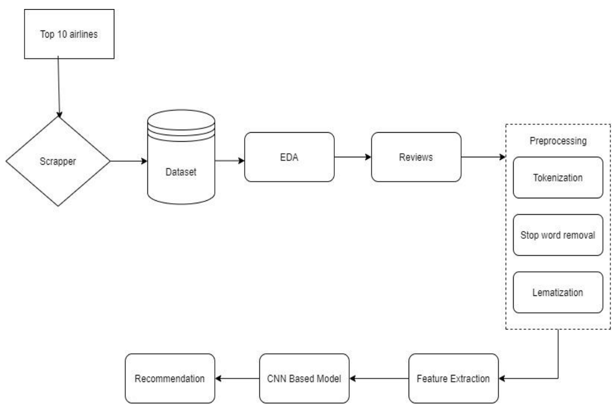
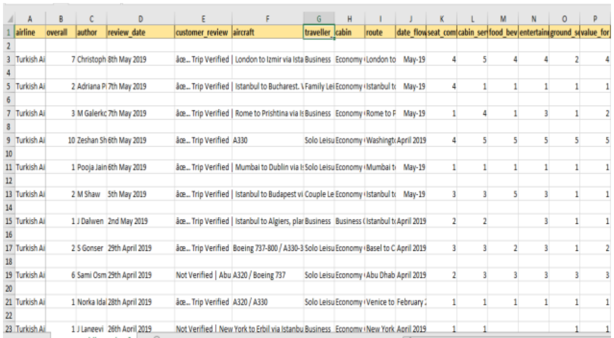


Fig. 1. Proposed Model

The process of acquiring data is known as data collection, measuring, and assessing correct insights for research following defined, accepted techniques. The data collection process is based on the answers to the following key question: What are customers' impressions of airline service quality, both in terms of the airline and the destination? In order to increase service quality and fulfil customer expectations, the airline sector must first define their problem. To do so, companies must first figure out which tourist locations cause unfavourable publicity for their enterprises. Data is scraped from top ten websites. These websites are as followings Cheapair, Priceline, TripAdvisor Flights, Kayak, Expedia, Orbitz, Skyscanner, Travelocity, Hotwire, Momondo.

### *A. Dataset*

collected information from many airline websites. The amount of data here is enormous for this recommender system. As a result, we're eliminating all the unnecessary information and simplifying it to meet our needs paper following the 14 parameters that are chosen based on the state of the art. In order to construct the dataset, we are using the information above. This dataset is shown as in figure below.



| | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | airline | overall | author | review_date | customer_review | aircraft | traveller_ | cabin | route | date_flow | seat_com | cabin_ser | food_bev | entertain | ground_s | value_for |
| 2 | | | | | | | | | | | | | | | | |
| 3 | Turkish Ai | 7 | Christoph | 8th May 2019 | âœ... Trip Verified \| | London to Izmir via Ista | Business | Economy | London to | May-19 | 4 | 5 | 4 | 4 | 2 | 4 |
| 4 | | | | | | | | | | | | | | | | |
| 5 | Turkish Ai | 2 | Adriana P | 7th May 2019 | âœ... Trip Verified \| | Istanbul to Bucharest. \ | Family Lei | Economy | Istanbul t | May-19 | 4 | 1 | 1 | 1 | 1 | 1 |
| 6 | | | | | | | | | | | | | | | | |
| 7 | Turkish Ai | 3 | M Galerkc | 7th May 2019 | âœ... Trip Verified \| | Rome to Prishtina via Is | Business | Economy | Rome to F | May-19 | 1 | 4 | 1 | 3 | 1 | 2 |
| 8 | | | | | | | | | | | | | | | | |
| 9 | Turkish Ai | 10 | Zeshan Sh | 6th May 2019 | âœ... Trip Verified | A330 | Solo Leisu | Economy | Washingt | April 2019 | 4 | 5 | 5 | 5 | 5 | 5 |
| 10 | | | | | | | | | | | | | | | | |
| 11 | Turkish Ai | 1 | Pooja Jain | 6th May 2019 | âœ... Trip Verified \| | Mumbai to Dublin via I | Solo Leisu | Economy | Mumbai t | May-19 | 1 | 1 | 1 | 1 | 1 | 1 |
| 12 | | | | | | | | | | | | | | | | |
| 13 | Turkish Ai | 2 | M Shaw | 5th May 2019 | âœ... Trip Verified \| | Istanbul to Budapest vi | Couple Le | Economy | Istanbul t | May-19 | 3 | 3 | 5 | 3 | 1 | 1 |
| 14 | | | | | | | | | | | | | | | | |
| 15 | Turkish Ai | 1 | J Dalwen | 2nd May 2019 | âœ... Trip Verified \| | Istanbul to Algiers, plar | Business | Business ( | Istanbul t | April 2019 | 2 | 2 | | 3 | 1 | 1 |
| 16 | | | | | | | | | | | | | | | | |
| 17 | Turkish Ai | 2 | S Gonser | 29th April 2019 | âœ... Trip Verified | Boeing 737-800 / A330-3 | Solo Leisu | Economy | Basel to C | April 2019 | 3 | 3 | 2 | 3 | 1 | 2 |
| 18 | | | | | | | | | | | | | | | | |
| 19 | Turkish Ai | 6 | Sami Osm | 29th April 2019 | Not Verified \| Abu | A320 / Boeing 737 | Solo Leisu | Economy | Abu Dhab | April 2019 | 2 | 3 | 3 | 3 | 3 | 3 |
| 20 | | | | | | | | | | | | | | | | |
| 21 | Turkish Ai | 1 | Norka Ida | 28th April 2019 | âœ... Trip Verified | A320 / A330 | Solo Leisu | Economy | Venice to | February | 1 | 1 | 1 | 1 | 1 | 1 |
| 22 | | | | | | | | | | | | | | | | |
| 23 | Turkish Ai | 1 | J Laneevi | 26th April 2019 | Not Verified \| New York to Erbil via Istanbu | | Business | Economy | New York | April 2019 | 1 | 1 | | | 1 | 1 |

Fig. 2. Dataset

### *B. Data scraping*

With the widespread use of the internet, the idea of utilizing its resources to aid in the recommending process has gained traction in recent years. On the web, you may find almost any type of information in a variety of formats. Web scraping is the process of extracting data from the internet. It is a sort of retrieval of web content. Web scraping has been present since the dawn of the Internet but scraping modern sites that rely heavily on new technology is far from simple. Web scrapping is need because we are collected dataset from different websites. This dataset is not refined form. Use of the csv, xls scrap the data. Almost all best airlines in the world are included to extract the data. This data is unstructured form. This dataset Different type of websites collected data and then scraped convert into a csv file. For data scrapping different libraries are used as beautiful soap 4, selenium.

*Scrap data convert into Dataset*

Data is scraped from several websites as described above, and then converted into a dataset format. Then create a CSV file. We are not using this junk data because it has various properties and raw materials, thus we are transforming it to a usable data format. Dataset contain 14 features and the target attribute is "Recommendation" the others are independent variables.as shown in figure below.

| No | Attributes | Description |
|---|---|---|
| 1 | Review Data | reviews are within or review or across |
| 2 | Customer reaction | ignore or select the airline |
| 3 | Aircraft Type | The Jets. Jets of light. Jets of med |
| 4 | Compartment | Good or bad |
| 5 | Route of plan | Airline routes, destination |
| 6 | Date of Flight | date of your booking flights |
| 7 | Seat comfort | comfort or safety |
| 8 | Cabin service | Good or bad |
| 9 | Food Beverage | fast food, soft drink, drink water |
| 10 | Entertainment | watch a film, write a journal, paly listen to a music |
| 11 | Traveler kind | socially conscious, education plat enjoy |
| 12 | Ground service | Executive<br>lounges cabin<br>presentation<br>baggage |
| 13 | Value for money | Business class or economy |
| 14 | Recommendation | Yes or no |

### *C. Data description*

The dataset used in this study comprises information on different websites. Dataset is not publicly available. Datasets contain many features like as airline, compartment, value of money, food beverage and many others shown as in figure below .It contain 131.895 records and 14 attributes. Description of features are as following.

Table 1 Table of Attributes

### *D. Data preprocessing*

The Pre-Processing step comes next. This is one of the most crucial aspects of the process. Data pre-processing is an important stage in developing a Machine Learning model, and the quality of the data depends on how well it was pre-processed. When data is pre-processed, machine learning models demonstrate enhanced categorization accuracy. Python's natural language toolkit was used to perform the pre-processing. Punctuation, stop words, and the use of lower- and capital words in reviews can all impair the model's capacity to learn. The data collected for any experiment is usually not in an appropriate format.

*Tokenization*

During this step, the text is subdivided into smaller sections. Sentence tokenization or word tokenization may be an option, depending on the details of the situation.

*Stemming*

The text standardization step is the process of stemming or reducing words to their root/base form. For example, the words 'programmer,' 'programming,' and 'programmed' will be stemmed to 'programmed.'

*Stop word removal*

Stop words are words that are removed from a document because they provide no value to the analysis. These words are ambiguous at best. A collection of terms deemed stop words in English is included in the NLTK library. I, we, they, you, would, now, should, our, yours, other, will, very, then, not, only, some, ourselves, so, he, yourself, don’t, can, nor most, myself, should have, own, she, you, will, such.

## E. *Data analysis*

Data analysis is the first and most critical stage in any forecast, recommendation (DA). DA is the process of identifying patterns, anomalies, testing ideas, and verifying assumptions using summary statistics and graphical representations. We are analysing the frequency or values of airlines shown as below.

Data analysis is the process of extracting usable information from data and then depending only on that data. In order to find relevant information for business decision-making, data must be cleaned, transformed, and modelled. This is nothing more than looking backwards or forwards in time and making judgments depending on our findings in our daily lives, when we make a decision, we examine what has happened previously or what will happen if we make that option. We accomplish this through recalling past events or speculating about the future. So that's the end of data analysis. An analyst currently does data analysis. For commercial goals

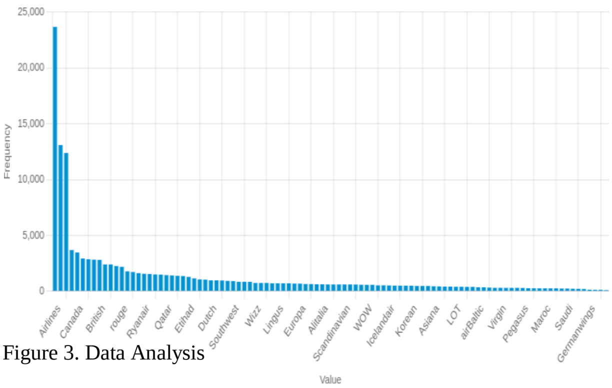



Figure 3. Data Analysis

## F. *EDA*

Exploratory dataset analysis (EDA). It's possible to utilize EDA in conjunction with or without a statistical methodology, although it's most often used to see what data can tell us without resorting to formal modelling or hypothesis testing. Exploratory data analysis is required while working with data. It's common for exploratory data analysis to include the use of statistical graphics and other data visual analytics to summarize essential aspects of a data collection. Exploratory Data Analysis can help with all of this. It removes abnormalities and extraneous values from data, allowing you to gain insights and make better sense of it.

Steps involve in EDA are as following:

- Assist in the preparation of your dataset for analysis.
- It also assists us in selecting a more appropriate machine learning model.
- This enables a machine learning model to better forecast our dataset.
- Provides more precise results.

*Data cleaning*

The process of cleaning data is known as data cleansing eliminating unwanted variables and values from your dataset, as well as any anomalies. Such anomalies can distort the data excessively, affecting the outcomes negatively. Cleaning data can include the following steps:

*Null values removal*

What is the meaning of a NULL value? A NULL value is a particular SQL marker that indicates that a data value in the database does not exist. In other words, it's only a placeholder for values that aren't known or are missing. In our dataset has many null values we are removing these values using the technique of cleansing values. In our dataset 65,947 null values.

| | index | AirLine | OverAll | Author | Review Date | |
|---|---|---|---|---|---|---|
| 0 | 7 | Turkish Airlines | 10.00 | Zeshan Shah | 6th May 2019 | â... Trip Verified \| Flew on … |
| 1 | 15 | Turkish Airlines | 2.00 | S Gonser | 29th April 2019 | â... Trip Verified \| Basel to … |
| 2 | 17 | Turkish Airlines | 6.00 | Sami Osman | 29th April 2019 | Not Verified \| Abu Dhabi to … |
| 3 | 19 | Turkish Airlines | 1.00 | Norka Idalia Orlando | 28th April 2019 | â... Trip Verified \| The exp… |
| 4 | 29 | Turkish Airlines | 2.00 | Trevor Khurana | 24th April 2019 | â... Trip Verified \| Houston … |
| 5 | 35 | Turkish Airlines | 5.00 | Gyan Fernando | 21st April 2019 | â... Trip Verified \| Nakhchi… |
| 6 | 37 | Turkish Airlines | 7.00 | Gyan Fernando | 21st April 2019 | â... Trip Verified \| London … |
| 7 | 41 | Turkish Airlines | 8.00 | S Tekin | 20th April 2019 | â... Trip Verified \| Izmir to … |
| 8 | 53 | Turkish Airlines | 10.00 | P Bartinescu | 15th April 2019 | â... Trip Verified \| Montrea… |
| 9 | 63 | Turkish Airlines | 9.00 | W Keale | 11th April 2019 | â... Trip Verified \| Manche… |
| 10 | 75 | Turkish Airlines | 6.00 | J Ngui | 30th March 2019 | â... Trip Verified \| Current… |
| 11 | 79 | Turkish Airlines | 1.00 | Jaylesh Gandhi | 28th March 2019 | â... Trip Verified \| Flew TK… |
| 12 | 81 | Turkish Airlines | 1.00 | Andrei Muravev | 20th March 2019 | â... Trip Verified \| I bought … |
| 13 | 87 | Turkish Airlines | 6.00 | S Railey | 13th March 2019 | â... Trip Verified \| Manila … |
| 14 | 89 | Turkish Airlines | 1.00 | Masudur Rahman | 13th March 2019 | â... Trip Verified \| Dhaka … |
| 15 | 91 | Turkish Airlines | 1.00 | Oscar Marcheggiani | 13th March 2019 | â... Trip Verified \| Hanoi … |
| 16 | 95 | Turkish Airlines | 4.00 | Sebastian Pietzsch | 12th March 2019 | Not Verified \| Boarding total… |
| 17 | 101 | Turkish Airlines | 8.00 | EY Genis | 10th March 2019 | â... Trip Verified \| If you … |
| 18 | 103 | Turkish Airlines | 9.00 | EY Genis | 10th March 2019 | â... Trip Verified \| Istanbul … |
| 19 | 105 | Turkish Airlines | 8.00 | Bobby Giannakopoulos | 8th March 2019 | â... Trip Verified \| This was … |
| 20 | 107 | Turkish Airlines | 4.00 | E Jenginen | 5th March 2019 | â... Trip Verified \| Dublin … |
| 21 | 123 | Turkish Airlines | 1.00 | Julie Fry | 24th February 2019 | â... Trip Verified \| Never … |
| 22 | 127 | Turkish Airlines | 4.00 | T Bartin | 20th February 2019 | â... Trip Verified \| Montrea… |
| 23 | 143 | Turkish Airlines | 8.00 | Manos Stakianakis | 13th February 2019 | â... Trip Verified \| Athens … |
| 24 | 157 | Turkish Airlines | 7.00 | L Meister | 5th February 2019 | â... Trip Verified \| Sharm … |
| 25 | 161 | Turkish Airlines | 10.00 | Feldman Pinchas | 31st January 2019 | â... Trip Verified \| Tel Aviv … |
| 26 | 163 | Turkish Airlines | 7.00 | N Darcatan | 28th January 2019 | â... Trip Verified \| Baku to … |

Figure 4. Null Values Removal

*Missing values*

In the EDA approach, we employ missing values. Missing data reduces the sample's generalization, which might lead to incorrect population inferences. There are three primary approaches to dealing with missing data in general: (1) imputation, which substitutes values for missing data, (2) omission, which excludes samples with erroneous data from further analysis, and (3) analysis,

which uses processes unaffected by missing values directly. Missing values are eliminated, as seen below.

Figure 5. Missing Values

*Outliers*

An outlier is a value that deviates significantly from other values in a random sample from a population. In some sense, this term outsourced to the analyst the judgement of what constituted abnormality (or a consensus procedure). Normal observations must be characterized before aberrant data may be discovered.

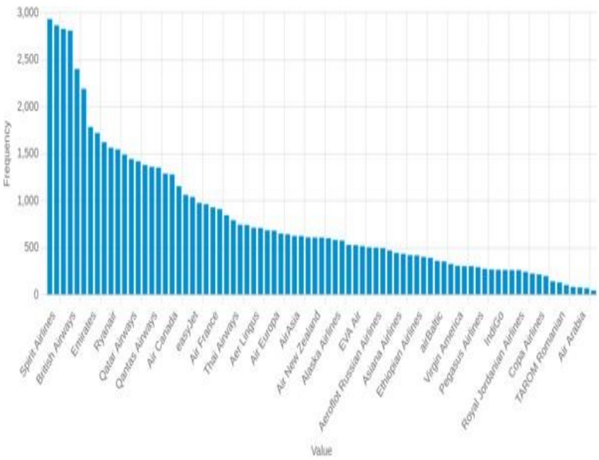


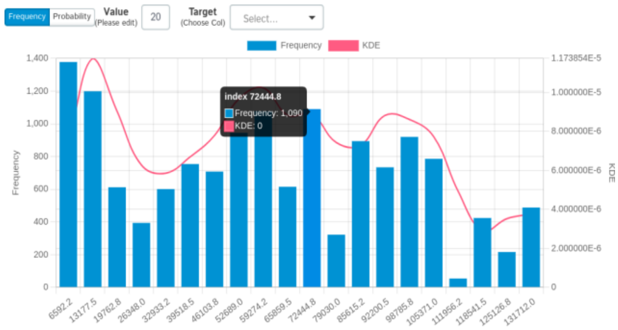

Figure 6. Outliers

*Descriptive Statistical Analysis*

Descriptive analytics and inferential statistics were distinct in some circumstances. Descriptive statistics simply describe what the data is or indicates. The mean, the standard deviation, and the frequency of a variable are examples of descriptive statistics. Descriptive statistics are used to define or summarize the features of a sample or data collection. Inferential statistics can help us grasp the collective properties of a data set's components. Inferential statistics are used to draw inferences based on incomplete data. For instance, on the basis of sample data, inferential statistics are used to try to determine what the overall population believes. To establish whether a difference between groups found in this study is reliable or just coincidental, we may apply inferential statistical techniques. In order to estimate general conditions from our data, inferential statistics are needed, while descriptive statistical analysis is only utilized to describe what is happening in the data.

*Heat map*

Visualizations of two-dimensional data using color to depict magnitude are known as "heat maps" or "heatmaps." The change in color could be in hue or intensity, and it would show the reader how the event is set up or how it happens over time. The size of each cell in a matrix is set, and each row and column represent a different phenomenon or category.




The order of rows and columns is both planned and random, with the goal of finding or showing clusters found by statistical analysis. This shows the relationship between each feature and each feature. There are two kinds of heat maps: the clustered heat map and the spatial heat map.

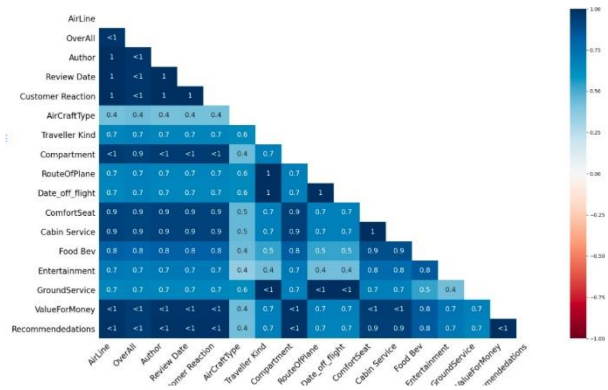
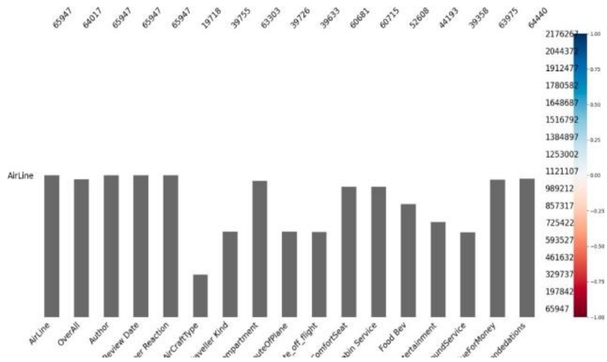



Figure 7. Heat Maps




*Bar chart*

The bar chart, which is also referred to as a bar graph, is a kind of chart or graph that utilizes rectangular bars with heights or lengths that match the values they represent in order to depict categorical data. Bar charts and bar graphs are commonly used interchangeably. There are two possible ways to layout the bars: horizontally or vertically. A vertical bar chart may also be referred to by its alternative term, a column chart. A bar chart, like the one that can be seen below, illustrates the relationship that exists between each characteristic.

Figure 8. Bar Chart

*Histogram*

Histograms are visual representations of a set of data points that have been divided into user-specified ranges by the user. Data series may be condensed into an easily understandable visual using the histogram, which resembles the bar graph.

Figure 9. Histogram

*Dataset after preprocessing*

Dataset after preprocessing look like as null values remove, nan, missing values etc. All raw data has been eliminated, and we are just keeping the true data, or the data that we have chosen for accuracy.

| | Index | AirLine | OverAll | Author | Review Date | |
|---|---|---|---|---|---|---|
| 0 | 7 | Turkish Airlines | 10.00 | Zeshan Shah | 6th May 2019 | âœ… Trip Verified \| Flew on Turkish Airlines IAD-IST-KHI and return KHI-IST-IAD. Turkish Airlines has consistently maintained its quality since I |
| 1 | 15 | Turkish Airlines | 2.00 | S Gonser | 29th April 2019 | âœ… Trip Verified \| Basel to Cape Town via Istanbul. When I arrived in Istanbul at 10pm we are informed that TK44 to Cape Town (departure |
| 2 | 17 | Turkish Airlines | 6.00 | Sami Osman | 29th April 2019 | Not Verified \| Abu Dhabi to Luxembourg via Istanbul. From AUH-IST, as the flight was at 1:30am, the flight was I comfortable due to the small |
| 3 | 19 | Turkish Airlines | 1.00 | Norka Idalia Orlando | 28th April 2019 | âœ… Trip Verified \| The experience with Turkish Airlines has been devastating one. First paid $200 Euros per luggage, totaling $400 Euros for |
| 4 | 29 | Turkish Airlines | 2.00 | Trevor Khurana | 24th April 2019 | âœ… Trip Verified \| Houston to Kiev via Istanbul. Fares seem competitive, but there is a catch in terms of layover time. Mine from Houston to |
| 5 | 35 | Turkish Airlines | 5.00 | Gyan Fernando | 21st April 2019 | âœ… Trip Verified \| Nakhchivan to Istanbul. The flight set off a little bit late and boarding was disorderly with too many passengers carrying o |
| 6 | 37 | Turkish Airlines | 7.00 | Gyan Fernando | 21st April 2019 | âœ… Trip Verified \| London Heathrow to Istanbul. The flight from LHR took off on time. Boarding was hassle free. Seating was 3x3x3. Legroo |
| 7 | 41 | Turkish Airlines | 8.00 | S Tekin | 20th April 2019 | âœ… Trip Verified \| Izmir to Istanbul. It was my first Boeing 777 experience. They gave a free hot sandwich with cheese and vegetables and a |
| 8 | 53 | Turkish Airlines | 10.00 | P Bartinescu | 15th April 2019 | âœ… Trip Verified \| Montreal to Bucharest via Istanbul. Delicious food, great choice of beverages. Cabin crew kind and careful. I would choose |
| 9 | 63 | Turkish Airlines | 9.00 | W Keale | 11th April 2019 | âœ… Trip Verified \| Manchester to Bangkok via Istanbul return in the last month. All flights were on time. Excellent food and drink service. Lo |
| 10 | 75 | Turkish Airlines | 6.00 | J Ngui | 30th March 2019 | âœ… Trip Verified \| Currently on flight Singapore to Ä°stanbul, sadly been trying to get a drink and was told later, trolley came by serving coff |
| 11 | 79 | Turkish Airlines | 1.00 | Jaylesh Gandhi | 28th March 2019 | âœ… Trip Verified \| Flew TK 5 / 22 MAR, Istanbul - Chicago in business class. Had requested AVML (Asian vegetarian Meal). TK had forgotten |
| 12 | 81 | Turkish Airlines | 1.00 | Andrei Muravev | 20th March 2019 | âœ… Trip Verified \| I bought and paid tickets 3 months before departure. I had pre-selected certain seats convenient to us. I am a self-moving |
| 13 | 87 | Turkish Airlines | 6.00 | S Railey | 13th March 2019 | âœ… Trip Verified \| Manila to Istanbul. If you ever flew with Emirates or Qatar Airways, you will miss their service if you fly Turkish Airlines. I c |
| 14 | 89 | Turkish Airlines | 1.00 | Masudur Rahman | 13th March 2019 | âœ… Trip Verified \| Dhaka to Copenhagen via Istanbul. This trip was full of disappointments. Already the airport lounge, Dhaka, demanded mo |
| 15 | 91 | Turkish Airlines | 1.00 | Oscar Marcheggiani | 13th March 2019 | âœ… Trip Verified \| Hanoi to Milano via Istanbul. As we 4 arrived this morning from Hanoi at 6 am (TK168) we learned that our connecting flig |
| 16 | 95 | Turkish Airlines | 4.00 | Sebastian Pietzsch | 12th March 2019 | Not Verified \| Boarding totally chaotic and far too long. Nice meal, creepy safety video which distracted the people more than giving them adv |
| 17 | 101 | Turkish Airlines | 8.00 | EY Genis | 10th March 2019 | âœ… Trip Verified \| If you fly economy with an A330-300 of Turkish Airlines, be prepared for a problematic seat pitch. I just could not became |
| 18 | 103 | Turkish Airlines | 9.00 | EY Genis | 10th March 2019 | âœ… Trip Verified \| Istanbul to Ankara. Great short-haul flight. A warm delicious panini was accompanied by a selection of drinks. Turkish's IFE |
| 19 | 105 | Turkish Airlines | 8.00 | Bobby Giannakopoulos | 8th March 2019 | âœ… Trip Verified \| This was my first time using Turkish Airlines and I was satisfied. Check in at Lyon airport was fine. The flight was full but I |
| 20 | 107 | Turkish Airlines | 4.00 | E Jangnen | 5th March 2019 | âœ… Trip Verified \| Dublin to Manila via Istanbul. This flight was both a disaster and a nightmare. This incident happened during Istanbul-Man |
| 21 | 123 | Turkish Airlines | 1.00 | Julie Fry | 24th February 2019 | âœ… Trip Verified \| Never again will I use Turkish Airlines. We booked the ticket to Cape Town from Malaga (via Istanbul) 6 months in advance |
| 22 | 127 | Turkish Airlines | 4.00 | T Barlin | 20th February 2019 | âœ… Trip Verified \| Montreal to Berlin via Istanbul. I was mostly disappointed with the services. I flew from Montreal to Istanbul and from the |
| 23 | 143 | Turkish Airlines | 8.00 | Manos Stakianakis | 13th February 2019 | âœ… Trip Verified \| Athens to Havana via Istanbul. Amazing crew service. On time. Great selection of music and movies on the IFE. The food |
| 24 | 157 | Turkish Airlines | 7.00 | L Meister | 5th February 2019 | âœ… Trip Verified \| Sharm el Sheikh to Frankfurt via Istanbul. Both outbound flights quite okay. Pleasant staff, my husband's normal meal was |
| 25 | 161 | Turkish Airlines | 10.00 | Feldman Pinchas | 31st January 2019 | âœ… Trip Verified \| Tel Aviv to Brussels via Istanbul It's not my first time I'm flying with Turkish , it's the airline I like the most to fly with. Mos |
| 26 | 163 | Turkish Airlines | 7.00 | N Darcatan | 28th January 2019 | âœ… Trip Verified \| Baku to Istanbul. Frequent traveler of Turkish Airlines and somehow I see that quality decreases over time. Still smiling cr |

Figure 10. Dataset after Pre-Processing

## IV. RESULTS

In this paper CNN model is used for good accuracy. Convolutional neural networks, or convnets, are one of the most intriguing new advancements in machine learning. By extracting features from photos and using them in neural networks, they have revolutionized image categorization and computer vision. They have the same qualities that make them valuable for image processing that also make them useful for sequence processing. A CNN can be thought of as a customized neural network capable of detecting certain patterns. Convolutional layers are hidden layers seen in CNNs.

*Training or testing Data*

The method of picking data from the real data set for splitting is the key distinction between all these strategies. The data for this model is created using a 5- fold cross validation procedure. The 5-fold approach divides the dataset into five equal- sized subgroups at random. A fold is a phrase used to describe a gathering of people. The first four folds are used to build the ensemble model, which is then applied to the fifth fold. This procedure is done five times, with each fold of the dataset serving as a validation set once. As a result, no patient is employed in the development and testing of the model.

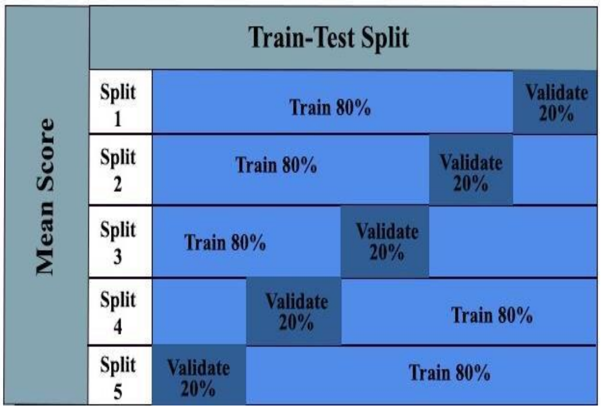



Figure 11. Training or Testing Data

Evaluation Matrix

Results are created through F measure, precision, Recall and find out the verification score,

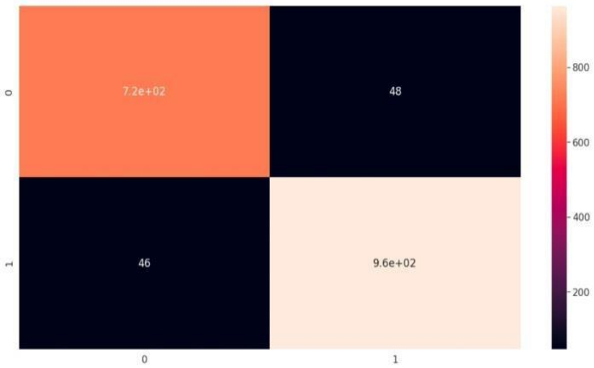



Figure 12. Evaluation Matrix

Using CNN model, the accuracy rate is 95%.

Table 2 Table of Results

| | precision | recall | F1-score |
|---|---|---|---|
| 0 | 0.94 | 0.94 | 0.94 |
| 1 | 0.95 | 0.95 | 0.95 |
| accuracy | | | 0.95 |
| macro avg | 0.95 | 0.95 | 0.95 |
| Weighted avg | 0.95 | 0.95 | 0.95 |

## V. CONCLUSION

Research purpose is to, as stated the reviews from different travelers and applying these , was to investigate the travelers' recommendation system. The CNN model is used to calculate the findings. F-measure, Precision, and Recall are examples of evaluation matrixes that are utilized to get better results. We looked at how the recommendation system works, how feature extraction works, and how the quantity of features extracted affects performance

## REFERENCES


[1] H. Arasli, M. B. Saydam, and H. Kilic, "Cruise travelers' service perceptions: a critical content analysis," Sustainability, vol. 12, no. 17, p. 6702, 2020.

[2] J. Kuljanin and M. Kalić, "Exploring characteristics of passengers using traditional and low-cost airlines: A case study of Belgrade Airport," Journal of Air Transport Management, vol. 46, pp. 12-18, 2015.

[3] A. Brochado, P. Rita, C. Oliveira, and F. Oliveira, "Airline passengers' perceptions of service quality: Themes in online reviews," International Journal of Contemporary Hospitality Management, 2019.

[4] M. Heidari and S. Rafatirad, "Using transfer learning approach to implement convolutional neural network model to recommend airline tickets by using online reviews," in 2020 15th International Workshop on Semantic and Social Media Adaptation and Personalization (SMA, 2020: IEEE, pp. 1-6.

[5] A. Khan, B. Baharudin, and K. Khan, "Sentiment classification from online customer reviews using lexical contextual sentence structure," in International Conference on Software Engineering and Computer Systems, 2011: Springer,pp. 317-331.

[6] R. P. de Oliveira, A. V. Oliveira, and G. Lohmann, "A Network-Design Analysis of Airline Business Model Adaptation in the Face of Competition and Consolidation," Transportation Science, vol. 55, no. 2, pp. 532-548, 2021.

[7] N. Donthu, S. Kumar, N. Pandey, N. Pandey, and A. Mishra, "Mapping the electronic word-of-mouth (eWOM) research: A systematic review and bibliometric analysis," Journal of Business Research, vol. 135, pp. 758-773, 2021.

[8] R. Filieri, S. Alguezaui, and F. McLeay, "Why do travelers trust TripAdvisor? Antecedents of trust towards consumer-generated media and its influence on recommendation adoption and word of mouth," Tourism management, vol. 51, pp. 174-185, 2015.

[9] Y. Yang et al., "Characteristics and formation mechanism of continuous hazes in China: a case study during the autumn of 2014 in the North China Plain," Atmospheric Chemistry and Physics, vol. 15, no. 14, pp. 8165-8178, 2015.

[10] A. N. Sulthana and S. Vasantha, "Influence of electronic word of mouth eWOM on purchase intention," International Journal of Scientific & Technology Research, vol. 8, no. 10, pp. 1-5, 2019.

[11] A. Audrezet and B. Parguel, "Using the Evaluative Space Grid to better capture manifest ambivalence in customer satisfaction surveys," Journal of Retailing and Consumer Services, vol. 43, pp. 285-295, 2018.

[12] R. Filieri and F. McLeay, "E-WOM and accommodation: An analysis of the factors that influence travelers' adoption of information from online reviews," Journal of travel research, vol. 53, no. 1, pp. 44-57, 2014.

[13] D. Wang, F. L. Weisstein, S. Duan, and P. Choi, "Impact of ambivalent attitudes on green purchase intentions: The role of negative moods," International Journal of Consumer Studies, vol. 46, no. 1, pp. 182-199, 2022.

[14] G. D. Moody, P. B. Lowry, and D. F. Galletta, "It's complicated: explaining the relationship between trust, distrust, and ambivalence in online transaction relationships using polynomial regression analysis and response surface analysis," European Journal of Information Systems, vol. 26, no. 4, pp. 379-413, 2017.

[15] L. Yang and H. R. Unnava, "Ambivalence, selective exposure, and negativity effect," Psychology & Marketing, vol. 33, no. 5, pp. 331-343, 2016.

[16] M. Zhang, B. Fan, N. Zhang, W. Wang, and W. Fan, "Mining product innovation ideas from online reviews," Information Processing & Management, vol. 58, no. 1, p. 102389, 2021.

[17] Y. Jin, A. Compaan, T. Bhattacharjee, and Y. Huang, "Granular gel support- enabled extrusion of three-dimensional alginate and cellular structures," Biofabrication, vol. 8, no. 2, p. 025016, 2016.

[18] H. Zhang, H. Rao, and J. Feng, "Product innovation based on online review data mining: a case study of Huawei phones," Electronic Commerce Research, vol. 18, no. 1, pp. 3-22, 2018.

[19] Z. Qiao, X. Zhang, M. Zhou, G. A. Wang, and W. Fan, "A domain oriented LDA model for mining product defects from online customer reviews," 2017.

[20] A. M. Evans, O. Stavrova, and H. Rosenbusch, "Expressions of doubt and trust in online user reviews," Computers in Human Behavior, vol. 114, p. 106556, 2021.

[21] A. Timoshenko and J. R. Hauser, "Identifying customer needs from user- generated content," Marketing Science, vol. 38, no. 1, pp. 1-20, 2019.

[22] M. Olmedilla, M. R. Martínez-Torres, and S. Toral, "Prediction and modelling online reviews helpfulness using 1D Convolutional Neural Networks," Expert Systems with Applications, p. 116787, 2022.

[23] Z. Xiang, Q. Du, Y. Ma, and W. Fan, "A comparative analysis of major online review platforms: Implications for social media analytics in hospitality and tourism," Tourism Management, vol. 58, pp. 51-65, 2017.

[24] S. Dooms, "Dynamic generation of personalized hybrid recommender systems," in Proceedings of the 7th ACM conference on Recommender systems, 2013, pp. 443-446.

[25] S. Bahulikar, V. Upadhye, T. Patil, B. Kulkarni, and D. Patil, "Airline recommendations using a hybrid and location based approach," in 2017 International Conference on Intelligent Computing and Control Systems (ICICCS), 2017: IEEE, pp. 972-977.

[26] R. Wagh and P. Punde, "Survey on sentiment analysis using twitter dataset," in 2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA), 2018: IEEE, pp. 208-211.

[27] M. Siering, A. V. Deokar, and C. Janze, "Disentangling consumer recommendations: Explaining and predicting airline recommendations based on online reviews," Decision Support Systems, vol. 107, pp. 52-63, 2018.

[28] A. Ajorlou, A. Jadbabaie, and A. Kakhbod, "Dynamic pricing in social networks: The word-of-mouth effect," Management Science, vol. 64, no. 2, pp. 971-979, 2018.

[29] S. Chatterjee, "Explaining customer ratings and recommendations by combining qualitative and quantitative user generated contents," Decision Support Systems, vol. 119, pp. 14-22, 2019.

[30] A. Dadoun, M. Defoin-Platel, T. Fiig, C. Landra, and R. Troncy, "How recommender systems can transform airline offer construction and retailing," Journal of Revenue and Pricing Management, vol. 20, no. 3, pp. 301-315, 2021.

[31] P. K. Jain, A. Patel, S. Kumari, and R. Pamula, "Predicting airline customers' recommendations using qualitative and quantitative contents of online reviews," Multimedia Tools and Applications, pp. 1-16, 2022.

[32] P. K. Jain, E. A. Yekun, R. Pamula, and G. Srivastava, "Consumer recommendation prediction in online reviews using Cuckoo optimized machine learning models," Computers & Electrical Engineering, vol. 95, p. 107397, 2021.

[33] B. Y. Liau and P. P. Tan, "Gaining customer knowledge in low cost airlines through text mining," Industrial management & data systems, 2014.

[34] F. R. Lucini, L. M. Tonetto, F. S. Fogliatto, and M. J. Anzanello, "Text mining approach to explore dimensions of airline customer satisfaction using online customer reviews," Journal of Air Transport Management, vol. 83, p. 101760, 2020.

[35] I. N. C. SN and R. D. Caytiles, "A Comparative study on Airline Recommendation System Using Sentimental Analysis on Customer Tweets," 2018.

[36] N. Korfiatis and M. Poulos, "Using online consumer reviews as a source for demographic recommendations: A case study using online travel reviews," Expert Systems with Applications, vol. 40, no. 14, pp. 5507-5515, 2013.

[37] L. T. T. Tran, "Online reviews and purchase intention: A cosmopolitanism perspective," Tourism Management Perspectives, vol. 35, p. 100722, 2020.

[38] P. Y. Michelle, "Electronic word of mouth influence on consumer purchase intention," Journal of Fundamental and Applied Sciences, vol. 10, no. 3S, pp. 126-141, 2018.

[39] F. Di Virgilio and G. Antonelli, "Consumer behavior, trust, and electronic word- of-mouth communication: Developing an online purchase intention model," in Social Media for Knowledge Management Applications in Modern Organizations: IGI Global, 2018, pp. 58-80.

[40] M. Salathé, D. Q. Vu, S. Khandelwal, and D. R. Hunter, "The dynamics of health behavior sentiments on a large online social network," EPJ Data Science, vol. 2, no. 1, pp. 1-12, 2013.

[41] J. Nicholas, A. S. Fogarty, K. Boydell, and H. Christensen, "The reviews are in: a qualitative content analysis of consumer

perspectives on apps for bipolar disorder," Journal of medical Internet research, vol. 19, no. 4, p. e7273, 2017.

[42] M. Liu, H. Lu, and J. Chen, "How Negative Online Reviews Affect Consumers' Purchase Intention," in 2021 International Conference on Big Data Analytics for Cyber-Physical System in Smart City, 2022: Springer, pp. 1029-1035.

[43] R. Tolety and S. K. Choudhary, "Text Analysis of American Airlines Customer Reviews," ed, 2018.

[44] H. Wang, R. Batra, and Z. Chen, "The moderating role of dialecticism in consumer responses to product information," Journal of Consumer Psychology, vol. 26, no. 3, pp. 381-394, 2016.

[45] Y. Hwang, S. Choi, and A. S. Mattila, "The role of dialecticism and reviewer expertise in consumer responses to mixed reviews," International Journal of Hospitality Management, vol. 69, pp. 49-55, 2018.

[46] J. Lee, D.-H. Park, and I. Han, "The effect of negative online consumer reviews on product attitude: An information processing view," Electronic commerce research and applications, vol. 7, no. 3, pp. 341-352, 2008.

[47] C. Ruiz-Mafe, K. Chatzipanagiotou, and R. Curras-Perez, "The role of emotions and conflicting online reviews on consumers' purchase intentions," Journal of Business Research, vol. 89, pp. 336-344, 2018.

[48] U. I. Siddiqi, J. Sun, and N. Akhtar, "The role of conflicting online reviews in consumers' attitude ambivalence," The Service Industries Journal, vol. 40, no. 13-14, pp. 1003-1030, 2020.

[49] E. Cornelis, N. Heuvinck, and A. Majmundar, "The ambivalence story: Using refutation to counter the negative effects of ambivalence in two-sided messages," International Journal of Advertising, vol. 39, no. 3, pp. 410-432, 2020.

[50] G.-H. Huang, N. Korfiatis, and C.-T. Chang, "Mobile shopping cart abandonment: The roles of conflicts, ambivalence, and hesitation," Journal of Business Research, vol. 85, pp. 165-174, 2018.

[51] C. A. Russell, D. W. Russell, and J. Klein, "Ambivalence toward a country and consumers' willingness to buy emblematic brands: The differential predictive validity of objective and subjective ambivalence measures on behavior," Marketing Letters, vol. 22, no. 4, pp. 357-371, 2011.

[52] C. A. Roster and M. L. Richins, "Ambivalence and attitudes in consumer replacement decisions," Journal of Consumer Psychology, vol. 19, no. 1, pp. 48-61, 2009.

[53] F. Van Harreveld, B. T. Rutjens, I. K. Schneider, H. U. Nohlen, and K. Keskinis, "In doubt and disorderly: Ambivalence promotes compensatory perceptions of order," Journal of Experimental Psychology: General, vol. 143, no. 4, p. 1666, 2014.

[54] F. Boukamcha, "The impact of attitudinal ambivalence on information processing and resistance to anti-smoking persuasion," Journal of Indian Business Research, 2017.

[55] W. Ran and M. Yamamoto, "Attitudinal ambivalence as a protective factor against junk food advertisements: A moderated mediation model of behavioral intention," Journal of Health Communication, vol. 20, no. 8, pp. 893-902, 2015.

[56] M. Conner, S. Wilding, F. van Harreveld, and J. Dalege, "Cognitive-affective inconsistency and ambivalence: Impact on the overall attitude–behavior relationship," Personality and Social Psychology Bulletin, vol. 47, no. 4, pp. 673-687, 2021.

[57] S. Farzadnia and I. R. Vanani, "Identification of opinion trends using sentiment analysis of airlines passengers' reviews," Journal of Air Transport Management, vol. 103, p. 102232, 2022.