T.C.

BAHCESEHIR UNIVERSITY

GRADUATE SCHOOL OF ENGINEERING

THE DEPARTMENT OF ARTIFICIAL INTELLIGENCE

RECOMMENDING ANCILLARY PRODUCTS IN AVIATION INDUSTRY: A COMPARATIVE STUDY ON RECOMMENDER SYSTEMS USING ONLINE CUSTOMER REVIEWS

MASTER'S THESIS

YAVUZ SELIM EMIR

ISTANBUL 2023

T.C.

BAHCESEHIR UNIVERSITY

GRADUATE SCHOOL OF ENGINEERING

THE DEPARTMENT OF ARTIFICIAL INTELLIGENCE

RECOMMENDING ANCILLARY PRODUCTS IN AVIATION INDUSTRY: A COMPARATIVE STUDY ON RECOMMENDER SYSTEMS USING ONLINE CUSTOMER REVIEWS OF AIRLINE COMPANIES

MASTER'S THESIS

YAVUZ SELIM EMIR

THESIS ADVISOR
ASSIST. PROF. TAMER UÇAR

ISTANBUL 2023

T.C.

BAHÇEŞEHİR UNIVERSITY

GRADUATE SCHOOL

...../...../.....

MASTER THESIS APPROVAL FORM

Program Name:	Artificial Intelligence Graduate Program
Student's Name and Surname:	Yavuz Selim Emir
Name of The Thesis:	Recommending Ancillary Products in Aviation Industry: A Comparative Study on Recommender Systems Using Online Customer Reviews of Airline Companies
Thesis Defense Date	08.09.2023

This thesis has been approved by the Graduate School which has fulfilled the necessary conditions as Master thesis.

……

Institute Director

This thesis was read by us, quality, and content as a Master's thesis has been seen and accepted as sufficient.

	Title/Name	**Signature**
Thesis Advisor's	Assist. Prof. Tamer UÇAR	
Member's	Assist. Prof. İsmail DURU	
Member's	Assist. Prof. Selami BAĞRIYANIK	

I hereby declare that all information in this document has been obtained and presentedin accordance with academic rules and ethical conduct. I also declare that, as requiredby these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Surname : Yavuz Selim Emir

Signature :

ABSTRACT

RECOMMENDING ANCILLARY PRODUCTS IN AVIATION INDUSTRY: A COMPARATIVE STUDY ON RECOMMENDER SYSTEMS USING ONLINE CUSTOMER REVIEWS

Artificial Intelligence Master's Program

August 2023, 47 pages

Increasing competition in the aviation industry forces airline companies to find new ways to increase their profitability. Airline companies try to do that by offering ancillary products beside their main service. However, there is the problem of offering a suitable product to the customer who really needs it. Recent improvements in the recommender systems area, especially in the e-commerce industry, raise the question of whether those systems are efficient for recommending ancillary products in aviation industry. In this study we aim to build a recommender system for ancillary products for the airline industry using online customer reviews. Customer reviews from various web sites were separated by their topics using BERTOPIC topic modelling algorithm. Expert labeled customer reviews were fed into the algorithms to build recommender system. The aim of the study is to compare recommender systems using different machine learning algorithms. As the result of the study Neural networks gave the highest accuracy results of 0.85.

Keywords: Recommender systems, Machine Learning, Deep Learning, SVM, Random Forest

ÖZ

HAVACILIKTA EK HİZMETLERIN ÖNERİLMESI: İNTERNET ORTAMINDAKİ MÜŞTERİ YORUMLARI KULLANILARAK ÖNERİ SİSTEMLERİ İLE İLGİLİ KARŞILASTIRMALI ÇALIŞMA

Emir, Yavuz Selim

Yapay Zekâ Yüksek Lisans Programı

Tez Danışmanı:

ASIST. PROF. TAMER UÇAR

Ağustos 2023, 47 sayfa

Havacılıkta artan rekabet havayolu firmalarını yeni yöntemler bulmaya zorlamaktadır. Havayolu firmaları verdikleri esas hizmetin yanında ek hizmetler vererek karlılıklarını artırmaya çalışmaktadırlar. Ancak burada doğru müşteriye doğru ek hizmeti önerme sorunu çıkmaktadır. Son yıllarda özellikle e-ticaret alanında gelişen öneri sistemleri havacılıkta yaygın bir şekilde kullanılmamaktadır. Bu çalışmada internetten alınan havayolları kullanıcı yorumları incelenerek ek hizmetler için öneri sistemleri oluşturulmaya çalışılmıştır. Çeşitli web sitelerinden toplanılan kullanıcı yorumları BERTOPIC isimli konu modelleme yöntemi kullanılarak konulara ayrılmıştır. Uzman tarafından etiketlenen kullanıcı yorumları öneri sistemine verilerek model oluşturulmuştur. Çalışmanın amacı farklı makine öğrenmesi metotlarını kullanarak, öneri sistemlerinin karşılaştırılmasıdır. Sonuç olarak yapay sinir ağı kullanılan yöntem ile 0.85 olan doğruluk sağlandı.

Anahtar Kelimeler: Öneri sistemleri, Derin Öğrenme, SVM, Random Forest, Makine Öğrenmesi

ACKNOWLEDGEMENTS

I wish to express my deepest gratitude to my supervisor Dr. Tamer Uçar for his guidance, advice, criticism, encouragements, and insight throughout the research.

I am extremely thankful to my friends, my family and my fiancé for their support and understanding during my study.

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

LIST OF ABBREVIATIONS

NLP	Natural Language Processing
TF-IDF	Term Frequency - Inverse Document Frequency
SVM	Support Vector Machines
IATA	The International Air Transport Association
NDC	New Distribution Capability
GBDT	Gradient Boosted Decision Tree
KNN	K-nearest Neighbors
BERT	Bidirectional Encoder Representations from Transformers
CNN	Convolutional Neural Network
LSTM	Long Short-Term Memory
RBF	Radial Basis Function
ReLU	Rectified Linear Unit
ANN	Artificial Neural Network
UMAP	Uniform Manifold Approximation and Projection
HDBSCAN	Hierarchical Density-Based Spatial Clustering of
Applications	with Noise
t-SNE	T-Distributed Stochastic Neighbor Embedding
XGBOOST	Extreme Gradient Boosting
NLTK	Natural Language Toolkit
RMSE	Root Mean Square Error
MSE	Mean Square Error
MAE	Mean Absolute Error
GRU	Gated Recurrent Unit

Chapter 1

Introduction

The airline industry is a highly competitive and dynamic market where airlines need to continuously improve their service quality to attract and retain customers. With a wide spectrum of options available for customers to choose from, it is obligatory for airlines to provide personalized and relevant recommendations to their customers. One effective approach to achieve this goal would be the use of recommender systems, which have been gaining attention in recent years due to their ability to analyze customer data and provide suggestions that reflect individual preferences, needs, and feedback.

Recommender systems are software tools that are widely used in various industries, including e-commerce, music, and movie streaming services. Increasing numbers of products and services make it harder for customers to decide what to buy. Even though customers made up their mind before going to a service provider it is still a burden to find products. It is a general practice in these industries to build personalized recommender systems using various machine learning approaches analyzing large amounts of data. Recommender systems shortens the decision process which takes a significant amount of time when the customer is not sure about the product they want, or they need.

In the airline industry, the use of recommender systems can be particularly useful for improving customer satisfaction and loyalty. By analyzing customer data, such as past purchases, flight destinations, and feedback, recommender systems can provide suggestions that are made for the individual needs and preferences of the customer. Analyzing customer behavior can give important insights to act on when looking for points to improve. If a customer expresses their dissatisfaction about a specific service of an airline company via an online review, it could be an opportunity for the company to improve its service.

The popularization of social media and online review platforms has generated a vast amount of customer feedback and ratings for airlines and their services.

Analyzing the feedback gives a unique opportunity to gain valuable information about the customer behavior and expectations, experiences, and their attitude towards the company that can be used to enhance the accuracy and effectiveness of recommender systems. However, the large volume and complexity of this data poses significant challenges for the design and implementation of such systems. Processing that amounts of data needs large processing power. Hence more effective models that require less processing power come in handy.

Despite these challenges, numerous studies have demonstrated the potential of using recommender systems with online customer reviews in the airline industry. For example, Aciar et al. (2007) proposed a recommendation model based on customer reviews that showed significant improvement in the accuracy of recommendations. Zhang et al. (2022) proposed a personalized recommendation system that considers customer preferences and feedback. Heidari et al. (2020) used a hybrid approach that combined customer reviews and flight itinerary data to provide personalized recommendations.

We aim to investigate the use of recommender systems recommending airline ancillary services in this study. We will analyze online customer reviews to recommend ancillary services to customers better. We will review the relevant literature and explore the challenges and opportunities presented by this approach. We will then propose a methodology that employs natural language processing (NLP) techniques to analyze customer reviews and to recommend an ancillary service. We will integrate different machine learning algorithms into our proposed method to compare their performance. Our analysis will focus on the following research questions:

- How can different recommendation algorithms be applied to improve the accuracy of recommendations using customer reviews?
- How effective are various machine learning algorithms perform when used for recommendation systems?
- What are the potential benefits and limitations of using recommender systems with online customer reviews in the airline industry?

This study will provide valuable information about the implementation of

recommender systems using online customer reviews in the airline industry. It will also contribute to the development of practical recommendations for airlines to improve their ancillary product sales by showing potential usage of recommender systems.

This paper consists of five chapters which are introduction, literature review, methodology, findings, discussion, and conclusion. There is a reference section at the end. A systematic literature review was conducted in the next chapter. The Methodology chapter gives information about the methods used in this study and illustrates the proposed method. Findings chapter shows the results measured by metrics which are defined in the methodology chapter. In the last chapter a brief discussion was made about the results gathered building the models using different machine learning algorithms.

Chapter 2

Literature Review

A literature was conducted after the problem definition. Recommender systems in general and their usage was searched. Then the search was narrowed down to recommender systems in the airline industry. The studies where online customer reviews had been analyzed to better understand the customer needs and opinions were searched. Recent studies in the area were in focus especially after the year 2015, however for the foundational definitions search went back as far as 1988. The main keywords used in this search were *recommender systems*, *recommender systems in airline industry*, *ancillary products*, *online customer review analysis for airline industry*, *airline industry ancillary product recommendations*.

Recommender systems have become increasingly popular in recent years, as they provide a valuable tool for helping customers navigate the vast and complex landscape of products and services available online. It might get too complicated to choose the desired product from a myriad of products when looking without a specific description. That is where recommender systems come in handy. Their benefit for the platforms was proven increasing customer loyalty and driving sales. (Keiningham et al. 2007) The increase in the number of online e-commerce and social platforms caused a hike in user generated review data. Such reviews give the opportunity of building recommender systems to the platform owners to provide better user experiences. (Lu et al. 2018).

Recommender systems are algorithms that use data on past behavior preferences, and feedback of customers to make personalized recommendations. These systems have been widely studied in various industries and have been found to be effective in providing personalized recommendations that can improve customer satisfaction and drive business. Online e-commerce, social media, streaming, tourism, and a lot of different platforms use recommender systems to build better user experiences. The purpose of building recommender systems is to assist customers in exploring relevant items which attunes to their needs or their preferences saving them the burden of navigating through huge number of products. (Adomavicius & Tuzhilin 2005).

Simply put, a recommender system is generally built in three sequential steps: information gathering, learning and recommendation. Information gathering stage is where systems get data to collect business insights from. The learning stage is where the system identifies patterns and relationships in the data that can be used to make accurate recommendations. The relationship might be found between users or products. Similar users might be recognized by looking into their past purchases, ratings they were given or reviews they wrote. The recommendation stage is where the model presents the user with a list of products or services based on the predictions of the model. These steps are repeated over time as the system continues to learn and make improvements. (Isinkaye et al. 2015).

There are three primary types of recommendation methods: content-based recommender systems, collaborative-based recommender systems, and hybrid methodology which uses both approaches. Recommender systems that are built using the content-based approach makes suggestions for a user analyzing their past behavior and interests, such as items they liked, bought, or viewed and tries to find similar items comparing characteristics of the items (Pazzani & Billsus, 2007). On the other hand, the collaborative filtering-based approach makes recommendations based on the similarities between the user and other users with similar preferences or interests in the past (Aciar et al., 2007). The hybrid approach combines more than one recommendation approach or algorithms into one recommender system where advantages of specific approaches would be suitable for the task in hand (Danilova & Ponomarev, 2016).

Most recommender systems that use user reviews employ NLP techniques such as sentiment analysis and opinion mining to gain business insights from them (Chelliah et al., 2017). Some systems only measure the positive or negative aspects of the reviews and make recommendations based on the polarity, while others use multi-label classification to build more complex models. Compared to models based only on one aspect of the review, more sophisticated techniques can be used to process user feedback and lead to better recommendations (Diao et al., 2014).

It is essential for companies to understand what customers expect from their services to provide high-quality service. Customer expectations play a major role in evaluating service quality and satisfaction. It is important for companies to listen to

the voice of their customers through various channels and incorporate their feedback into their design process. It is a well-known practice in different industries to implement techniques to experiment with design, quality function deployment, and value engineering. After customers receive service, it is important for an airline company to measure if they could have satisfied the expectations of their customers using methods which are built for specific scenarios. The quality of service is measured as the gap between the expectations and perceptions. The quality is considered high if there is no or small gap. Perception is the subjective response of a customer to the quality of a service. Therefore, it is relatively hard to measure. SERVQUAL method was developed by Parasuraman et al to fill this satisfaction measurement gap in the airline industry. It is a method which measures ten dimensions to evaluate the service quality of an airline company. These dimensions are responsiveness, reliability, tangibles, communication, security, credibility, courtesy, competence, access and understanding/knowing the customer. (Parasuraman et al., 1988)

Pakdil and Aydin (2007), made a 35-question survey which extends the SERVQUAL method to measure the quality of airline service. They added new dimensions to measure airline service quality. They calculated SERVQUAL points with new weights that they determined via factor analysis. Their questions measure the expectations and perceptions of the passengers about the airline service quality using a 5-point Liker scale. Ratings for expectations are starting from 1 = unimportant to 5 = very important. Ratings for perceptions are starting from 1 = strongly disagree to 5 = strongly agree. The findings showed that the past experiences of a given customer were the most important factor in selecting an airline. This indicates the importance of holding onto the old customers. By improving their service quality according to user feedback airline companies increase their revenues. They also discovered that none of the perceptions of the passengers met their expectations. Responsiveness was the highest scored dimension in the expectations section whereas availability was the lowest. Similarly in the perceptions section responsiveness and availability took the same positions. Furthermore, the educational level of the passenger seems to affect both expectations and perceptions in various dimensions. This study shows the importance of analyzing user reviews for airline companies to gain insights about their customers and finding areas to

improve their quality of service.

Baker (2013) conducted a study looking at the data from the top 14 airlines in the US from 2007 to 2011 to see if there was any difference between low-cost carriers and legacy carriers in terms of quality of service and customer satisfaction. Their data was collected from Department of Transportation Air Travel Reports. The results showed that there was a problem in the airline industry delivering exceptional services in general although it has been getting better from 2007 to 2011. According to a study, low-cost carriers that fly between large destinations tend to perform better whereas legacy airlines are placed in the middle. Regional airlines that usually fly with small planes took the last place. Another similar study conducted by Tusar et al. (2021) analyzed Twitter data of customers who flew with 6 major U.S. airlines in February 2015. They used NLP techniques when pre-processing and vectorizing the data. They then used diverse types of machine learning algorithms such as SVM, Naive Bayes, Random Forest, and Logistic Regression to make a sentimental analysis on the textual data. NLP techniques such as Bag-of-Words and TF-IDF were also used. Finally, they compared the measured performance of different algorithms and NLP techniques to determine the best approach. They found similar results that the airline industry was struggling to provide exceptional services.

In the airline industry, recommender systems have been used to help customers find the best flights based on their preferences and needs, considering factors such as destination, travel dates, and budget. However, the performance of these systems can be improved by incorporating additional information, such as online customer reviews, into the recommendation process.

Dadoun et al. (2021) discussed possible use cases of introducing recommender systems in the business processes of the airline industry. They claimed it is possible to find new ways of implementing recommender systems since New Distribution Capability (NDC) was introduced by the International Air Transport Association (IATA). NDC will give the airlines the ability to offer products in a more customer centric means. Current distribution systems do not allow airlines to offer real-time products to the customers. However, the arrival of NDC changes that and gives the airline companies to offer in real-time and gives the ability to differentiate the pricing functionality using a greater number of parameters.

A comparative study conducted by Khaturia et. al (2018) compares NaiveBayes algorithm and sentimental analysis approaches when building recommender system using online customer tweets. Many flights are made daily by airline companies, and they generate vast amounts of data. Authors aimed to understand the possible solutions to build such a system that will ease out the decision-making process for the customers. They used Twitter data of 6 major U.S. airlines for this task. They separated the tweets into categories of positive, negative, and neutral. They chose NaiveBayes algorithm to provide a benchmark to their sentimental analysis method. Scores for every airline in the dataset are calculated via NaiveBayes and sentimental analysis. Recommendations are based on the score each airline got from the algorithms. They state that the accuracy of sentimental analysis on a single sentence like an online movie review text never got above %80 for the 7 years prior to their publication data. The authors of the paper achieved %20 increase in accuracy, which is a significant improvement. They predict that demand for recommendation systems will increase with improvements in techniques such as data mining and sentiment analysis.

Lu H. et al (2017) proposed a custom recommender system method to suggest ancillary services based on Gradient Boosting Decision Tree (GBDT) and named it IR-GBDT. They combined personal characteristics of passengers and details of their order history to infer features for passengers using order data of flights for 6 months from a large online travel company. Passenger features include gender, age, sensitivity to price, purchasing history etc. Since ancillary services do not have features, they considered flight features such as flight time, takeoff time, flight punctuality etc. Their model was first built to calculate the independent possibilities of a passenger buying a specific service. They then calculated the interrelationship between two ancillary services using GBDT. Their interrelationship algorithm is a 3-step operation. In the first step the influence of the impact of the flight's feature on the passenger's decision is calculated. In the second step interrelationship is calculated for every pair of ancillary services. In the third step the probability of buying service b given service a has been bought is calculated. Ancillary services are recommended if their probability passes the thresholds. Because they take interrelationships into account, they can recommend a bundled service package instead of a single service. Five ancillary services were used in this study which are

rapid screen channel, VIP lounge, airport transportation service, security insurance and delay insurance.

Mottini, A. et al (2018) suggested that recommender systems were not utilized in the airline industry and said that the reason behind it was the decision process the customer undergone when deciding on a product has always been seen as a black box. They decided to compare three different choice modelling techniques to see which one would be the best choice for recommender systems. They found that deep learning algorithms perform better with higher accuracy rates. However, there are disadvantages to using deep learning algorithms such as low interpretability, many hyperparameters and requirement of high computing power. Hidden layered structure of deep learning algorithms makes it hard to interpret. These disadvantages make other algorithms feasible. The choice modelling method was simple and interpretable, but it required feature engineering and it had limitations handling large amounts of data. Machine learning algorithms on the other hand are easily interpretable and able to handle big data but they also may require feature engineering and they need independence of samples. They concluded that it is best to choose an algorithm based on the specific requirements of the problem.

P. K. Jain et al (2019) used K Nearest Neighbors (KNN), SVM and decision tree to analyze customer generated feedback collected from airlinequality.com to predict if user will give positive or negative feedback. The motivation behind their study is the increasing number of online transactions and reviews and the possibility of gaining business insights from them. Their data covered the flight which took place in the dates between January 2011 and December 2015 with four different airline classes which are economy, business, premium economy and first-class. They mostly compared economy and business classes. It consists of 2000 reviews and is balanced with 1186 positive and 814 negative reviews. They applied natural language processing to the review texts before applying predictive analysis. They found that SVM works better with the accuracy score of 82.75% whereas KNN and decision tree performed 64.75% and 74.75% respectively. Their study showed customers of business class cared more about the food served on the plane and friendliness of the flight attendants whereas economy class customers cared about legroom and seat comfort. They also found that recommendations based on ratings perform better compared to recommendations based on reviews.

Srinivas, S. et al (2020) conducted a study on unsupervised text analytics of online passenger reviews. Their aim was to classify user reviews into groups of topics and sentiments and gain business intelligence from them, provide a summary on review and identify the key points affecting a passenger's experience using the airline company. Dataset used in this study is collected from online travel websites such as TripAvisor, Kayak etc. They first preprocessed the reviews by spelling correction, lowercasing, stop word removal, stemming etc. They then built a custom methodology which uses topic extraction and sentimental analysis and root cause detection to gain insight on the user reviews. For the topic extraction part, they used Latent Dirichlet Allocation (LDA) over Probabilistic Latent Semantic Analysis(pLSA) because pLSA is prone to overfitting and there is no easy way to introduce new documents for modelling which are not part of the training data. The sentimental analysis method is chosen by comparing four different popular methods which are SentiStrength, VADER, AFINN and E-SA. After they labeled the reviews by topics and sentiments, they tried to gain business insight from them. They found most of the negative reviews is about ticketing services and seating comfort.

Heidari, M. et al (2020) promoted a method implementing a convolutional neural network (CNN) model using online reviews and flight data to recommend airline tickets. Online reviews data about four major US Airlines was collected from Twitter. Each airline has more than 3000 user reviews. Flight data was collected from various websites such as Google Flights, Kayak, SkyScanner, Hotels.com. Flight data consists of departure and arrival times, price, flight class and airline company. Their proposed method uses Bidirectional Encoder Representations from Transformers (BERT) to classify sentiments of online reviews. For vectorization process they chose GloVe instead of word2vec because GloVe is faster and more scalable when used with large amount of data. They combine flight data, sentiment and vectorized reviews and feed them into CNN. Their CNN is sequential model with two-dimensional convolution layer with two dense layer and flattening layer. They chose their hyperparameters according to semantic data extraction for sentences guideline of the Microsoft. They then recommend tickets as economical ones or not economical ones according to output of CNN. Their new model performed better compared to other methods where there was no sentiment analysis or flight information. It performed with 0.92 F-1 score whereas CNN without BERT

sentiment scores performed with 0.81 and CNN with the best flight information performed with 0.81.

Dundar, B. et al (2021) proposed a recommendation system that analyzes passenger evaluations of airports. They utilized a method using linguistic summarization techniques. The dataset used in the study includes 2960 passenger evaluation information about airports. For airports nine of the related facilities/evaluations were chosen among others which are departure, arrival and travel times, cleanliness, passing the security and screening and getting to the airport. For passengers two attributes were chosen which are age and income. They summarized the evaluations using two types of fuzzy sets which are trapezoidal and triangular. They used 3,5 and 7 linguistic labels and 3 types of quantifiers such as most, about half, a few etc. They put passenger attributes and airport facilities together to get fuzzy quantified sentences. To extract the top three linguistic summaries from text they used Zadeh scalar method and gradient descent fuzzy cardinality methods. After extracting summaries, they compared labels according to 3, 5 and 7 linguistic terms. They found that most of the low-income passengers find it difficult to get to the airport and young and low-income passengers find it difficult to pass the security and screening. They conclude that these insights might be helpful for airlines to improve their services and implement better recommender systems. Their study might also be applied to other industries.

Zhang, Y. et al (2022) proposed a contextualized recommender system to solve the cold start problem and provide dynamic recommendations. They used travel-related factors such as time, location, weather etc. to add context to the air passenger portrait based collaborative filtering model recommending ancillary services. The proposed method constructed a contextualized passenger portrait which consists of four tag sets. Tags they chose included attributes of passengers, attributes of ancillary services, attributes of interaction of passenger and contextual attributes. Attributes of passengers were age, gender, city, education etc. Attributes of ancillary services were type and price. The attributes of passenger interaction were divided into two subcategories which are behavior of passenger like browsing, searching etc. and interaction with ancillary services like seat preference, early boarding etc. After building the passenger portrait the similarity between portraits were measured in three different aspects which are similarity of passenger attributes, similarity of

service scenario and similarity of interactions of passenger. They randomly picked 100 passengers to measure the performance of their methodology. Their proposed algorithm performed better with a 4.3 percent higher accuracy rate and 2.62 percent higher recall rate than content filtering-based algorithm. The results showed that their proposed algorithm also decreased the effect of sparsity in the data.

Table 1 shows the literature map which illustrates the work done by various scientists in the recommender systems area and its usage in the airline industry. It also shows the methods the scientists used in their studies. There are studies which analyses user reviews or build recommender systems for ancillary services. However, there is no study conducted on building recommender systems for ancillary services using online user reviews. This study aims to fill the gap between user review analysis and ancillary service recommender systems in the airline industry. Online user reviews collected from various web sites will be used to do sentiment analysis and topic extraction and then they will be fed into a machine learning model to recommend an ancillary service to the customer.

Table 1

List of Studies Examined in the Literature Review Section

Year	Title	Author(s)	Method(s)
2018	Coevolutionary Recommendation Model: Mutual Learning between Ratings and Reviews	Lu, Y., Dong, R. & Smyth, B.	CF, Matrix Factorization
2005	Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions	Adomavicius & Tuzhilin	
2015	Recommendation systems: Principles, methods, and evaluation	Isinkaye et al	
2007	Content-Based Recommendation Systems	Pazzani & Billsus	

Table 1 (cont'd)

Year	Title	Author(s)	Method(s)
2007	Informed recommender: basing recommendations on consumer product reviews	Aciar et al	
2014	Jointly Modeling Aspects, Ratings and Sentiments for Movie Recommendation (JMARS)	Diao et al	
2017	Product Recommendations Enhanced with Reviews	Chelliah, M. and Sarkar, S.	
2016	Hybrid Recommender Systems: The Review of State-of-the-Art Research and Applications	Danilova & Ponomarev	
1988	SERVQUAL: a multiple item scale for measuring consumer perceptions of service quality	Parasuraman et al	Survey
2007	Expectations and perceptions in airline services: An analysis using weighted SERVQUAL scores	Pakdil and Aydin	Survey
2013	Service quality and customer satisfaction in the airline industry: A comparison between legacy airlines and low-cost airlines	Baker	
2013	Using online consumer reviews as a source for demographic recommendations: A case study using online travel reviews	Korfiatis, Nikos & Poulos, Marios.	k-means Clustering, Collaborative Filtering
2021	How recommender systems can transform airline offer construction and retailing	Dadoun et al	
2021	A Comparative Study of Sentiment Analysis Using NLP and Different Machine Learning Techniques on US Airline Twitter Data	Tusar et al	SVM, Logistic Regression
2017	Auxiliary Service Recommendation for Online Flight Booking	Lu, H., Cao, J., Tan, Y., Xiao, Q.	Gradient Boosting Decision Tree
2018	A Comparative study on Airline Recommendation System Using Sentimental Analysis on Customer	Khaturia et al	Sentiment Analysis, Naive Bayes

Table 1 (cont'd)

Year	Title	Author(s)	Method(s)
	Tweets		
2018	Understanding Customer Choices to Improve Recommendations in the Air Travel Industry	Mottini, A., Lhéritier, A., Acuna-Agost, R., & Zuluaga, M. A.	Discrete Choice Modelling, RF, DL
2019	Airline recommendation prediction using customer generated feedback data	P. K. Jain, R. Pamula, S. Ansari, D. Sharma, and L. Maddala	KNN, SVM, DT
2020	Using Transfer Learning Approach to Implement Convolutional Neural Network model to Recommend Airline Tickets by Using Online Reviews	M. Heidari and S. Rafatirad,	CNN, BERT
2020	Unscrambling Customer Recommendations: A Novel LSTM Ensemble Approach in Airline Recommendation Prediction Using Online Reviews	P. K. Jain, R. Pamula, S. Ansari, D. Sharma, and L. Maddala	LSTM, SVM
2020	Discovering airline-specific business intelligence from online passenger reviews: an unsupervised text analytics approach	Srinivas, S., & Ramachandiran, S	Topic Extraction, Sentiment Analysis
2021	Linguistic Descriptions of Airport Evaluation Data for Recommendation Strategies	B. Dundar, D. Akay and S. Ozdemir,	Fuzzy Linguistic Summarization
2022	Contextualized Recommendation of Aviation Ancillary Services Based on Passenger Portraits	Y. Zhang, W. Luo, M. Li, and T. Chen	Context Aware Collaborative Filtering

Chapter 3

Methodology

In this chapter the data is explained in detail and the proposed model design will be illustrated along with the methods used. Data was explained in detail describing columns and giving statistical information. To measure how effective various machine learning methods perform, support vector machines (SVM), extreme gradient boosting (XGBoost) and artificial neural networks (ANN) were chosen. Chosen algorithms were explained after the data section. Sentiment analysis method BERT, topic extraction method BERTopic and artificial neural network frameworks Tensorflow and Keras which were all used in proposed method also explained. Later the proposed method was illustrated using these three machine learning algorithms. Definitions of evaluation metrics which were used to measure performance of the algorithms were given at the end of this chapter.

3.1. Data

We used web scraping method to gather the publicly available user reviews. The data was collected from travel websites where customers write their experiences on airline review pages to use in our study. They write about both good and bad experiences using an airline company and give a rating point regarding the service quality. About 300.000 unique reviews were collected. The data collected consists of 5 columns which are rating, review title, review text, route, and flight class. Rating is where users gave score from 1 to 5 according to their general experience with the flight where 1 means the worst and 5 means the best. Review title is the title they give for their review. Review text is where users wrote about their experience flying with the airline company. The route is written as the departure city and destination city together in one column where cities are separated by "- ". The last column is Class which shows the fare class the customer has bought such as economy, business etc. The mean rating was 3.5 for all airlines used in this dataset. Table 2 shows the statistical information of the airline ratings. 81% of fare classes were economy class. The route column was divided as 'from city' and 'to city' and hence two new columns were generated. Another column was created from 'from city' and 'to city' which states the distance between two cities. Distance was calculated using geopy

library which takes city names as input parameter and return the distance based on latitude and longitude values. The last column was named ‘distance’.

Table 2

Statistical Information of Ratings

	Rating
count	308418
mean	3.362
std	1.516
min	1.000
25%	2.000
50%	4.000
75%	5.000
max	5.000

The training data for the recommender system is a subset of reviews data with an extra label of ancillary product column. Training data consists of 600 rows. Featured product column was introduced to pair reviews and products. Reviews were labelled by an expert in the airline industry via reading the review text and choosing the suitable ancillary product. Four of the available ancillary products were used in this study which are Exit Seat, Special Meal, Priority Boarding and Extra Baggage. The data labeling process was done by choosing appropriate keywords for each ancillary product. For example, legroom, cramped and narrow keywords were chosen for Exit Seat product to find related reviews. After finding reviews that contain keywords, they were read manually to check it was correct label. Table 3 shows a few of the reviews labeled by the expert.

Table 3

Labeled Review - Ancillary Product Matchings

Rating	Review Text	Route	Class	Product
1	Issues began from buying the ticket. In the last buying step price was raised. They wrote that the last ticket for that price was sold at that moment. I would believe it if the same situation would not happen for several of my friends:) The boarding both ways were totally unorganized, and stuff was confused. Boarding both ways was late. In general, unpleasant experience	Vilnius - Munich	Economy	Priority Boarding
2	We were told to be at the airport 2 hours before our flight 779 boarding was at 6.35 the majority like myself were there at 4pm the Check in did not open until 5pm it took 3 check-in stalls 90 minutes to check us in that's the equivalent of 4 hours by the time we were checked in we had to board so did not get to go through duty free and our 4 hour flight only had chocolate bars and tea/coffee. If I wanted this type of treatment, I would have flown airline for a cheaper fare in future I just might! No ear buds on sale for flight entertainment state-of-the-art and possible extensions	Lanzarote - Dublin	Economy	Priority Boarding

Table 3 (cont'd)

Table 3 (cont'd)

Rating	Review Text	Route	Class	Product
	flight entertainment from Chicago to Manchester. Awful experience.			
2	Our flight didn't take off on time, was delayed by over an hour because the gate agents couldn't get their act together. No formal reason was given for the delay. We had a group of students travelling on our flights and it sounded like some students were randomly identified for additional screening? The flight was full but there was enough space for one carry on item per passenger, yet they were coercing passengers to gate check their carry on. This was an international flight with no TV screens or Wi-Fi. Snack offered was a ham sandwich, no vegetarian options!	New Delhi - Vancouver	Economy	Special Meal
2	I am a very avid and frequent traveler on airlines, and used to be blindly loyal to the airline, but I cannot understand why, lately the airline has chosen to reduce the legroom and seat pitch this much. It is now worse than airline, airline, and most low-cost airlines (this is true to all my flights I have taken with them in the past year. That, in addition to the uncomfortable seats, is making me shift my loyalty elsewhere to another airline, where passenger comfort is valued. The legroom is so bad, such that my knees hurt after the flight from being squeezed against the seat. I am 185 cm tall. Passengers won't stay loyal passenger of airline, wake up. I would recommend if prices were	Amman - Nice	Economy	Exit Seat

Table 3 (cont'd)

Rating	Review Text	Route	Class	Product
	comparable to low-cost airlines because the experience is almost the same as flying with such airlines.			
1	I don't understand how legroom and seat comfort gets 4 out of 5 in the overview. It's ridiculous and offensive to common sense. Our knees were bruised and our back still hurts due to the nonexistent space between seats. There should be a warning that passengers that exceed 1.80 m in height cannot be accommodated to their current setting and require to purchase more space. As far as COVID 19 is concerned, there are few words to describe the joke of it as we were packed like poultry. Way too many passengers around us didn't fit their masks correctly and flight attendants not only didn't do anything to correct it but some of them also didn't wear their masks properly. Of course, everyone took their masks off to drink and eat. You call these measures against COVID 19?	Thessaloniki - Larnaca	Economy	Exit Seat
1	Recently experienced the worst flight I have ever had. From check in to arrival it was a nightmare.!. Told at check in if we wanted to put any hand luggage in the hold we could free of charge. There were 9 of us including 2 x 6-month-old twins. We decided to put most in the hold and kept 3 of which one had the baby's milk, nappies etc. in, the others had snacks, iPad etc. We tried to argue but to no avail. Fortunately, a lady at	London - Madrid	Economy	Extra Baggage

Table 3 (cont'd)

Rating	Review Text	Route	Class	Product
	another desk said she would allow the baby's case. We then had to carry all iPad etc. in our hands as no bag to carry on place. his caused a few problems as was a security breach. The worst was still to come. We had booked 2 rows behind each other as we could only have 1 baby in each row due to drop down masks. The family said they had been moved as they had a baby and their row had only 3 masks. This obviously left 1 of our group with no seat. We asked why they cannot find another row with 4 masks and asked people to move so could go back to original 1 baby on row, but no one listened until we had to get very annoyed, so they called the captain who agreed with us that was the solution. We had left a double buggy at the entrance to the plane and when arrived in Madrid were told it would be on the belt with luggage. After waiting about an hour, we had to queue again for another hour at the airline enquiry desk. They could not locate it, so we were frantic in Madrid with 2 very tired babies by this time and nowhere to sleep. Eventually we spoke to another security guard who suggested we try excess baggage and thank goodness it was there. No thanks to Iberia. We eventually got to our hotel 4 hours later than planned			

Table 3 (cont'd)

Rating	Review Text	Route	Class	Product
2	Had it not been for the excess baggage charge, this review would have been completely different. As it is, the extra charge ruined our experience. On the outbound journey, 3 days earlier, we took the same - single piece - of hold luggage which was a little over 23 kilos. However, presumably as this was between the two of us, no charge was made. Conversely - not to mention inconsistently - on our return journey we were charged for the same piece of luggage. The desk clerk simply would not budge, even when we pointed out that our personal weights, together with the bag weight, were well under the maximum allowed. This cost us 25,000 dollars and we think it is very unfair indeed.	Easter Island - Santiago	Economy	Extra Baggage

3.2. Methods

In this section the methods, libraries and metrics that are used in our proposed methodology will be explained in detail.

3.2.1 Support vector machines. Introduced by Vapnik et al in 1995 SVM falls into the category of supervised machine learning algorithms. It is popular in solving regression and classification problems. (1995). The primary idea behind the algorithm is to draw possible boundary or hyperplane that can maximize the distance between the different classes in the dataset. The hyperplane is called the decision boundary. The purpose of the SVM is not finding a random decision boundary but finding the optimal decision boundary. The margin is described as the distance between the hyperplane and the nearest observed data points of each class. Those data points are also called the support vectors. Figure 1 shows the optimal hyperplane with maximized margin and support vectors. Blue and red dots are data

points and the closest data points to the optimal hyperplane form the support vector. The equation of hyperplane is shown as $wx + b = 0$ where w denotes weight for each feature, x is input vector and b is bias. In the cases where the data is not linearly separable the kernel trick comes in handy. Kernel trick works by mapping the data which is in a lower dimension into another space with higher dimension. The goal is to have classes linearly separable in a higher dimension. There are five hyperparameters to set when training to tune the model according to the dataset.

- C: C parameter is called the regularization parameter. It adds a penalty when there is a misclassified data point. If a high value is chosen for C then SVM will minimize the number of misclassified data points because of the high penalty it will give for each misclassification. If C is small, then the decision boundary will be with a large margin because of the high tolerance for the misclassified data points.
- Kernel: This parameter specifies the kernel function which will be used in the model {'linear', 'poly', 'RBF', 'sigmoid', 'precomputed'}
- Degree: If kernel method is chosen as polynomial, this parameter specifies the degree of the polynomial function. Must be non-negative number.
- Gamma: When the kernel function is chosen as RBG the gamma parameter comes in handy. The gamma parameter specifies the distance of influence around a single data point. When a higher gamma value is chosen the model becomes more complex, and it tries to fit the training dataset exactly hence overfitting may occur. Furthermore, the lower value for gamma parameter makes the area around the data points more generalized, and this may cause underfitting.

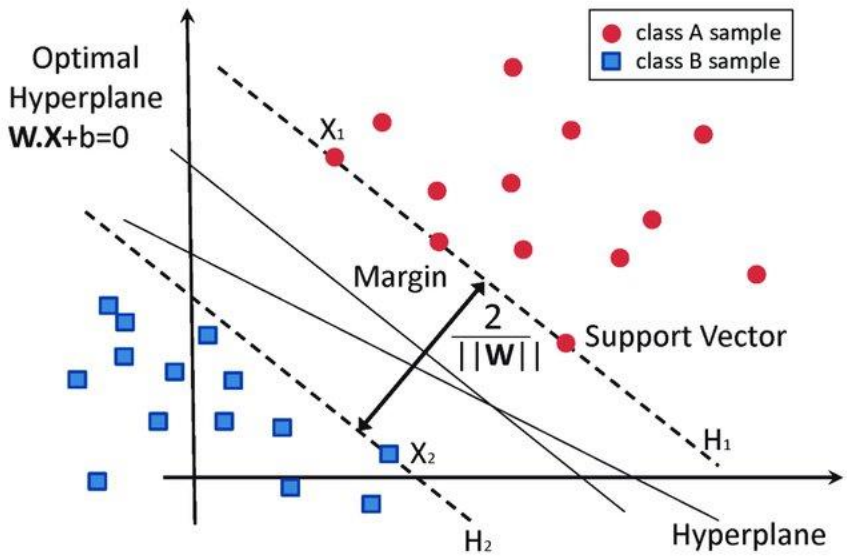

Figure 1. Classification of data by support vector machine (SVM)on the margin are called the support vectors.

3.2.2 Extreme gradient boosting. XGBoost is an example of supervised machine learning algorithms which was introduced by Chen et al. (2016). It uses multiple decision trees to make predictions. It was developed using basic principles of gradient boosting. It is an ensemble technique which aims to build a stronger learning model by assembling multiple smaller models. Figure 2 shows the evolution of tree-based algorithms starting from decision trees to XGBoost in the end. A decision tree is a tree structured supervised machine learning algorithm that consists of root node, branches for decision rules, internal nodes for features, and leaf nodes outcomes. They are easy to interpret. Figure 3 depicts the general structure of the XGBoost algorithm where it works by adding decision trees iteratively to a model. Every tree takes a shape to fix the errors of the previous. It has advantages over gradient boosting by using regularization to penalize building complex models. Another advantage is it can work with missing values by providing new branches when encountered with missing values.

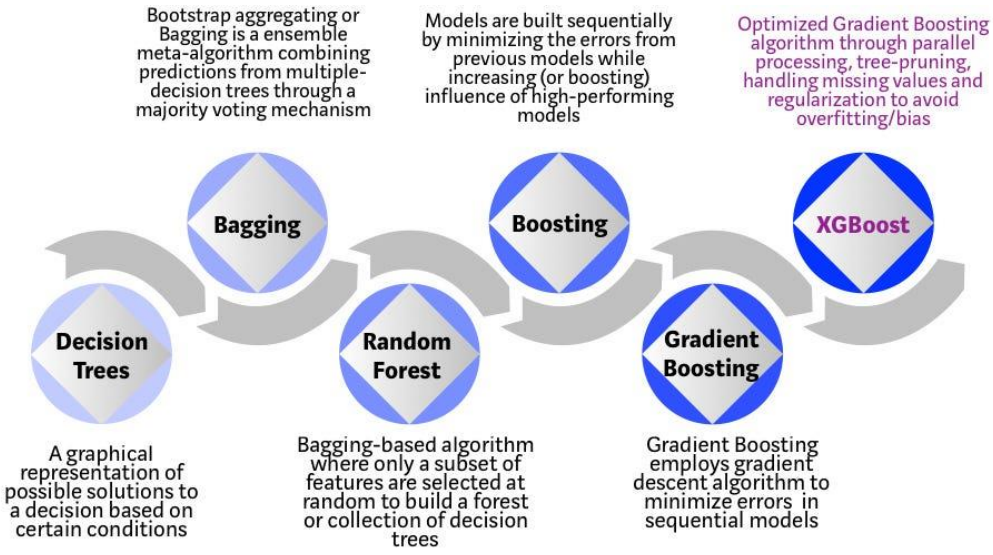

Figure 2. The Evolution of XGBoost Algorithm (Morde, 2021)

There are some parameters that could be used to tune the XGBoost model to achieve higher accuracy rates.

- max_depth: Specifies the maximum depth each tree can grow. Higher values increase the complexity and may cause overfitting. When chosen 0 there is no limit to the depth. l Default value for it is 6. Deep trees may consume large amounts of memory. The exact tree method requires non-zero value. range: [0, ∞]
- learning rate. Regularization parameter to define step size shrinkage used in update to prevent overfitting. After each boosting step, we can directly get the weights of new features, and to make the boosting process more conservative the learning rate shrinks the feature. It takes values between 0 and 1.
- n_estimators. Number of runs XGBoost algorithm will try to learn.
- colsample_bytree. Randomly selected features that will be used to train each tree.
- Gamma: Regularization parameter to prune the tree. It specifies the minimum loss reduction required to make a further partition on a leaf node of the tree.

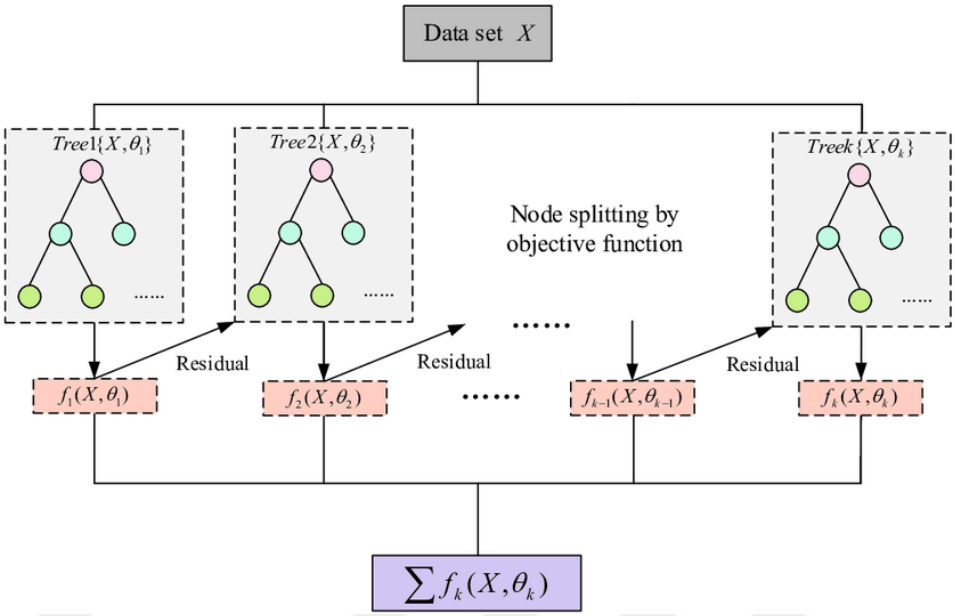

Figure 3. Flow chart of XGBoost.

3.2.3 Artificial Neural Networks. ANNs inspired by the learning ability of the human brain consist of interconnected nodes, or neurons, organized into layers. Figure 4 illustrates the layered architecture of ANNs. The input layer receives input data, and its size depends on the attributes of the input data. Hidden layers can be one or more layers where the transformations are made and the relationships between features are determined by each node. Node number in the output layer is given according to the number of features which will be classified or predicted. The weighted sum of its input is fed into a neuron, which is then passed through an activation function to produce an output. The decision that whether a specific neuron should be activated is made in the activation function which is a non-linear function. The flexible structure of neural network allows modelling complex relationships between the input and output data. There are different common activation functions, such as the sigmoid, softmax, and ReLU functions. ANN training process involves adjusting the weights of the connections according to the data between neurons to minimize a cost or loss function. Weights are typically adjusted by an algorithm called backpropagation, which calculates the gradients of the cost function with respect to the weights and updates them using gradient descent or a related optimization algorithm. Flexible structure of ANNs make them suitable to be used for a wide variety of tasks, including classification, regression, and clustering. Despite the hard to interpret nature of ANNs, they are particularly well-suited for

tasks that involve substantial amounts of data or complex relationships between the input and output variables.

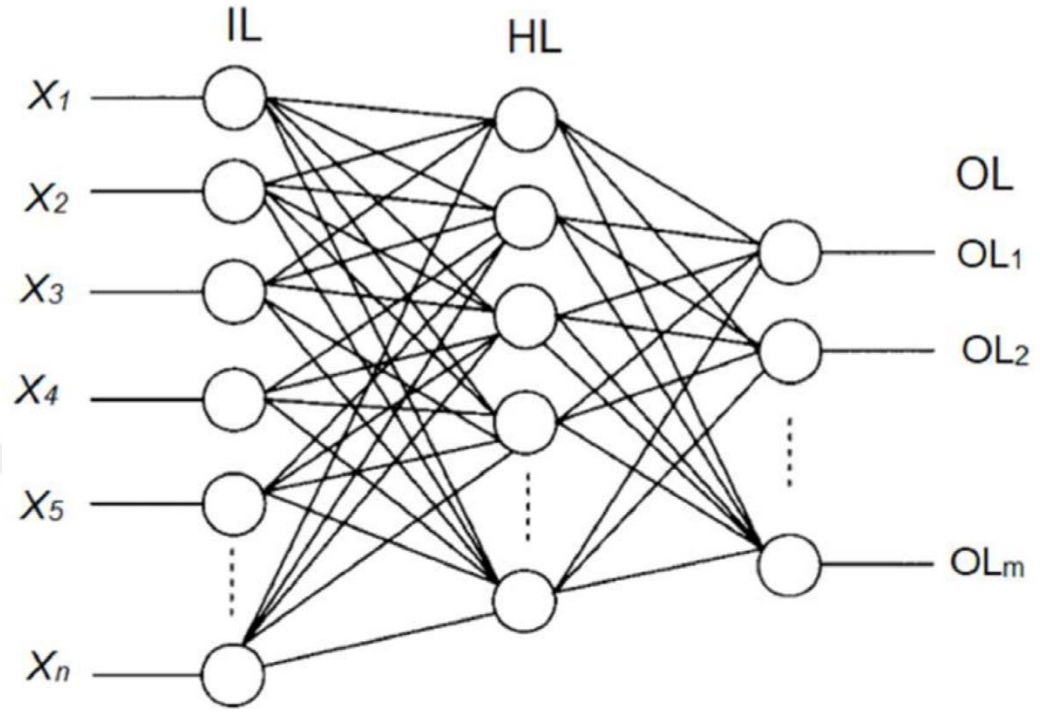

Figure 4. Artificial neural network architecture (Bekesiene, 2021)

3.2.4 Bidirectional encoder representations from transformers (BERT). Introduced by Google BERT is an NLP technique. It is a deep learning algorithm that can be used for various tasks such as question-answering, sentiment analysis, and language translation. BERT is unique in its ability to pre-train a bidirectional Transformer encoder, which can then be fine-tuned on a specific NLP task. Figure 5 shows the basic structure of BERT with tokens as inputs and class label as output. Inside the structure there are transformer encoders. The Transformer encoder is a neural network architecture that processes input sequences in parallel using self-attention mechanisms. The self-attention mechanism allows BERT to consider the context of each word in the input sequence in relation to all the other words, both before and after it.

During the pre-training phase, BERT is trained with unannotated text data to learn a general understanding of natural language. This is done using two tasks: masked language modeling and next sentence prediction. Masked language

modelling tries to find the masked tokens which are randomly masked previously. Prediction regards the context of the words as well. In the next sentence prediction, the model is trained to find the successor and premise of a given sentence.

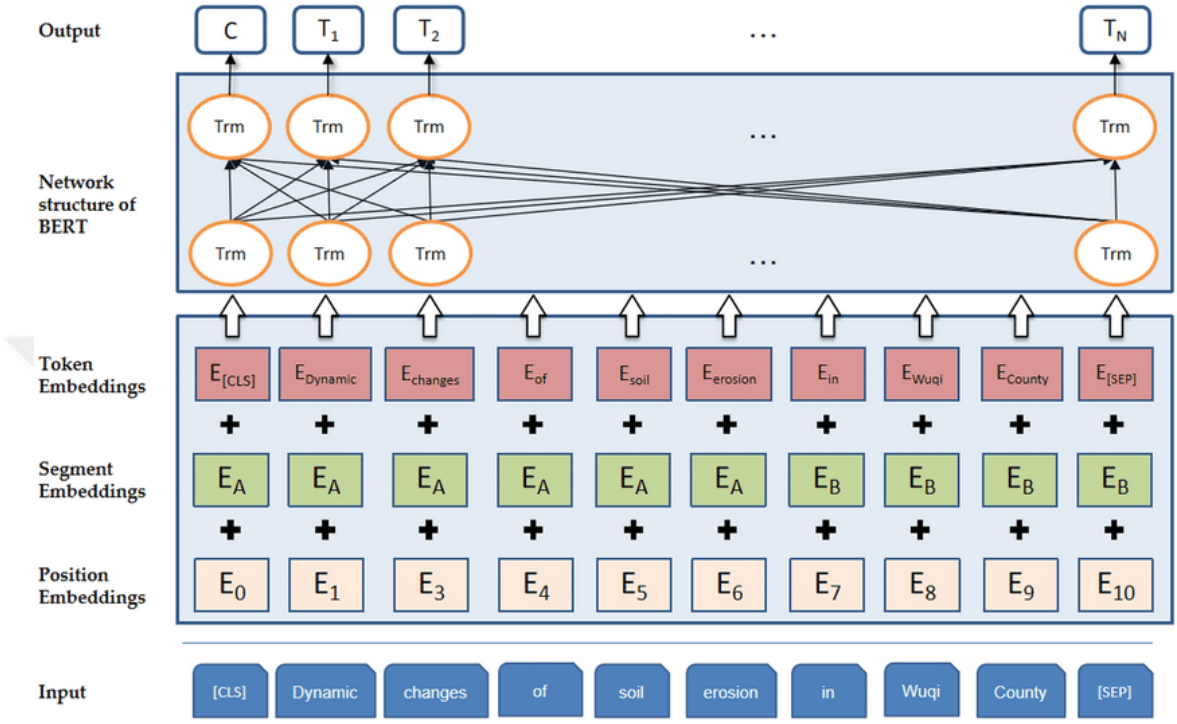

Figure 5. BERT (Sun J, 2022)

3.2.5 BERTopic. Inspired by BERT, BERTopic is a topic modeling technique that uses transformers and c-TF-IDF to create clusters to produce topics that are simple to understand while preserving key terms from the topic descriptions. It uses BERT to identify topics in text documents. BERTopic is based on the idea that topics can be represented as clusters of similar documents in high-dimensional space. BERTopic achieves this by first embedding each document in the collection into a low-dimensional space using the BERT model, which is pre-trained on large amounts of unannotated text. It requires dimensionality reduction because the clustering model works better for low-dimension data than high-dimension data. We need to reduce the dimensionality of embeddings before passing them into clustering model because the document embeddings tend to have hundreds of dimensions. BERTopic uses UMAP (Uniform Manifold Approximation & Projection) as the default algorithm for dimension reduction. UMAP conserves the local and global

structure of the data when reducing the dimensionality which is very important when representing the semantics of the text data. That is the reason UMAP was chosen as the dimensionality reduction algorithm over other ones such as PCA (Principal Component Analysis). Another advantage of the UMAP algorithm is its flexibility about customized hyperparameters. The document embeddings with reduced dimensionalities are then clustered using a technique called Hierarchical Density Based Spatial Clustering of Applications with Noise (HDBSCAN). HDBSCAN is a clustering algorithm that can identify clusters of varying sizes and densities. By applying HDBSCAN to the document embeddings, BERTopic can identify clusters of similar documents that represent different topics in the collection. BERTopic also includes several features to improve the interpretability of the resulting topics. For example, it can summarize each topic by selecting the most representative documents and words for each cluster. It can also visualize the topics in two- or three-dimensions using techniques such as t-SNE.

3.2.6 Tensorflow. Created by the Google Brain team in 2015, Tensorflow is an open-source software library for building numerous machine learning and artificial intelligence models. It comes with a variety of machine learning and deep learning models. It gives the developers the abstraction when building machine learning models. Its major feature is auto differentiation which is a process of the gradient vector calculation of a model with respect to each of its parameters automatically.

3.2.7 Keras. Keras is an open-source high level deep learning API to the Tensorflow library. It was created to be user-friendly and modular. It was written in Python language and intended to be used in Python environment. Its broad adoption gives it the advantage of wide support.

3.2.8 Evaluation Metrics. The definitions of the evaluation metrics used to measure model performance are given in this section.

• True Positives (TPs): The count of positive records which the model correctly estimated.

• True Negatives (TNs): The count of negative records which the model correctly estimated.

• False Positives (FPs): The count of negative records which the model incorrectly estimated as positive.

• False Negatives (FNs): The count of positive records which the model incorrectly estimated as negative

- $Accuracy = \frac{(\mathrm{TP} + \mathrm{TN})}{(\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN})}$
- $Recall = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}$
- $Precision = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}$
- $\mathrm{F} - 1\ \mathrm{Score} = 2 * \frac{(\mathrm{Recall} * \mathrm{Precision})}{(\mathrm{Recall} + \mathrm{Precision})}$

3.3 Proposed Method

After determining the problem, we moved on to build a structure to use online reviews to recommend ancillary products to customers. Figure 6 shows our proposed method where we process reviews and review product matchings to make recommendations. We used a pretrained BERT model and fine-tuned it with our reviews data to make better sentiments analysis on our review texts. BERT model uses 12 transformer blocks with 768 hiddens. BERTopic method was used to build a model to extract topics from review texts.Then negative reviews, topic extraction model and product review matchings were all fed into the recommender system to provide a product recommendation.

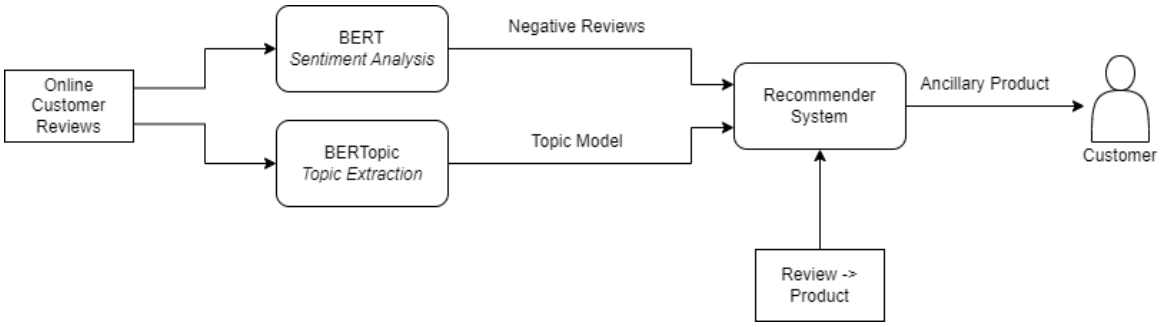

Figure 6. Proposed Method

Reviews are preprocessed before they are fed into the system. Preprocessing started with lowercasing the review text. Then all the punctuation, links and html code were removed using regular expression code. Stop words were removed from the text. We used pre-defined stop words from Natural Language Toolkit (NLTK) for English language. Then we added some new words to our stop words dictionary we thought would not add value when building our topic extraction model. Then reviews were lemmatized using WordNetLemmatizer from NLTK library. Lemmatization is

a process of grouping together the inflected forms of a word so that they can be analyzed as a single item, based on their common root or lemma. For instance, the words happiness and happy both become ‘happi’ after lemmatization. Lemmatized text then tokenized using same library. Tokenization is a process used in natural language processing to split sentences into smaller units that can be given meaning easier.

Fine-tuned BERT model was used to determine negative reviews. Fine tuning process was done by labeling reviews which have ratings lower than 3 as negative. The same preprocessing procedure was used for reviews before sentiment analysis.

BERTopic method was used to extract topics from given reviews. The text column from whole reviews dataset was fed into the algorithm to build the topic extraction model. The UMAP algorithm was used to reduce the dimension to have better clustering performance. The most important parameter n_neighbors determine the size of the local neighborhood for UMAP. A low value means UMAP will focus more on the local neighborhood and may lose insights into the big picture while reducing the dimensionality. With higher values the algorithm looks at the broader neighborhoods and may lose fine details on local structure. When focused too much on details, this usually results in a larger cluster size. The default value for it is 15 and different values for n_neighbors were tried to find the best cluster. Min_dist controls the minimum distance between points allowed to pack together in the lower dimension. It was set to zero since the goal of dimension reduction in this case was to build clustering models. Figure 7 shows scores of top five words for every topic. After we got the top five words for our topics in the first iteration, we added new words to our stop words dictionary. Some of those words are 'one', 'mr', 'bbc', 'image', 'getty', 'de', 'en', 'caption', 'also’ etc. which does not make sense in the ancillary service context.

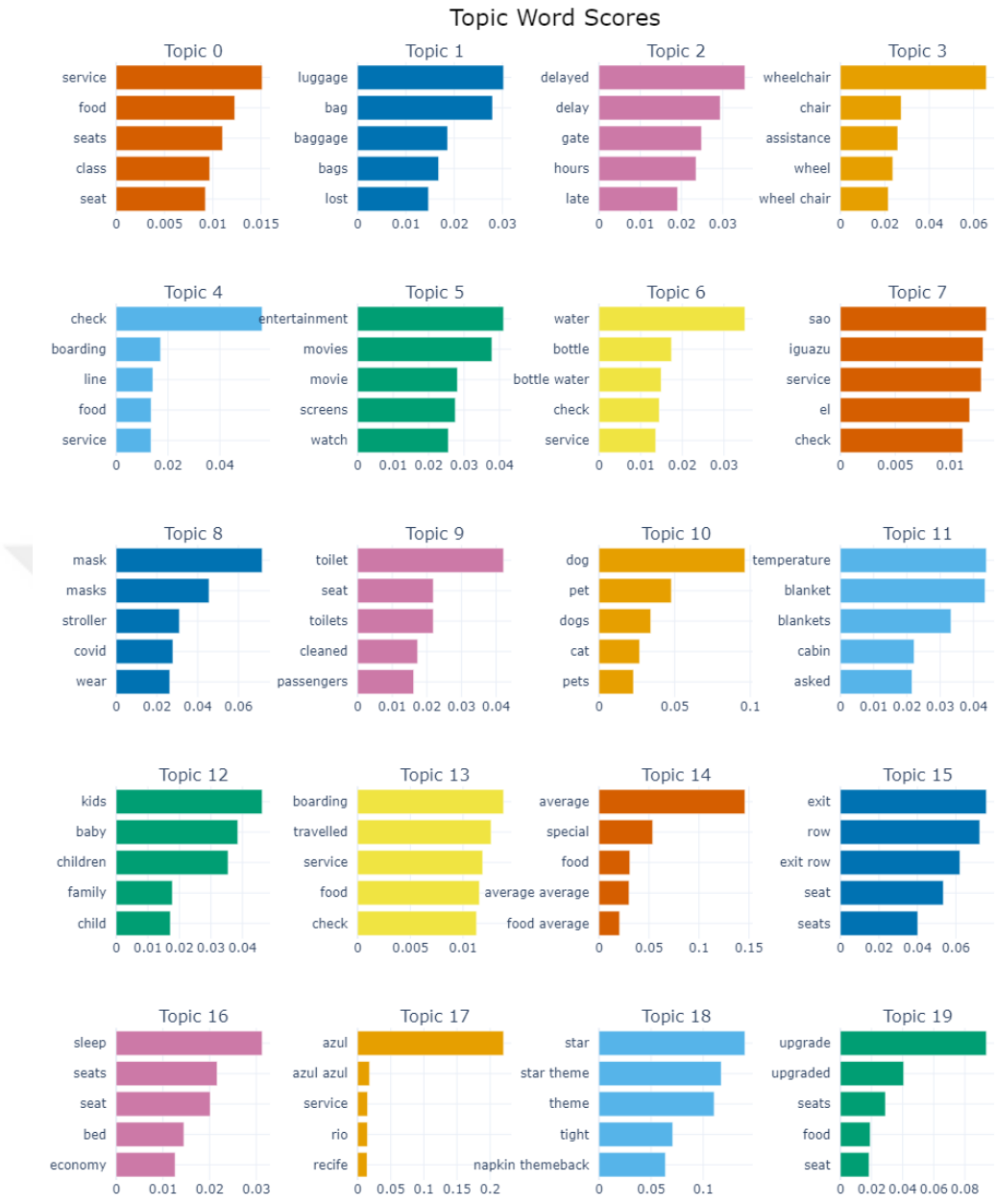

Figure 7. Topics

Once the sentiment analysis and topic extraction models were ready the dataset which contains review and product matchings was fed into the topic extraction model. Then a vector for each review was generated containing topics assigned for that review. Distance information was also added to the vector.

Data vector was scaled using standard scaler to have better results. Scaled data

was fed into the support vector classifier of sklearn library. Different kernel methods were used such as linear, polynomial, and radial basis function. Different hyperparameters were used to get higher accuracy measurements. One of the vital parts of improving the overall performance and workings of the machine learning algorithms is hyperparameter tuning. Hyperparameters are set before the learning process begins. There is a method called GridSearchCV which is used to find the optimal hyperparameters. It takes a dictionary that describes the parameters with different values that could be tried on a model to train it. It starts with initial values and trains the model with all the parameters provided in the dictionary until it finds the optimal parameters.

XGBoost algorithm was used second to see how efficient tree-based methods would perform. After running the algorithm with initial values, the hyperparameters were tuned to get better prediction results. Max_depth value was initially set to 3 and increased to 10. Learning_rate was set to 0.01 in the beginning and increased to 0.1. n_estimators was initially set to 100 and increased to 1000. Colsample_bytree was initially 0.3 and incremented to 0.7. Gamma value was set between 0.00001 and 200. The parameters were compared according to change in accuracy and RMSE scores of the model.

The last method was artificial neural network to recommend ancillary products. Tensorflow library was used to build a network with hidden layers. Different hyperparameter configurations were tried to find the optimal structure. Since the data vector had 20 members the input layer had 20 nodes. Then for the second layer and third layer different combinations were used. For the output layer argmax and softmax activation functions were tried to get better results. Softmax function gave better results. Output layer has 4 nodes because there are 4 classes in the target variable. Table 4 depicts the final model structure used to build neural networks. The hidden layer has 128 nodes, and the output has softmax activation function. Sparse categorical cross entropy function used as loss function because target value has integers 0,1,2,3. Adam function was chosen as optimizer.

Table 4

Neural Network Model

Layer(type)	**Output Shape**	**Parameters Number**
Input Layer (Dense)	(20)	420
Hidden Layer (Dense)	(128)	2688
Output Layer (Dense)	(4)	516
Total Params: 3,624		
Non-trainable params: 0		

Chapter 4

Findings

In this chapter results which are gathered using SVM, XGBoost and ANN are explained using evaluation metrics. The accuracy rate of SVM method was 0.79 using linear kernel and C value with 0.5. Table 5 illustrates the classification report of SVM showing precision, recall and F-1 score of every class.

Table 5

SVM Classification Report

Accuracy	**Precision**	**Recall**	**F- 1 Score**
0.79	0.79	0.79	0.79

The second method we used to recommend ancillary products was XGBoost algorithm. Table 6 shows the highest accuracy rates we achieved using XGBoost. It also shows precision, recall and F-1 scores.

Table 6

XGBoost Classification Report

Accuracy	**Precision**	**Recall**	**F-1 Score**
0.79	0.81	0.80	0.79

Lastly using the ANN methods in our model gave the highest accuracy rate of 0.85. Figure 8 shows the training and validation accuracies over time. Figure 9 shows loss rate over time.

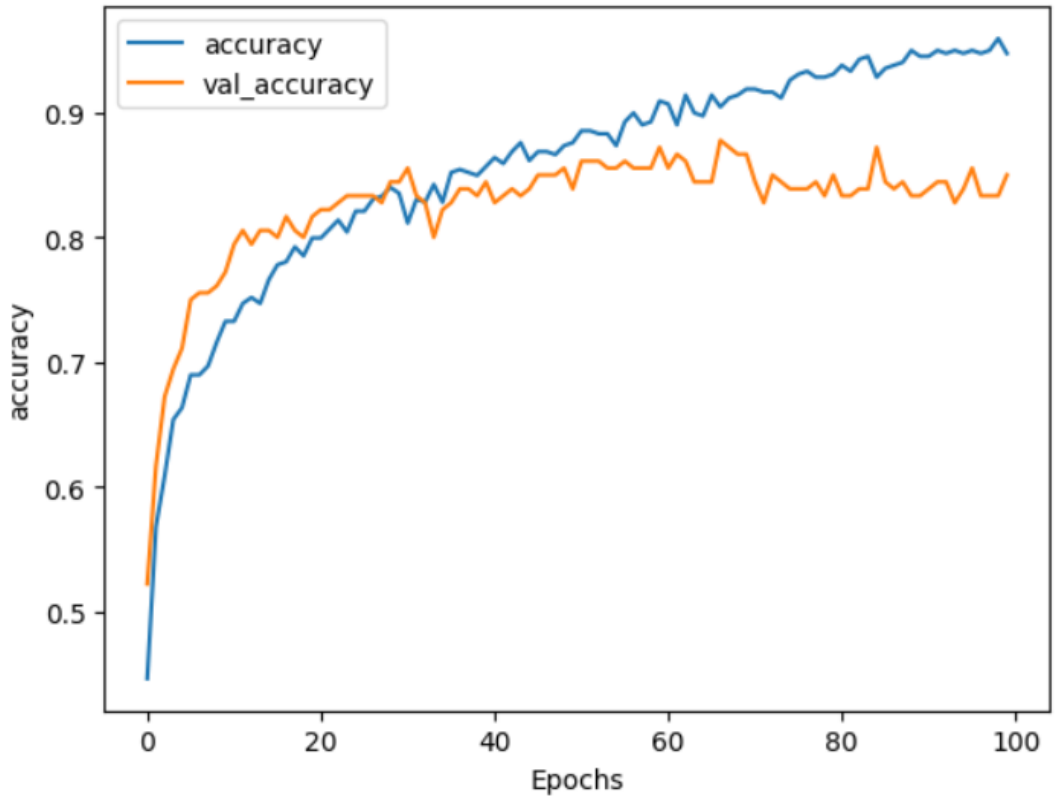

Figure 8. Accuracy Graph of NN

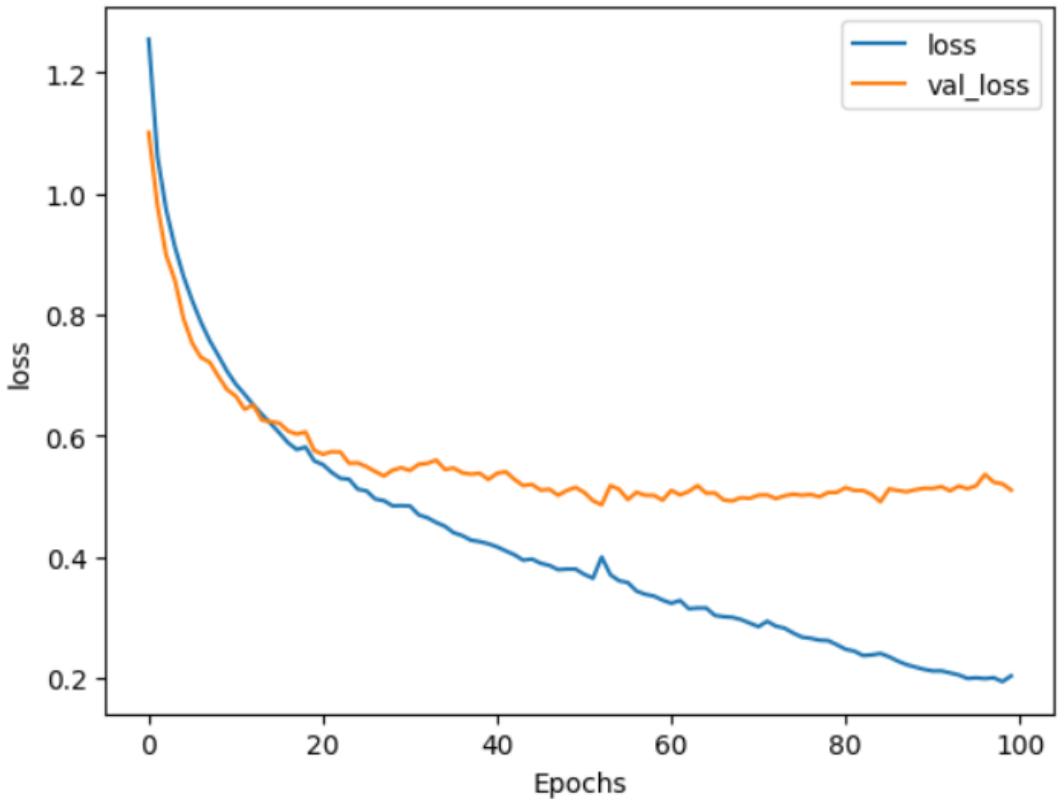

Figure 9. Loss Value Over Time

Chapter 5

Discussion and Conclusion

The use of recommender systems based on online reviews in the airline industry can provide many benefits for both airlines and customers. These systems can make personalized recommendations for flights, destinations, and other services, which can help to improve customer satisfaction and drive business. Additionally, they can provide valuable insights into customer preferences and feedback, which can help airlines to improve their products and services. However, there are also challenges associated with using these systems. One key challenge is the need to accurately analyze and interpret customer feedback from online reviews. NLP techniques can be used to help with this task, but applying these techniques effectively can be difficult, and there may be limitations to their accuracy and effectiveness.

Another challenge is the need to collect and manage large amounts of data from online reviews. This can require significant resources and expertise and may require the use of specialized tools and platforms. Additionally, there may be privacy and ethical considerations to consider when using customer data from online reviews. Despite these challenges, many airlines are already using recommender systems based on online reviews to improve customer satisfaction and drive business. Overall, the results of these and other cases suggest that recommender systems based on online reviews can be effective in the airline industry. By using these systems, airlines can provide personalized recommendations that can improve customer satisfaction and drive business.

In this study we tried to investigate the use of recommender systems with online customer reviews to better recommend ancillary products to customers. There was a gap in the literature where there was no linkage between online reviews and ancillary service products recommendations in the airline industry. Table 7 shows the performances of three models comparing their accuracy, RMSE, recall and F-1 scores. We found that ANN performed the highest accuracy rate of 0.85 with our

data when used in our recommender system. SVM and a decision tree-based algorithm XGBoost gave similar accuracy scores of 0.79. The biggest challenge we faced in our study was the manual labelling of our data by reading the text content. Since there was no dataset available containing the ancillary service choice of customers who wrote reviews, we had to label the review set using our domain knowledge. Manual labor of labelling of the data made it hard to have more data. This might have affected the performance of deep learning algorithms.

Table 7

Comparison of Model Performances

Model	**Accuracy**	**RMSE**	**Recall**	**F-1 Score**
SVM	0.79	0.23	0.79	0.79
XGBOOST	0.79	0.32	0.80	0.80
ANN	0.85	1.7	0.76	0.42

In conclusion, we tried one different machine learning algorithm from classification, one from tree based and one from neural network algorithms using 600 rows of manually labelled review ancillary product data which was gathered from online airline customer reviews. Then we gathered results using evaluation metrics to see how they affect the accuracy of our recommendation model. The artificial neural network model using Tensorflow and Keras performed better in terms of accuracy.

For the future works the ancillary products could include other services such as Wi-Fi internet on board, lounge access etc. using larger dataset. To have a share of ancillary products market enhancing recommendations could be a big opportunity for airline companies. Other possible extension of this study can be done via adding translation ability into the proposed pipeline to process languages other than English. For the industry part airlines can implement such systems to increase the quality of both their pre-flight and in-flight experiences for customers.

REFERENCES

Lu, Y., Dong, R. & Smyth, B. 2018. Coevolutionary Recommendation Model: Mutual Learning between Ratings and Reviews. Proceedings of the 2018 World Wide Web Conference, hlm. 773-782.

G. Adomavicius and A. Tuzhilin, "Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions," in *IEEE Transactions on Knowledge and Data Engineering*, vol. 17, no. 6, pp. 734-749, June 2005, doi: 10.1109/TKDE.2005.99.

Isinkaye, Folasade & Folajimi, Yetunde & Ojokoh, Bolanle. (2015). Recommendation systems: Principles, methods, and evaluation. Egyptian Informatics Journal. 16. 10.1016/j.eij.2015.06.005.

Pazzani, M. J. & Billsus, D. 2007. Content-Based Recommendation Systems. Dlm. (pnyt.). *The Adaptive Web*, hlm. 325-341. Springer.

Dadoun, A., Defoin-Platel, M., Fiig, T. *et al.* How recommender systems can transform airline offer construction and retailing. *J Revenue Pricing Manag* **20**, 301–315 (2021) https://doi.org/10.1057/s41272-021-00313-2

Aciar, S., Zhang, D., Simoff, S. and Debenham, J. (2007) Informed recommender: basing recommendations on consumer product reviews. *IEEE Intell. Syst.*, 22, 39–47.

Diao, Q., Qiu, M., Wu, C., Smola, A.J., Jiang, J. and Wang, C. (2014) Jointly Modeling Aspects, Ratings and Sentiments for Movie Recommendation (JMARS). In *Proc. SIGKDD2014, New York, NY, USA, August 24–27 2014*, pp. 193–202. ACM, New York, USA.

Chelliah, M. and Sarkar, S. (2017) Product Recommendations Enhanced with Reviews. In *Proc. RecSys2017, Como, Italy, August 27–31*, pp. 398–399.

ACM, New York, USA.

Danilova, V. & Ponomarev, A. 2016. Hybrid Recommender Systems: The Review of State-of-the-Art Research and Applications. *PROCEEDING OF THE 20TH CONFERENCE OF FRUCT ASSOCIATION*

A. Parasuraman, A. Zeithaml, V., L. Berry, 1988. SERVQUAL: a multiple item scale for measuring consumer perceptions of service quality. Journal of Retailing 64, 12–40.

F. Pakdil, O. Aydin, Expectations, and perceptions in airline services: An analysis using weighted SERVQUAL scores, Journal of Air Transport Management 13 (4) (2007) 229-237 doi:10:1016/j:jairtraman:2007:04:001.

D. M. A. Baker, "Service quality and customer satisfaction in the airline industry: A comparison between legacy airlines and low-cost airlines," Am. J. Tour. Res., vol. 2, no. 1, 2013.

Korfiatis, Nikos & Poulos, Marios. (2013). Using online consumer reviews as a source for demographic recommendations: A case study using online travel reviews. Expert Systems with Applications. 40. 5507-5515. 10.1016/j.eswa.2013.03.046.

T.H.K. Tusar, T. Islam, A Comparative Study of Sentiment Analysis Using NLP, and Different Machine Learning Techniques on US Airline Twitter Data. CoRR abs/2110.00859 (2021)

Khaturia, Divisha & saxena, Aditi & Basha, Muzamil & Iyenger, N Ch Sriman Narayana & Caytiles, Ronnie. (2018). A Comparative study on Airline Recommendation System Using Sentimental Analysis on Customer Tweets. International Journal of Advanced Science and Technology. 111. 107-114. 10.14257/ijast.2018.111.10.

Lu, H., Cao, J., Tan, Y., Xiao, Q. (2017). Auxiliary Service Recommendation for Online Flight Booking. In:, *et al.* Web Information Systems Engineering – WISE 2017. WISE 2017. Lecture Notes in Computer Science (), vol 10570.

Springer, Cham. https://proxy.bau.edu.tr:2102/10.1007/978-3-319-68786-5_35

Mottini, A., Lhéritier, A., Acuna-Agost, R., & Zuluaga, M. A. (2018) Understanding Customer Choices to Improve Recommendations in the Air Travel Industry. *RecTour 2018*

P. K. Jain, R. Pamula, S. Ansari, D. Sharma, and L. Maddala, "Airline recommendation prediction using customer generated feedback data," *2019 4th International Conference on Information Systems and Computer Networks (ISCON)*, Mathura, India, 2019, pp. 376-379, doi: 10.1109/ISCON47742.2019.9036251.

Srinivas, S., & Ramachandiran, S. (2020). Discovering airline-specific business intelligence from online passenger reviews: an unsupervised text analytics approach. arXiv preprint arXiv:2012.08000.

M. Heidari and S. Rafatirad, "Using Transfer Learning Approach to Implement Convolutional Neural Network model to Recommend Airline Tickets by Using Online Reviews," *2020 15th International Workshop on Semantic and Social Media Adaptation and Personalization (SMA*, Zakynthos, Greece, 2020, pp. 1-6, doi: 10.1109/SMAP49528.2020.9248443.

Y. Zhang, W. Luo, M. Li, and T. Chen, "Contextualized Recommendation of Aviation Ancillary Services Based on Passenger Portraits," in *IEEE Transactions on Aerospace and Electronic Systems*, vol. 58, no. 6, pp. 5078-5088, Dec. 2022, doi: 10.1109/TAES.2022.3193894.

Morde, V. (2021). "XGBoost Algorithm: Long May She Reign!". Towards data science. Retrieved from https://towardsdatascience.com/https-medium-comvishalmorde-xgboost-algorithm-long-she-may-rein-edd9f99be63d.

Bekesiene, S., Smaliukiene, R., & Vaicaitiene, R. (2021). Using Artificial Neural Networks in Predicting the Level of Stress among Military Conscripts. *Mathematics*, *9*(6), 626. https://doi.org/10.3390/math9060626

Sun, J., Liu, Y., Cui, J. *et al.* Deep learning-based methods for natural hazard named entity recognition. *Sci Rep* 12, 4598 (2022). https://doi.org/10.1038/s41598-022-08667-2

Tianqi Chen and Carlos Guestrin. 2016. XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (KDD '16). Association for Computing Machinery, New York, NY, USA, 785–794. https://doi.org/10.1145/2939672.2939785

Timothy L Keiningham, Bruce Cooil, Lerzan Aksoy, Tor W Andreassen, and Jay Weiner. 2007. The value of different customer satisfaction and loyalty metrics in predicting customer retention, recommendation, and share-of-wallet. *Managing Service Quality: An International Journal* 17, 4 (2007), 361–384.