Sentiment Analysis on Bangladesh Airlines Review Data using Machine Learning

by

Khan Md. Hasib
20266015

A thesis submitted to the Department of Computer Science and Engineering
in partial fulfillment of the requirements for the degree of
M.Sc. in Computer Science

Department of Computer Science and Engineering
BRAC University
February 2022

Declaration

It is hereby declared that

1. The thesis submitted is my/ our own original work while completing degree at BRAC University.

2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.

3. The thesis does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.

4. I have acknowledged all main sources of help.

Student's Full Name & Signature:

Khan Md. Hasib
20266015

Approval

The thesis/project titled "Sentiment Analysis on Bangladesh Airlines Review Data using Machine Learning" submitted by

1. Khan Md. Hasib (20266015)

of Spring, 2022 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of M.Sc. in Computer Science and Engineering on February 20, 2022.

Examining Committee:

Supervisor:
(Member)

Md. Golam Rabiul Alam, PhD
Associate Professor
Department of Computer Science and Engineering
BRAC University
Email: rabiul.alam@bracu.ac.bd

Examiner:
(External)

Md. Mahbubur Rahman, PhD
Professor
Department of Computer Science and Engineering
Military Institute of Science and Technology (MIST)
Email: mahbub@cse.mist.ac.bd

Examiner:
(Internal)

Md. Ashraful Alam, PhD
Assistant Professor
Department of Computer Science and Engineering
BRAC University
Email: ashraful.alam@bracu.ac.bd

Examiner:
(Internal)

Muhammad Iqbal Hossain, PhD
Assistant Professor
Department of Computer Science and Engineering
BRAC University
Email: iqbal.hossain@bracu.ac.bd

Program Coordinator:
(Member)

Amitabha Chakrabarty, PhD
Associate Professor
Department of Computer Science and Engineering
BRAC University
Email: amitabha@bracu.ac.bd

Head of Department:
(Chair)

Sadia Hamid Kazi, PhD
Chairperson and Associate Professor
Department of Computer Science and Engineering
BRAC University
Email: skazi@bracu.ac.bd

Ethics Statement

This thesis was carried out in complete compliance with research ethics norms, and the codes and practices set by BRAC University. I have ensured that all our sources have been cited. As the author of this thesis, I take full responsibility for any ethics code violations.

Abstract

A common means of transportation in our everyday lives is air travel. As a result, it's no surprise that more and more customers are posting their airline reviews online. However, in the age of machine learning, it would be much easier to extract millions of pieces of information and knowledge from them if a model was used to polarize and comprehend them. Sentiment analysis may be used to understand people's attitudes or sentiments by utilizing sites that provide opinion-rich data. In this work, we worked on a customized dataset including online reviews for 4 major Bangladesh Airlines, performed a multiclass sentiment analysis, and compared the classifiers. Alongside sentiment analysis, topic modeling is also done to get better decisions based on the actual experiences of other customers who have flown with airlines. This method begins with pre-processing procedures used to clean the reviews and balance the review data using the Pegasus model's oversampling mechanism. The analysis was carried out 3 different machine learning (Decision Tree, Random Forest, and XGBoost) and 3 different deep learning classification strategies (CNN, LSTM, BERT). The test set's output is the review sentiment (positive/negative/mixed) using a three-class dataset, and the performance in terms of accuracy is calculated. Based on the results, we have achieved the best accuracy 83% in terms of BERT. The accuracies were determined to compare each categorization technique, and the total sentiment count for all four airlines of Bangladesh was displayed in terms of domestic route, international route and overall route. We comprehend the results acquired from USA airlines Tweets data and demonstrate that our framework is more efficient than the earlier model. Therefore, it is essential to consider whether a sentiment makes a particular prediction. Thus, we then train an interpretable LIME model for the sentiments and the construction of explainable sentiments can have a major advantage.

Keywords: Bangladesh Airlines, Online Review, Sentiment Analysis, Topic Modeling, Machine Learning, Deep learning, LIME

Dedication

I would like to dedicate this research to my parents who have brought me to this world and nurtured me to become an adult.

Acknowledgement

First and foremost, praises and thanks to the Almighty, for His showers of blessings throughout my research work to complete the research successfully.

Secondly, I would like to express my deep and sincere gratitude to my research supervisor, Dr. Md. Golam Rabiul Alam, Associate Professor, Department of Computer Science and Engineering, BRAC University for his constant guidance, faith and help throughout my thesis project. Also, my special thanks go to Nurul Akter Towhid, Lead Data Scientist at COEL Incorporation for helping to generate some part of the dataset of the thesis.

Thirdly, I would like to thank all the faculty members, staff, our peers and other stakeholders related to BRAC University for providing us an environment in which we were able to develop ourselves, learn to our fullest extent and conduct our research properly.

Table of Contents

List of Figures

List of Tables

Nomenclature

The next list describes several symbols & abbreviation that will be later used within the body of the document

$BERT$ Bidirectional Encoder Representations from Transformers

CNN Convolutional Neural Network

DT Decision Tree

$eWOM$ Electronic Word-of-Mouth

$IATA$ The International Air Transport Association

$LIME$ Local Interpretable Model-Agnostic Explanations

$LSTM$ Long Short Term Memory

RF Random Forest

SVM Support Vector Machine

Xgb XGBoost

Chapter 1

Introduction

1.1 Overview

This chapter contains the thesis's initial descriptions. To begin, it offers a broad introduction of sentiment analysis, topic modeling, and the goals of our study field. It also provides the problem definition, motivation, accessible research fields, and research limits. Finally, the research methods and analytic framework are presented.

Our research book consists of 6 chapter. The topics discussed throughout the chapters are listed below:

Chapter 1 presents a complete summary of sentiment analysis, topic modeling, the reasoning behind our theory, our goals, and contributions to the subject.

Chapter 2 demonstrates some relevant studies done on our issues. In addition, we conduct a survey on sentiment analysis and topic modeling in the airline industry.

Chapter 3 briefly addressed the relevant theory required for our research projects. Some significant machine learning and deep learning algorithms are explained.

Chapter 4 contains the stages and architecture of our proposed model, the datasets utilized in the experiments, techniques to processing the datasets, preprocessing, and experimental examples.

Chapter 5 gives information on the custom-created datasets we utilized, as well as the performance matrix and experimental outcomes.

Chapter 6 explains the conclusion of our study effort and presents potential future studies relating to this research.

1.2 Background

Social media's rapid expansion has changed the world, and we are no longer constrained by the knowledge of those in our immediate social circles. Because of the intangibility of travel services, word-of-mouth marketing has become increasingly important because clients tend to be skeptical of new service providers. Online reviews are a great source of information about the consumer experience that may be mined. Airline service providers have to deal with a lot of client feedback regarding their products and services, therefore they need to analyze it. Traditional techniques for gathering consumer feedback from airline service firms include organizing as well as gathering surveys, which are tedious and unreliable. It will require a great deal of effort to distribute and collect surveys from consumers, as well as to record and file those questions, given the number of passengers that fly daily. Many consumers don't bother to fill out surveys, and this results in a lot of noise in the data that is used for sentiment analysis. When it comes to analyzing airline customer feedback, social media and the web are far superior to surveys.

1.3 Problem Statement

Sentiment Analysis (SA) and opinion mining have been more popular study fields in recent years [2]. NLP, computational linguistics, and text mining are used to separate public thoughts and feelings about an event, product, or other [19]. EWOM has given social media a new life of its own, and it's taking over the world. There are others who believe that "social buzz" can be good. In addition, it increases sales and improves the efficacy of marketing and public relations campaigns. Others argue that social media may harm reputation by causing a "firestorm" of high-volume social media interactions. Understanding, predicting, and projecting the behavior of passengers and the consequences of online word-of-mouth (WOM) is critical to the airline business because it offers airlines with chances for service improvement, expansion of brand identity, and profit generation.

In machine learning, sentiment analysis has taken the lead. Airline firms rely heavily on client input to enhance the quality of their services and amenities. Traditional customer satisfaction questionnaires and forms are used to do sentiment analysis in the airline business. When assessing these operations, it is crucial to take into account that they are tedious and demand a lot of people, which adds up to a high price. There is also a lot of inconsistency in the information acquired via surveys. There's a chance that consumers aren't taking their input seriously, which leads to extraneous information being included in sentiment analysis. Also the interpretability provides a crucial part in the realm of sentiment analysis which is an unique explanation approach that develops a regional explainable prototype around the expectation in order to explain the classifier's outputs during an explainable and reliable manner which is more important in terms customer satisfaction in aviation industry.

1.4 Motivation

Social media's fast growth has changed the world, and we are no longer constrained by the knowledge of those in our immediate social circles. Online reviews are a great source of information about the consumer experience that may be mined. Tourism and hospitality are only two examples of fields where sentiment analysis has been used. Customer happiness and real-world experience are indeed the two major essential elements in determining the quality of flight facilities.

Humans are now in the twenty-first generation, having a worldwide demographic of near about 8 billion inhabitants. The number of clients going by airline is rapidly increasing on a daily basis. In order to suit the needs of clients, the number of airline firms has grown significantly. When there are several firms offering the same function, it is difficult for individuals to decide which one to choose. As a result, it is critical for travelers to understand which airline is ideal for them based on their needs and budget. It may be determined by the honest evaluations offered by passengers who have shared their past experience. The reviews offered not only assist customers in selecting the best airline, but also assist airline firms in identifying flaws and improving service quality. Because airline reviews data is available online, and there is a large volume of data, managing things manually becomes tough.

Like many other countries for which airline is best to travel and how the service of visiting these airlines for both domestic and international flight in Bangladesh, our research will get the output. The objective is to draw appropriate findings from a study of customer satisfaction and dissatisfaction. We want to utilize the ratings and evaluations that passengers have supplied to predict whether or not they are satisfied with their airline or airport choice. A secondary goal of the project is to provide an opportunity to assist airlines regarding overall organizational processes and judgment activities, as well as sensitive files and quantitative methodologies that may be used to improve the utilization.

1.5 Research Challenges

Research on the demands of Bangladeshi airline passengers was conducted using a combination of topic modeling and sentiment analysis of big data. A customer's propensity to purchase utilizing Bangladesh airlines was evaluated by examining internet evaluations of four airlines: Biman Bangladesh Airlines, NOVOAIR, Regent Airways, and US-Bangla Airlines. Using data from this inquiry, the following theoretical and practical conclusions were drawn. There are number of research challenges we faced:

- Finding the online reviews from the social media as Bangladeshi People hardly post any airline reviews in social media like Twitter, Facebook etc.
- Dealing with highly imbalanced and noisy dataset.
- Finding the best approach to design a model that can analyze imbalanced dataset.
- Finding the appropriate preprocessing step convenient to our work.
- Determining the approach which would help us to get features from our dataset.
- Selecting the correct model to classify our data.

1.6 Contributions

To our knowledge, Our research is the first to recommend the best Bangladesh Airline for the people to travel countries in terms of customer online reviews. The primary goal of this study is to inform Bangladesh's airline sector on customer attitudes toward the corresponding carriers. A full perspective of their clients' sentiments is depicted in order to effectively meet their demands. This study includes the following contributions:

- A roadmap for the processing of Bangladesh Airline Service related customized dataset from online reviews is provided.
- Identifying the best airline of Bangladesh in terms of domestic route, international route and overall route passengers online reviews.
- A statistical survey for sentiment analysis and for topic modeling for different years in terms of airline services is provided.
- Balanced the unstructured dataset using the paraphrased technique which is called the PEGASUS model - a oversampling mechanism for better outcome.
- A proposed framework incorporating a variety of text representations and methods to machine learning and deep learning in online reviews is offered. When the proposed model generates increased functionality and various data collecting techniques, high classification progress has been obtained.
- Executing topic modeling on online review posts, which have a wide range of interests and engagement from people who have used or want to use it for Bangladesh airlines.
- Understanding the sentiments prediction of positive, negative and mixed reviews interpretation with a 3-class dataset using explainable AI.

1.7 Research Strategy

The following approaches were used to achieve the goals of our research (figure 1.1) :

Research & Survey:
We examined many current methods for sentiment classification and topic modeling on various types of datasets, discovered flaws in existing approaches, and attempted to solve them by developing our own strategies and solutions.

Data Acquisition:
In data acquisition step, we have used customized datasets named Bangladesh-Air-Data in terms of domestic, international and overall route regarding Bangladesh Airlines for our research.

Pre-Processing:
To balance the final dataset, we use the oversampling techniques using the PEGASUS pretrained model. First, the less sample of negative and mixed reviews in the

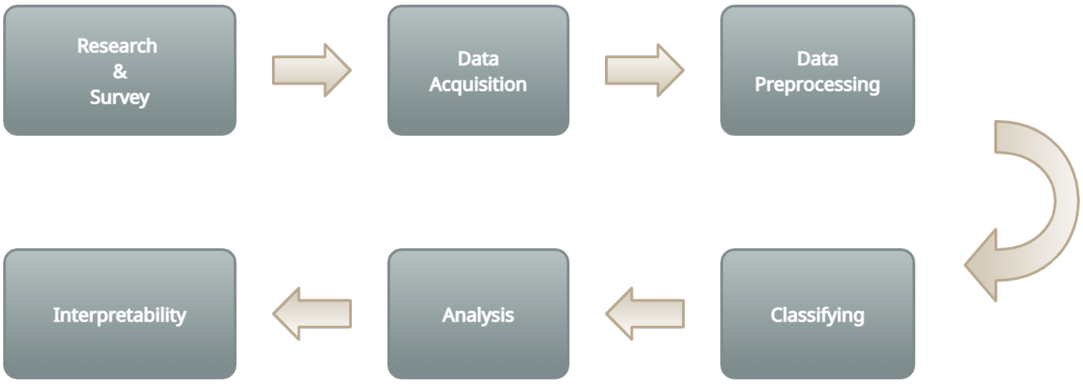

Figure 1.1: Overall Research strategy

dataset are paraphrased to upsampling the data. Then we merge the newly negative and mixed reviews in the final dataset. After merging, we got then 1047 reviews after oversampling. We also divide the datasets into domestic and international route category for finding the specific customer satisfaction along with best airline of Bangladesh.

Classifying:
After data preprocessing, we divided the input data into two steps: Phase-I: Machine Learning (ML) Approach and Phase-II: Deep Learning (DL) Approach. We describe the algorithms regarding machine learning and deep learning.

Results:
In this approach, we applied sentiment classification, topic modeling to get the result. Also, we compare our result with existing USA Airline Tweets data to validate our model. Finally, frequency analysis and word cloud is used for more visualizations.

Interpretability:
We used LIME for understanding the sentiments that predict positive, negative, and mixed responses from the passengers of Bangladesh airlines.

Chapter 2

Literature Review

2.1 Overview

We have described the reviews of numerous papers connected to our research in this chapter. We've done some reviews of various study papers in Section 2.2. We also created year-by-year statistics survey tables for both sentiment analysis and topic modeling on airline industries that were useful in our research in both Section 2.3 and Section 2.4.

2.2 Review of the Related Works

Sentiment analysis is a method for extracting emotions from textual data such as online contents, consumer evaluations, film ratings, and Tweets data. Tweets information often includes details on a people's opinions on a wide range of issues. One of the most popular subjects on Twitter right now is air travel. Passengers on planes frequently tweet about their travel experiences in all over the world. This data might become useful whether analyzed using machine learning techniques, since it may give facts about the passenger's level of comfort during the journey. The field of sentiment analysis has such a rich and diverse literature. A number of scholars are interested in analyzing social media. Unstructured text can be difficult to analyze using traditional approaches like lexicons and rule-based procedures since they need human labeling of documents and have a limited scope [59]. Traditional machine learning and machine-learning hybrids were both reported to increase categorization performance over time by various studies. Online reviews become more popular and so a powerful and complete source of information is available. However, in reality, internet evaluations include a wealth of useful hidden information [21]. Websites such as the Skytrax portal and TripAdvisor, described above, are therefore essential in analyzing how customers rate the quality of airline service.

Again, a connection between passenger pleasure and profitability was established by Heskett et al. [1] and research on quality of airline service was a crucial issue for the airlines industry. As a result, the authors of [6] argue that in order to compete in the airline business, data on traveler satisfaction and the way it connects to the level of service given are essential for continuous collection and evaluation. One recent study in this area is detailed in [20], in which the authors mined review data from the Skytrax site on airlines' in-flight services. The airline sector also employed machine learning and deep learning classification for sentiment categorization. In order to determine the airline service's feelings in the US, Rane et al. [45] suggested many methods of pre-processing, following the seven algorithms. A new deep learning model, which successfully integrates different word integration with deep learning methods, was developed by Hasib et al. [85] for evaluating data from six important US Airlines tweets, and for multi class sentiment analyzes, with 91% precision. Furthermore, Rajat et al. [79] developed five distinct categorization methods for evaluating the emotions included within Indian Air Asia Service tweets. AdaBoost (Ensemble), which integrates many different algorithms to generate a strong predictor with the efficiency of 84.5% [60], is one of the most often used classification algorithms. Several ANN-based models for prequalified, educated, and hybrid word integration were first developed and validated using SVM [61]. In addition, IMDb comments were subjected to sentiment analysis using CNN and LSTM [35]. The proposed framework comprises a modern approach of emotional mining, which involves a large quantity of CNN-LSTM layers linked with kernels allowing assessment [84].

Dutta Das et al. [27], analyzed airline Twitter data for sentiment analysis employing the Naive Bayes approach on 200 tweets directed towards Emirates and Jet Airways. They used R and Rapid Miner to fine-tune the categorization system and classify messages as positive, negative, or neutral. They stated the findings achieved utilizing Naive Bayes models looked interesting for a greater number of tweets in the study. Hakh et al. [28] used the SMOTE approach to address the dataset's unbalanced problem and machine learning techniques to assess a collection of tweets about six airline firms found in the United States They discovered that feature selection and over-sampling strategies are both necessary for achieving finer findings. They then used sentiment classification algorithms (AdaBoost, Decision Tree, Linear SVM, Naive Bayes, Random Forest, KNN, and Kernel SVM).

Sentiment analysis has long been one of the most popular deep learning research areas. In [51], a novel deep learning architecture with mixed CNNs and BiLSTM characteristics (H2CBi) is proposed, combining the power of CNNs and BiLSTM. Authors used two distinct pre-trained word embeddings to produce different input images, which were then input into the LSTM. Aside from word embedding characteristics, a research [33] identified individual as well as evidenced automated characteristics from a sample obtained from the Chinese online community Sina Weibo. The characteristics derived from their research are sent into the LSTM network for classification purposes. Their results show that their suggested hybrid technique outperforms simple LSTM or traditional machine learning. The complexity of the Lithuanian language, with its rich vocabulary, vast quantity of varieties, and problems in morphological evaluation, is the fundamental reason for the achievement of machine learning technologies. On the basis of machine learning, a hybrid strategy has been presented. Their hybrid approach combines expert system, SVM, and the semantic perception strategy. Another study [49] employed a multilayer network of CNN and LSTM to analyze sentiment on a dataset using the Tibetan social media application. The features were retrieved using a three-layer CNN network. The acquired features are fed into a two-layer LSTM network. It was discovered that the hybrid deep learning model outperformed CNN and LSTM. A hybrid bidirectional recurrent CNN attention model was presented [66]. In order to classify text, our model used Word2Vec and an attention mechanism to integrate BiLSTM and CNN. They employed LSTM, CNN, and attention layers sequentially in their hybrid model. One CNN layer and two stacked LSTM layers were employed to analyse Indian tweets sequentially by Kaladevi and Thyagarajah [89]. In their hybrid experiments, CNN features are used as an input to the LSTM network. Like earlier hybrid experiments, theirs employed CNN to extract characteristics. Using both CNN and BiLSTM for feature extraction, our method varies from that of the authors of the previous work. When these investigations are analyzed, a single data format such as Word2Vec, Glove, or character-level embedding is used for hybrid deep learning. Furthermore, a variety of deep learning techniques were applied to a single dataset.

Topic modeling, on the other hand, is based on a process probability distribution model and uses a vast amount of documentation to identify potentially valuable themes. Several studies have mimicked various types of unstructured texts, such as SNS and online reviews [92]. Using the latent Dirichlet allocation (LDA) model, Lucini et al. [72] found 27 satisfaction variables from 55000 evaluations of 400 airlines and consumers from 170 countries. Full-service and low-cost carriers were evaluated using LDA Theme Modelling, which was applied to a significant number of passengers' online evaluations [70]. Use of wide-ranging text data from more than 100,000 Korean consumer online reviews by Sutherland and colleagues was used to uncover topics of interest to visitors [77]. There were 231 abstracts of blockchain-related publications published in the last five years that were analyzed using the Word2vec-based, innovative subject modeling approach [71]. For the second time, the LDA model developed by Sun et al. [34] sums up 50 important topics. They demonstrate the representativeness and relevance of these usual topics, particularly in terms of established transport study subfields. Internet reviews were utilized by Korfiatis et al. [55] to demonstrate how topic modeling of unstructured data can be used to better assess customer satisfaction and thereby service quality.

In order to measure emotions, sentiment analysis uses emotional phrases and polarities to develop a sentiment lexicon. The appropriate interpretation of sentiments requires the use of a number of terminology, including a sentiment vocabulary. By extracting 4,783 negative and positive words across time, Liu et al. [11] created the English Opinion Lexicon in 2006 to undertake an emotional analysis. Sentiment Dictionary by Wibe et al. [3] expresses emotion and sensitivity in roughly 10,000 lines in accordance with the purpose of creating emotional vocabulary. SentiWordNet was developed by Esuli and Sebastiani [5] using preexisting WordNet synonyms to distinguish between three levels of sensitivity: positive, neutral, and negative. According to this webpage[1], a recent research in this field collected review data from the Skytrax site for airlines' in-flight services. The scientists showed that deductions may be derived to explain when passengers assess in-flight services employing content and additional article models.

Once the preceding study was evaluated, different machine learning and deep learning approaches were applied on a single data representation for the aviation business of other nations. Again, there have been few experiments concerning the first passenger experience in converting data from the "Airline" to the time-line corresponding

[1] https://www.airlinequality.com/,_(Skytrax)

to the first voyage of each passenger. Furthermore, we did fundamental extrapolation models and prior studies in terms of Bangladesh airlines to give comparative analysis from customer online reviews.

2.3 Survey of Related Works on Sentiment Analysis for Airline Services (Year-wise Statistics)

There has been a dramatic increase in the amount of text data created by consumers over the past few decades. This includes text messages, tweets, emails, postings and blogs made by customers [41]. Aside from the fact that social media has become a tool that can be utilized by those who need it, people nowadays exchange, use, and seek out information on the internet on a regular basis [54]. In terms of airline industry, a better customer experience and more profits are both possible as a result of these inputs from the aviation industry. As a result, collecting and analyzing consumer input from online sources such as social media and websites is critical.

We studied some articles based on sentiment analysis for airline industries. These papers helped us to get some knowledge to work on sentiment analysis. Most of the work on this section have been done in 2006, 2011, 2012, 2014 - 2021 and still working. We listed the comparison review of some of the papers published in different journals over the years and this list is sorted by year. In figure 2.1, we represents the number of research papers based on different years.

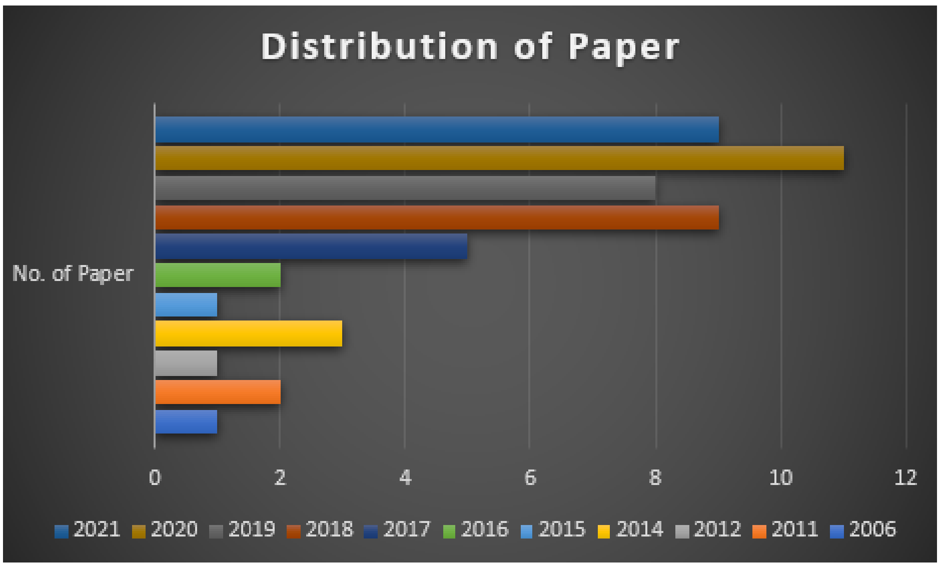

Figure 2.1: Distribution of papers on Sentiment Analysis in Airline Industry based on Years

The survey consists of 52 articles and 9 papers for 2018 and 2021 were included in it. A record amount of publications were examined by scholars from 2020. According to table 2.1, sentiment analysis for airlines is the subject of a statistical survey.

After reviewing the publications on sentiment analysis in the airline business, we discovered that the most frequent ways to classifying sentiment about different nations'

Table 2.1: **Year wise statistics of Sentiment Analysis on Airline Industry**

Year	Reference	Title	Airline	Data Source	ML/DL Techniques/ Method	Citation
2021	Jain et al. [88]	A Hybrid CNN-LSTM: A Deep Learning Approach for Consumer Sentiment Analysis Using Qualitative User-Generated Contents	186 various airlines around the globe and U.S. airlines	Airlinequality, Twitter	LR, DT, NB, SVM, LSTM, CNN, CNN-LSTM	4
2021	Verma et al. [98]	Implicit Aspect-Based Opinion Mining and Analysis of Airline Industry Based on User-Generated Reviews	16 airlines	TripAdvisor, Airline Ratings	SVM, DT, RF, XGBoost	*
2021	kumari et al. [91]	Collaborative Classification Approach for Airline Tweets Using Sentiment Analysis	U.S. airlines	Twitter	RF, LR, KNN, NB DT, XGBoost, SGD	1
2021	Hasib et al. [85]	A Novel Deep Learning based Sentiment Analysis of Twitter Data for US Airline Service	6 U.S. major airlines	Twitter	CNN, DNN	2
2021	Piccineli et al. [93]	Air-travelers' concerns emerging from online comments during the COVID-19 outbreak	Italian airlines	Italian National Consumer Union website	Automated textual analysis	6
2021	Kwon et al. [92]	Topic Modeling and Sentiment Analysis of Online Review for Airlines	Asia-Pacific Airlines	Skytrax	Statistical analysis	10
2021	Jain et al. [87]	Content-Based Airline Recommendation Prediction Using Machine Learning Techniques	Not indicated	Airlinequality	LR, SGD, RF	1
2021	Tusar et al. [97]	A Comparative Study of Sentiment Analysis Using NLP and Different Machine Learning Techniques on US Airline Twitter Data	6 U.S. airlines	Twitter	SVM, LR, MB, RF	*

Year	Reference	Title	Airline	Data Source	ML/DL Techniques/ Method	Citation
2021	Al-Qahtani et al. [94]	Predict Sentiment of Airline Tweets Using ML Models	U.S. airlines	Twitter	LR, NB, BERT, CNN, XLNET, ALBERT	*
2020	Park et al. [92]	Understanding the dynamics of the quality of airline service attributes: Satisfiers and dissatisfiers	20 U.S. airlines	TripAdvisor	Statistical analysis	20
2020	Ahmed et al. [80]	Significant Labels in Sentiment Analysis of Online Customer Reviews of Airlines	Iberia airline	TripAdvisor	Regression analysis	2
2020	Sharmna et al. [75]	Testing loss aversion and diminishing sensitivity in review sentiment	20 U.S. airlines	TripAdvisor	Statistical analysis	12
2020	Yadu et al. [79]	Sentimental Classification Method of Twitter Data for Indian Air Asia Services Analysis	Air Asia Airline services	Twitter	DT, RF, SVM LR, NB	1
2020	Zheng et al. [82]	Analysis of Airline Sentiment Data: Does the social media image reflect real performance?	6 U.S. airlines	Twitter	Statistical analysis	*
2020	Heidari et al. [69]	Using Transfer Learning Approach to Implement Convolutional Neural Network model to Recommend Airline Tickets by Using Online Reviews	4 major U.S. airlines	GoogleFlight, Kayak, Skyscanner, Twitter	BERT, CNN	19
2020	Song et al. [76]	Analyzing passengers' emotions following flight delays-a 2011–2019 case study on SKYTRAX comments	14 big commercial airlines	Skytrax	Statistical analysis	9
2020	Bezek et al. [67]	Analysis of Airline Tweets by Using Machine Learning Methods	7 different airlines	Twitter	SVM, DT, RF, Gradient Boosting, NB, ANN	*

Year	Reference	Title	Airline	Data Source	ML/DL Techniques/ Method	Citation
2020	Shahbaznezhad et al. [74]	Exploring firms' fan page behavior and users' participation: evidence from airline industry on Twitter	Different airlines around the globe	Twitter, Skytrax	K-Means	2
2020	PARK et al. [73]	Understanding the dynamics of the quality of airline service attributes: Satisfiers and dissatisfiers	20 U.S. airlines	TripAdvisor website	Statistical Analysis	20
2020	Lucini et al. [72]	Text mining approach to explore dimensions of airline customer satisfaction using online customer reviews	419 airlines from 171 countries	Skytrax	NB, LR	60
2019	Utama et al. [65]	Sentiment Analysis in Airline Tweets Using Mutual Information for Feature Selection	U.S. airlines	Twitter	NB, SVM, SVM-RBF, LR, DT	2
2019	Prabhakar et al. [60]	Sentiment Analysis of US Airline Twitter Data using New Adaboost Approach	U.S. airlines	Skytrax, Twitter	SVM, DT, RF, AdaBoost	12
2019	Monika et al. [57]	Sentiment Analysis of US Airlines Tweets Using LSTM/RNN	U.S. airlines	Twitter	RNN, LSTM	2
2019	Tian et al. [64]	A new approach of social media analytics to predict service quality: evidence from the airline industry	12 U.S. airlines	Twitter	Statistical analysis	11
2019	Anitsal et al. [53]	Is your business sustainable? A sentiment analysis of air passengers of top 10 US-based airlines	10 U.S. airlines	Skytrax	Statistical analysis	3
2019	Sezgen et al. [62]	Voice of airline passenger: A text mining approach to understand customer satisfaction	50 airlines around the globe	TripAdvisor	Latent semantic analysis	77

Year	Reference	Title	Airline	Data Source	ML/DL Techniques/ Method	Citation
2019	Kumar et al. [56]	A machine learning approach to analyze customer satisfaction from airline tweets	12 airlines around the globe	Twitter	SVM, ANN, CNN	36
2019	Tao et al. [63]	Social Media Data-Based Sentiment Analysis of Tourists' Air Quality Perceptions	195 Chinese airlines	Sina Weibo	ANN	21
2018	Rane et al. [45]	Sentiment Classification System of Twitter Data for US Airline Service Analysis	6 major U.S. airlines	Twitter	DT, RF, SVM, KNN, LR, NB, AdaBoost	77
2018	Punel et al. [44]	Using Twitter network to detect market segments in the airline industry	Air Newzealand	Twitter	Clustering analysis	22
2018	Siering et al. [47]	Disentangling consumer recommendations: Explaining and predicting airline recommendations based on online reviews	different airlines	Airlinequality	NB, SVM, NN	105
2018	Asi et al. [38]	Pre-trained Word Embeddings for Arabic Aspect-Based Sentiment Analysis of Airline Tweet	National carrier Airline of Saudi Arabia	Twitter	SVM	17
2018	AL-Sharuee et al. [46]	Sentiment Analysis: An Automatic Contextual Analysis and Ensemble Clustering Approach and Comparison	4 Australian Airlines	www.productreview.com.au website	K-Means	27
2018	Sternberg et al. [48]	Analysing Customer Engagement of Turkish Airlines Using Big Social Data	Turkish Airlines	Facebook Page	Naïve Bayes	13
2018	Tiwari et al. [50]	Sentiment Analysis for Airlines Services Based on Twitter Dataset	U.S. airlines	Twitter	K-Means	9

Year	Reference	Title	Airline	Data Source	ML/DL Techniques/ Method	Citation
2018	Khan et al. [42]	Airline Sentiment Visualization, Consumer Loyalty Measurement and Prediction using Twitter Data	18 Airlines around world	Twitter	RF, DT, LR, K-Means	10
2018	Adarsh et al. [36]	An Effective Method of Predicting the Polarity of Airline Tweets using sentimental Analysis	Indigo Airlines, Emirates Airlines, Qatar Airlines	Twitter	Statistical analysis	7
2017	Chumwatana et al. [26]	Using Social Media Listening Technique for Monitoring People's Mentions from Social Media: A Case Study of Thai Airline Industry	Thai airlines	Pantip Webboard	Statistical analysis	5
2017	Hu et al. [30]	Analyzing users' sentiment towards popular consumer industries and brands on Twitter	Not indicated	Twitter	RNN, LSTM	22
2017	Joshi et al. [31]	Aspect based sentiment analysis for United States of America Airlines	U.S. airlines	Twitter, Skytrax	DT, SVM, RF, Bagging, Boosting,, SLDA, Maximum Entropy	1
2017	Seyfioglu et al. [32]	A Hierarchical Approach for Sentiment Analysis and Categorization of Turkish Written Customer Relationship Management Data	Turkish Airline	Private Airline Company	XGBoost	12
2017	Das et al. [27]	Sentimental Analysis for Airline Twitter data	Emirates and Jet Airways	Twitter	Naïve Bayes	15
2016	Chen et al. [22]	Big Data Analytics On Aviation Social Media: The Case of China Southern Airlines on Sina Weibo	China Southern Airlines	Sina Weibo	Statistical Analysis	15
2016	Lacic et al. [24]	High Enough? Explaining and Predicting Traveler Satisfaction Using Airline Reviews	All around globe	Skytrax	NB, C4.5, RF, CART, Hoeffding Tree algorithm	30

Year	Reference	Title	Airline	Data Source	ML/DL Techniques/ Method	Citation
2015	Park et al. [18]	Disentangling consumer recommendations: Explaining and predicting airline recommendations based on online reviews	South Korean airlines service	Real time data	Structural equation modeling method	41
2014	Misopolous et al. [16]	Uncovering customer service experiences with Twitter: the case of airline industry	4 airlines	Twitter	Statistical analysis	137
2014	Adeborna et al. [13]	An Approach to Sentiment Analysis – The Case of Airline Quality Rating	Three major airlines - AirTran Airways, Frontier, SkyWest Airlines	Twitter	Correlated Topics Models (CTM) method, Variational Expectation-Maximization (VEM) algorithm method	37
2014	Liau et al. [15]	Gaining customer knowledge in low cost airlines through text mining	AirAsia, Berjaya Air, FireFly, MASwings, Malindo Air	Twitter	K-Means	142
2012	Chiou et al. [10]	Service quality effects on air passenger intentions: a service chain perspective	Spring Airlines	Real time data	Latent semantic analysis	49
2011	Jen et al. [7]	Managing passenger behavioral intention: an integrated framework for service quality, satisfaction, perceived value, and switching barriers	Taiwan airlines	Real time data	Structural Equation Modeling (SEM)	181
2011	Mikulic et al. [8]	What drives passenger loyalty to traditional and low-cost airlines? A formative partial least squares approach	Lufthansa, Croatia Airlines, Germanwings	Real time data (Zagreb Airport)	Statistical analysis	112
2006	Park et al. [4]	Modelling the Impact of Airline Service Quality and Marketing Variables on Passengers' Future Behavioural Intentions	Australian airlines	Real time data (Sydney Airport)	Structural equation modelling	276

airlines include various machine learning and deep learning methods. The sentiment of three and two class datasets was then classified using statistical analysis, latent semantic analysis, and structural equation modeling (SEM) hybrid methodologies. The majority of the papers focused on the fact that many of the data sources for the airline industries come from the web, such as Twitter, Skytrax, and TripAdvisor. Some projects concentrate on real-time sentiment data from aircraft passengers. According to this survey, the United States (18) is one of the major nations exploring sentiment analysis in the aviation industry.

2.4 Survey of Related Works on Topic modeling for Airline Services (Year-wise Statistics)

Natural language processing (NLP) is a large subject of mathematical and linguistic techniques to interpreting, evaluating, and producing human text and speech [12]. 'Topic modeling' is a term used in the NLP sector to represent numerous unsupervised ways to discovering latent associations between words in a collection of textual texts. Table 2.2 shows the number of research articles published in each year regarding topic modeling in aviation services.

We looked at various publications about subject modeling in the aviation industry. These publications assisted us in gaining some expertise to work on subject modeling. The majority of the work on this segment was completed in 2017, 2019, 2020, and 2021, and it is currently ongoing. We have compiled a comparison evaluation of some of the articles published in various journals over the years, and this list is organized by year. In figure 2.2, we represent the number of research papers based on different years.

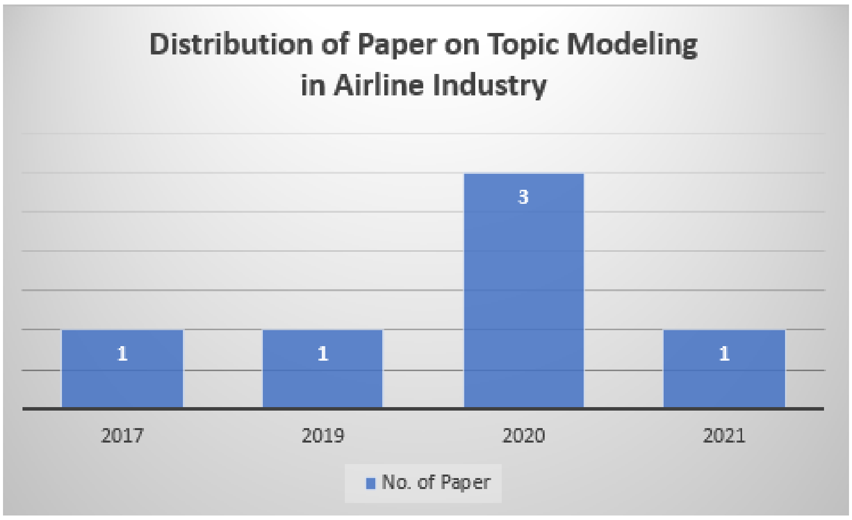

Figure 2.2: Distribution of papers on Topic Modeling in Airline Industry based on Years

We learned from the above survey which strategies for topic modeling in airline industries have been frequently employed in recent years, as well as information about their (methods/algorithm) success in satisfying consumers. We also obtained

Table 2.2: **Recent research utilizing topic modeling analysis**

Year	Author	Reference	Title	Summary	Citation
2021	Kwon, H. J., Ban, H. J., Jun, J. K. & Kim, H. S.	[92]	Topic Modeling and Sentiment Analysis of Online Review for Airlines	14,000 reviews covered from 27 airlines of Asia evaluated using LDA model and uncover the key concerns of customer satisfaction using topic modeling and sentiment analysis	9
2020	Lucini, F.R., Tonetto, L.M., Fogliatto, F.S., Anzanello, M.J.	[72]	Text mining approach to explore dimensions of airline customer satisfaction using online customer reviews	55,000 evaluations from 400 airlines and passengers from 170 countries were examined using the latent Dirichlet allocation (LDA) model which discovered 27 satisfaction variables	60
2020	Kim, S., Park, H., Lee, J.	[70]	Word2vec-based latent semantic analysis (W2V-LSA) for topic modeling: A study on blockchain technology trend analysis	That report examines both customer level for comprehensive carriers versus reduced price carriers using LDA topic modeling and a large number of online passenger reviews	35
2020	Sutherland, I., Sim, Y.; Lee, S.K.; Byun, J., Kiatkawsin, K.	[77]	Topic Modeling of Online Accommodation Reviews via Latent Dirichlet Allocation	An empirical method was employed in this essay to determine which subjects of interest visitors deem essential by analyzing vast amounts of textual input with no structure from 104,161 digital evaluation of Korean hotel cconsumers	23
2019	Lim, J., Lee, H.C.	[71]	Comparisons of service quality perceptions between full service carriers and low cost carriers in airline travel	This work suggested a novel topic modeling approach known as Word2vec-based Latent Semantic Analysis to conduct a yearly pattern investigation of cryptocurrency studies by region and period for 231 summaries of cryptocurrency studies produced during last five years	29
2017	Sun, L., Yin, Y.	[34]	Discovering themes and trends in transportation research using topic modeling	This study used an LDA model to infer 50 main topics from article abstracts. The author demonstrate that the defined themes are both representative and significant, with the majority of them matching to well-established subfields in transportation research	156

a good understanding of our study from this poll, which included sentiment analysis and topic modeling.

Chapter 3

Background Study

3.1 Overview

In this chapter firstly, we described the characteristics of sentiment analysis and it's importance in customer satisfaction criteria at Section 3.2. Then also discussed about the text classification in Section 3.3.

After that, discussed classification techniques that are used in this research at Section 3.4 including machine learning approaches at 3.4.1 and deep learning approaches at 3.4.2 respectively. Finally, we depicted the topic modeling in terms of airline industry at Section 3.5 and Explainable AI in brief at Section 3.6.

3.2 Sentiment Analysis

Sentiment analysis is a technique for evaluating the degree of positivity or negativity associated with a piece of data. It is also known as opinion mining and emotion AI. Sentiment analysis is a powerful tool for extracting insights from large amounts of text data quickly and easily. It is, in essence, a categorization system which seeks the categorize an authoritative viewpoint as well as its disposition yet likewise underlining relevant facts [17]. Customers' feelings are analyzed utilizing natural language processing, text analysis, and statistics in the form of sentiment analysis. Customers' feelings, including what they say, how they say it, and what they mean, are critical to corporate success.

From social media posts, reviews, and other places where your business is spoken about, you may learn about customer sentiment. Through the help of a computer program, we may analyze these feelings. It's essential for today's engineers and business executives to have access to this technology. Sentiment analysis has been propelled to the forefront of cutting-edge algorithms by deep learning, like many other fields. The most significant advantage of sentiment analysis is assessing user opinions about products or service providers. From figure 3.1, we can analysis the area of sentiment analysis for evaluating the different fields. In the airline business, sentiment analysis can assist customers in determining which airline is the best by evaluating other customer opinions from online comments posted on reviews or microblogging sites.

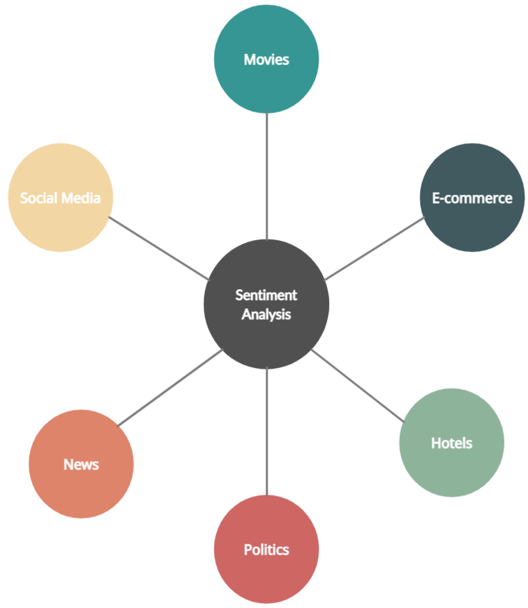

Figure 3.1: Sector of Sentiment Analysis

People categorize their emotions on a regular basis. As it turns out, there are a total of six universally recognized feelings that transcend language and culture. Anger, Disgust and Displeasure are all examples of these emotions. It is hypothesized that written texts may represent the authors' feelings at the time they were penned since these feelings are universally acknowledged [14]. One of the most important costs of categorization is the time required for the annotation of a learning set. For any company or organization that is interested in using their own classification, reducing the amount of time needed to intuitively annotate a learning set will be a huge benefit. This led us to create and test a mechanism for automatically annotating tweets using emoji or hashtag tags.

Customers are increasingly using social media to provide evaluations and concerns about service problems. Consumer Reports, Amazon reviews, and epinion.com are popular product review sites; Angie's list, Insider pages, Judy's book, Better Business Bureau, and Facebook are popular business review sites; Yelp and Zomato are popular restaurant review sites; and hotels.com and TripAdvisor are popular travel review sites. Skytrax, like the other consumer forums mentioned above, is an essential customer forum for airline feedback. Sentiment Analysis is the process of determining a piece of text's evaluative character. For example, a product review can communicate a positive, negative, or neutral mood. Tracking sentiment toward goods, movies, politicians, etc., refining customer relation models, and detecting pleasure and well-being are only some of the uses of sentiment analysis in text. Sentiment analysis mostly depends on the users of different topics which is collected from social media and various online reviews website. After crawling the data, some steps included data aggregation, data preprocessing and text embedding needed for get the sentiments of the user.

A major advantage of sentiment analysis is that it allows you to analyze consumer

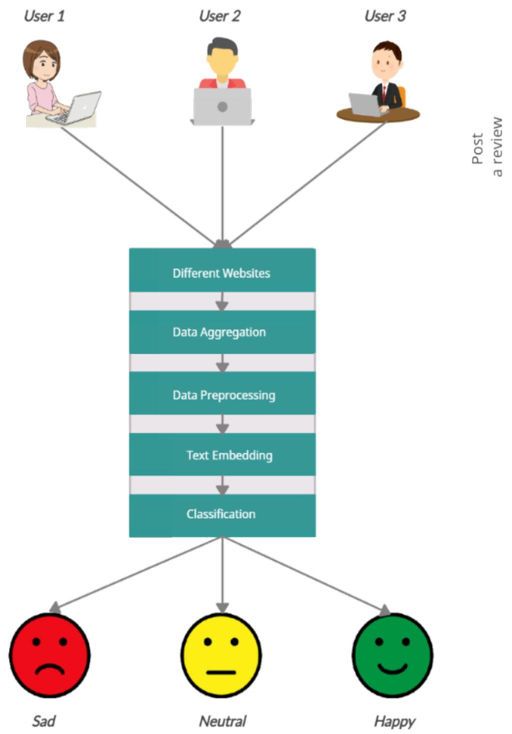

Figure 3.2: The fundamentals of social media sentiment analysis.

comments on products and services. In sentiment analysis, the sub-processes depicted in figure 3.2 are carried out by analyzing user-generated material that is posted on social media.

3.3 Text Classification

Assigning predetermined categories to free text documents is known as text categorization (also known as text classification). Spam filtering, email routing, language identification, subject classification, and sentiment classification are all examples of text classification applications[1]. The amount of computerized word documents has increased as digital and communications techniques have evolved to the point that individuals can no longer process them manually. Natural Language Processing (NLP) approaches such as text categorization have faced both obstacles and possibilities as a result of this shift. Statistical or probabilistic algorithms can be used to use parallel computing science to classify large digitized textual archives.

Speech analytics is also a subfield of information categorization. On the other hand, presents certain distinct challenges not seen in traditional data categorization. Text classification applications, unlike ordinary data classification systems, focus on text data, which includes letters, words, and phrases, rather than digits or nominal qualities. Most commonly, regular data classification approaches may be used in conjunction with regular text classification by transforming text data into regular numeric data. We may, for example, convert each text dataset into an identifier as well as single word article into a binary value vector that represent actual amount of occur-

[1]http://www.scholarpedia.org/article/Text_categorization,_(Text_Categorization)

rences of the term in the article. Classification tasks will still be impossible because of the high dimensionality of the digital information that was changed. There are more than a thousand different words in even the tiniest text datasets, not to mention the phrases and longer sentences. The "curse of dimensionality" is the name given to this situation.

Among the least prevalent methods of information classification is to use supervised or unsupervised classification. A pre-labeled dataset is supplied, and classification models are trained on the labeled datasets. Pre-labeled data are not required for unsupervised categorization. Cooperation between the techniques for representing texts and their algorithms is critical in many text categorization tasks. For high-performance classification, it is necessary to use the right text representation and classification method. There are both advantages and disadvantages to the method of converting text to numerical format. To solve issues and achieve high accuracy, for sentiment analysis, it is critical to combine the best speech presentation with the best deep learning (i.e., the sentiment classification problem). Text-based data characteristics may be extracted using a variety of ways to improve classification accuracy. The numerical representation of a user's emotion may be incomplete in any text representation approach (i.e., word embedding, character-level embedding). In response to this issue, we are attempting to combine the strengths of several text representation approaches and various deep learning models.

3.4 Classification Techniques

In this section, we discuss a wide range of algorithms depending on various classification methods. In addition to text categorization, these algorithms may also be used to analyze the sentiment of online reviews for Bangladesh airlines. In addition to machine learning, certain deep learning approaches were also discussed.

3.4.1 Machine Learning (ML) Approaches

Here we discuss the machine learning systems which been employed to forecast the sentiments of Bangladesh Airlines. After text information to sequence, we combine new feature text data. Lastly, we apply three machine learning algorithms such as: Decision Tree, Random Forest, and XGBoost.

Decision Tree (DT)

A decision tree classifier is a fundamental and frequently used data classification tool [39]. Root node, leaf node, and expanding are all part of the structure. As a system, the decision tree of each branch and leaf node imply an attribute requirement, which is shown by the trial's results. A base component seems to be the biggest node in the tree. The next step is to purchase a defining device that determines if a corporate customer is willing to purchase a machine. The attribute inspection identifies every subnode inside the main nodes. The nodes on each leaf indicate a class. As seen in figure 3.3, a decision tree breaking the data into two or more sets is what an algorithm accomplishes.

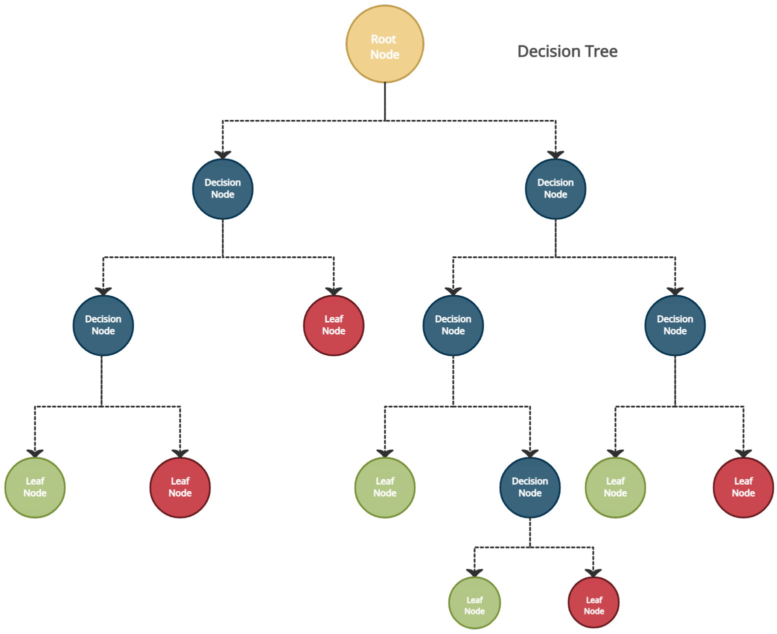

Figure 3.3: Structure of Decision Tree

In order to create as many separate classes as feasible, information gain and entropy perform an essential part in the process for separating. As indicated in the calculation 3.1, entropy evaluates the purity of the result class in any D dataset with the properties of $\mathrm{q_j}$.

$$Z(q1, q2, \ldots qf) = \sum_{j=1}^{f} \left(q_j \log \left(\frac{1}{q_j} \right) \right) \tag{3.1}$$

Information gain is measured as the difference between the entire dataset entropy and the splitting attribute entropy as shown in the formula 3.2:

$$\text{Gain}(D, L) = Z(D) - \sum_{r=1}^{f} q\left(D_j\right) Z\left(D_j\right) \tag{3.2}$$

It is possible to implement the decision tree method without scaling the data. It is not essential to normalize the data while using Decision Tree method. As complexity increases, the training time for a decision tree model increases as well. The decision tree model's repeatability is extremely sensitive, since even slight changes in the data can have significant effects on the tree's structure [99].

Random Forest (RF)

A random forest classifier is a classification method that uses ensemble learning [68]. This approach is capable of maintaining massive databases as well as hundreds of source factors without removing any of them. That technique could manage with data point errors. If you're looking to sort or regress data, Random Forest can help with that too. Installing a number of decision-making trees is now underway. Using this Random Forest Technique, you may utilize it for classification or regression purposes, which is the most major advantage of this algorithm. If you're looking for the best performance possible, Random Forest Algorithm is your best option.

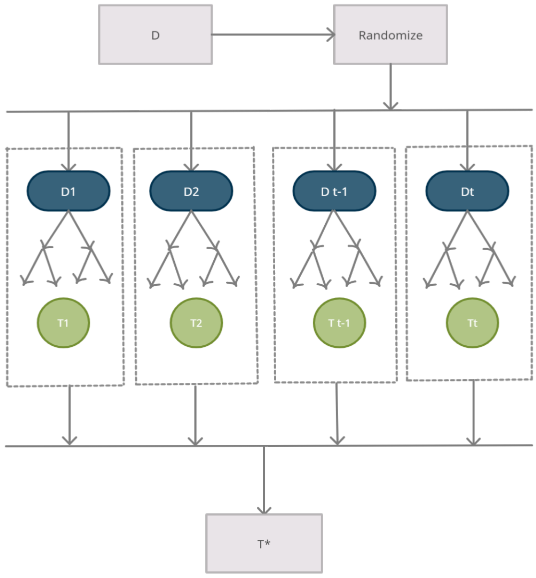

Figure 3.4: Structure of Random Forest

When learning a new algorithm, a random forest or random decision forests generate several decision trees, as shown in figure 3.4. There is a considerable reduction in the likelihood of over-fitting when numerous trees are averaged together. It works well with huge datasets, although it's less precise than Boosting in most cases [97].

XGBoost

The XGBoost is a learning algorithm with two advantages: separate learning units that eliminate the need for engineering.
Gradient boosting is a supervised learning technique that attempts to properly predict an objective variable by combining the predictions of many weaker models [23]. XGBoost is another tree model, which is a popular data mining technique with great speed and performance as shown in figure 3.5. In such case, it's possible that we won't have to create any functional interactions by hand. Additive tree boosting with two derivatives is used to build the XGBoost model. Gradient $\mathrm{g_i}$ and hessian $\mathrm{h_i}$, on their own, build a booster tree to deal with the problem of class imbalance. Regularization for training features and aim is given as follows by Chen and Guestrin's equation 3.3 for a tree set with K trees:

$$\hat{y}_i = \sum_{g=1}^{g} t_g\left(y_j\right), t_g \in \mathcal{F} \tag{3.3}$$

Where $\mathrm{t_g}$ is the practical field and F is the collection of potential classification and regression trees (CART). Optimized regularized target equation 3.4:

$$O(\theta) = \sum_{j}^{n} l\left(x_j, \hat{x}_j\right) + \sum_{g=1}^{g} \Omega\left(t_g\right) \tag{3.4}$$

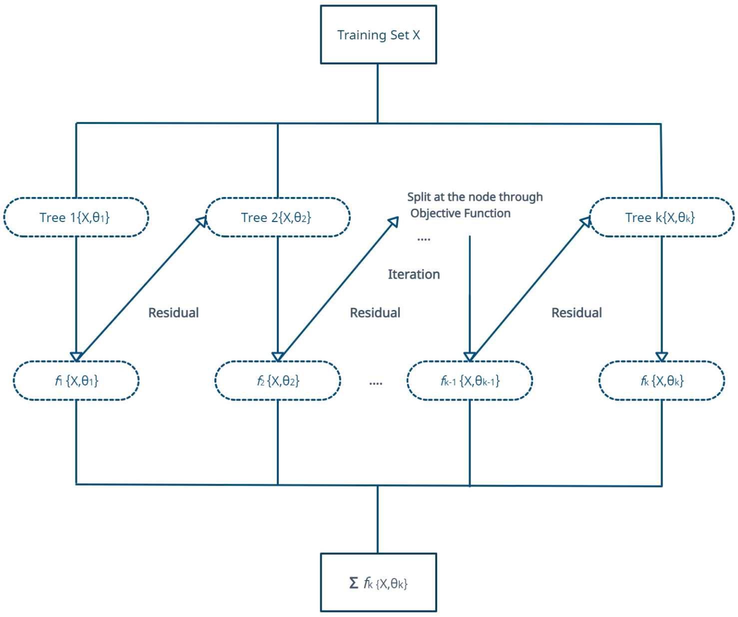

Figure 3.5: XGBoost Model generation

Now consider the additive tree boosting preparation, the optimized objective function is defined as in Eqn. 3.5:

$$O = \sum_{j=1}^{m} l\left(x_j, \hat{x}_j^{(t)}\right) + \sum_{i=1}^{f} \Omega\left(t_j\right) \tag{3.5}$$

It's a well-executed strategy. When it comes to large datasets and complex data, it's a good fit. Data that is scarce or widely spread is less amenable to this approach, and it can even cause problems. On the other hand, it performs better than the majority of supervised learning algorithms when faced with these kinds of data. The major drawback is the fact that it's a black box. XGBoost will not provide you with effect sizes if you request them (though some adaboost-type algorithms can give that to you). This is something you'd have to design and implement yourself. XGBoost isn't the best method for those circumstances because of the existing models (like penalized GLMs) [91].

3.4.2 Deep Learning (DL) Approach

In this section, we will go through the deep learning models that were utilized to forecast Bangladesh Airlines' feelings. Despite the fact that reviews are transformed to word embedded vectors, embedding is generated based on all of the individual characters in the background. Finally, we employ three deep learning algorithms: CNN, LSTM, and BERT.

Convolutional Neural Network (CNN)

Artificial Intelligence researchers employ Convolutional Neural Networks (CNNs) for picture feature extraction. However, it has proven its ability to handle data sets of one, two, and three dimensions. Convolutional, pooling and fully constructed layers are all hidden layers in the usual CNN design [78]. The network's foundation is made up of these layers, which are responsible for feature learning, down sampling, and classifying.

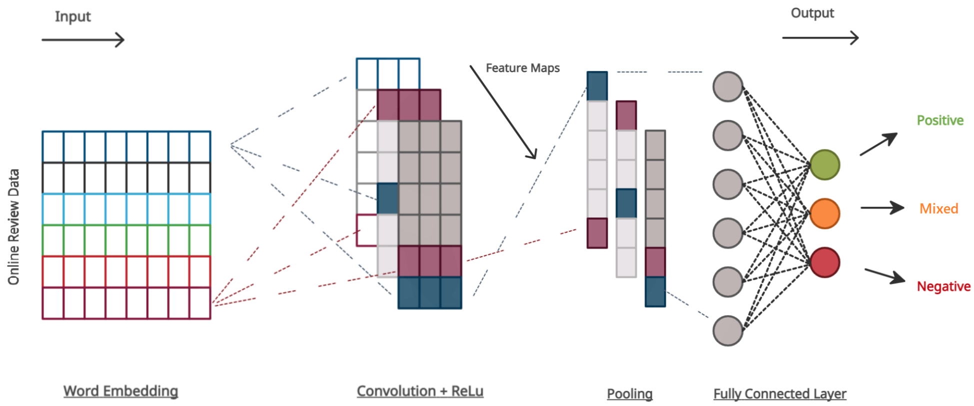

Figure 3.6: Stages of CNN architecture

As well as having an output layer that delivers predictions for a given job, the network has an input layer that receives input data. First, CNN was studied as a kind of network that uses back-propagation for error analysis and modifying network parameters. [37]. Figure 3.6 depicts the conventional architecture of a convolutional neural network. An artificial neural network's most significant component is its neurons. The neuron model includes multiplication, summation, and activation functions. The weight assigned to each input will be multiplied by it. The summing function combines all of the weighted inputs and biases together, and the output is generated by running the combined inputs through an activation function [56]. A CNN deep learning network, like a multilayer neural network, contains convolution layers and fully connected layers and the completely connected layer makes the final forecasts.

$$\begin{aligned} F(n) &= z(n) * k(n) \\ Fi(n) &= z(n)k(1) + z(n-1)k(2) + \ldots + z(0)k(n) \end{aligned} \tag{3.6}$$

The input signal and kernel filter are denoted by z and k, respectively. After the convolution operation, the convoluted signal is subsampled. As a result, the feature set's size will be lowered. The fundamental benefit of CNN over its predecessors is that it automatically discovers significant traits without the need for human intervention. The disadvantages are that huge amounts of training data are required, and the location and orientation of the object are not encoded [95].

Long Short-Term Memory (LSTM)

LSTM is another common RNN network model, with the architecture designed for the vanishing gradient problem [40]. They are highly useful for analyzing time series data trends. Because the forget gate, input gate, and output gate are all inserted inside this block of memories, the LSTM's capabilities are greatly enhanced. Using these gate architectures, it is possible to alter both old and new data in the memory cell. The cell state and the hidden state are sent to the next cell by the present cell.

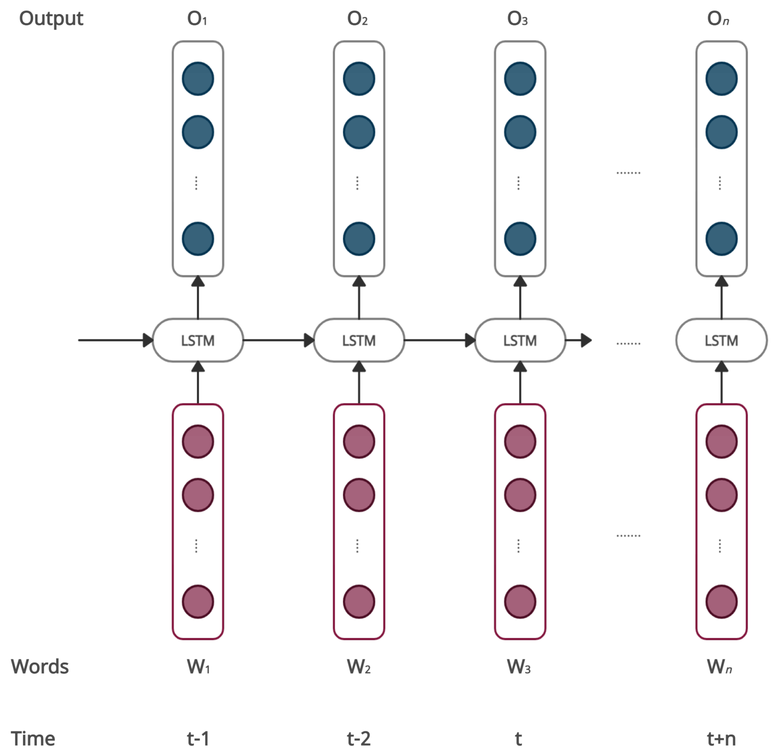

Figure 3.7: Unrolled standard LSTM node

LSTM is another prominent RNN network model, with an architecture designed for the vanishing gradient problem. The unrolled LSTM in the time series is depicted in the figure 3.7. Here, At time t, w_t is the input value, while o_t is the output value.

Again, The design of the LSTM network node is made up of three fundamental gates: the input gate i_t, the output gate o_t, and the forget gate f_t which is shown in the following figure 3.8. In contrast, the input and output gates indicate data entering and exiting the node at time t, respectively. The forgetting gate chooses whether or not to forget information based on the prior status information (h_{t-1}) and the current input (x_t). These three gates determine how to update the current memory cell c_t and latency h_t values.

To enhance the LSTM network's performance, forget gates can remove unneeded information from the cell state using a sigmoid function and an applied filter. The following sentence 3.7 is used to describe this procedure:

$$t_v = \varphi\left(E_b \cdot [x_{n-1}, y_n] + j_b\right) \tag{3.7}$$

where t_v refers to the forget gate, y_n refers to the new input and x_{n-1} refers to the previous hidden state. E_b and j_j are weights and bias respectively, which are learnable parameters, and φ mentions to the sigmoid function. To guarantee that

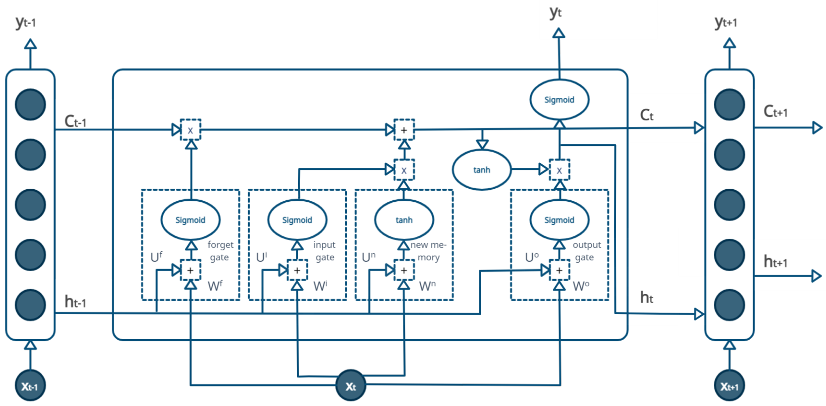

Figure 3.8: Connections of gates of standard LSTM

hardly meaningful material is supplied towards the unit condition, both the sigmoid and tanh functions are used. There are a few ways to express 3.8:

$$\begin{aligned} u_r &= \varphi\left(E_i \cdot [x_{n-1}, y_n] + j_i\right) \\ \tilde{D}_p &= \alpha\left(E_u \cdot [x_{t-1}, y_n] + j_u\right) \\ D_p &= q_c \circ D_{n-1} + u_r \circ \tilde{D}_p \end{aligned} \tag{3.8}$$

where it refers to the input gate and E_i and J_i are weights and bias of the input gate. Approximately D_p refers to the potential update vector, D_p is the new cell state and D_{n-1} is the previous cell state.

In addition, it also uses the *tanh* function to create a vector, similar to the input gate. Expressions 3.9 are as follows:

$$\begin{aligned} a_q &= \sigma\left(E_o \cdot [x_{n-1}, y_n] + j_o\right) \\ f_q &= o_q * \tanh\left(D_p\right) \end{aligned} \tag{3.9}$$

where a_q refers to the output gate, f_q refers to the new hidden state, and E_o and j_o denotes to the weights and biases of the output gate are discussed here. We've used a conventional LSTM network because of their ability for analysing as well as projecting temporal sequence statistics. It is possible to encounter the vanishing gradient problem when training conventional RNNs. This is why LSTMs were created. LSTMs outperform RNNs, hidden Markov models, and other sequence learning algorithms in a variety of ways. The problem of overfitting makes it difficult to execute dropout in LSTMs [30].

BERT

In order to learn extensive bidirections via unidentified materials through relying on both left and right contexts at all levels concurrently [58]. BERT (Bidirectional Encoder Representations from Transformers) is an embedding layer. In a manner similar to the vanilla encoder in the transformer, BERT receives a string of words as input and transmits it up to the following encoder unit. Self-awareness is applied to each encoder layer. A feed-forward network is then used to distribute the findings.

The feedforward network's output is subsequently fed into the following encoder. For each activity, BERT uses a fine-tuning strategy that does not require any specific architecture. An intelligent agent should employ as little human knowledge as possible when creating a model. When it comes to data, it should instead be taught.

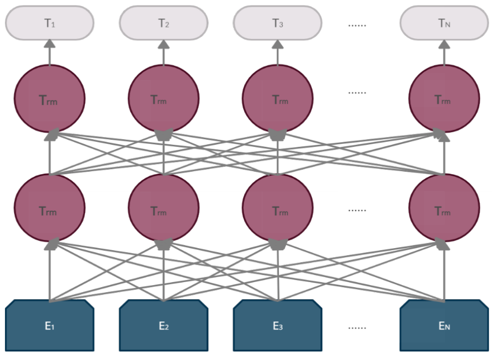

Figure 3.9: BERT Architecture

A pre-trained BERT model may be fine-tuned with just one extra layer to get cutting-edge outcomes in a variety of NLP applications, where E_n is the n-th input token, T_{rm} is the Transformer block, and T_n is the matching output embedding which is shown in the figure 3.9. With the support of surrounding text, BERT is meant to help computers decipher confusing language in texts. An initial training set of questions and answers were used to fine-tune the BERT architecture.

3.5 Topic Modeling

Topic models are a common and helpful method for analyzing the unstructured datasets. However, topic models aren't flawless, and existing models and frameworks are typically a "take it or leave it" proposition for many users in computational social science, digital humanities, and information studies who aren't machine learning professionals. For the past several years, topic modeling has been a popular tool for researchers who want to uncover hidden trends in the data they collect. Sentiment analysis, a text mining component, is known as opinion mining because it detects whether or not individuals have expressed positive or negative opinions in text data [86]. Understanding enormous amounts of unstructured material is still a challenge. Unlike information retrieval, when users know exactly what they want, users may need to comprehend a corpus' high-level themes and explore articles of interest. Topic models are a formalism for presenting a collection's topics, and they've been used to help people find information [11]. Outside of text, topic models have been used to learn natural scene categories in computer vision, find patterns in population genetics, and better comprehend the relationship between Bayesian models and cognition [90].

The unsupervised nature of topic models makes them particularly useful. They do not necessitate any pronouncement. A topic model requires just a quantity of articles as well as any range of topics you want the model to find as inputs. In addition, there are models and heuristics for automatically determining the number of topics [83]. Using topic modeling, we can organize, interpret, and summarize vast amounts of textual material quickly and efficiently. It aids in the discovery of previously unnoticed thematic patterns in the collection. Document annotations based on these subjects. Unstructured textual data can be discovered via topic modeling, which identifies patterns in the data. Inaccurate subject descriptions and a decrease in the effectiveness of text mining classification findings might arise from this inaccuracy. Bangladesh Airlines online passenger reviews were subject-classified into six broad categories, and from this topic modeling, we were able to extract key words from the evaluations. We then used these key phrases to create a topic model for the entire document.

3.6 Explainable AI

Non-interpretable concepts could bring to a host of problems. According to [52], there are still few studies in which models explain their decision-making process, and the task of detecting rumors is no exception. As a result, we use LIME to try to make the forecasts more understandable.

3.6.1 LIME

In the field of artificial intelligence, LIME stands for Local Interpretable Model-agnostic Explanations. Its purpose is to make machine learning predictions intelligible to people. As a local explanation approach, it is suited for describing individual cases. To produce fake data that has just a portion of the actual properties, LIME manipulates the incoming data. Text data may be used as a case study in this regard: a random selection is made from the original text and eliminated from each of the new copies. Artificial data is then categorized into various subcategories (classified).

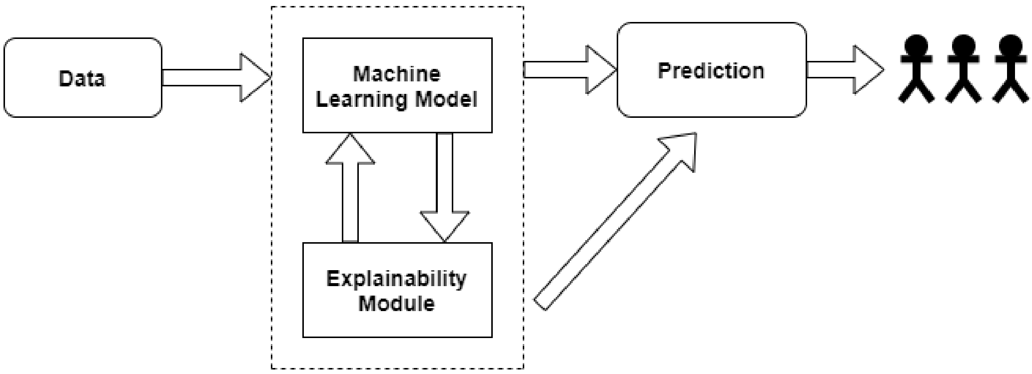

Figure 3.10: Working Procedure of LIME

As a result, the presence or lack of specific keywords may be shown to alter the categorization of the text that is being studied. Using LIME, you can get a list of explanations for how each feature affected the final forecast. For any datasets include a wide range of options and it will focus on just a few of LIME's most useful features to make things easier to understand [25]. Again, each classifier's predictions are properly described by approximating them to an interpretable model locally using LIME. There are certain classifiers that employ unintuitive representations. For example, Lime uses words to describe these classifications, despite the fact that this is not the actual representation utilized by the classifier. Lime also considers the constraints of human attention span, thus the explanations are kept to a minimum. Explanation and selection mechanism for a representative set of interpretations [SP-LIME] to verify that the model behaves consistently while mimicking human thinking. [SP-LIME]. In figure 3.10, a flowchart of LIME is shown for understanding the model. LIME uses a sampling strategy that takes into account both occurrences that are close to and distant from the original entry's interpretable form. These samples are represented in a way that is understandable, and LIME shows their forecasts while reducing losses and complexity.

Chapter 4

Proposed Model and Methodology

4.1 Overview

In this chapter, we explain our suggested model for sentiment categorization and topic modeling on Bangladesh Airlines in order to achieve customer satisfaction.

In Section 4.2, we described our models' broad concept proposition in terms of sentiment analysis and topic modeling. Then, in Section 4.3, we showed the data preparation stages.

Again, the section 4.4 describes the classification methods used in sentiment analysis. In Section 4.5, we explained the topic modeling process as well as LIME.

4.2 Idea Proposition

In this work, we worked on a customized dataset named Bangladesh-Air-Data including online reviews for 4 major Bangladesh Airlines, performed a multiclass sentiment analysis, and compared the classifiers. This method begins with pre-processing procedures used to clean the reviews and balanced the review data using the Pegasus model's oversampling mechanism. System choices use various machine learning techniques to integrate feature engineering and word embedding for deep learning. The analysis was carried out 3 different machine learning (Decision Tree, Random Forest, and XGBoost) and 3 different deep learning classification strategies (CNN, LSTM, BERT). The test set's output is the online review sentiment (positive/negative/mixed) using a three-class dataset, and the performance in terms of accuracy is calculated. Based on the results, we have achieved the best accuracy 83% in terms of BERT. The accuracies were determined to compare each categorization technique, and the total sentiment count for all four airlines of Bangladesh was displayed in terms of domestic route, international route and all route. Again, on based of passenger online reviews from our Bangladesh Airline review data, we also find our the best services airline of Bangladesh regarding domestic route, international route and overall route. We also compare our result along with existing USA Airline Tweets data [45] to validate our model. This study work on both 519 online reviews for domestic route and 528 online reviews for international route passengers online review and lastly, work on overall 1047 online reviews of Bangladesh Airlines.

Users that having employed or plan to utilize carriers' rating systems, such as Skytrax[1] or Tripadvisor[2] , are interested in the topic modeling and sentiment analysis of posts on these sites. With the help of websites like Skytrax and Tripadvisor, clients may learn from the experiences of others who have travelled with the same airline. Customers who have previously taken flights with Bangladesh airlines have provided us their online feedback. Finally, we apply LIME to understand the sentiment predictions of 3-class dataset using XGBoost model. This study work on increasing 1095 online reviews of Bangladesh Airlines.

4.3 Data Preprocessing

Since raw data contains noise, duplication which is unsuitable for study, for that reason data pre-processing is a critical step in any type of data analysis. There are many various ways to classify data, but one of the most important is to shift the information towards a shape that could be evaluated. It additionally demands information labeling when the categorization techniques involve automated learning categorization strategies as well as system assessments are required. In addition, true digital information is frequently contaminated with typos, abbreviations, and symbols, making classification results questionable. "THX," "Thanks," "Thenks," and other abbreviations are commonly used to express gratitude in social media posts, whereas the final one is a typo. Even though it's obvious to humans that both names signify exact identical item, which provides a tremendous difficulty for machine learning techniques to tackle.

In our model, the final text-based data generated by websites, users have a diverse and unbalanced set of material. To balance the final dataset that is Bangladesh-Air-Data, we use the oversampling techniques using the PEGASUS pretrained model [81]. First, the less sample of negative and mixed reviews in the dataset are paraphrased to upsampling the data. Then we merge the newly negative and mixed reviews in the final dataset. After merging, we got then 1047 reviews after oversampling.

4.3.1 PEGASUS Model

Extractive summaries may be constructed using PEGASUS, that eliminates crucial phrases out of an intake manuscript as well as creates a single return order from the remainder. Several full sentences are omitted from documents in this pre-training strategy, and the model is tasked with finding them. We apply the PEGASUS model for balancing our initial dataset which is one kind of oversampling technique.

Then we remove some stop word from our dataset, which is repetitive among three classes like "br", "href", "https" etc. as they do not have contributions in the reviews. We also do some feature engineering after removing the stop word. For example, we find out the food quality, stuff behaviour features from the dataset which are positive or negative from customer reviews. Regarding food quality, "nice food" and "less food" words are counted for reviews using feature engineering. A

[1]http://airlinequality.com/(Skytrax_Web)
[2]http://tripadvisor.com/(Tripadvisor_Web)

hashtag-based crawl was used to find reviews published in English, and then we selected the best ones [96].

Again, we clean the text data removing any HTML decoding. Then all the characters in the dataset were converted to lowercase. Because website links in the dataset do not offer any emotional context, the link information in the dataset is removed. Furthermore, the numerous spaces between words were condensed to a single space. In the online reviews, we also deleted punctuation, numerals, and undefined characters. After that, we extracted key phrases in the cleaning data using bi-gram and tri-gram to find the common phrases in the text data and comb the phrases into one token. Finally, we convert the text data into the sequence to vectorize the data. The Complete Outline of our framework is illustrated in figure 4.1.

After data preprocessing, we separated the input data into two phases:

- Phase-I: Machine Learning (ML) Approach
- Phase-II: Deep Learning (DL) Approach

4.4 Classification

In our proposed methodology for sentiment classification, we applied three machine learning algorithm such as Decision Tree (DT), Random Forest (RF) and XGBoost and three deep learning algorithm such as CNN, LSTM and BERT.

4.4.1 Machine Learning (ML) Approach

Phase-I examines the machine learning models that were used to predict the sentiments of Bangladesh Airlines. After text data to sequence, we combine new feature text data. Lastly, we apply three machine learning algorithms such as: Decision Tree (DT), Random Forest (RF), and XGBoost (Xgb).

Combining new feature text data In this process, we add some feature data while feature engineering in data preprocessing with text data to sequence for increasing the number of features. Now we applied the machine learning algorithms on the new feature based data.

Decision Tree (DT) A decision tree classifier is a fundamental and frequently used data classification tool. The Decision Tree is a tree-like structure with core nodes which reflect the class labels. This categorization method asks well prepared questions regarding the test data set's characteristics [99]. We used the 'DecisionTreeClassifier' algorithm to construct the model. Within the algorithm, we've listed 'min_sample_leaf' to be '1' and the prediction comes from the value of "y_pred". An online review document is examined using a decision tree in order to determine its mood. It is possible to pass the root node to the branch of no. in an online review like "The planes always delay". Afterwards, the tweet travels via the "delay" node, and it is rated as a negative.

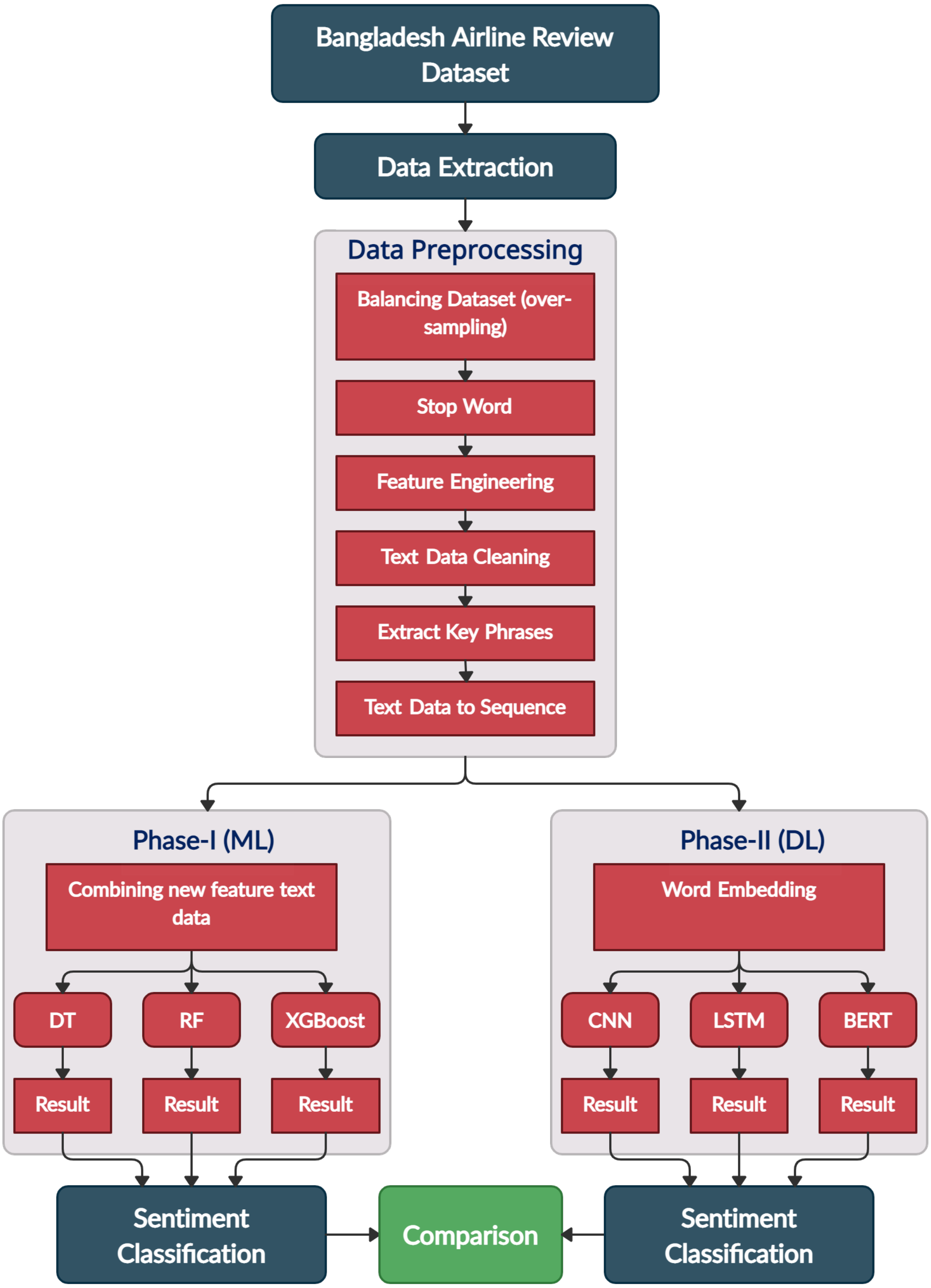

Figure 4.1: The Complete Outline of our framework.

Decision trees are a common knowledge discovery approach since they don't require domain expertise or parameter configuration. In addition, Decision Trees may be shown and the physics underlying them are intuitively understood by individuals. This classifier may be developed in a minimal space of period and have a significant rate of success.

Random Forest (RF) A random forest classifier is a classification method that uses ensemble learning [68]. This method is capable of managing huge datasets and thousands of input variables without deleting any of them. This model can deal with data point overfitting. To classify an instance, the Random Forest classifier uses the classification decisions of each Decision Tree in the Random Forest classifiers to determine which class it belongs to. In our model, we developed using the 'RandomForestClassifier' algorithm and we listed n_estimators of 250 and n_jobs of 4 for constructing model.

XGBoost The XGBoost is a learning algorithm with two advantages: separate learning units that eliminate the need for engineering. Gradient boosting is a supervised learning technique that attempts to properly predict an objective variable by combining the predictions of many weaker models [98]. XGBoost is another tree model, which is a popular data mining technique with great speed and performance. In our framework, we designed the model using the 'XGBClassifier' algorithm given by the xgboost bundle. We stated the n_estimators of 250 and learning rate of 0.15 and finally equipped and stored the expected values in the "y_pred".

4.4.2 Deep Learning (DL) Approach

Phase II is meant to help emotionally categorize the message by swiftly identifying the link between the sentences in the online assessments. Despite the fact that reviews are converted to word embedded vectors, embedding is produced based on all individual characters in the backdrop. Then we use three deep learning models: CNN, LSTM, and BERT.

Word Embedding Word Embeddings is a technique in which each word is represented with a single vector while its semantic significance is taken into account. Word integration that works effectively can support the performance of text encoding and categorization. Using this strategy, the dataset may be represented using a variety of approaches [69]. Here, we use here 200 dimension embedding layer for vector representation.

Convolutional Neural Network (CNN) Now the embedded features were extracted using CNN. The Convolutional Neural Network (CNN) is a neural network with several layers feedforward model that is one of the most significant deep learning designs [95]. CNN is made up of three layers: convolution, pooling, and a fully linked layer. Unlike traditional artificial neural networks, a convolution layer automatically extracts features and a pooling layer reduces them [94]. In CNN model, visual display that accepts text as an input and displays linkage between CNN model layers to determine the class of the text.

Long Short-Term Memory (LSTM) LSTM is another common RNN network model, with the architecture designed for the vanishing gradient problem [88]. In the LSTM network design, the first word to the last word of a phrase is used as an input. This entails tracing the origins and relationships of individual words all the way through to their final form. Human linguistic as well as electromagnetic databases, for example, contain lengthy connections. LSTM may extract useful features. For NLP challenges including solving semantic dependencies among words, LSTM delivers excellent results. These approaches are useful because they can help solve sentiment categorization issues. The categorization findings show that this is the case.

Pre-processing removes material such as URLs, emojis, stop words, and other non-essential information. However, the user's thinking is still reflected in the information that has been deleted. Using a variety of text representations will help convey a more accurate sentiment.

BERT To develop complex bidirectionals using unidentified words, training of the left and right contexts at multiple levels at the same time is used [69], BERT (Bidirectional Encoder Representations from Transformers) is an embedding layer. A pre-trained BERT system may be perfectly alright using hardly single additional level to get cutting-edge results in a variety of NLP applications. We use BERT to assign a polarity score to each online review, categorizing them as either positive, negative, or mixed based on their sentiment.

4.5 Topic Modeling and LIME

Here, we discuss about Lime and Topic Modeling. Using Topic Modeling in our approach, Skytrax and Tripadvisor users may make better decisions based on the actual experiences of other consumers who have traveled with airlines. We gathered input from Bangladesh airlines passengers who have previously taken the airline's flights. Topic modeling and sentiment analysis were used to identify important terms in the online reviews based on the collected data.

4.5.1 Data Collection

Skytrax and Tripadvisor online data were used in this part. After combining the Skytrax and Tripadvisor datasets, we arrived with 1095 review records for Bangladesh Airlines. Selenium may be used for this first stage, crawling each review site's page-by-page HTML unstructured data. In order to extract structured text data from HTML, we utilized beautifulsoup and the pandas package of Python. There were four Bangladesh airlines that they researched: Biman Bangladesh Airlines, NOVOAIR, Regent Airways, and US-Bangla Airlines.

When it comes to airline and airport reviews and rankings, you can't beat Skytrax, a UK-based consultant. Complaints were lodged with the UK Advertising Standards Authority by KwikChex in 2012. All five complaints were upheld by the authorities, and Skytrax agreed to make a few wording changes to its promotional materials as a result. Again, Tripadvisor is an American digital tourism service which operates a site as well as smartphone software containing consumer material as well as a comparison buying site.

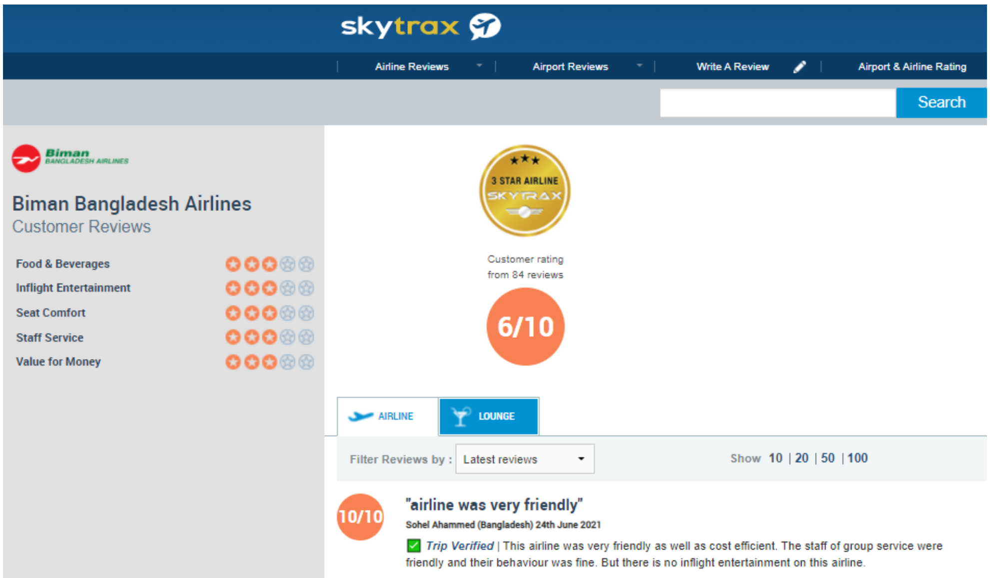

Figure 4.2: The Skytrax Biman Bangladesh airline review and rating portal [86]

We utilize data from Skytrax's Bangladesh airline review site to show how the amenities lead to traveler satisfaction. It's possible for Skytrax customers to leave feedback in the form of a review, as seen in the figure 4.2. To help others evaluate the quality of airlines and airports, a rising number of customers (i.e. passengers) provide online evaluations of their travel experiences.
This time around, Tripadvisor data from Bangladesh airlines is being used to show how the qualities of the airlines contribute to the overall experience of the travelers. US-Bangla airline review and rating scores are depicted in the figure 4.3. When it comes to the airline industry, sentiment analysis and topic modeling techniques are becoming more popular as research tools since internet evaluations and customer comments have a considerable influence on consumers' purchasing intentions.

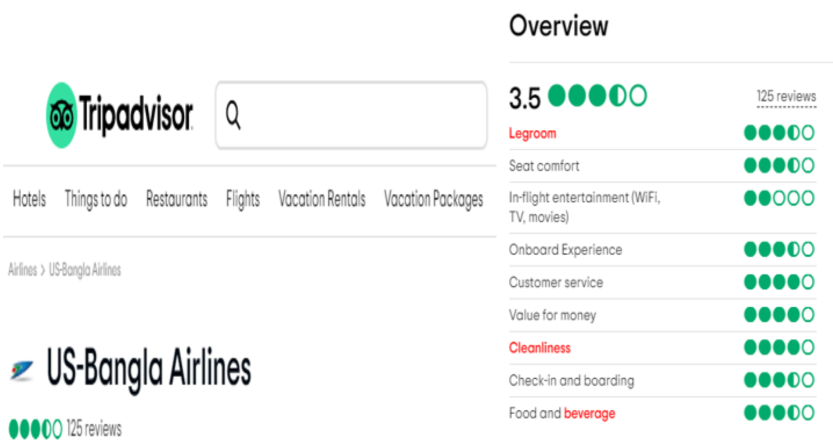

Figure 4.3: The Tripadvisor US-Bangla airline review and rating portal [86]

4.5.2 Data Preprocessing

Using data preprocessing, you may change basic information towards a shape that can be utilized. The final text-based data offered by websites contains a wide range of material that is unevenly distributed among its users. We remove a number of stop words from our dataset that appear in three different classifications as part of our data pre-processing. HTML decoding is omitted from the text data that we collect. The letters in the sample were then all lowercased. All source data has been removed. from the dataset [85] since it provides no emotional context for the website linkages. There was also a reduction in the number of spaces between words. Remove punctuation, digits and other characters that were not clearly stated the online evaluations. Stemming words were used to create a tokenization approach that treated words as if they were one word.

4.5.3 Data Analysis

The top 100 words based on frequency analysis are presented in this section. Then, using the dataset, we created word clouds to create the entire graphic. Using data mining techniques, we were able to estimate the chance that certain themes will appear across the whole text in a topic modeling research on structurally modified manuscripts. Finally, sentiment analysis was used to deduce good, negative, and mixed phrases based on the emotional words that customers used in the content of the article. The framwork study depicted in figure 4.4. The explanation of this study discussed in chapter 5.

4.5.4 LIME

In the field of artificial intelligence, LIME stands for Local Interpretable Model-agnostic Explanations. Its purpose is to make human predictions intelligible by machine learning algorithms. As a local explanation approach, it is suited for describing individual cases. Using the input data, LIME produces false data that only contains a portion of the actual properties. For example, in the case of textual data, multiple copies on any actual content being created, each with a certain quantity of randomly selected letters. are omitted from the original. The new fake data is then categorized into many subcategories (classified). As a result, we can observe the impact of specific keywords on the categorization of the selected text. As a list of explanations, LIME shows how each attribute contributed to the final forecast, which can be seen in the output. There are a lot of options in our datasets. We'll focus on a couple of LIME's most popular features to make things clearer and easier to grasp. Here, we apply Xgb model for predicting sentiment Bangladesh airline reviews in terms of clarity and ease of comprehension.

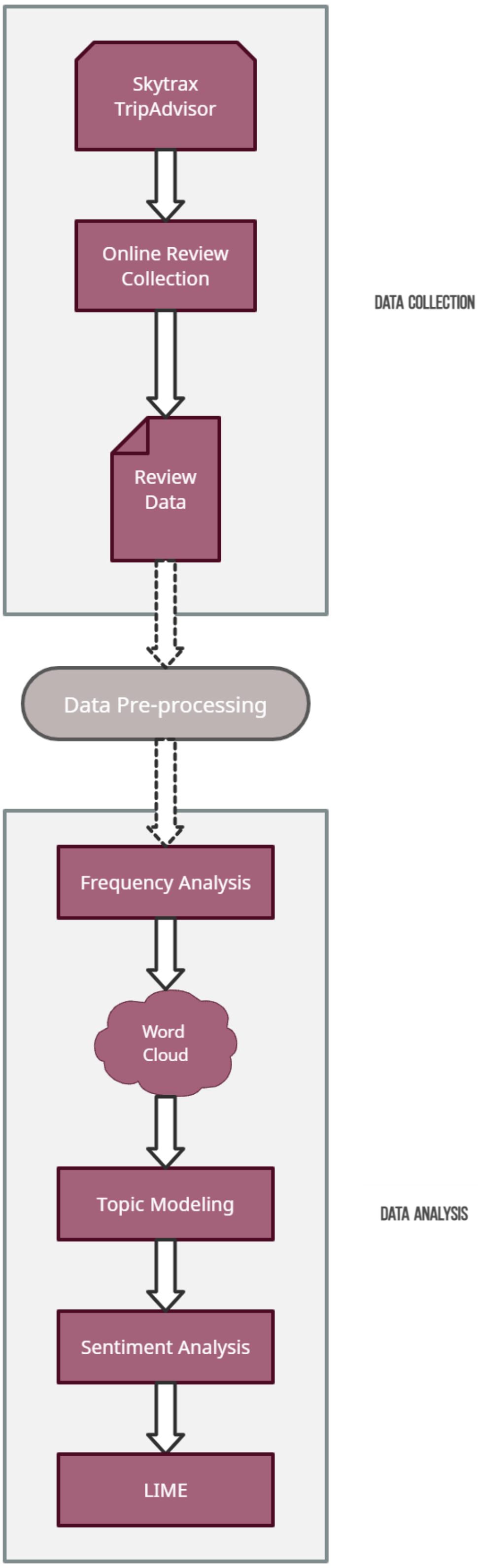

Figure 4.4: Framework of this study regarding Topic Modeling and LIME

Chapter 5

Experimental Results and Empirical Evaluation

5.1 Overview

There are two sections, Section 5.2 and Section 5.3 in this chapter that deals with the datasets, we utilized in our research and the categorization method we employed are described precisely. Section 5.4 shows the performance matrix for our proposed methodology. In Section 5.5, we describe the performance regarding topic modeling and LIME in terms of Bangladesh Airlines dataset.

5.2 Bangladesh Airline Review Dataset

The dataset named Bangladesh-Air-Data gives a view of customer online reviews in terms of Bangladesh Airline for all route that covers all the reviews from September 2018 to July 2021. It has created from Skytrax[1] and Tripadvisor sites[2] and preprocessed by us.

	review	target	title	airline
0	Worst experience at Dhaka airport. Delayed fli...	Negative	Worse Flying experience	Biman_Bangladesh_Airlines
1	Good value for money. Somethimes you might con...	Positive	Good value for price and food was excellent.	Biman_Bangladesh_Airlines
2	Very good airlines ever i see. Good service fo...	Positive	excellent	Biman_Bangladesh_Airlines
3	I naver used Biman Bangladesh and it was first...	Positive	Fantastic food and service	Biman_Bangladesh_Airlines
4	I have travelled from London to Dhaka direct f...	Negative	Rude flight supervisor at the airport desk	Biman_Bangladesh_Airlines
...	...	...	...	...
747	Excellent round-trip between DAC-ZYL recently....	Positive	Regent Airways customer review	Regent_Airways
748	Cox's Bazar to Dhaka non-stop and one hour fli...	Mixed	Regent Airways customer review	Regent_Airways
749	DAC-CGP. Online booking available. On arrival ...	Positive	Regent Airways customer review	Regent_Airways
750	Saidpur to Dhaka. Worse airline ever. Late of ...	Negative	"Worse airline ever"	NOVOAIR
751	Jashore to Dhaka. A not well-known domestic ai...	Mixed	"Not a bad airline!"	NOVOAIR

752 rows × 4 columns

Figure 5.1: Initial Sample of Bangladesh-Air-Data

[1] http://airlinequality.com/(Skytrax_Web)
[2] http://tripadvisor.com/(Tripadvisor_Web)

We have 6 title such as: characteristics, review, evaluation, travel_date, airline, rating from Tripadvisor information where the sample size is 638. We also have 19 Skytrax attributes and the sample size is 114. Our initial and merged final dataset has a common set of 3 attributes, such as the title, review and airline, and an additional target feature for sentiment prediction based on the reviews. The following figure 5.1 visualised the initial sample of our dataset. Here, we can see the sample size is 752 and column size is 4.

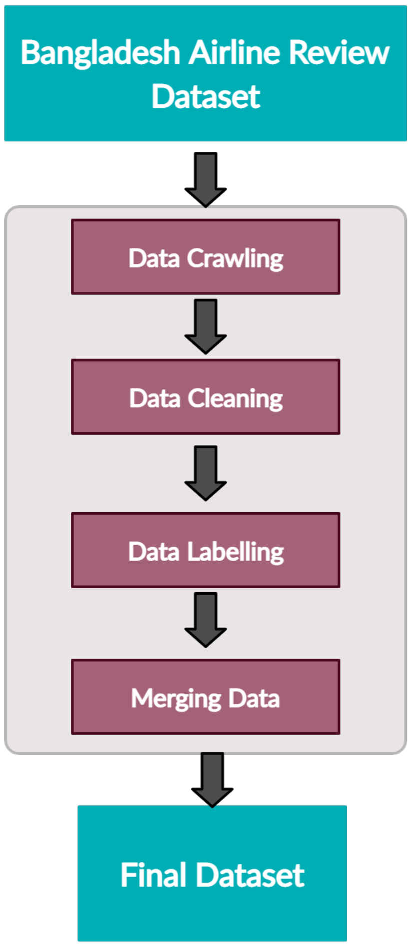

Figure 5.2: Processing of Bangladesh Airline Review Dataset

To process our Bangladesh Airline review dataset, we take some necessary steps. At first crawl, page-by-page HTML unstructured data from online review websites using selenium library to download the whole page. Then for cleaning the HTML unstructured data, we used beautifulsoup and pandas library to get the structured text data. For data labelling, 10 annotators work on the data to label the dataset and two researchers observe this dataset to correct them. Then finally, we merge the two source dataset of Tripadvisor and Skytrax using pandas library to get the final dataset. In the Tripadvisor dataset, we got 638 reviews, and for the Skytrax dataset, we got 114 reviews. After merging, we got a total of 752 reviews as our final dataset. The figure 5.2 depicts the processing of Bangladesh Airline Review Dataset.

5.3 Data Extraction

In this work, we customized a dataset of Bangladesh Airline Service. A total of 752 online reviews were extracted, which formed the final initial dataset for sentiment classification. The online reviews were collected for four major Bangladesh Airlines: Biman Bangladesh Airlines, NOVOAIR, Regent Airways and US-Bangla Airlines. The online reviews were divided into three categories: positive, negative, and mixed. Here, a mixed review means a mixture of positive and negative. The sentiment distribution of online reviews is shown in the Table 5.1 below.

Table 5.1: Sentiment Distribution of Online Reviews for Initial Dataset

Sentiment	**Reviews Count**
Positive	349
Negative	264
Mixed	139

From the following figure 5.3, we see the class distribution of three class dataset where the positive class is 46% of 100% and others are 35% and 19% respectively which denotes negative class and mixed class.

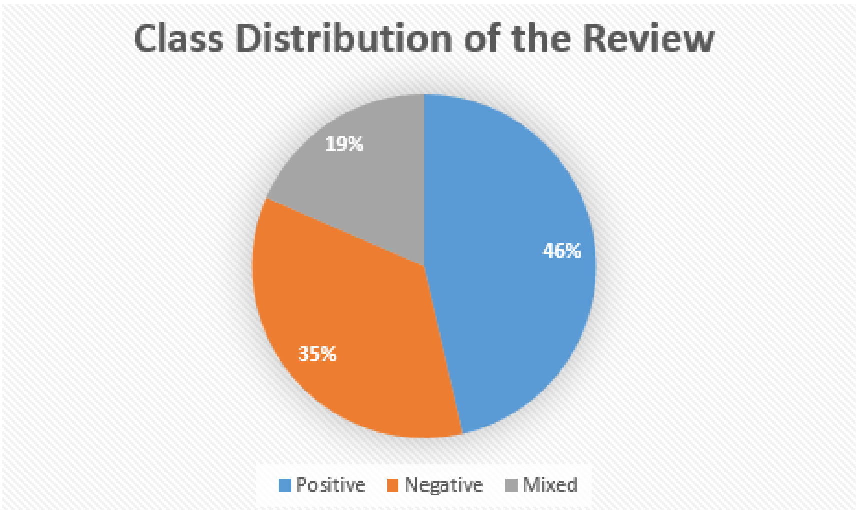

Figure 5.3: Graphical View of Sentiment Distribution of Online Reviews for Initial Dataset

We also find out the sentiments of each review in all four airlines and identify the best one. The figure 5.4 depicts the number of sentiment in each airline. From the

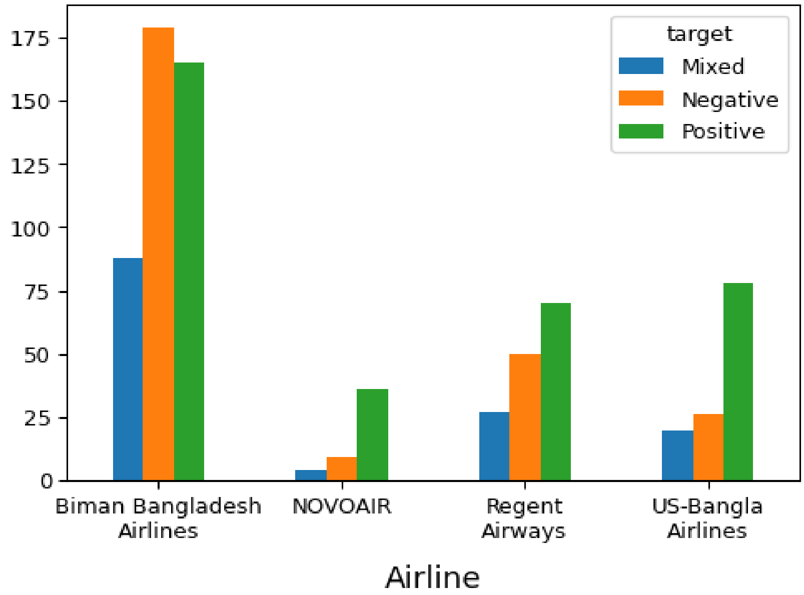

Figure 5.4: Distribution Graph of each Airline group by sentiments for all route

figure, we visualized that Biman Bangladesh Airline received most negative reviews of 179 than positive and mixed.

Again, in terms of NOVOAIR, we got the positive reviews of 36 outperforms the other. For Regent Airways, we also got the most positive reviews of 70. Finally, for US-Bangla, a positive review is higher with 78 numbers than mixed and negative. Comparatively, US-Bangla is the best Airline of Bangladesh in terms of customer online reviews for all the route.

Table 5.2: Sentiment Distribution of Online Reviews for four Bangladesh Airline for all route

Airline	**Positive Review**	**Negative Review**	**Mixed Review**
Biman Bangladesh Airlines	165	179	88
NOVOAIR	36	9	4
Regent Airways	70	50	27
US-Bangla Airlines	78	26	20

From the following Table 5.2, we see that the highest number of positive review comes from US-Bangla Airlines than negative and mixed reviews. Again, other airlines of Bangladesh got higher negative reviews than positive and mixed. From on that experiment, we say that US-Bangla is the most satisfied airline from customer reviews considering all the route.

Domestic Route Data

Again, we distributed the initial sample of Bangladesh-Air-Data into two category: domestic route and international route. In terms of domestic route, we got 350 online reviews from domestic route passengers and from the following figure 5.5 visualised the initial sample of our dataset in terms of domestic route.

	title	review	Route	Target	airline
0	"airline was very friendly"	This airline was very friendly as well as cost...	Dhaka to Rajshahi	Mixed	US_Bangla
1	"significantly improved recently"	Biman Bangladesh have significantly improved ...	Chittagong to Dhaka	Mixed	US_Bangla
2	"The space was enough"	Chattogram to Cox's Bazar by BG1433. The boar...	Chattogram to Cox's Bazar	Positive	US_Bangla
3	"happy to use Biman again"	Dhaka to Rajshahi. Nice meal. Good crew. The ...	Dhaka to Rajshahi	Positive	US_Bangla
4	"There were no announcements"	I reached Dhaka from Chittagong at 3 pm and m...	Dhaka to Chittagong	Negative	US_Bangla
5	"happy to fly with Biman"	Dhaka to Rakshahi, the flight was on time. Th...	Dhaka to Rajshahi	Positive	NOVOAIR
6	"food was terrible"	heathrow to Sylhet with Biman Bangladesh and w...	Sylhet to Dhaka	Negative	NOVOAIR

Figure 5.5: Initial Sample of Bangladesh Airlines review data for domestic route

The initial sample of domestic route has five column named: title, review, Route, Target, airline. The route column denotes the domestic destination of Bangladesh. Normally Dhaka, Chattogram, Sylhet, Cox's Bazar, Saidpur, Jashore, Rajshahi, Barishal etc. districts are covered by four major airlines of Bangladesh in domestic route. Regent_Airways covered only four district from Dhaka named Chattogram, Cox's Bazar, Jashore, Saidpur whereas other airlines covered most of the district in this domestic route of Bangladesh. We also find out the sentiments of each review in all four airlines and identify the best one in terms of domestic route flights. The figure 5.6 depicts the number of sentiment in each airline for domestic route.

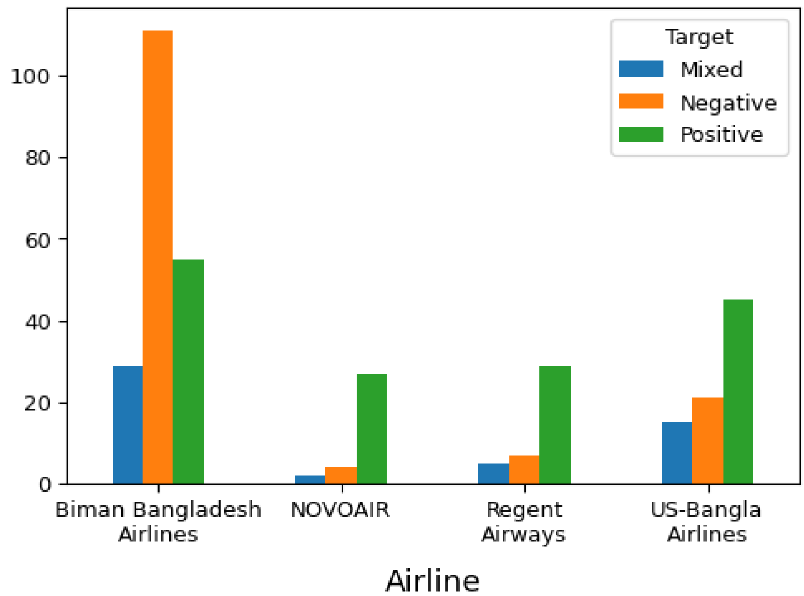

Figure 5.6: Distribution Graph of each Airline group by sentiments for domestic route

From the figure, we visualized that Biman Bangladesh Airline received most negative reviews of 111 than positive and mixed. In case of Regent_Airways, we got 29 positive reviews whereas for US_Bangla, we got 45 positive reviews in the domestic route from passengers online reviews. Again, for NOVOAIR, we got 27 positive reviews which outperforms the negative and mixed reviews.

Table 5.3: Sentiment Distribution of Online Reviews for four Bangladesh Airline for domestic route

Airline	**Positive Review**	**Negative Review**	**Mixed Review**
Biman Bangladesh Airlines	55	111	29
NOVOAIR	27	4	2
Regent Airways	29	7	5
US-Bangla Airlines	45	21	15

Again, from the following Table 5.3, we see that the highest number of positive review comes from NOVOAIR and US-Bangla Airlines than negative and mixed reviews. Again, Biman Bangladesh Airlines got higher negative reviews than positive and mixed in this domestic route. From on that experiment, we say that NOVOAIR and US-Bangla is the most satisfied airline from customer reviews considering domestic route of Bangladesh.

International Route Data

In terms of international route, we got 402 online reviews from domestic route passengers and from the following figure 5.7 visualised the initial sample of our dataset in terms of international route.

	title	review	Route	target	airline
0	"an excellent and comfortable journey"	It was really an excellent and comfortable jo...	Dhaka to Dubai	Positive	US_Bangla
1	"the service quality has been improved"	It is assumed from my last journey Dhaka to L...	Dhaka to London	Positive	Biman_Bangladesh_Airlines
2	"employees are very supportive and friendl"	This is my last journey from Dhaka to London....	Dhaka to London	Positive	Biman_Bangladesh_Airlines
3	"kind and friendly"	It was really brilliant to fly on Boeing 787. ...	London to Dhaka	Positive	Biman_Bangladesh_Airlines
4	"Good food, good seating"	I flew Dhaka to Bangkok on 27 August 2019 with...	Dhaka to Bangkok	Positive	US_Bangla
5	"they were completely disregarded and ignored"	Manchester to Sylhet. One of the worst experi...	Manchester to Sylhet	Negative	Biman_Bangladesh_Airlines
6	"Worst customer service"	Have a flight canceled from Myanmar to Bangla...	Yangon to Dhaka	Negative	Biman_Bangladesh_Airlines
7	"Good food, good seating"	Dhaka to Bangkok in August 2019. Good food, g...	Dhaka to Bangkok	Positive	US_Bangla
8	"It was a pleasant experience"	Dhaka to Doha. It was a pleasant experience f...	Dhaka to Doha	Positive	US_Bangla
9	" The staff could be more polite"	Doha to Dhaka. The line was quite long in the...	Doha to Dhaka	Mixed	US_Bangla
10	"service is extremely poor"	London to Sylhet, the service is extremely po...	London to Sylhet	Negative	Biman_Bangladesh_Airlines

Figure 5.7: Initial Sample of Bangladesh Airlines review data for international route

The initial sample of international route has five column named: title, review, Route, Target, airline. The route column denotes the international destination of Bangladesh. Normally Dhaka, Bangkok, Dammam, Doha, Kathmandu, Kolkata, Kuala Lumpur, Muscat, Singapore, London, Manchester, Yangon, Kuwait, Riyadh, Jeddah, Abu Dhabi, Chennai, Male etc. cities all over the world are covered by four major airlines of Bangladesh in international route. NOVOAIR covered only one city from Dhaka named Kolkata whereas other airlines covered most of the cities all over the world in this international route. We also find out the sentiments of each review in all four airlines and identify the best one in terms of international route flights. The figure 5.8 depicts the number of sentiment in each airline for international route.

From the figure, we visualized that Biman Bangladesh Airline received most positive reviews of 110 than negative and mixed. In case of Regent_Airways, we got 35 positive reviews whereas for US_Bangla, we got 45 positive reviews in the international route from passengers online reviews. Again, for NOVOAIR, we got only 5 positive review and the number of negative and mixed reviews are 4 and 5 accordingly as this airline visit only Kolkata.

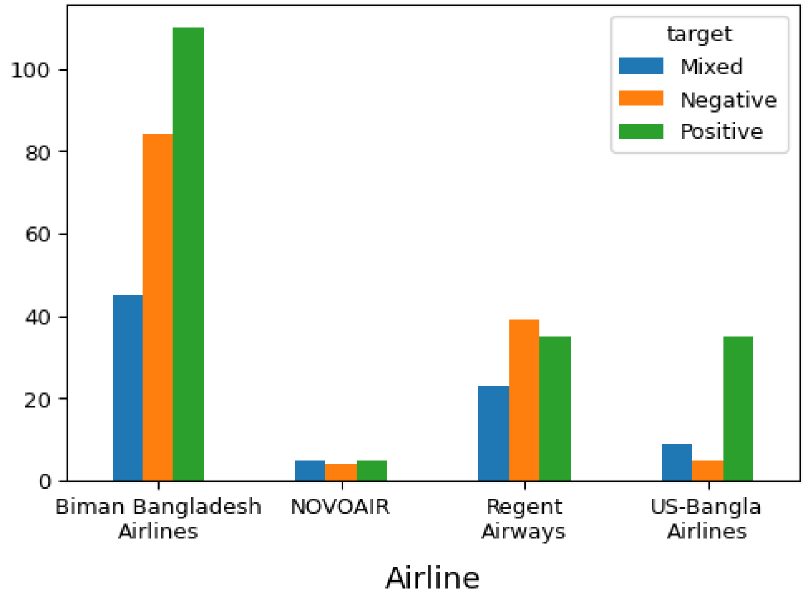

Figure 5.8: Distribution Graph of each Airline group by sentiments for international route

Again, from the following Table 5.4, we see that the highest number of positive review comes from Biman Bnagladesh Airlines and US-Bangla Airlines than negative and mixed reviews. Again, Regent_Airways got higher negative reviews than positive and mixed in this international route. From on that experiment, we say that Biman Bnagladesh Airlines and US-Bangla is the most satisfied airline from customer reviews considering international route.

Balancing Dataset

For balancing the dataset, we apply PEGASUS model (an oversampling technique) to increase the number of dataset in terms of sentiment classification. The initial sample of our Bangladesh Airline review data is 752 considering all route where the dataset is imbalanced. After applying the PEGASUS model, the number of online reviews in case of negative and mixed reviews are increased as they are in

Table 5.4: Sentiment Distribution of Online Reviews for four Bangladesh Airline for international route

Airline	Positive Review	Negative Review	Mixed Review
Biman Bangladesh Airlines	110	84	45
NOVOAIR	5	4	5
Regent Airways	35	39	23
US-Bangla Airlines	35	5	9

less number compare to positive reviews. From the following Table 5.5, we see that the total number of online reviews are 349 among three class dataset where initially the positive class has 349 reviews and now for all three class, we get the number in common. The model worked on negative and mixed reviews as they are in small number and increase the number of those reviws.

Table 5.5: Sentiment Distribution of Online Reviews for Final overall Dataset

Sentiment	Reviews Count
Positive	349
Negative	349
Mixed	349

Before applying PEGASUS model, the number of online reviews are 350 and 402 accordingly for domestic route and international route. In terms of domestic and international data, we got 519 data and 528 data respectively for domestic and international route after applying PEGASUS model.

We have employed python programming language which is a very extensively used high-level programming language and helpful to our task. The python application was ran in Google Colab with pandas, scikit-learn library. For data preparation and data manipulation, we utilize pandas and NLTK package. Again, for machine learning models and for embracing deep learning models, we utilize a science learning library.

5.4 Performance Evaluation

Predictive accuracy is commonly used as a criterion for evaluating machine learning algorithms. As a result, adopting the appropriate assessment metrics is critical for dealing with the data imbalance problem. Accuracy, precision, recall, and f1-score values are calculated from the confusion matrix for three class samples [43].
The equation Eqn. 5.1 may be used to determine the accuracy.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{5.1}$$

Precision is the accurately anticipated average approximation of each class's labels. Eqn. 5.2 is used to calculate precision figures.

$$\text{Precision} = \frac{TP}{TP + FP} \tag{5.2}$$

The recall value is the weighted average of properly stated points across all classes. This value is defined in Eqn 5.3.

$$\text{Recall} = \frac{\text{TP}}{TP + FN} \tag{5.3}$$

The F_1 measure is specified in Eqn 5.4, and the F_1 value is close to 1 for a good measure.

$$F_1 \ score = \frac{2 * \ \text{Precision} \ * \ \text{Recall}}{\text{Precision} \ + \ \text{Recall}} \tag{5.4}$$

In the following Eqn. 5.1- Eqn 5.3, TP represents true positive, TN refers to true negative, FP and FN means false positive and false negative respectively.

5.4.1 Word Cloud

When it comes to conveying the subject and attributes of a piece of writing, this strategy is often employed. By showing the content in relation to its distribution, the word cloud gives a more accurate picture of the article's properties [9]. The following figure 5.9 displays the outcome of the word cloud in result of 1047 review dataset. Bangladesh Airline Review word cloud analysis shows that excluding 'time', 'good', 'food' are more frequently mentioned for positive sentiment.

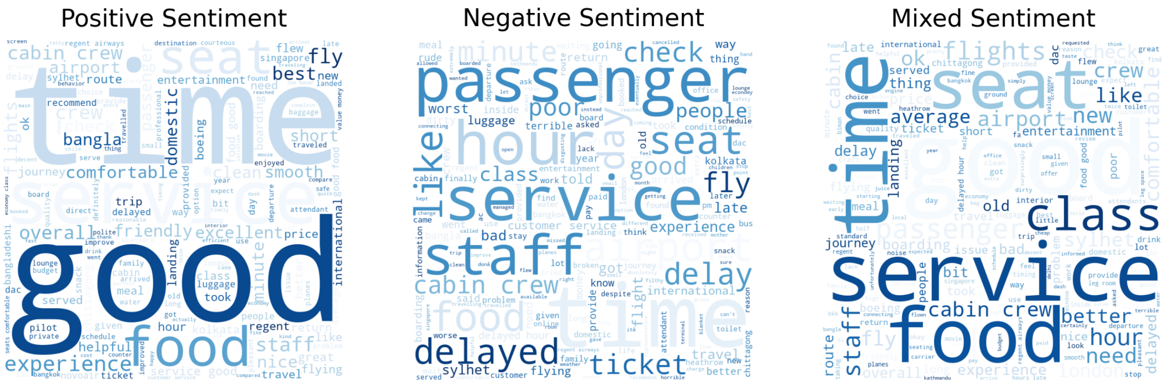

Figure 5.9: The Result of the Word Cloud for three class Dataset in Sentiment Classification

Again, for negative sentiment, we visualize that 'delay', 'poor' are more frequently mentioned. In case of mixed sentiment, we determine that 'service', 'class', 'seat' are mentioned often. This indicates that while picking an airline, customers appreciate the airline's brand, service, in-flight food, and speed.

5.4.2 Results and Comparisons

The data were collected from 752 online reviews for all the route. We categorized the data with domestic and international route and found 350 and 402 online reviews respectively. As the dataset is imbalanced, we apply an oversampling technique PEGASUS model for balancing the dataset. After balancing the dataset, we got overall 1047 online reviews with 349 each class of the dataset named positive, negative and mixed class where the number of domestic route is 519 and for international route, we got 528 review data. Then we conduct a train-test divide in accordance with 75-25 rules, with 75% of the data for training and 25% for testing purposes.

The objective of the variance classifiers should be to minimize or enhance the value of the confusion matrix by reducing the percentages of true positive, true negatives, false positives or false negatives.To correlate our results in this study, we chose the accuracy value as the main success metric regarding ML and DL methods.

To compare our proposed model, we analyse six classifier from machine learning and deep learning criteria such as: Decision Tree (DT), Random Forest (RF), XGBoost (Xgb), Convolutional Neural Network (CNN), Long Short-term Memory (LSTM) and Bidirectional Encoder Representations from Transformers (BERT). Learning from data without being explicitly programmed is what machine learning is all about. Deep learning is based on the human brain's intricate structure of algorithms. Unstructured data, such as documents, photos, and text, may be processed using this method. We want to compare this different algorithm to show the variety of our working procedure and make ensure that our work make impacts in near future in this research field.

Results for Domestic Route

In result section, we compare the six classifiers for domestic route of Bangladesh Airline review data at first. In terms of accuracy calculation, we take precision, recall, and F1-score into account to determine the classifier's overall accuracy. The total classification accuracy result (in percentage) is shown in Table 5.6, along with precision, recall, and f1-score.

From the table, we visualised that deep learning model BERT outperformed all the other machine learning and deep learning classifiers with accuracy of 99%. In terms of positive class, BERT outperforms the other classifier of precision, recall, f1-score evaluation with an accuracy of 98%, 97% and 99% accordingly. CNN got the closest accuracy and for specific precision result, we got 98% which outperforms the other classifier results.

If we consider only machine learning classifiers, we depicts from the Table that XGBoost outperformed other classifiers with an accuracy of 93% while Random Forest is the closest one with an accuracy of 92%. In case of specific classes, we see that for positive class, Random Forest got better result just in case of precision and recall with the result of 95% and 92%. On other hand, XGBoost got the better result most of the cases for negative and mixed class accordingly.

Table 5.6: Experimental Results of Domestic Route Airlines

Classifier	Accuracy	Positive Class			Negative Class			Mixed Class		
		Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score
Decision Tree	74%	76%	70%	73%	76%	86%	80%	70%	65%	67%
Random Forest	92%	95%	89%	92%	89%	95%	92%	93%	93%	93%
XGBoost	93%	91%	91%	91%	95%	93%	94%	93%	95%	94%
CNN	98%	99%	99%	**99%**	99%	99%	98%	97%	96%	98%
LSTM	96%	98%	92%	94%	90%	96%	93%	91%	98%	96%
BERT	**99%**	**99%**	**99%**	98%	**98%**	**99%**	**98%**	**98%**	**99%**	**99%**

We also visualised the result from the figure 5.10 and analysed that the BERT model (green color border line) with an accuracy of 99% outperformed the other models with different color border line overall. From this graphical results, we also depicted that in case of machine learning classifiers, XGBoost with gray border line outperforms the other Random Forest (orange color border line) and Decision Tree(light blue color border line) with an accuracy of 93%.

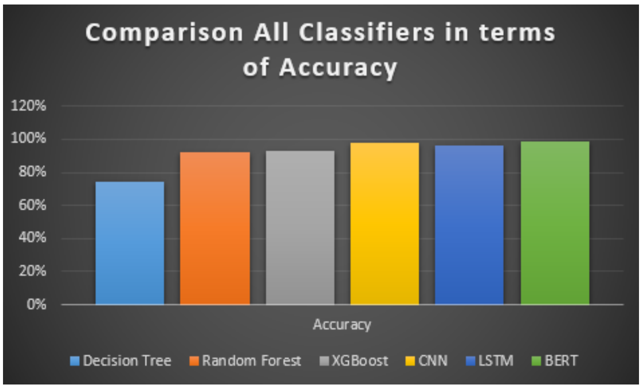

Figure 5.10: Graphical Representation for Domestic Route Classifiers Result

Results for International Route

Now, we compare the six classifiers for international route of Bangladesh Airline review data. In terms of accuracy calculation, we take precision, recall, and F1-score into account to determine the classifier's overall accuracy. The total classification accuracy result (in percentage) is shown in Table 5.7, along with precision, recall, and f1-score.

From the table, we visualised that deep learning model BERT outperformed all the other machine learning and deep learning classifiers with accuracy of 98%. In terms of positive class, BERT outperforms the other classifier of precision, recall evaluation with an accuracy of 99%, 99%. CNN got the closest accuracy in terms of positive class and for specific f1-score, we got 99% which outperforms the other classifier.

If we consider only machine learning classifiers, we depicts from the Table that XGBoost outperformed other classifiers with an accuracy of 91% while Random Forest is the closest one with an accuracy of 90%. In case of specific classes, we see that for mixed class, Random Forest got better result just in case of precision with

Table 5.7: Experimental Results of International Route Airlines

Classifier	Accuracy	Positive Class			Negative Class			Mixed Class		
		Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score
Decision Tree	83%	84%	82%	83%	80%	82%	81%	84%	84%	84%
Random Forest	90%	91%	93%	92%	89%	84%	89%	90%	93%	89%
XGBoost	91%	91%	93%	92%	95%	84%	89%	89%	93%	91%
CNN	97%	97%	96%	96%	95%	97%	97%	**98%**	96%	**98%**
LSTM	95%	92%	94%	95%	93%	97%	91%	93%	97%	97%
BERT	**98%**	**98%**	**97%**	**99%**	**98%**	**97%**	**98%**	97%	**97%**	**98%**

the result of 90%. On other hand, XGBoost got the better result most of the cases for negative and mixed class accordingly.

We also visualised the result from the figure 5.11 and analysed that the BERT model (green color border line) with an accuracy of 98% outperformed the other models with different color border line overall. From this graphical results, we also depicted that in case of machine learning classifiers, XGBoost with gray border line outperforms the other Random Forest (orange color border line) and Decision Tree(light blue color border line) with an accuracy of 91%.

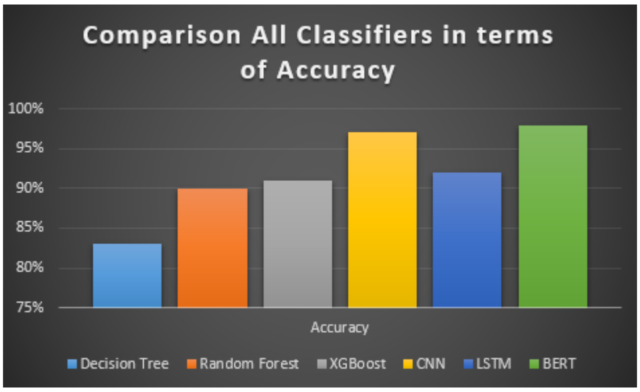

Figure 5.11: Graphical Representation for International Route Classifiers Result

Experimental Results for Overall Dataset and Comparison

Now we focus on our overall dataset considering all routes. Similarly, we compare the six classifiers Bangladesh Airline review data and also compare the results with existing USA airlines Tweets data [45]. In terms of accuracy calculation, we take precision, recall, and F1-score into account to determine the classifier's overall accuracy.

For overall dataset, we consider to visualise the confusion matrix of all the classifiers. For confusion matrix in terms of positive review, negative review and mixed review, we denote the column number, where column 0 denotes negative sentiment, column 1 denotes mixed sentiment and column 2 denotes positive sentiment. In our method, we try to compare the machine learning method and deep learning method at a glance while reviewing confusion matrix.

In case of Decision Tree (DT) and Convolutional Neural Network (CNN) in figure 5.12, for positive sentiment (column 2), 44 reviews correctly classified and other 24 and 15 reviews are missclassified for Decision Tree classifier. Again, for mixed sentiment, 43 reviews are correctly classified and 21 and 28 reviews are missclassified. For negative sentiment, we see that the column 0 denotes that 52 reviews are classified correctly while 20 and 15 reviews are missclassified.

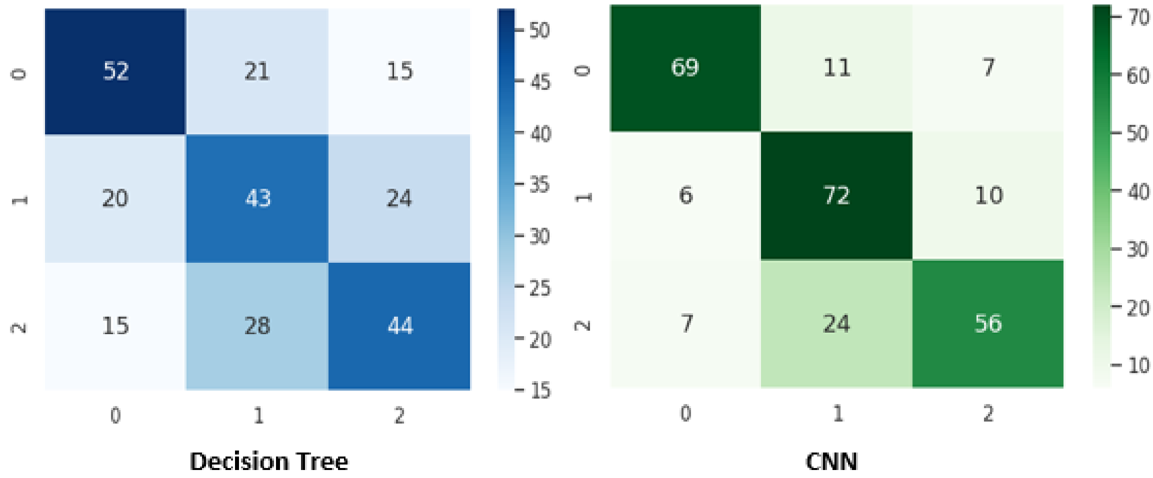

Figure 5.12: Confusion Matrix of DT and CNN

In contrary, for negative sentiment (column 0) of CNN classifier, 69 reviews are correctly classified and 6 and 7 reviews are missclassified. In terms of positive sentiment (column 2) denotes that 56 reviews are correctly classified and the number of missclassification is 10 and 7 reviews. For mixed sentiment (column 1) visualizes that 72 reviews are classified perfectly while missclassifying the other 11 and 24 reviews respectively.

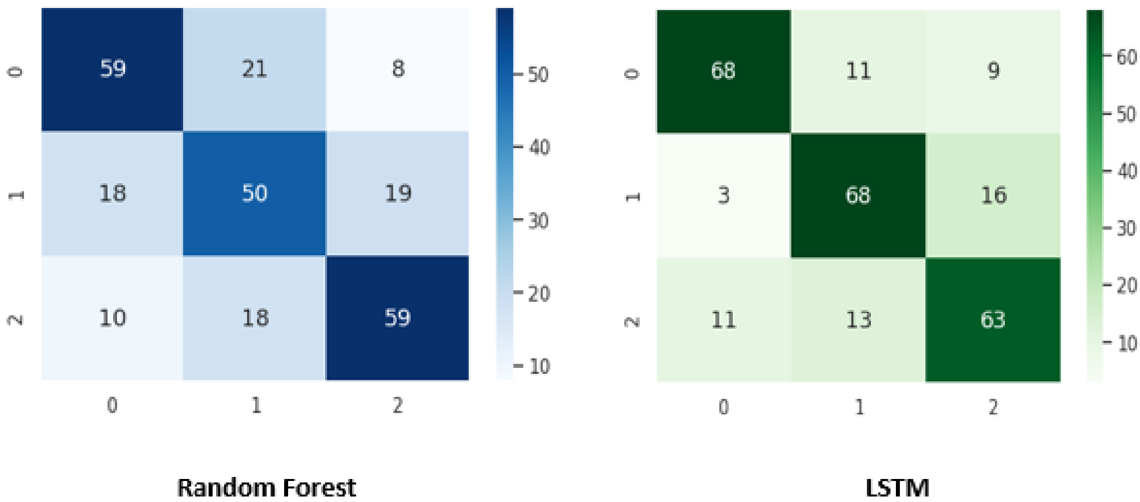

Figure 5.13: Confusion Matrix of RF and LSTM

Again in the following figure 5.13, we see that for Random forest (RF), mixed sentiment are (column 1) correctly classify 50 reviews and are missclassified 21 and 18 reviews respectively. For positive sentiment (column 2), we see that 59 reviews are classified correctly and 8 and 19 reviews are missclassified. Finally, for negative sentiment (column 0), we analysis from the figure that 59 reviews are correctly classified and 18 and 10 reviews are missclassified.

On other hand, for LSTM classifier, positive sentiment (column 2) are correctly classified 63 reviews and missclassified the other 9 and 16 reviews. In case of negative sentiment (column 0), we see that 68 reviews are classified perfectly and other reviews that is 3 and 11 are missclassified. For mixed reviews (column 1), we see that 68 reviews are classified correctly and 11 and 13 reviews are missclassified.

In final confusion matrix of XGBoost (Xgb) and BERT in the following figure 5.14, we analyzed that for XGBoost model, mixed sentiment (column 1) are correctly classified 52 reviews and 21 and 18 reviews are missclassified. In terms of positive sentiment (column 2), we see that 57 reviews are correctly classified and 14 and 9 reviews are classified in wrong way. Finally, for negative sentiment which is column 0 denotes that 58 reviews are correctly classified and 21 and 12 reviews are missclassified.

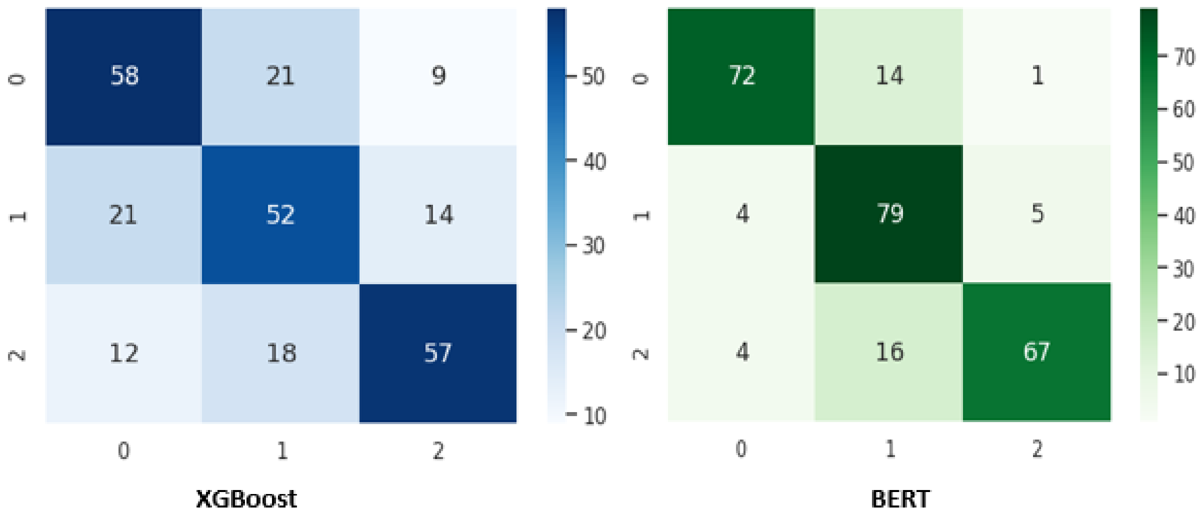

Figure 5.14: Confusion Matrix of XGBoost and BERT

On other hand, for BERT model, negative sentiment (column 0) are missclassified 4 and 4 reviews while correctly classified 72 reviews. For positive sentiment (column 2), we see that 67 reviews are classified correctly and 1 and 5 reviews are missclassified. Finally for mixed sentiment (column 1), we analysed that 79 reviews are classified correctly and 14 and 16 reviews are missclassified.

Regarding accuracy calculation, we consider precision, recall, and F1-score to determine the classifier's overall accuracy. The Table 5.8 depicts The total classification

Table 5.8: Comparison of Classification Results with Existing Data

Datasets	Classifier	Accuracy	Positive Class			Negative Class			Mixed Class		
			Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score
Bangladesh Airlines Review Data	Decision Tree	53%	53%	51%	52%	**60%**	**59%**	**59%**	**47%**	**49%**	**48%**
	Random Forest	64%	69%	68%	68%	**68%**	**67%**	**67%**	**56%**	**57%**	**57%**
	XGBoost	64%	71%	66%	68%	**64%**	**66%**	**65%**	**57%**	**60%**	**58%**
	CNN	**75%**	77%	64%	70%	**84%**	**79%**	**82%**	**67%**	**82%**	**74%**
	LSTM	**76%**	72%	72%	72%	**83%**	**77%**	**80%**	**74%**	**78%**	**76%**
	BERT	83%	**92%**	77%	84%	**90%**	**83%**	**86%**	**72%**	**90%**	**80%**
USA Airlines Tweets Data	Decision Tree	55%	71%	68%	69%	45%	47%	46%	39%	42%	40%
	Random Forest	67%	70%	93%	80%	52%	47%	46%	52%	37%	46%
	XGBoost	69%	71%	93%	81%	55%	32%	41%	65%	22%	33%
	CNN	73%	84%	82%	83%	49%	51%	50%	63%	68%	65%
	LSTM	73%	85%	83%	81%	49%	58%	53%	66%	62%	64%
	BERT	84%	90%	91%	90%	71%	65%	68%	75%	82%	78%

accuracy result (in percentage) is shown, along with precision, recall, and f-score and also compare the result with USA airlines Tweets data to validate our model. From the Table, we showed that deep learning model BERT outperform the other classifier with 83% accuracy in terms of considering Bangladesh airlines review data. In terms of positive class considering Bangladesh airlines review data, BERT outperforms the other classifier of precision, recall and f1-score evaluation. For mixed class, LSTM model performs better in terms of precision value of 74% and BERT outperforms other ML and DL models with recall value of 90% and f1-score value of 80%.

The USA airlines data containing 14640 tweets which is more polished than our dataset. When we compared our dataset with existing USA airlines Tweets data, we visualised that for CNN, LSTM we got better result with an accuracy of 75% and 76% respectively. For specific classes, we depicted that our dataset considering all the classifiers of Bangladesh airlines review data outperformed most of the cases in negative and mixed classes compare to USA airlines Tweets data. For positive class, we see that only precision result of BERT model outperforms the USA airlines data with an accuracy of 92%.

5.5 Performance of Topic Modeling and LIME

Research trends analysts are increasingly turning to topic modeling, a technique that uses observable words and hidden meanings to derive latent themes and extract them for better understanding of the entire pattern. Sentiment analysis, a subset of text mining, is known as opinion mining for its ability to identify positive and negative sentiments expressed in text data [11].

5.5.1 Framework for topic modeling

Generally speaking, subject modeling strategies fall into three categories. The initial step is to gather the data that will be used for topic modeling. As part of the second phase, the unstructured data acquired was preprocessed to make it suitable for topica modelling. Data analysis is the final phase of a project. We already described the framework of this study in chapter 4. The entire structure of topic modeling on online-based reviews, as well as the actual forecast for feelings, is visualised here. This study works on merging dataset of 1095 reviews where 343 sample size from skytrax and previous customized 752 online review dataset are merged. In the following figure 5.15, we see the number of sample size is 1095 in terms of topic modeling and the column size is 2.

	review	target
0	like flights comfortable, strange, panic, angr...	Mixed
1	good flight schedule, comprehensive online boo...	Mixed
2	average airline fly, recently maintaining time...	Mixed
3	airlines maintain time. inform passengers even...	Mixed
4	traveled london bangladesh. biman needs direct...	Mixed
...	...	...
1090	excellent round-trip dac-zyl recently. booked ...	Positive
1091	cox's bazar dhaka non-stop one hour flight bom...	Mixed
1092	dac-cgp. online booking available. arrival dom...	Positive
1093	saidpur dhaka. worse airline ever. late hours...	Negative
1094	jashore dhaka. well-known domestic airline ban...	Mixed

1095 rows × 2 columns

Figure 5.15: Merging of Bangladesh Airline Review Dataset

5.5.2 Frequency Analysis

Frequency analysis was utilized to identify positive and negative themes in the ratings and reviews [29]. Weight of topic (WOT) is defined in Eqn. 5.5 -

$$WOT_{x,y} = \frac{t_{x,y}}{\sum_j t_{v,y}} \tag{5.5}$$

where $WOT_{x,y}$ weight of topic x in dcuments y, $t_{x,y}$ number of times a word in document y is assigned to the topic x. Events and values can be counted on the basis of their frequency. A frequency table is a table in which objects are listed and their frequency is shown. We utilize a frequency table in our research because it is a simple way to explore data and get a sense of how variables are related. One or more category variables may be counted in a frequency table.

In our study, we count the most repetitive word from the cleaning customized dataset of 1095 reviews. The following Table 5.9 lists the most frequently used terms in on-

Table 5.9: The result of frequency analysis

Rank	Word	Frequency	Rank	Word	Frequency	Rank	Word	Frequency
1	airline	726	35	also	147	69	way	99
2	biman	664	36	take	146	70	price	96
3	good	643	37	ticket	143	71	best	95
4	time	621	38	travel	137	72	due	95
5	service	611	39	sylhet	136	73	last	93
6	food	471	40	comfortable	136	74	luggage	92
7	dhaka	442	41	new	132	75	could	91
8	hour	373	42	first	131	76	average	89
9	crew	371	43	day	130	77	got	87
10	bangladesh	370	44	people	129	78	clean	86
11	plane	364	45	late	129	79	lot	86
12	seat	360	46	better	126	80	thing	85
13	staff	319	47	bad	124	81	inflight	84
14	cabin	290	48	old	118	82	nice	83
15	passenger	277	49	boarding	117	83	chittagong	83
16	class	248	50	took	117	84	kolkata	83
17	one	245	51	well	115	85	though	82
18	airport	242	52	london	115	86	hotel	82
19	aircraft	234	53	regent	115	87	year	81
20	delayed	229	54	return	113	88	two	81
21	business	225	55	really	113	89	flew	80
22	would	222	56	meal	111	90	make	80
23	flight	215	57	customer	110	91	money	79
24	like	202	58	check	110	92	trip	79
25	fly	200	59	never	109	93	great	78
26	experience	184	60	flying	108	94	terrible	78
27	get	175	61	route	105	95	provided	77
28	even	166	62	landing	105	96	helpful	77
29	entertainment	153	63	much	104	97	given	76
30	minute	151	64	domestic	104	98	airway	76
31	delay	150	65	international	103	99	short	76
32	need	149	66	friendly	103	100	attendant	75
33	back	147	67	journey	103			
34	air	147	68	poor	100			

line reviews. The first step in our task is to determine the three categories of positive, negative, and mixed. Then, each class's word count is tallied and the stop words are removed. Thus, the top terms for airline services were used to develop positive expression keywords: 'entertainment,' 'new,' 'better,' 'well,' 'much,' 'friendly,' 'money,' 'trip,' and 'provided'. On the other side, the keywords for negative expressions were 'back,' 'need,' 'price,' 'last,' 'though,' and 'terrible'. For mixed reviews, the keywords are 'cabin,' 'like,' 'minute,' 'air,' 'travel,' 'people,' 'check,' 'journey'.

Along with positive and negative expressions, mixed expressions contain several frequent terms. In this frequency study, positive terms exceed negative ones. Again, Bangladesh Airlines' online review has a higher ratio of favorable to negative evaluations. Additionally, various terminology denoting areas were prominent, including 'seat,' 'class,' 'fly,' 'even,' 'first,' 'london,' and 'regent.' As can be observed from the review, this was heavily emphasized due to Bangladesh Airlines' strong importance in this region. Several keywords relating to 'service' were developed for various aspects of each airline, including 'seat,' 'food,' 'crew,' 'cabin,' 'class,' 'staff,' 'meal,' 'check,' 'inflight,' 'flew,' and 'attendant.' The fact that multiple keywords linked to airline services were generated demonstrates that many customers see services as critical criteria.

5.5.3 Word Cloud

This kind of visualization is widely used since it provides for a more natural depiction of the text's subject and features. The word cloud provides a more comprehensible representation of the document's properties by displaying the corpus in proportion to its frequency [9]. The word cloud is a technique for identifying top phrases, having the advantage of quickly presenting major words with keywords that are difficult to see in the table at first sight.

The figure 5.16 illustrates the result of our research's word cloud. The Bangladesh Airline Review's word cloud analysis suggests that favorable attitudes are frequently expressed, including the following: 'nice,' 'comfortable' and 'friendly' as well. Again, the negative emotion we see is more frequently described as 'delaye,' 'paid' and 'left'. In mixed feelings we determined that 'service' is often discussed, 'hour' and 'regent' is often mentioned. This indicates that travelers value the airline's brand, service, food, and speed while selecting an airline.

5.5.4 Topic Modeling

The entire document of the online passenger review employing Bangladesh Airlines was classified by subject in terms of the document's primary subjects, and we discovered that the most relevant issues were separated into six categories. The subject modeling technique enables us to repeat the number of subjects numerous times and pick the best descriptive subset of them. As a result, the subject group identifies the subjects that the text best explains. The following figure 5.17 demonstrates the results of the topic modeling analysis.

The large word count is the most crucial factor in this subject since it is used to

Figure 5.16: The overall result of the word cloud

select the topic name for each of the six graphs. Topic 1 was a sample theme of 10 keywords called 'Flight schedule' that may be explained by terms like 'airline' and 'service' as well as 'time.' Topic 2 has been dubbed 'Food and beverage' with such key-words that might depict the inside of the jet as 'food,' 'seat' and 'plane' The topic 3 was entitled 'Stuff service' with key phrase for the rating of the stuffs, such as 'crew,' 'good,' 'time,'. Topic 4 was titled 'Interior and seat' as a topic name, and included keywords such as 'service', 'staff' etc. The next topic name which is number 5 for the keywords is 'Luggage,'and the keywords are named 'biman,' 'airline,' and 'service'. Topic 5 reflected the very low word count than the others. The last subject 6 was referred to as 'Entertainment,' due to the difference in the word of 'bangladesh,' 'time' and 'hour'. As seen above, concealed concerns were classified into six categories based on entire texts from Bangladesh Airlines Online Review. Bangladesh Airlines expert customers may determine that 'Food and beverage' are the main reasons for travelling. The following things are more likely to impact their shopping intentions as 'Entertainment,' 'Seat,' and 'Staff'.

5.5.5 Sentiment Analysis

Emotions in texts may be categorized or measured and turned into objective data using sentiment analysis [52]. People use language to express their thoughts and feelings. A text-mining strategy that estimates the "attitude contained in text" can be used if a text-mining approach identifies the "objective covered by text" in the theme modeling described earlier. Emotional analysis of a text is similar to topic

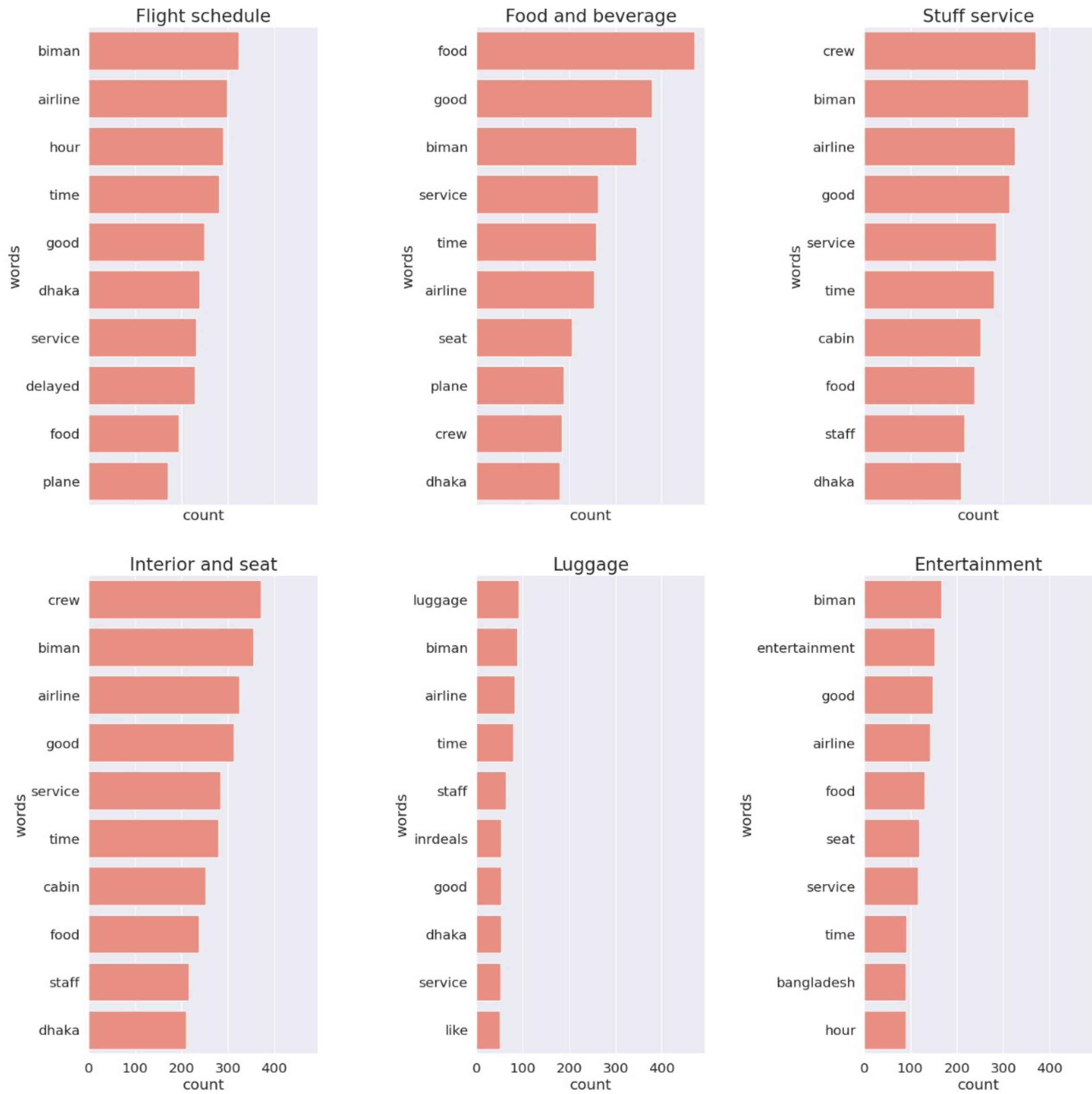

Figure 5.17: The result of topic modeling

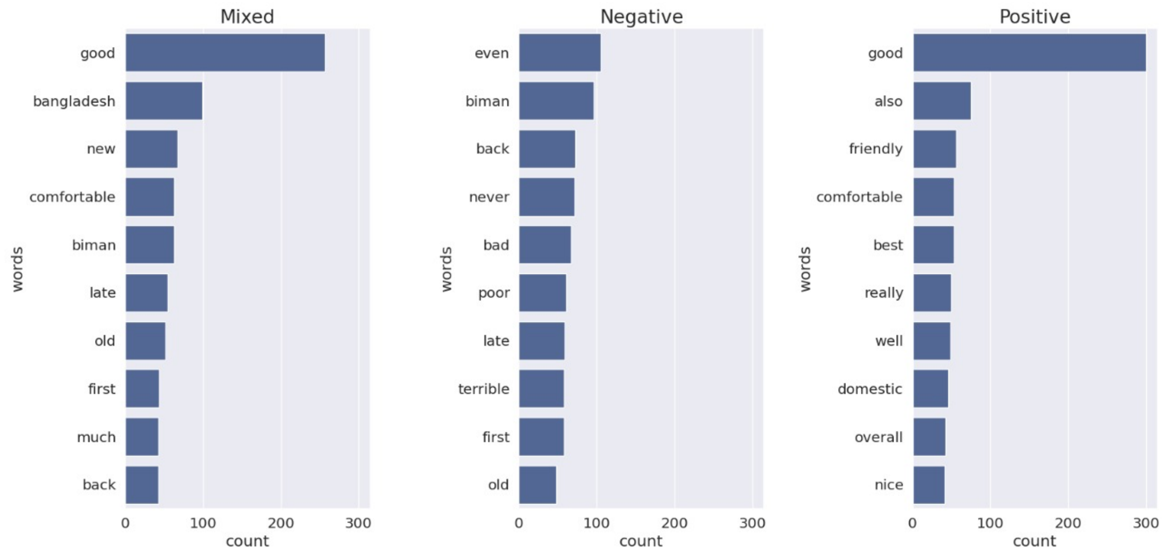

Figure 5.18: The result of sentiment analysis

modeling, which eliminates words that include themes that are expected to be part of its content. The findings revealed that the terms conveying negatives were 'never', 'bad', 'poor', 'late' and 'terrible,' in the following figure 5.18. It can therefore be established that when a problem occurs, a negative emotion occurs when the service is delayed or time is delayed.

Words like 'friendly', 'comfortable,' 'pleasant,' and 'good' were used to express positive feelings. These are words that may be used to express happiness after using them. Good connotes happy feelings when the airline service is excellent and customers love the food flavor as one they expect of the airline. Again, for mixed sentiment, we get the word like 'new,' 'biman', 'old' and 'much,'. It is a term that communicates both positive and negative feelings. When it comes to 'service,' the customer is looking for the best possible experience while the plane is in use. As a result, our customers may regard service as a positive emotion. Since airline passengers expect to receive high-quality meals, the term 'food' has come to signify an enjoyable experience. In future, airlines should emphasize in the training of flight attendants and it may be mentioned.

5.5.6 LIME

LIME builds variations of the prediction we wish to explain, and then gathers the models predictions. Based upon exactly the initial information is closed matches, samples are given weight. It is then used to train a new model that is easier to understand and less complex by employing the weights connected to each variant. It is therefore possible to explain the prediction by means of a local interpretable model [25]. We use graphs to depict LIME's descriptions of feelings as a visual aid. To get the sentiments of positive, negative and mixed online reviews, we used XGBoost model to explain the results.

In the figure 5.19, the green bars represent the positive sentiment. On other hand, blue and orange denotes mixed and negative sentiment in prediction probabilities section.

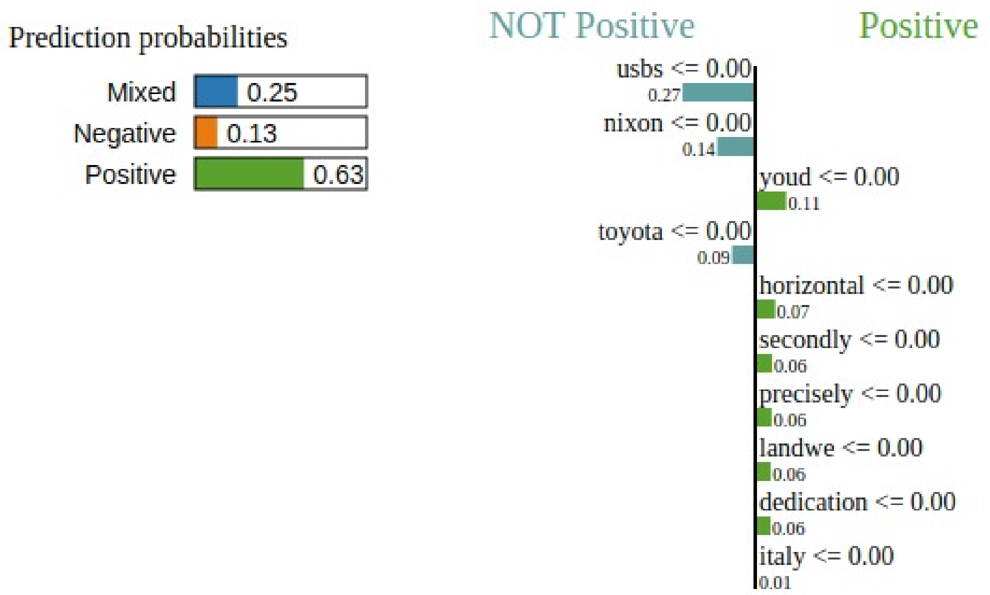

Figure 5.19: Explainable data for positive sentiment

In case of positive sentiment, LIME predicts for ‘horizontal’, ‘secondly’, ‘precisely’, ‘dedication’ word related reviews are positive by the passengers of different airlines. On the other hand, ‘nixon’, ‘toyota’ word related reviews are mixed and negative that is predicted by LIME. The featuring word ‘youd’ from the reviews hold the strongest relevance inside the framework forecast of a content connected to positive sentiment and the featuring word ‘usbs’ from the reviews contain minimal significance suggesting this characteristic works against the group to be regarded non-positive.

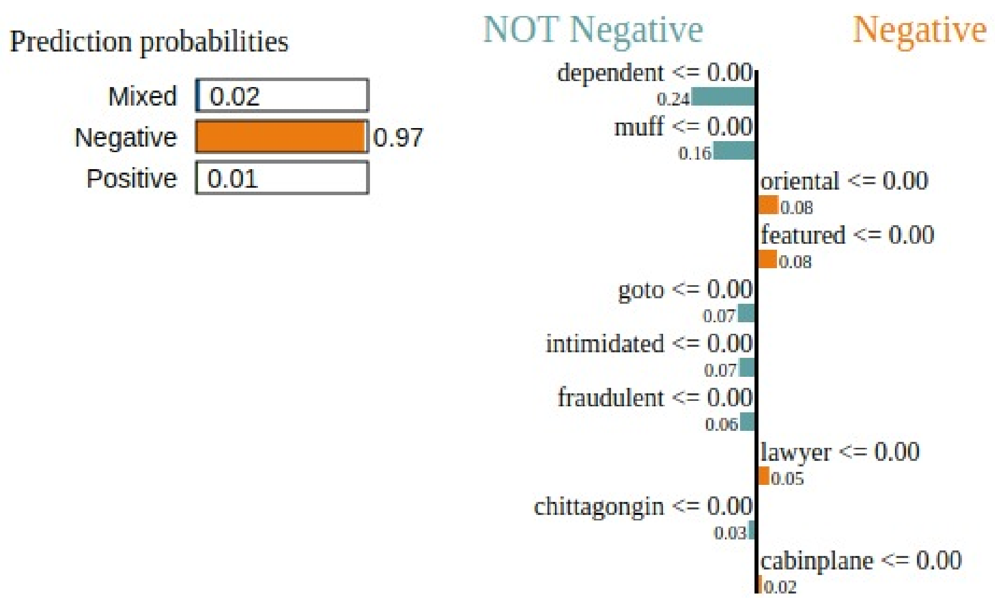

Figure 5.20: Explainable data for negative sentiment

In the figure 5.20, the orange bars represent the negative sentiment. On other hand, both white denotes mixed and positive sentiment in prediction probabilities section. In case of negative sentiment, LIME predicts for 'oriental', 'featured', 'lawyer' word related reviews are negative by the passengers of different airlines. On the other hand, 'dependent', 'goto' word related reviews are mixed and positive that is predicted by LIME. The featuring word 'oriental' from the reviews hold the strongest relevance inside the framework forecast of a content connected to negative sentiment and the featuring word 'dependent' from the reviews contain minimal significance suggesting this characteristic works against the group to be regarded non-negative.

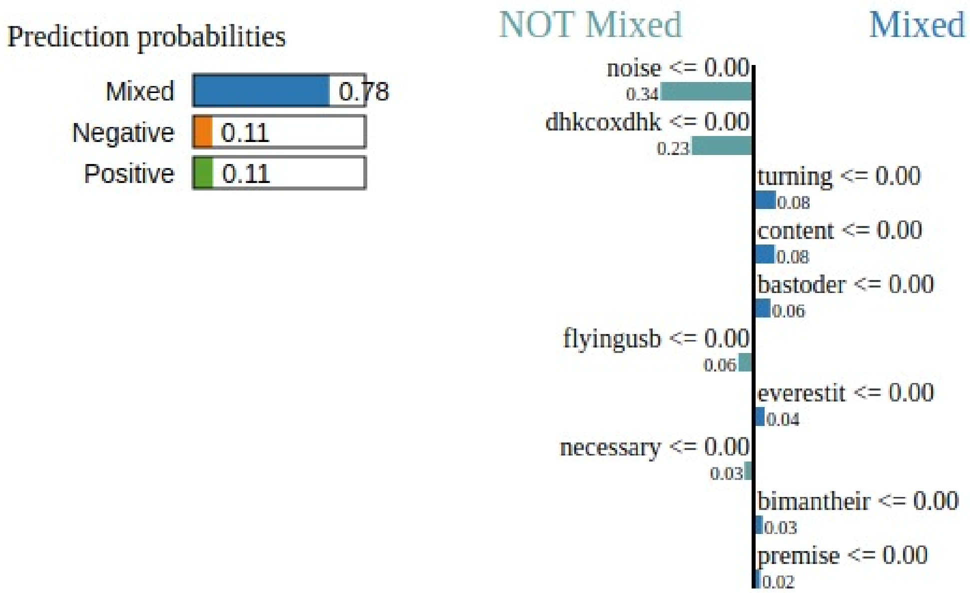

Figure 5.21: Explainable data for mixed sentiment

In the figure 5.21, the blue bars represent the mixed sentiment. On other hand, both orange and green denotes negative and positive sentiment in prediction probabilities section. In case of negative sentiment, LIME predicts for 'turning', 'content', 'premise' word related reviews are mixed by the passengers of different airlines. On the other hand, 'noise', 'necessary' word related reviews are mixed and positive that is predicted by LIME. The featuring word 'turning' from the reviews hold the strongest relevance inside the framework forecast of a content connected to mixed sentiment and the featuring word 'noise' from the reviews contain minimal significance suggesting this characteristic works against the group to be regarded non-mixed. As a result of LIME, the Bangladesh airline review linked postings containing these exact terms have a larger likelihood of being either good, negative, or mixed, according to the predictions.

Chapter 6

Conclusion

6.1 Overview

The entertainment industry has seen a surge in the consumption of digital network as well as online applications as a repository of personal information. These programs have spawned an immense amount of feedback from its users. This study conducted experiments on the Bangladesh Airline Review dataset with six classification methods to predict customer sentiment prediction on online data. Three machine learning algorithms (Decision Tree, Random Forest, XGBoost) and three deep learning algorithms (CNN, LSTM, BERT) were used for prediction. We investigated how the model transform, feature extraction, and the number of classes influence classification outcomes. In dataset preprocessing, we also balanced the unstructured data using the PEGASUS model for better prediction. In terms of comparison, BERT gives the best accuracy of 83% among all those classifiers. The accuracies were determined to compare each categorization technique, and the total sentiment count for all four airlines of Bangladesh was displayed in terms of domestic route, international route and overall route. We comprehend the results acquired from USA airlines Tweets data and demonstrate that our framework is more efficient than the earlier model. We also visualised that the best airlines in terms of domestic route, international route and overall route. The first sentiment classification deployed in 1047 online reviews. Then in second phase, we increase the number of online reviews of 1095 and then apply Topic modeling and LIME to get the topic from the dataset and then train an interpretable LIME model for the sentiments and the construction of explainable sentiments can have a major advantage. Bangladesh Airline Review data was used in this thesis for sentiment classification and topic modeling. When it comes to the Bangladesh Airline Online Review dataset, our research is the first to examine the consumer satisfaction level. In the second chapter, we conduct a study of 52 articles related to sentiment analysis in the airline industry and 6 papers connected to topic modeling. The proposed approach was shown in the chapter 4 of the research. Chapter 5 describes the modified dataset and its outcomes.

6.2 Future Works

This research still has room for improvement in future, as the main limitation is the small amount of online review utilized to train the model. We can create a stronger model and improve categorization accuracy by increasing the number of online reviews. Then, we will compare with other nations result with us when we will increase the number of online reviews regarding the Bangladesh Airlines dataset. Airlines of Bangladesh can utilize the approach outlined in this research to evaluate customer satisfaction.

Bibliography

[1] J. L. Heskett, T. O. Jones, G. W. Loveman, W. E. Sasser, L. A. Schlesinger, *et al.*, "Putting the service-profit chain to work," *Harvard business review*, vol. 72, no. 2, pp. 164–174, 1994.

[2] H. Yu and V. Hatzivassiloglou, "Towards answering opinion questions: Separating facts from opinions and identifying the polarity of opinion sentences," in *Proceedings of the 2003 conference on Empirical methods in natural language processing*, 2003, pp. 129–136.

[3] J. Wiebe, T. Wilson, and C. Cardie, "Annotating expressions of opinions and emotions in language," *Language resources and evaluation*, vol. 39, no. 2, pp. 165–210, 2005.

[4] J.-W. Park, R. Robertson, and C.-L. Wu, "Modelling the impact of airline service quality and marketing variables on passengers' future behavioural intentions," *Transportation Planning and Technology*, vol. 29, no. 5, pp. 359–381, 2006.

[5] A. Esuli and F. Sebastiani, "Sentiwordnet: A high-coverage lexical resource for opinion mining," *Evaluation*, vol. 17, no. 1, p. 26, 2007.

[6] G. C. Saha *et al.*, "Service quality, satisfaction, and behavioural intentions: A study of low-cost airline carriers in thailand," *Managing Service Quality: An International Journal*, 2009.

[7] W. Jen, R. Tu, and T. Lu, "Managing passenger behavioral intention: An integrated framework for service quality, satisfaction, perceived value, and switching barriers," *Transportation*, vol. 38, no. 2, pp. 321–342, 2011.

[8] J. Mikulić and D. Prebežac, "What drives passenger loyalty to traditional and low-cost airlines? a formative partial least squares approach," *Journal of Air Transport Management*, vol. 17, no. 4, pp. 237–240, 2011.

[9] R. Atenstaedt, "Word cloud analysis of the bjgp," *British Journal of General Practice*, vol. 62, no. 596, pp. 148–148, 2012.

[10] Y.-C. Chiou and Y.-H. Chen, "Service quality effects on air passenger intentions: A service chain perspective," *Transportmetrica*, vol. 8, no. 6, pp. 406–426, 2012.

[11] B. Liu, "Sentiment analysis and opinion mining," *Synthesis lectures on human language technologies*, vol. 5, no. 1, pp. 1–167, 2012.

[12] A. Clark, C. Fox, and S. Lappin, *The handbook of computational linguistics and natural language processing*. John Wiley & Sons, 2013.

[13] E. Adeborna and K. Siau, "An approach to sentiment analysis-the case of airline quality rating.," in *PACIS*, 2014, p. 363.

[14] S. Kiritchenko, X. Zhu, and S. M. Mohammad, "Sentiment analysis of short informal texts," *Journal of Artificial Intelligence Research*, vol. 50, pp. 723–762, 2014.

[15] B. Y. Liau and P. P. Tan, "Gaining customer knowledge in low cost airlines through text mining," *Industrial Management & Data Systems*, 2014.

[16] F. Misopoulos, M. Mitic, A. Kapoulas, and C. Karapiperis, "Uncovering customer service experiences with twitter: The case of airline industry," *Management Decision*, 2014.

[17] X. Fang and J. Zhan, "Sentiment analysis using product review data," *Journal of Big Data*, vol. 2, no. 1, pp. 1–14, 2015.

[18] E. Park, S. Lee, S. J. Kwon, and A. P. Del Pobil, "Determinants of behavioral intention to use south korean airline services: Effects of service quality and corporate social responsibility," *Sustainability*, vol. 7, no. 9, pp. 12 106–12 121, 2015.

[19] K. Ravi and V. Ravi, "A survey on opinion mining and sentiment analysis: Tasks, approaches and applications," *Knowledge-based systems*, vol. 89, pp. 14–46, 2015.

[20] I. Yakut, T. Turkoglu, and F. Yakut, "Understanding customers' evaluations through mining airline reviews," *arXiv preprint arXiv:1512.03632*, 2015.

[21] R. K. Bakshi, N. Kaur, R. Kaur, and G. Kaur, "Opinion mining and sentiment analysis," in *2016 3rd international conference on computing for sustainable global development (INDIACom)*, IEEE, 2016, pp. 452–455.

[22] S. Chen, Y. Huang, and W. Huang, "Big data analytics on aviation social media: The case of china southern airlines on sina weibo," in *2016 IEEE Second International Conference on Big Data Computing Service and Applications (BigDataService)*, IEEE, 2016, pp. 152–155.

[23] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 2016, pp. 785–794.

[24] E. Lacic, D. Kowald, and E. Lex, "High enough? explaining and predicting traveler satisfaction using airline reviews," in *Proceedings of the 27th ACM Conference on Hypertext and Social Media*, 2016, pp. 249–254.

[25] M. T. Ribeiro, S. Singh, and C. Guestrin, "" why should i trust you?" explaining the predictions of any classifier," in *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, 2016, pp. 1135–1144.

[26] T. Chumwatana and I. Chuaychoo, "Using social media listening technique for monitoring people's mentions from social media: A case study of thai airline industry," in *2017 2nd International Conference on Knowledge Engineering and Applications (ICKEA)*, IEEE, 2017, pp. 103–106.

[27] D. D. Das, S. Sharma, S. Natani, N. Khare, and B. Singh, "Sentimental analysis for airline twitter data," in *IOP Conference Series: Materials Science and Engineering*, IOP Publishing, vol. 263, 2017, p. 042 067.

[28] H. Hakh, I. Aljarah, and B. Al-Shboul, "Online social media-based sentiment analysis for us airline companies," *New Trends in Information Technology*, p. 176, 2017.

[29] H. Hao, K. Zhang, W. Wang, and G. Gao, "A tale of two countries: International comparison of online doctor reviews between china and the united states," *International Journal of Medical Informatics*, vol. 99, pp. 37–44, 2017.

[30] G. Hu, P. Bhargava, S. Fuhrmann, S. Ellinger, and N. Spasojevic, "Analyzing users' sentiment towards popular consumer industries and brands on twitter," in *2017 IEEE International Conference on Data Mining Workshops (ICDMW)*, IEEE, 2017, pp. 381–388.

[31] S. Joshi, "Aspect based sentiment analysis for united states of america airlines," Ph.D. dissertation, Dublin, National College of Ireland, 2017.

[32] M. S. Seyfioğlu and M. U. Demirezen, "A hierarchical approach for sentiment analysis and categorization of turkish written customer relationship management data," in *2017 Federated Conference on Computer Science and Information Systems (FedCSIS)*, IEEE, 2017, pp. 361–365.

[33] S. Shi, M. Zhao, J. Guan, Y. Li, and H. Huang, "A hierarchical lstm model with multiple features for sentiment analysis of sina weibo texts," in *2017 International Conference on Asian Language Processing (IALP)*, IEEE, 2017, pp. 379–382.

[34] L. Sun and Y. Yin, "Discovering themes and trends in transportation research using topic modeling," *Transportation Research Part C: Emerging Technologies*, vol. 77, pp. 49–66, 2017.

[35] A. Yenter and A. Verma, "Deep cnn-lstm with combined kernels from multiple branches for imdb review sentiment analysis," in *2017 IEEE 8th Annual Ubiquitous Computing, Electronics and Mobile Communication Conference (UEMCON)*, IEEE, 2017, pp. 540–546.

[36] M. Adarsh and P. Ravikumar, "An effective method of predicting the polarity of airline tweets using sentimental analysis," in *2018 4th International Conference on Electrical Energy Systems (ICEES)*, IEEE, 2018, pp. 676–679.

[37] A. M. Alayba, V. Palade, M. England, and R. Iqbal, "A combined cnn and lstm model for arabic sentiment analysis," in *International cross-domain conference for machine learning and knowledge extraction*, Springer, 2018, pp. 179–191.

[38] M. M. Ashi, M. A. Siddiqui, and F. Nadeem, "Pre-trained word embeddings for arabic aspect-based sentiment analysis of airline tweets," in *International Conference on Advanced Intelligent Systems and Informatics*, Springer, 2018, pp. 241–251.

[39] A. Bayhaqy, S. Sfenrianto, K. Nainggolan, and E. R. Kaburuan, "Sentiment analysis about e-commerce from tweets using decision tree, k-nearest neighbor, and naıve bayes," in *2018 international conference on orange technologies (ICOT)*, IEEE, 2018, pp. 1–6.

[40] Y. Guo, W. Li, C. Jin, Y. Duan, and S. Wu, "An integrated neural model for sentence classification," in *2018 Chinese Control And Decision Conference (CCDC)*, IEEE, 2018, pp. 6268–6273.

[41] A. Humphreys and R. J.-H. Wang, "Automated text analysis for consumer research," *Journal of Consumer Research*, vol. 44, no. 6, pp. 1274–1306, 2018.

[42] R. Khan and S. Urolagin, "Airline sentiment visualization, consumer loyalty measurement and prediction using twitter data," *International Journal of Advanced Computer Science and Applications*, vol. 9, no. 6, pp. 380–388, 2018.

[43] M. H. Popel, K. M. Hasib, S. Ahsan Habib, and F. Muhammad Shah, "A hybrid under-sampling method (husboost) to classify imbalanced data," in *2018 21st International Conference of Computer and Information Technology (ICCIT)*, 2018, pp. 1–7. DOI: 10.1109/ICCITECHN.2018.8631915.

[44] A. Punel and A. Ermagun, "Using twitter network to detect market segments in the airline industry," *Journal of Air Transport Management*, vol. 73, pp. 67–76, 2018.

[45] A. Rane and A. Kumar, "Sentiment classification system of twitter data for us airline service analysis," in *2018 IEEE 42nd Annual Computer Software and Applications Conference (COMPSAC)*, IEEE, vol. 1, 2018, pp. 769–773.

[46] M. T. AL-Sharuee, F. Liu, and M. Pratama, "Sentiment analysis: An automatic contextual analysis and ensemble clustering approach and comparison," *Data & Knowledge Engineering*, vol. 115, pp. 194–213, 2018.

[47] 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.

[48] F. Sternberg, K. H. Pedersen, N. K. Ryelund, R. R. Mukkamala, and R. Vatrapu, "Analysing customer engagement of turkish airlines using big social data," in *2018 IEEE international congress on big data (BigData Congress)*, IEEE, 2018, pp. 74–81.

[49] B. Sun, F. Tian, and L. Liang, "Tibetan micro-blog sentiment analysis based on mixed deep learning," in *2018 International Conference on Audio, Language and Image Processing (ICALIP)*, IEEE, 2018, pp. 109–112.

[50] P. Tiwari, P. Yadav, S. Kumar, B. K. Mishra, G. N. Nguyen, S. P. Gochhayat, J. Singhk, and M. Prasad, "Sentiment analysis for airlines services based on twitter dataset," *Social Network Analytics: Computational Research Methods and Techniques*, p. 149, 2018.

[51] Z. Z. Wint, Y. Manabe, and M. Aritsugi, "Deep learning based sentiment classification in social network services datasets," in *2018 IEEE international conference on big data, cloud computing, data science & engineering (BCD)*, IEEE, 2018, pp. 91–96.

[52] C. Zucco, H. Liang, G. Di Fatta, and M. Cannataro, "Explainable sentiment analysis with applications in medicine," in *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, IEEE, 2018, pp. 1740–1747.

[53] M. M. Anitsal, I. Anitsal, and S. Anitsal, "Is your business sustainable? a sentiment analysis of air passengers of top 10 us-based airlines," *Journal of Global Scholars of Marketing Science*, vol. 29, no. 1, pp. 25–41, 2019.

[54] F. Hemmatian and M. K. Sohrabi, "A survey on classification techniques for opinion mining and sentiment analysis," *Artificial Intelligence Review*, vol. 52, no. 3, pp. 1495–1545, 2019.

[55] N. Korfiatis, P. Stamolampros, P. Kourouthanassis, and V. Sagiadinos, "Measuring service quality from unstructured data: A topic modeling application on airline passengers' online reviews," *Expert Systems with Applications*, vol. 116, pp. 472–486, 2019.

[56] S. Kumar and M. Zymbler, "A machine learning approach to analyze customer satisfaction from airline tweets," *Journal of Big Data*, vol. 6, no. 1, pp. 1–16, 2019.

[57] R. Monika, S. Deivalakshmi, and B. Janet, "Sentiment analysis of us airlines tweets using lstm/rnn," in *2019 IEEE 9th International Conference on Advanced Computing (IACC)*, 2019, pp. 92–95. DOI: 10.1109/IACC48062.2019.8971592.

[58] M. Munikar, S. Shakya, and A. Shrestha, "Fine-grained sentiment classification using bert," in *2019 Artificial Intelligence for Transforming Business and Society (AITB)*, IEEE, vol. 1, 2019, pp. 1–5.

[59] U. Naseem, S. K. Khan, I. Razzak, and I. A. Hameed, "Hybrid words representation for airlines sentiment analysis," in *Australasian Joint Conference on Artificial Intelligence*, Springer, 2019, pp. 381–392.

[60] E. Prabhakar, M. Santhosh, A. H. Krishnan, T. Kumar, and R. Sudhakar, "Sentiment analysis of us airline twitter data using new adaboost approach," *International Journal of Engineering Research & Technology (IJERT)*, vol. 7, no. 1, pp. 1–6, 2019.

[61] F. Rustam, I. Ashraf, A. Mehmood, S. Ullah, and G. S. Choi, "Tweets classification on the base of sentiments for us airline companies," *Entropy*, vol. 21, no. 11, p. 1078, 2019.

[62] E. Sezgen, K. J. Mason, and R. Mayer, "Voice of airline passenger: A text mining approach to understand customer satisfaction," *Journal of Air Transport Management*, vol. 77, pp. 65–74, 2019.

[63] Y. Tao, F. Zhang, C. Shi, and Y. Chen, "Social media data-based sentiment analysis of tourists' air quality perceptions," *Sustainability*, vol. 11, no. 18, p. 5070, 2019.

[64] X. Tian, W. He, C. Tang, L. Li, H. Xu, and D. Selover, "A new approach of social media analytics to predict service quality: Evidence from the airline industry," *Journal of Enterprise Information Management*, 2019.

[65] H. Utama, "Sentiment analysis in airline tweets using mutual information for feature selection," in *2019 4th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE)*, IEEE, 2019, pp. 295–300.

[66] J. Zheng and L. Zheng, "A hybrid bidirectional recurrent convolutional neural network attention-based model for text classification," *IEEE Access*, vol. 7, pp. 106 673–106 685, 2019.

[67] U. Bezek and P. Güneş, "Analysis of airline tweets by using machine learning methods," *International Journal of Engineering Research and Applications*, vol. 10, pp. 42–45, 2020.

[68] K. M. Hasib, M. Iqbal, F. M. Shah, J. A. Mahmud, M. H. Popel, M. Showrov, I. Hossain, S. Ahmed, O. Rahman, *et al.*, "A survey of methods for managing the classification and solution of data imbalance problem," *arXiv preprint arXiv:2012.11870*, 2020.

[69] 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*, IEEE, 2020, pp. 1–6.

[70] S. Kim, H. Park, and J. Lee, "Word2vec-based latent semantic analysis (w2v-lsa) for topic modeling: A study on blockchain technology trend analysis," *Expert Systems with Applications*, vol. 152, p. 113 401, 2020.

[71] J. Lim and H. C. Lee, "Comparisons of service quality perceptions between full service carriers and low cost carriers in airline travel," *Current issues in Tourism*, vol. 23, no. 10, pp. 1261–1276, 2020.

[72] 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. 101 760, 2020.

[73] S. Park, J.-S. Lee, and J. L. Nicolau, "Understanding the dynamics of the quality of airline service attributes: Satisfiers and dissatisfiers," *Tourism Management*, vol. 81, p. 104 163, 2020.

[74] H. Shahbaznezhad and M. Rashidirad, "Exploring firms' fan page behavior and users' participation: Evidence from airline industry on twitter," *Journal of Strategic Marketing*, pp. 1–22, 2020.

[75] A. Sharma, S. Park, and J. L. Nicolau, "Testing loss aversion and diminishing sensitivity in review sentiment," *Tourism Management*, vol. 77, p. 104 020, 2020.

[76] C. Song, J. Guo, and J. Zhuang, "Analyzing passengers' emotions following flight delays-a 2011–2019 case study on skytrax comments," *Journal of Air Transport Management*, vol. 89, p. 101 903, 2020.

[77] I. Sutherland, Y. Sim, S. K. Lee, J. Byun, and K. Kiatkawsin, "Topic modeling of online accommodation reviews via latent dirichlet allocation," *Sustainability*, vol. 12, no. 5, p. 1821, 2020.

[78] A. Yadav and D. K. Vishwakarma, "Sentiment analysis using deep learning architectures: A review," *Artificial Intelligence Review*, vol. 53, no. 6, pp. 4335–4385, 2020.

[79] R. Yadu and R. Shukla, "Sentimental classification method of twitter data for indian air asia services analysis," EasyChair, Tech. Rep., 2020.

[80] A. Zaki Ahmed and M. Rodrıguez-Dıaz, "Significant labels in sentiment analysis of online customer reviews of airlines," *Sustainability*, vol. 12, no. 20, p. 8683, 2020.

[81] J. Zhang, Y. Zhao, M. Saleh, and P. Liu, "Pegasus: Pre-training with extracted gap-sentences for abstractive summarization," in *International Conference on Machine Learning*, PMLR, 2020, pp. 11 328–11 339.

[82] H. Zheng, "Analysis of airline sentiment data: Does the social media image reflect real performance?," pp. 1–6, 2020.

[83] N. Deraman, A. Buja, K. Samah, M. Jono, M. Isa, and S. Saad, "A social media mining using topic modeling and sentiment analysis on tourism in malaysia during covid19," in *IOP Conference Series: Earth and Environmental Science*, IOP Publishing, vol. 704, 2021, p. 012 020.

[84] P. Ghosh, S. Azam, K. M. Hasib, A. Karim, M. Jonkman, and A. Anwar, "A performance based study on deep learning algorithms in the effective prediction of breast cancer," in *2021 International Joint Conference on Neural Networks (IJCNN)*, IEEE, 2021, pp. 1–8.

[85] K. M. Hasib, M. A. Habib, N. A. Towhid, and M. I. H. Showrov, "A novel deep learning based sentiment analysis of twitter data for us airline service," in *2021 International Conference on Information and Communication Technology for Sustainable Development (ICICT4SD)*, IEEE, 2021, pp. 450–455.

[86] K. M. Hasib, N. A. Towhid, and M. G. R. Alam, "Topic modeling and sentiment analysis using online reviews for bangladesh airlines," in *2021 IEEE 12th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*, 2021, pp. 0428–0434. DOI: 10.1109/IEMCON53756.2021.9623155.

[87] P. K. Jain and R. Pamula, "Content-based airline recommendation prediction using machine learning techniques," in *Machine Learning Algorithms for Industrial Applications*, Springer, 2021, pp. 185–194.

[88] P. K. Jain, V. Saravanan, and R. Pamula, "A hybrid cnn-lstm: A deep learning approach for consumer sentiment analysis using qualitative user-generated contents," *Transactions on Asian and Low-Resource Language Information Processing*, vol. 20, no. 5, pp. 1–15, 2021.

[89] P. Kaladevi and K. Thyagarajah, "Retracted article: Integrated cnn-and lstm-dnn-based sentiment analysis over big social data for opinion mining," *Behaviour & Information Technology*, vol. 40, no. 9, pp. XI–XIX, 2021.

[90] S. Kiliç and T. O. Çadirci, "An evaluation of airport service experience: An identification of service improvement opportunities based on topic modeling and sentiment analysis," *Research in Transportation Business & Management*, p. 100 744, 2021.

[91] M. V. Kumari *et al.*, "Collaborative classification approach for airline tweets using sentiment analysis," *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, vol. 12, no. 3, pp. 3597–3603, 2021.

[92] H.-J. Kwon, H.-J. Ban, J.-K. Jun, and H.-S. Kim, "Topic modeling and sentiment analysis of online review for airlines," *Information*, vol. 12, no. 2, p. 78, 2021.

[93] S. Piccinelli, S. Moro, and P. Rita, "Air-travelers' concerns emerging from online comments during the covid-19 outbreak," *Tourism Management*, vol. 85, p. 104 313, 2021.

[94] R. Al-Qahtani and P. N. bint Abdulrahman, "Predict sentiment of airline tweets using ml models," EasyChair, Tech. Rep., 2021.

[95] S. Sakib, K. M. Hasib, I. K. Tasawar, A. K. Tanzeem, M. F. Arefin, S. Islam, and M. S. Alam, "A data-driven hybrid optimization based deep network model for short-term residential load forecasting," in *2021 IEEE 12th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*, 2021, pp. 0187–0193. DOI: 10.1109/IEMCON53756.2021.9623080.

[96] M. I. H. Showrov, V. K. Dubey, K. M. Hasib, and M. A. Shameem, "News classification from microblogging dataset using supervised learning," in *2021 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS)*, IEEE, 2021, pp. 53–57.

[97] M. Tusar, T. H. Khan, M. Islam, *et al.*, "A comparative study of sentiment analysis using nlp and different machine learning techniques on us airline twitter data," *arXiv preprint arXiv:2110.00859*, 2021.

[98] K. Verma and B. Davis, "Implicit aspect-based opinion mining and analysis of airline industry based on user-generated reviews," *SN Computer Science*, vol. 2, no. 4, pp. 1–9, 2021.

[99] F. Javed Mehedi Shamrat, R. Ranjan, K. M. Hasib, A. Yadav, and A. H. Siddique, "Performance evaluation among id3, c4. 5, and cart decision tree algorithm," in *Pervasive Computing and Social Networking*, Springer, Singapore, 2022, pp. 127–142.