Fırat Üniversitesi Müh. Bil. Dergisi
37(2), 909-918, 2025

Araştırma Makalesi
https://doi.org/10.35234/fumbd.1708787

A Generative AI-Driven Analysis of Airline Passenger Feedback: Revealing What Matters Most

Hakan KOÇAK[1*]
[1] Meteoroloji Genel Müdürlüğü, Uzaktan Algılama Şube Müdürlüğü, Ankara, Türkiye
[*1] hkocak@gmail.com

(Geliş/Received: 29/05/2025; Kabul/Accepted: 29/09/2025)

Abstract: The airline industry, characterized by intense competition, relies heavily on customer satisfaction to assess strengths and weaknesses. Online passenger reviews provide a rich source of data, capturing customers' opinions, expectations, and emotions. Analyzing this feedback helps airlines identify areas for improvement and understand what matters most to passengers. This study employs a zero-shot prompting approach using Google Gemini to interpret Turkish Airlines reviews from Trip Advisor in 2024, demonstrating the model's effectiveness without domain-specific fine-tuning. The findings highlight factors influencing perceived service quality, performance, and value, illustrating the potential of generative AI in specialized customer sentiment analysis and its practical applications in the airline industry.

Key words: Customer review analysis, natural language processing, large language models, generative AI, google gemini.

Üretken Yapay Zekâ Destekli Bir Havayolu Yolcu Geri Bildirimi Analizi: En Önemli Olanı Ortaya Çıkarmak

Öz: Yoğun rekabetin yaşandığı havayolu endüstrisi, müşteri memnuniyetine büyük ölçüde dayanarak güçlü ve zayıf yönlerini değerlendirir. Çevrimiçi yolcu yorumları, müşterilerin görüşlerini, beklentilerini ve duygularını yansıtan zengin bir veri kaynağı sunar. Bu geri bildirimlerin analizi, havayollarının geliştirilmesi gereken alanları belirlemesine ve yolcular için en önemli olanı anlamasına yardımcı olur. Bu çalışma, 2024 yılında Trip Advisor web sitesinden toplanan Türk Hava Yolları yorumlarını yorumlamak için Google Gemini'nin sıfırdan yönlendirme (zero-shot prompting) yaklaşımını kullanarak, alan spesifik ince ayara gerek kalmadan modelin etkinliğini göstermektedir. Bulgular, algılanan hizmet kalitesini, performansını ve değeri etkileyen faktörleri ortaya koymakta ve üretici yapay zekânın uzmanlaşmış müşteri duygu analizindeki potansiyelini ve havayolu endüstrisindeki uygulamalarını göstermektedir.

Anahtar kelimeler: Müşteri yorumları analizi, doğal dil işleme, büyük dil modelleri, üretken yapay zekâ, google gemini.

1. Introduction

In the airline industry, customer reviews provide direct insights into passenger experiences, expectations, and perceptions of service quality. Analyzing these reviews allows airlines to pinpoint operational strengths and weaknesses, improve service delivery, enhance customer satisfaction, and make informed strategic decisions [1,2]. The volume of feedback makes manual analysis impractical, highlighting the need for automated, scalable Sentiment Analysis solutions.

Sentiment Analysis (SA), a branch of Natural Language Processing (NLP), automatically identifies whether text expresses positive, negative, or neutral sentiment. Also known as Opinion Mining (OM), SA transforms large volumes of unstructured data from social media, review platforms, and forums into actionable insights [3].

Early approaches based on bag-of-words or lexical features often lacked accuracy. Machine learning classifiers, such as SVM and Naive Bayes, improved predictions, while modern Large Language Models (LLMs) like BERT and GPT enable high-accuracy, large-scale sentiment analysis [4,5]. Yet, few studies have applied these models specifically to airline customer feedback [6].

This study evaluates Google Gemini, an advanced LLM, for analyzing airline passenger reviews. Using a zero-shot prompting strategy, it assesses Gemini's ability to interpret sentiment without domain-specific fine-tuning. The research demonstrates how generative AI can turn customer feedback into actionable insights, support operational improvements, enhance passenger satisfaction, and provide a competitive edge, offering airlines practical guidance for data-driven decision-making in a customer-focused industry.

[*] Sorumlu yazar: hkocak@gmail.com Yazarların ORCID Numarası: [1] 0000-0003-2491-327X

2. Related Work

Customer reviews are critical for airlines' reputation, operations, and strategic decisions. Several studies have applied machine learning and text mining to analyze such feedback. Wu and Gao [7] employed SVM on Kaggle's 2015 Twitter dataset, achieving 91.86% accuracy by excluding neutral entries. Patel et al. [8] compared Naïve Bayes, SVM, Decision Tree, and BERT, with BERT outperforming others at 83% accuracy, precision, recall, and F1 score. Sezgen et al. [9] used SVD on over 5,000 TripAdvisor reviews, demonstrating that LSA can identify factors influencing airline choice. Similarly, Siering et al. [10] analyzed airlinequality.com reviews using Naïve Bayes, SVM, and neural networks, reporting accuracies of 71–75% depending on sampling. Kumar and Zymbler [11] examined airline tweets using SVM, ANN, and CNN, finding CNN achieved the highest accuracy of 92%. Lucini et al. [12] combined LDA and logistic regression on 55,000+ reviews across 400 airlines, predicting customer recommendations with 79% accuracy. More recently, large language models have shown promise in sentiment analysis: ChatGPT effectively interprets customer sentiment [13] and performs comparably to BERT and other state-of-the-art models [14].

3. Literature Review

Sentiment analysis classifies text as positive, negative, or neutral, usually beginning with preprocessing steps such as removing punctuation, symbols, stop words, and stems. The processed text is then represented numerically, often through TF-IDF, to form a vector space model. Analysis can be performed at the document, sentence, or aspect level, enabling researchers to capture overall polarity, classify shorter texts, or extract opinions on specific features.

Two main approaches are commonly applied: lexicon-based and machine learning. Lexicon-based techniques use predefined polarity dictionaries such as SentiWordNet [5, 15] or TF-IDF weighting [16], but they often struggle with linguistic subtleties like slang, sarcasm, and negation. Machine learning methods, including Naïve Bayes and Support Vector Machines, generally deliver higher accuracy but depend heavily on large, high-quality datasets [17–18]. While smaller datasets reduce processing time, they tend to compromise performance [19]. These methods may be supervised, semi-supervised, or unsupervised, typically employing Bag-of-Words features, with effectiveness measured through precision, recall, and F-measure [5].

More recently, transformer-based models such as GPT and Google Gemini have shown strong potential by analyzing large volumes of unstructured text, including financial news, to inform decisions in areas like investment strategies and risk management [20–23]. Across both traditional and modern techniques, sentiment analysis generally follows five key stages: collecting and preprocessing data, preparing text, detecting subjectivity, classifying polarity, and presenting results through visualizations, which translate unstructured opinions into actionable insights [24].

4. Methodology

The code for this study was developed in Python 3 and executed on Kaggle Notebook. To utilize Large Language Models (LLMs) in Google Gemini Pro, an API key was obtained through Google AI, with a pay-as-you-use pricing model applying. Google Gemini was chosen for its advanced generative and contextual understanding capabilities, enabling zero-shot sentiment analysis and topic standardization with nuanced handling of domain-specific language, outperforming or matching BERT and GPT in similar tasks. Gemini Pro features seamless integration with topic standardization, leveraging representative documents from BERTopic.

While BERT excels at contextual embeddings and GPT at text generation, Gemini uniquely combines deep understanding of text with generative abilities, making it ideal for zero-shot sentiment and topic classification

The study followed a series of steps, including obtaining an online dataset from TripAdvisor, cleaning and preprocessing the data using NLTK and Spacy, conducting sentiment analysis with VADER, performing topic modeling and classification with BERTopic, re-generating topic labels and classifications using Google Gemini Pro 1.5 LLMs based on BERTopic outputs, and finally visualizing and interpreting the results.

4.1 Dataset

For this study, Turkish Airlines customer reviews from TripAdvisor (Jan-Dec 2024) were collected using a Google Chrome web scraping extension tool (Instant Data Scraper). The dataset includes 1,139 unique reviews, each containing a published date, username, route, flight type, title (or subject), date flown, and review text, all stored in a CSV file for analysis. The dataset was then cleaned by removing entries with missing information or reviews written in languages other than English, as well as eliminating unnecessary fields not relevant to the study.

4.2 Pre-processing

The preprocessing process includes tokenization, part-of-speech tagging, stop-word removal, stemming, noun extraction, and noun filtering. Tokenization identifies individual words, which are then tagged with syntactic roles (e.g., verb, subject, adjective). Stop words such as “a,” “an,” “and,” “but” “the,” “that,” “of,” and “from” are removed. Afterward, prefixes and suffixes are eliminated from words, leaving only their stems. Stemming consolidates word derivatives with similar meanings into a single concept, such as reducing “flying” and “flew” to “fly.”

The removal of stop words and lemmatization of each review was performed using the Natural Language Toolkit (NLTK) and the Spacy libraries in Python. After removing non-alphabetic characters, stop words, and white spaces at the end of the reviews, all the review texts were converted into documents for further analysis.

4.3 Sentiment Analysis of Customer Reviews

VADER (Valence Aware Dictionary for Sentiment Reasoning) is employed in this study for sentiment analysis. VADER is a simple, rule-based model designed for general sentiment analysis, assessing both the polarity and intensity of emotions in unlabeled text data. Integrated into the NLTK package for Python, VADER improves upon traditional sentiment lexicons like LIWC (Linguistic Inquiry and Word Count) by offering greater adaptability across various domains and a heightened sensitivity to sentiment in social media contexts. VADER is an effective tool for analyzing sentiment in online customer reviews, news headlines, and social media posts.

The pre-processed reviews are then analyzed using the VADER library (vaderSentiment) in Python to generate sentiment scores for each review, categorizing them as negative, neutral, or positive. Sentiment scores in VADER range from -1 to +1, where -1 represents the most negative sentiment, 0 denotes neutrality, and +1 signifies the most positive sentiment. These sentiment scores are added to the original dataset for further analysis.

4.4 Topic Modeling

For sentiment classification, the topic modeling method BERTopic, as implemented in Python, was used to generate individual topic classes and assign each review to a specific topic class. The classification process begins by fine-tuning BERTopic on a pre-labeled dataset to tailor it for sentiment analysis. After fine-tuning, the model is applied to unlabeled customer review data to compute sentiment scores.

BERTopic is an advanced library that enhances traditional topic modeling by leveraging machine learning algorithms to cluster documents based on deep semantic representations. It excels at handling noisy tokens, such as stop words, and uses them to improve the representation of customer experiences.

Sentence embeddings were generated using Google’s Universal Sentence Encoder – Multilingual (USE) sentence transformers for reviews with VADER sentiment scores less than +0.1 to identify the most common customer complaints. This process resulted in 881 reviews being selected, representing the issues that are troubling customers the most.

VADER assigned scores between -1 and +1. Reviews scoring below 0.1 were selected for topic modeling to capture the most dissatisfied customers, as scores near 0 are indicative of neutral to slightly negative sentiment. This slightly relaxed threshold captures both clearly negative and borderline negative reviews, ensuring that subtle expressions of dissatisfaction or service-related concerns are included in the analysis. This threshold made it possible to capture predominantly negative experiences for topic modeling and issue prioritization.

To classify the 881 reviews, they were embedded into BERTopic using Google’s Universal Sentence Encoder (USE) to further assess the areas (i.e., topics) where customers were most dissatisfied. Reviews with sentiment scores of less than +0.1 were then clustered by creating and optimizing a BERTopic model using the USE embeddings generated in the previous step. BERTopic clustered 881 negative/neutral reviews into 21 topics. This process resulted in the identification of 21 distinct topics through the trained model.

Table 1. BERTopic-identified clusters of negative airline reviews, including topic names, representative words, and sample documents.

#	Topic	Count	Name	Representation	Representative_Docs
0	-1	379	-1_airline bad_bad	['airline bad', 'bad	[bad airline experience ve 40 year wife busine...
1	0	113	0_airline istanbul	['airline istanbul	[good airline good compliment inflight staff e...
2	1	110	1_airline terrible	['airline terrible	[staff airline rude mannerless staff come pay ...
3	2	33	2_terrible customer	['terrible customer	[bad customer service need help customer servi...
4	3	29	3_airline change_airline	['airline change',	[bad customer service imagine book advance feb...
5	4	29	4_horrible delay	['horrible delay hour',	[delay result miss connect 8 hour delay arriva...
6	5	24	5_customer service	['customer service	[bad airline bad customer service ground airpo..
7	6	23	6_pay seat	['pay seat selection',	[purchase seat 39 app seat single seat plan no...
8	7	20	7_luggage instanbul	['luggage instanbul	[terrible customer service lose bag receive ti...
9	8	19	8_meal airline	['meal airline horrible',	[horrible food disgusting cabin crew unfrienly...
10	9	12	9_luggage chase	['luggage chase	[airline rent car review objective frustration..
11	10	11	10_meal terrible	['meal terrible	[truly unpleasant staff sullen unhelpful ear p..
12	11	11	11_horrible baggage	['horrible baggage	[recent trip australia hear airline vote good ..
13	12	10	12_pathetic airline	['pathetic airline bad',	[arrive airport time waste lot time try find r...
14	13	10	13_disappointing	['disappointing	[subject extremely disappointing experience ai..
15	14	10	14_bad airline	['bad airline	[change transfer suddenly schedule ruin compla...
16	15	8	15_airline ticket	['airline ticket office',	[family plan trip year 1st problem cancel book..
17	16	8	16_point airline	['point airline cancel',	[good september 2023 cancel ticket family medi...
18	17	8	17_pay seat	['pay seat choice',	[pay select seat 74 year old mom knee surgery ..
19	18	7	18_refundable bad	['refundable bad	[bad airline change outbound ticket multiple t...
20	19	7	19_bad airline	['bad airline run',	[bad airline change outbound ticket multiple t...

Table 1 presents sample output from the BERTopic model used in this study. The Topic field represents the unique ID assigned to each topic. Count indicates the number of documents assigned to that topic. Name is the automatically generated topic name, while Representation lists the key words that summarize the topic. Representative_Docs provides a subset of documents that best exemplify the topic, illustrating the semantic content captured by the model. Figure 1 shows sentence embeddings by topic labels with each topic label represented with a different color.

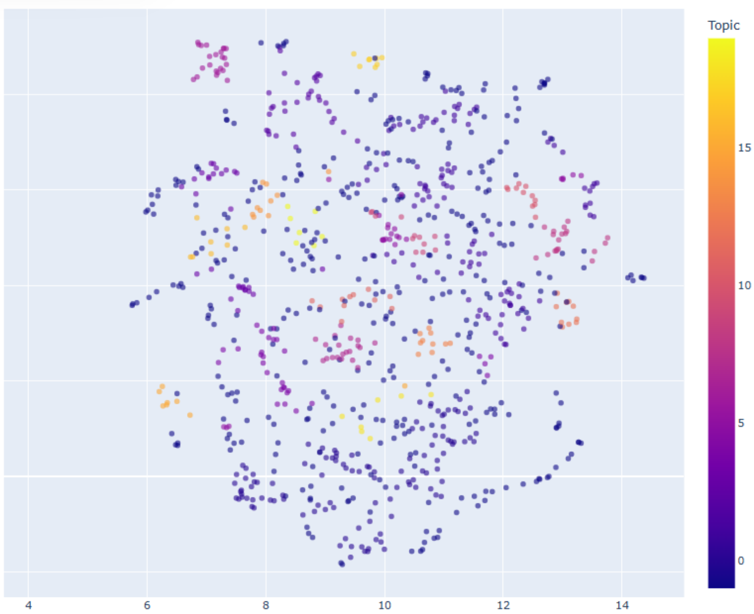

Figure 1. Sentence embeddings by topic labels colored by topic labels.

4.5 Re-generating Topic Labels with Gemini Pro LLM, Based on Bertopic Labels

The topic labels generated by the BERTopic model were often not meaningful or business-friendly. To enhance interpretability, they were standardized using Google Gemini Pro, which incorporated domain knowledge from both the BERTopic outputs and representative documents. Gemini Pro, an advanced large language model (LLM), is designed to capture the nuances of human language, making it well-suited for tasks such as sentiment analysis and topic labeling. In this study, Gemini Pro was engaged solely through natural language instructions, without any task-specific fine-tuning or additional training.

The input was prepared by combining topic representative words with the top 200 tokens from representative documents into a Pandas data frame, which was then used with the designed prompts for Gemini Pro. Example prompts include:

'''

As a Senior Customer Experience Executive in aviation, your task is to identify customer pain points from the provided reviews.

Here are your instructions:

Create one topic label (maximum 3 words) that reflects the themes in the input list.
Give priority to the first 10 topic words generated by BERTopic.
Output only the topic headline labels in list format.
Do not include supporting keywords.
Do not rephrase or alter the original label wording.
'''

The model then generated more coherent, business-oriented topic labels that were directly applicable to decision-making processes, addressing the limitations of the initial BERTopic outputs. For airlines in particular, clear and interpretable topic labels are essential, as they enable managers to identify customer concerns more efficiently, prioritize operational improvements, and develop strategies that enhance passenger satisfaction. While the use of Gemini Pro incurs costs due to API token consumption, the insights gained significantly outweigh these expenses by providing actionable intelligence for business applications.

The usefulness of the Gemini-generated labels was validated through a manual review by the author. A subset of topics and documents was examined to ensure that (1) the labels accurately reflected the semantic content of the clustered reviews, (2) the labels were clear, interpretable, and business-friendly, and (3) the insights were easy to communicate to airline management teams for strategic decision-making. This evaluation confirmed that the Gemini-generated labels aligned with human judgment and improved the overall interpretability of results.

5. Results

As an initial step, descriptive graphs were generated to explore the dataset. Figure 2 presents a histogram of sentence lengths in the reviews, showing that most sentences contain 10–50 words, with a few shorter than 10 or exceeding 200 words.

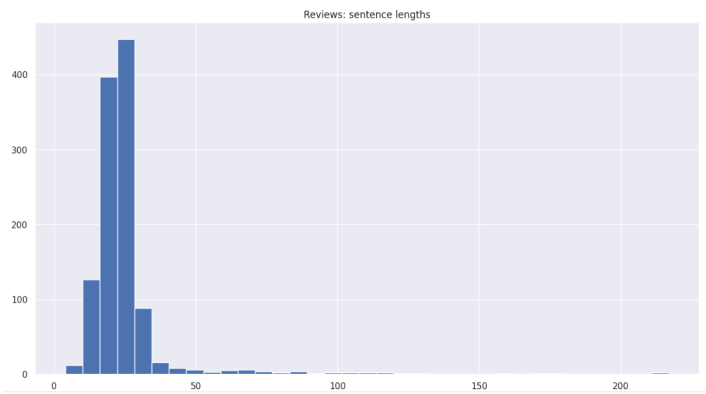

Figure 2. Histogram of review sentence lengths, showing most sentences are 10–50 words.

Figure 3 shows the distribution of flight types among customers in the dataset. Most reviews are from passengers on international flights, while domestic flight comments are relatively few. The plot also indicates that international flights outside Europe and North America considerably outnumber those associated with these regions.

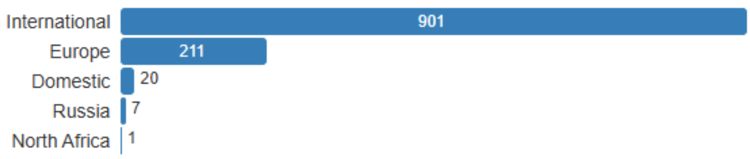

Figure 3. Distribution of flight types; majority are international, highlighting dataset bias.

A word cloud was generated to highlight the most frequent terms in the processed text. As shown in Figure 4, most comments pertain to flight experience and service issues. The most common words include: "flight," "service," "seat," "customer," "change," "bad," "good," "gate," "time," "cancel," "ticket," "hour," "terrible," "delay," "baggage," "food," "experience," "luggage," "boarding," "staff," "miss," and "lose."

Figure 4. Word cloud of frequent terms; common words relate to service, flight, seat, and baggage.

Following the descriptive statistics, customer sentiment analysis was performed, and the distribution of positive, negative, and neutral sentiments is shown in Figure 5. Negative sentiments dominate the dataset, comprising 70.4% of the total, while positive and neutral sentiments account for 23.6% and 6.0%, respectively, as illustrated in the pie chart.

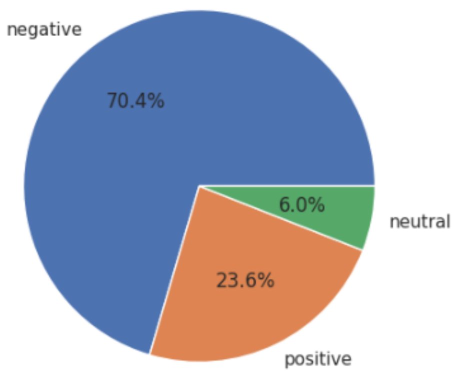

Figure 5. Pie chart of sentiment distribution (Negative 70.4%, Positive 23.6%, Neutral 6%).

Figure 6 presents a bar chart of binned sentiments ranging from -1 to +1 in 0.25 increments. The highest frequency occurs in the most negative bin, and as sentiment scores increase, their frequency decreases. Beyond 0.25, frequencies begin to rise again on the positive side, but at a slower rate compared to the negative side.

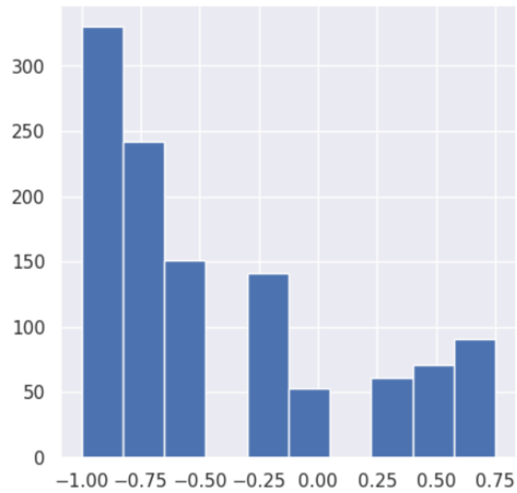

Figure 6. Histogram of VADER compound scores, showing skew toward negative sentiment.

BERTopic labels were standardized using Google Gemini Pro, as the initial model output was incoherent. The LLM processed the representative documents and a seed list of domain-specific topics to generate clear, business-friendly labels. Through prompt fine-tuning, each BERTopic label was matched with a standardized, AI-generated topic.

Table 2. Example Customer Reviews with BERTopic Topics and Gemini Pro LLM Labels.

Document	Topic	Representation	Probability	Representative document	LLM input	Gemini topic label
terrible deliberately delay luggage know come ir..	1	[staff airline rude mannerless staff come pay ...	0.663	False	airline terrible customer terrible customer se..	Customer Service Issues
istanbul delay 15 hour run denpasar lose luggage..	7	[luggage istanbul compensation, luggage leave.	0.982	False	luggage istanbul compensation luggage leave..	Missing Luggage Issues
good seat good lounge good regular food b...	-1	[bad airline experience ve 40 year wife busine...	0.000	False	airline bad bad airline experience airline air...	Customer Service Issues
buy ticket attempt modify return date website ...	16	[good september 2023 cancel ticket family medi...	0.950	True	point airline cancel lose return ticket cancel...	Cancellations Issues

Table 2 presents sample output from BERTopic topic modeling and subsequent re-labeling using Google Gemini Pro's LLM. It includes the Document, representing the original text or customer review being analyzed, and the Topic, which is the initial topic ID assigned by BERTopic. Representation field lists the key words that summarize the topic, while Representative_Docs shows a subset of documents that best exemplify the topic. Probability reflects the model's confidence that the document belongs to the assigned topic. The llm_input field contains the prompt provided to Gemini Pro to generate business-friendly labels, and gemini_topic_labels shows the final topic label output by the LLM, designed to be coherent, interpretable, and actionable for business decision-making.

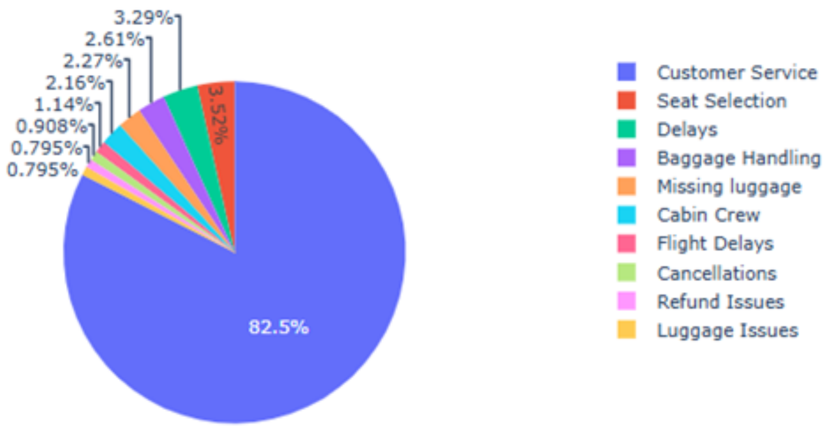

Figure 7. Turkish Airlines customer pain-points

Figure 7 shows customer reviews categorized by topic labels from Google Gemini, summarizing Turkish Airlines' main pain points. The 21 topic labels identified by BERTopic were consolidated into 10 standardized topics.

The largest share of complaints, 82.5%, relates to Customer Service, highlighting it as the primary source of customer dissatisfaction. Luggage issues follow at 3.52%, covering lost or damaged baggage. Flight delays account for 3.29%, flight cancellations 2.61%, seat selection issues 2.27%, and cabin crew complaints 2.16%, often related to service or attitude. Refund issues are the least frequent at 1.14%.

These results indicate that improving customer service should be Turkish Airlines' top priority. Enhanced training and process improvements are necessary to address complaints effectively. Luggage handling also requires attention to reduce lost or damaged baggage. Additionally, minimizing flight delays and cancellations, along with better communication and compensation, can improve overall passenger satisfaction. Analyzing the specific causes of customer service complaints can provide targeted insights for further improvements.

6. Conclusion and Discussion

This study demonstrates the practical potential of AI-powered sentiment analysis for understanding airline customer feedback. By integrating VADER sentiment scoring, BERTopic topic modeling, and Google Gemini Pro's zero-shot relabeling, the study successfully identified key customer pain points, particularly those related to service quality, responsiveness, and communication. These findings highlight the ability of generative AI to extract actionable insights from unstructured text without requiring domain-specific fine-tuning.

The effectiveness of Google Gemini Pro can be attributed to its extensive pre-training on diverse text corpora, enabling nuanced understanding of context, idiomatic expressions, and subtle sentiment cues. This capability allowed the model to capture customer concerns that may be overlooked by traditional sentiment analysis methods, emphasizing the importance of LLMs in specialized applications such as airline service evaluation.

Compared to previous approaches relying solely on lexicon-based or supervised methods, this zero-shot, topic-driven approach offers a scalable and flexible alternative for analyzing large volumes of customer feedback. The findings suggest that airlines can leverage generative AI not only to monitor sentiment trends but also to prioritize service improvements and enhance customer experience proactively.

It is worth noting that although the use of Google Gemini Pro involves token-based costs, these costs are modest relative to the value of actionable insights generated. The ability to produce coherent, interpretable, and business-oriented topic labels enables airlines to prioritize improvements efficiently, making the benefits of using the model substantially greater than the associated expenses. Therefore, cost does not pose a significant barrier to practical application, even for larger datasets or repeated analyses.

However, the study has several limitations. The analysis focused on a single airline's dataset, which may limit generalizability to other carriers or regions. Additionally, while zero-shot methods reduce the need for labeled data, their performance can vary depending on the complexity and specificity of the feedback. Human validation and cross-model comparisons remain essential to ensure robustness and reliability.

Future research could expand this framework to multi-airline datasets, incorporate comparative evaluations with other LLMs, and explore human-in-the-loop validation strategies. Further investigation into the model's ability to detect subtle emotional cues could inform more personalized customer engagement and proactive service interventions.

The analysis revealed that the broad 'Customer Service' category encompassed a variety of issues, including staff attitude, responsiveness, communication breakdowns, and problem resolution challenges. While this study aggregated these issues into a single category to illustrate the methodology, a more granular, aspect-based sentiment analysis could provide deeper, more actionable insights for airlines. Future research should consider disaggregating customer service complaints to identify specific operational areas that require targeted improvement. Moreover, longitudinal datasets would be highly valuable to capture evolving trends, seasonal effects, and the impact of operational changes, enabling airlines to monitor service quality over time and respond proactively to shifting customer expectations.

Overall, the study illustrates that generative AI, particularly large language models like Google Gemini Pro, can significantly enhance the analysis of customer feedback, providing airlines with actionable insights and a foundation for data-driven service improvements.

References

[1] Al-Otaibi S T, Rasheed A. Review and Comparative Analysis of Sentiment Analysis Techniques. Informatica. 2022; 46 (1): 33–44

[2] Hai Z, Kim K C J ve Yang C. Identifying features in opinion mining via intrinsic and extrinsic domain relevance. IEEE Trans Knowl Data Eng. 2014; 26 (3): 634.

[3] Cui J, Wang Z ve Cambria E. Survey on sentiment analysis: evolution of research methods and topics. Artificial Intelligence Review. 2023; (56): 8469–8510.

[4] Taherdoost E ve Madanchian M. Artificial Intelligence and Sentiment Analysis: A Review in Competitive Research. Computers, 2023; 12 (37).

[5] Khan M T, Durrani M ve Ali A. Sentiment analysis and the complex natural language. Complex Adapt Syst Model. 2016; 4 (2).

[6] Ramadhan F, Sitanggang A, Wibawa J ve Radliya N. Implementation of Digital Marketing Strategy with Chatbot Technology. Int. J. Artif. Intell. Res. 2023; 7 (2): 132.

[7] Wu S ve Gao Y. Machine Learning Approach to Analyze the Sentiment of Airline Passengers' Tweets. Transp. Res. Rec. 2023; 2678 (1).

[8] Patel A, Oza P ve Agrawal S. Sentiment Analysis of Customer Feedback and Reviews for Airline Services using Language Representation Model. Procedia Comput. Sci. 2023; 218: 2459–2467.

[9] Sezgen E, Mason K J ve Mayer R. Voice of airline passenger: A text mining approach to understand customer satisfaction..J. Air Transp. Manag. 2019; 77: 65–74..

[10] Siering M, Deokar A V ve Janze C. Disentangling consumer recommendations: Explaining and predicting airline recommendations based on online reviews. Decis. Support Syst. 2018; 107: 52–63.

[11] Kumar S ve Zymbler M. A. Machine learning approach to analyze customer satisfaction from airline tweets. J. Big Data. 2019; 6 (62).

[12] Lucini F, Tonetto L M, Fogliatto F S ve Anzanello M J. Text mining approach to explore dimensions of airline customer satisfaction using online customer reviews. J. Air Transp. Manag. 2020; 83 (4).

[13] Sudirjo F, Diantoro K, Al-Gasawneh J -A, Azzaakiyyah H K ve Ausat A M A. Application of ChatGPT in Improving Customer Sentiment Analysis for Businesses. Jurnal Teknologi Dan Sistem Informasi Bisnis. 2023; 5 (3): 283–288.

[14] Wang Z, Xie Q, Ding Z, Feng Y ve Xia R. Is ChatGPT a good sentiment analyzer? A preliminary study. arXiv preprint. 2023. arXiv 2304.04339.

[15] Agarwal B, Mittal N, Bansal P ve Garg S. Sentiment Analysis Using Common-Sense and Context Information. Journal of Computational Intelligence and Neuroscience. 2015; 6:715730.

[16] Das B ve Chakraborty S. An Improved Text Sentiment Classification Model Using TF-IDF and Next Word Negation. 2018; 10.48550/arXiv.1806.06407.

[17] Hassan A U, Hussain J, Hussain M, Sadiq M ve Lee S. Sentiment Analysis of Social Networking Sites (SNS) Data Using Machine Learning Approach for the Measurement of Depression. In: International Conference on Information and Communication Technology Convergence (ICTC). Jeju, South Korea: IEEE, 2017.

[18] Akter A ve Aziz M T. Sentiment Analysis on Facebook Group Using Lexicon Based Approach. In: 3rd International Conference on Electrical Engineering and Information Communication Technology (ICEEICT); September 2016.

[19] Dhaoui C, Webster C M ve Tan L P. Social Media Sentiment Analysis: Lexicon Versus Machine Learning. Journal of

Consumer Marketing. 2017; 34 (6): 480-488.

[20] George A S ve George A H. A review of chatgpt ai's impact on several business sectors. Partners Universal International Innovation Journal, 2023; 1 (1): 9–23.

[21] Radford A, Narasimhan K, Salimans T ve Sutskever I. Improving language understanding by generative pre-training. pre-print, 2018.

[22] Chen H, De P, Hu Y J ve Hwang B H. Wisdom of crowds: The value of stock opinions transmitted through social media. The Review of Financial Studies. 2014; 27: 1367–1403.

[23] Tetlock P C. Giving content to investor sentiment: The role of media in the stock market. The Journal of Finance. 2007; 62 (3): 1139–1168.

[24] Singh N ve Jaiswal U C. Sentiment Analysis Using Machine Learning: A Comparative Study. ADCAIJ (Advances in Distributed Computing and Artificial Intelligence Journal) Regular Issue. 2023; 12 (1): 2255-2863.