Araştırma Makalesi | *Research Article* DOI: 10.29157/etusbed.1711500

Sentiment Analysis of Online Customer Reviews Using Modern Natural Language Processing Approaches

Modern Doğal Dil İşleme Yaklaşımları Kullanarak Çevrimiçi Müşteri Yorumlarının Duygu Analizi

Kutan KORUYAN[1]
(Sorumlu Yazar-Corresponding Author)

[1] *Dokuz Eylül Üniversitesi, Yönetim Bilişim Sistemleri Bölümü, Enformasyon Teknolojisi Anabilim Dalı, İzmir, Türkiye.*
kutan.koruyan@deu.edu.tr

Geliş Tarihi/Received: 01.06.2025
Kabul Tarihi/Accepted: 09.09.2025

Atıf /Cite this article:
Koruyan, K. (2025). Sentiment Analysis of Online Customer Reviews Using Modern Natural Language Processing Approaches. *Erzurum Teknik Üniversitesi Sosyal Bilimler Enstitüsü Dergisi* (23), 22-40.

Abstract

In this study, sentiment analysis was conducted using Turkish online customer reviews of four seafood restaurants based in İzmir. Modern natural language processing approaches were employed as part of the analysis, including a Turkish sentiment analysis model for BERT and multilingual models within the framework of zero-shot text classification. In addition, large language models (LLMs) such as OpenAI 4o, Gemini 2.0 Flash, and DeepSeek V3 were evaluated. Model performance was assessed using evaluation metrics, including accuracy, precision, recall, and F1 score. The findings indicate that LLMs—particularly DeepSeek V3—demonstrated high performance and could effectively process contextual representations even in unlabelled datasets.
Furthermore, sentiment trends towards restaurants over a four years were analysed to track temporal changes in customer satisfaction. This approach revealed temporal performance differences among the restaurants over time and enabled the development of sustainable improvement strategies aligned with customer expectations. The proposed method offers a fast and data-driven solution to assist managers in monitoring customer satisfaction, evaluating service quality, and identifying underlying causes of dissatisfaction, supporting strategic decision-making processes and contributing to corporate image management.

Keywords: Sentiment Analysis, BERT, Zero-Shot Text Classification, Large Language Models, Natural Language Processing.

Öz

Bu çalışmada, İzmir'de faaliyet gösteren dört deniz ürünleri restoranına yapılan Türkçe çevrimiçi müşteri yorumları kullanılarak duygu analizi gerçekleştirilmiştir. Analiz sürecinde modern doğal dil işleme yaklaşımlarından yararlanılmış; bunlar arasında BERT için kullanılan bir Türkçe duygu analizi modeli ile sıfır atışlı metin sınıflandırma yöntemi kapsamında kullanılan çok dilli modeller yer almaktadır. Bunlara ek olarak, OpenAI 4o, Gemini 2.0 Flash ve DeepSeek V3 gibi Büyük Dil Modelleri (BDM) de değerlendirmeye dahil edilmiştir. Model performansları doğruluk, kesinlik, duyarlılık ve F1 puanı metrikleri aracılığıyla ölçülmüştür. Elde edilen bulgular, özellikle DeepSeek V3 başta olmak üzere BDM'lerin yüksek performans sergilediğini ve bağlamsal temsilleri etiketlenmemiş veri kümelerinde dahi etkili biçimde işleyebildiğini ortaya koymaktadır.
Ayrıca müşteri memnuniyetindeki zamansal değişimleri izleyebilmek amacıyla restoranlara yönelik dört yıllık bir dönemdeki duygu eğilimleri analiz edilmiştir. Bu yaklaşım sayesinde restoranların zaman içerisindeki performans farklılıkları ortaya konmakta ve müşteri beklentilerine yönelik sürdürülebilir iyileştirme stratejilerinin geliştirilmesine olanak tanınmaktadır. Önerilen yöntem, yöneticilerin müşteri memnuniyetini izlemeleri, hizmet kalitesini değerlendirmeleri ve olası memnuniyetsizlik nedenlerini tespit etmeleri açısından hızlı ve veri odaklı bir çözüm sunmakta; böylece stratejik karar alma süreçlerini desteklemekte ve kurumsal imaj yönetimine katkı sağlamaktadır.

Anahtar Kelimeler: Duygu Analizi, BERT, Sıfır Atış Metin Sınıflandırma, Büyük Dil Modelleri, Doğal Dil İşleme.

Introduction

In the modern era of intense competition, long-term success for businesses means more than developing better quality products and services. Organisations must constantly meet customer expectations and implement innovative measures to respond. In this framework, it is of the utmost importance for businesses to analyse feedback that reflects customers' attitudes and sentiments regarding their products and services. This importance stems from the fact that customer feedback serves not only as a source of information for remedial actions regarding current offerings but also as critical input in new product development and continuous improvement initiatives (Fundin & Bergman, 2003; Wirtz et al., 2010). In this context, Wirtz and Tomlin (2000) posit that integrated systems based on systematic customer feedback analysis are essential for institutionalising a learning paradigm focused on customer requirements. Similarly, Hudson (2008) highlights the importance of feedback as a strategic tool in developing innovative solutions aligned with user needs, especially in product development processes. Shah and Rai (2022) also highlight the strategic role of customer feedback in enabling sustainable business success by developing service quality and strengthening customer satisfaction, loyalty, and brand equity. These outcomes underscore the necessity of establishing robust mechanisms that transform feedback into actionable insights. An effective feedback mechanism thus allows organisations to realise customers' current expectations while looking ahead and visualising future needs in advance (Celuch et al., 2015; Kotsonis et al., 1989).

With the ubiquitous spread of digitalisation, online customer reviews (OCRs) have been one of the prevalent types of customer feedback. OCRs can be any form of content posted by customers on various online platforms—notably e-commerce websites, microblogs, independent product or service review websites, and social media—that encapsulate their opinions regarding products and services (Eshkevari et al., 2022; Katole, 2022). This content may be text, photos, videos, or reviews, singly or in combination (Pocchiari et al., 2024). OCRs are typically available to the public on the websites. In this way, these websites have a part to play in establishing confidence and trust between sellers and buyers, thereby assisting buyers in making their decisions (Hennig-Thurau et al., 2004). This is because customers consider information obtained from other customers to be more influential, as they view it as credible and trusted (Willemsen et al., 2012). In addition, OCRs directly and indirectly influence customers' purchase intentions via customer trust (Tanuwijaya et al., 2023). From a business perspective, OCRs enhance online presence, hence their part in customer retention and gaining new customers (Faizi & El Fkihi, 2019; Lohse & Kemper, 2019).

Customers openly share their experiences and reactions to products and services online, revealing genuine, experience-based sentiments (Chen & Farn, 2020). Therefore, OCRs are a rich source of textual data about products and services. Today, the growing volume of this data source renders manual analysis methods inadequate, prompting businesses to adopt advanced data processing techniques to extract meaningful insights from large-scale datasets. In this regard, artificial intelligence and natural language processing (NLP) techniques facilitate the methodical and purposeful evaluation of OCRs. Specifically, with sentiment analysis, one can determine the positive or negative emotional states included in customer reviews, which enables organisations to gauge customer satisfaction levels more objectively using large data sets.

Sentiment analysis is not only a way to measure customer satisfaction but also a tool that offers businesses a strategic edge in the marketplace. Thus, it is widely acknowledged that sentiment analysis can derive actionable insights from customer feedback, thereby influencing business strategy. Specifically, through the utilisation of machine learning (ML) and deep learning (DL) models, businesses can classify customer sentiments effectively with high accuracy, thereby facilitating more informed decision-making in product development, marketing strategies, and customer relationship management (Akter et al., 2025). In addition, sentiment analysis is also highlighted as an essential technique in competitive analysis, where the strengths and weaknesses of competing businesses can be systematically exposed by using customer feedback (Taherdoost & Madanchian, 2023). Sentiment analysis of customer feedback helps organisations enhance their products, resulting in higher customer satisfaction and loyalty, ultimately making them more competitive (Al-Barrak & Al-Alawi, 2024).

Traditional ML and NLP methods, founded on mathematical and statistical approaches, have been extensively applied in text classification and sentiment analysis tasks. These methods often need large amounts of manually labelled data and bring about operational challenges concerning time and computational cost when processing large data sets (Yang et al., 2024). Furthermore, since these approaches fail to consider the positional context of words within the text, they are limited in their effectiveness for processing complex linguistic structures (Almalis et al., 2022). Essentially, popular representation techniques in academic studies—such as Bag-of-Words (BoW), Term Frequency–Inverse Document Frequency (TF-IDF), and n-grams—are not sufficiently robust in capturing semantic relations, word order and contextual representation in the text; this limitation, therefore, lowers the performance of sentiment analysis (Kalaivani & Kuppuswami, 2023; Lebret & Collobert, 2014).

This research compares the performance of modern NLP methods based on contextual representation for sentiment analysis of restaurant OCRs. In this paradigm, various approaches were evaluated, including Bidirectional Encoder Representations from Transformers (BERT), the Zero-Shot Text Classification (ZSTC) method, and large language models (LLMs) such as OpenAI 4o, Gemini 2.0 Flash, and DeepSeek-V3. The results show that LLMs achieved higher accuracy than other approaches. This indicates that LLMs offer a strong alternative for overcoming the contextual limitations of traditional methods. Thanks to the vast pre-trained language representations of LLMs, it is possible to achieve high performance even with unlabelled data, raising the potential for cost-effective, flexible, and scalable sentiment analysis systems.

However, despite the recent advances in LLMs, sentiment analysis studies leveraging these models for domain-specific applications such as restaurant OCRs remain limited. This gap highlights the need for more empirical evidence to understand their effectiveness in real-world contexts and contributes to defining the research problem addressed in this study.

Moreover, the research also examines trends with positive and negative sentiments across four years of customer reviews. This analysis revealed changes in customer perception, and the feedback progress over time could be evaluated. This study contributes to developing data-driven decision support systems that enable faster, more accurate, and more meaningful analysis of OCRs in domains where customer feedback holds strategic importance, such as restaurant chains, online food delivery platforms, and tourism-oriented hospitality services.

1. Conceptual Framework

Understanding the foundations of sentiment analysis and the evolution of computational approaches is essential to contextualize the present study. This section outlines the core concepts and models that underpin the analysis, including sentiment analysis and recent advancements such as transformer-based architectures, LLMs, and the ZSTC method. These components collectively form the conceptual basis upon which this research's methodological choices and empirical evaluations are built.

1.1. Sentiment Analysis

Sentiment analysis is usually performed using ML and NLP techniques to identify and classify sentiments expressed in textual data (Henrickson et al., 2019). The classification often involves dividing sentiments into positive and negative categories; in some cases, a neutral category is also added. With the popularity of web platforms today, users freely review various products and services, locations, and events (Hauthal et al., 2020; Orea-Giner et al., 2022). As a result, such reviews have become a valuable data source for analysing sentiment.

Sentiment analysis enables the evaluation of individuals' sentiment inclinations regarding a specific subject with relatively little effort (Zhang et al., 2019). This method is mainly used in customer service and market research to understand customer feedback and social media trends (Chakraborty et al., 2023). Moreover, sentiment analysis is also employed to examine public perception, e.g., analysis of political discourse and hate speech (Singh et al., 2022; Swamy et al., 2022).

Numerous sentiment analysis studies have been conducted in the literature using lexicon-based, ML-based, and DL-based methods from the past to the present. In the lexicon-based approaches (e.g., SentiWordNet, AFINN, VADER), predefined word lists containing positive and negative terms are used to determine the sentiment within a text. The approach offers broader term analysis and low computational overhead since they do not require dataset training. However, lexicons are unsuitable for context-level classification due to their limited vocabulary, fixed sentiment scores, and dependency on strong, often unavailable linguistic resources (Cernian et al., 2015; Rajalakshmi et al., 2017).

In ML-based approaches to sentiment analysis, supervised ML algorithms such as Naïve Bayes (NB), Support Vector Machine (SVM) and Logistic Regression (LR) are often used for classification tasks. These algorithms are usually based on training data and feature engineering (e.g., BoW, TF-IDF). However, since these methods address context only at the word level in a limited manner, they may fail to capture deeper semantic relationships (Nicholls & Song, 2010; Stine, 2019).

Sentiment analysis methods based on DL can acquire contextual relations and complicated linguistic structures of text data. These models can identify key features from text automatically without manual feature engineering. Notably, DL architectures such as Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) achieve high accuracy when applied to large datasets. Nevertheless, such models necessitate many hyperparameters and large amounts of training data, thereby entailing high computational costs (Masood et al., 2020; Tsiligaridis, 2024).

In the literature, it has been noted that the above approaches have certain limitations in sentiment analysis. In this regard, issues such as linguistic ambiguities (Deng et al., 2017), language complexities including sarcasm and irony (Dheemonth et al., 2024), as well as text processing issues in multilingual datasets (Martina Jose Mary et al., 2014), have been proven to be essential aspects that adversely influence the performance of these techniques. Furthermore, comprehensive data preprocessing is required to use these methods effectively, such as removing stop words, punctuation, and emojis from the text, lowercasing, and stemming or lemmatisation, particularly in agglutinative languages such as Turkish. These linguistic and structural challenges can result in increased processing time and additional workload.

In NLP, some authors categorise these methods as traditional methods (Kapočiūtė-Dzikienė et al., 2019; Revathi et al., 2023). In recent years, however, the development of DL-based models has brought architectures such as BERT and LLMs—and techniques such as ZSTC that rely on these models—to the forefront of sentiment analysis research. These methods share a common foundation in the transformer architecture, which enables the effective processing of contextualised language representations.

1.2. Modern Methods in Sentiment Analysis

1.2.1. BERT Model

BERT is a language model developed by Devlin et al. (2018). It is based on the transformer architecture introduced by Vaswani et al. (2017) and is particularly effective at modelling contextual relationships in language due to its self-attention mechanism. The most distinctive feature of BERT is its ability to represent words bidirectionally based on context. This advantage enables the model to derive meaning considering both the words before and after (Sabharwal & Agrawal, 2021). By leveraging Masked Language Modelling and Next Sentence Prediction tasks, BERT can capture both intra-sentence and inter-sentence contextual relationships (Mickus et al., 2019). The ability of BERT to embed contextual representations allows it to have higher classification accuracy than traditional ML models. As a result, its performance improves particularly in sentences with grammatically complex structures (Garrido-Merchan et al., 2023; Sabiri et al., 2023; Saragih & Manurung, 2024).

Fine-tuning, which has become especially popular with BERT, is the process of retraining a pre-trained model with a smaller dataset for a new task (Ruiz-Millán et al., 2022). This method is used to make large models more efficient for a specific domain or task. Numerous fine-tuned versions of BERT have been presented in the literature to tackle different domain-related problems. A few notable examples include BioBERT for biomedical texts (Lee et al., 2020), SciBERT for scientific articles (Beltagy et al., 2019), and the Bert-Base Turkish Sentiment Model (Yildirim, 2024). In subsequent years, several advanced models—most of which are based on the BERT architecture—such as RoBERTa (Liu et al., 2019), DistilBERT (Sanh et al., 2019), and mBERT, have also been made available to researchers.

1.2.2. Large Language Models

LLMs are DL-based artificial intelligence systems with billions of parameters pre-trained on large-scale data (Viet & Vinh, 2024). They are trained using transformer architectures and self-attention, which enables them to consider the relation of each word with every other word in a sequence to effectively model contextual relationships in text (Xue, 2024).

LLMs are trained based on self-supervised learning techniques over enormous text corpora such as Wikipedia, GitHub, and Reddit. Self-supervised learning enables the model to learn from the structural patterns within the data itself without the need for externally labelled input (Liu et al., 2019). Thus, the models learn context by predicting the next word in a sequence (Radford et al., 2019). The diversity of the training data significantly affects the model's overall performance and ability to adapt across different domains (Matarazzo & Torlone, 2025). Through this process, the model can produce responses based only on representational input prompts, without even preliminary training for the target task (zero-shot learning), or produce improved outcomes when instructed with a limited number of examples (few-shot learning) (Brown et al., 2020). In this case, prompts generally refer to text-based instructions that lead the model to generate the required output (Phoenix & Taylor, 2024).

Apart from their capacity to understand human language, produce text, and acquire sophisticated linguistic patterns, LLMs can be utilised in a wide range of applications, such as text classification, summarisation, translation, and sentiment analysis (Brown et al., 2020; Mohan et al., 2024).

For all their strengths, these models also have notable weaknesses. Specifically, the literature has identified ethical and operational challenges, including hallucination (the production of information that seems plausible but is incorrect or made up), reinforcement of social biases, lack of transparency, and excessive energy consumption (Bender et al., 2021; Weidinger et al., 2021).

In the past few years, numerous LLMs have been developed in the context of the broader class of generative AI, e.g., Google Gemini, OpenAI's ChatGPT, DeepSeek, Alibaba's Qwen, and Meta's LLaMA, open source as well as proprietary models. The models are generally made available with chatbot interfaces or used through APIs.

1.2.3. Zero-Shot Text Classification Method

Zero-shot learning (ZSL) aims to build a model to classify objects from unseen classes based on semantic information to transfer knowledge learned from seen classes (Liu et al., 2023; Pourpanah et al., 2022). A two-stage procedure of training and inference dominates ZSL. In the training phase, attribute knowledge is acquired, and in the inference phase, this knowledge is used to classify instances into a new set of classes (Çelik & Dalyan, 2023). ZSL is frequently used in computer vision (Sarma, 2023) and NLP domains.

ZSTC is based on pre-trained language models (PLMs). Some models like BERT and RoBERTa, when fine-tuned on natural language inference (NLI) corpora, can effectively be utilised in ZSTC (Gera et al., 2022). NLI is a fundamental NLP task that attempts to determine the relationship between a premise and a hypothesis-provided pair of sentences. This association can be in any of three forms: entailment (the hypothesis is logically entailed by the premise), contradiction (the hypothesis contradicts the premise), or neutrality (there is no direct relation between the premise and the hypothesis) (Harsha et al., 2022). ZSTC enables classification tasks, such as sentiment analysis, to be carried out without having access to labelled data (Birim et al., 2021).

2. Literature Review

Today, OCRs have become an important data source for measuring user satisfaction with products and services. Analysing the sentimental content in these documents is crucial for ascertaining customer behaviour and supporting decision-making systems. In this case, advancements in NLP, PLMs, LLMs, and unsupervised classifiers have become widespread in sentiment analysis research. This section initially provides examples of sentiment analysis research founded on traditional methodologies and then shifts to examples of research utilising newer approaches.

Within the paradigm of sentiment analysis conducted through traditional ML techniques, customer reviews were examined to determine their applicability. Laksono et al. (2019) compared TripAdvisor restaurant reviews through the NB algorithm and obtained higher accuracy than TextBlob. The accuracy measure of the NB was stated to be 0.72. Alrehili and Albalawi (2019) analysed Amazon customer reviews and combined NB, SVM, Random Forest (RF), Bagging and Boosting algorithms with a voting-based ensemble approach. It was observed that the best accuracy of up to 0.90 in certain instances was obtained with RF, whereas in specific scenarios, the ensemble voting outperformed the base classifiers. Tuzcu (2020) conducted sentiment analysis on book reviews posted on an online platform using Multi-Layer Perceptron (MLP), NB, SVM, and LR algorithms. Among the applied methods, the highest classification performance was achieved with the MLP algorithm. In their study, Hemalatha and Velmurugan (2020) evaluated more than 90,000 grocery and gourmet food reviews, comparing the effectiveness of LR, NB, SVM, and Artificial Neural Networks (ANN). In this evaluation, the SVM model achieved a 0.95 accuracy score with a sensitivity and specificity score of 0.96 and 0.10, respectively. Nazar & Bhattasali (2021) evaluated LR on Amazon food product reviews using BoW and TF-IDF. TF-IDF worked better than BoW, with an F1 score of 0.95. Lin (2021) analysed e-commerce reviews for women's clothing products with LR, SVM, RF, XGBoost, and LightGBM algorithms, with evaluation results obtained by utilising LightGBM (Accuracy: 0.98, Precision: 0.97, Recall: 0.97, F1: 0.97, AUC: 0.96). In a newer study, Mingo (2024) analysed customer reviews of software products on Amazon, showing that the combination of LR and TF-IDF outperformed lexicon-based approaches such as TextBlob and VADER. The evaluation metrics are accuracy: 0.68, precision: 0.66, recall: 0.68, and F1 score: 0.66. Karuna et al. (2024) compared LR and RF algorithms on hotel reviews at TripAdvisor and concluded that the LR model outperformed RF with an accuracy of 0.94, while precision, recall, and F1 measures were also 0.94. Panduro-Ramirez (2024) implemented sentiment analysis on OCRs from certain e-commerce websites to support product recommendation systems. The study emphasised the significance of ML-based models in acquiring customer satisfaction and optimising recommendations. The K-Nearest Neighbours (KNN), LR, RF, and CatBoost Classifier (CBC) algorithms were tried out, with LR providing the best performance (Accuracy: 0.90, Precision: 0.85, Recall: 0.78, F1 Score: 0.81). In their examination of Turkish online reviews of the iPhone 11 (128 GB) on the Trendyol website, Kayakuş et al. (2024) emphasised the usefulness of these techniques in customer satisfaction measurement and brand reputation management. The research employed the SVM algorithm and noted the ensuing metrics: 0.50 in accuracy, 0.50 in precision, 0.96 in recall, and an F1 score of 0.66.

The above research confirms that sentiment analysis allows the quantitative measurement of customer satisfaction and brand reputation. Further, these studies highlight that it is not only about classifying customer reviews but also that the sentiments derived from them are essential for marketing, product development, and strategic decision-making. In particular, tourism and e-commerce-focused studies argue that reviews can directly influence business decisions.

LR, RF, NB, and SVM algorithms, along with some lexicon-based methods, are predominantly used in these studies. In these studies, sentiment analysis was typically based on positive and negative classes. While the algorithms showed high accuracy in some cases, they performed poorly in others. The poor performance was primarily attributed to class imbalances within the datasets. In general, traditional approaches have demonstrated the capability to deliver competitive performance in classification tasks, but with some shortcomings stemming from their finite capacity to interpret contextual meaning. Further, although traditional ML algorithms can offer a certain degree of competence in classification tasks, data preprocessing requirements and training data labelling require extra workloads and time expenses. This situation restricts efficiency, especially when handling large-scale datasets.

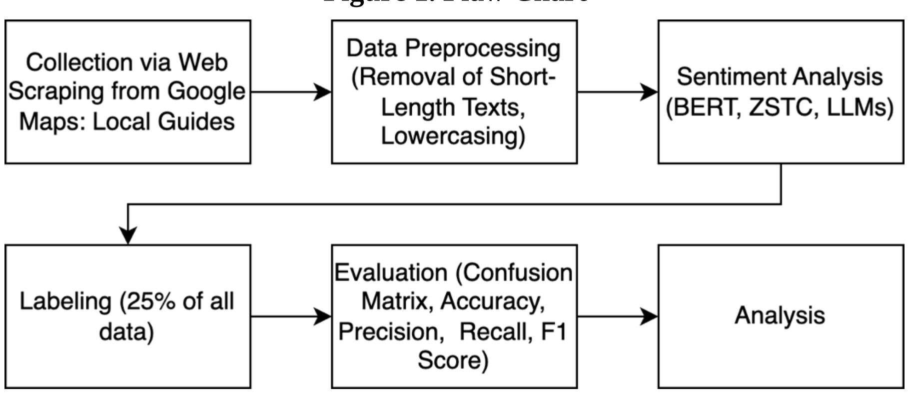
Recent research in sentiment analysis has witnessed the transition from traditional ML approaches to DL-based architectures and LLMs. Although these methods have been relatively recently available, research in sentiment analysis for classifying customer reviews is still limited. Masarifoglu et al. (2021) employed NB, SVM, and LR classifiers, and the BERTurk and multilingual BERT models, as well as the ZSL strategy (XLM-RoBERTa-large-XNLI), to analyse Turkish bank customer reviews. They attained excellent performance with limited labelled data, with BERTurk having a weighted F1 score of 0.91. In another paper, Mostafa & AlSaeed (2022) compared KNN, Decision Tree (DT), NB, RF, Bidirectional LSTM (Bi-LSTM), Gated Recurrent Unit (GRU), and the BERT model on Amazon product reviews. BERT was the best-performing model with an accuracy of 0.94. Patra et al. (2023) also compared LR, DT, and BERT using Amazon product reviews, where BERT demonstrated its superiority in contextual syntax comprehension and more stable results than the traditional algorithms. In this study, BERT's performance included 0.89 accuracy, 0.88 precision, 0.89 recall, and 0.88 F1 score. Manias et al. (2023) conducted a comparison of multilingual Twitter datasets with mBERT (cased), mBERT (uncased), XLM-RoBERTa (with classification head), DistilBERT, and the ZSL (XLM-RoBERTa-large-XNLI). Based on their results, multilingual BERT models achieved high accuracy, whereas the ZSL method presented scalability benefits. The highest performance was obtained with the XLM-RoBERTa model with a classification head (Accuracy: 0.76, Precision: 0.77, Recall: 0.76, F1 score: 0.76). Kyritsis et al. (2023) explored ZSL sentiment analysis using the BART-large (BART-large-MNLI) and DistilBART (DISTILBART-MNLI-12-1) models. They used the US Airline Sentiment Dataset, and the best-performing model was BART-large, with 0.73 accuracy and 0.88 F1 score. In his study, Polatgil (2024) employed the BERTurk (128k uncased) model, fine-tuned with the BounTi dataset, to conduct sentiment analysis on approximately 30,000 Turkish-language YouTube user comments related to the TOGG automobile. The sentiment labels generated by the transformer model were subsequently used as ground truth for supervised learning, and their reproducibility was tested using ANN, KNN, and RF algorithms. The ANN model achieved the highest performance, with an accuracy of 0.85. In Shen & Zhang (2024), comparisons between GPT-3.5-turbo and GPT-4o were made with FinBERT on both zero-shot and few-shot versions. The performance of the GPT-based models was compared without fine-tuning and elicited by prompt engineering. Nevertheless, FinBERT being superior (Accuracy: 0.88, Precision: 0.85, Recall: 0.89, F1 score: 0.80), the results showed that GPT-4o, under few-shot learning and well-designed prompts, could achieve FinBERT-like accuracy. Chaudhary (2024) compared GPT-3.5 and Gemini Pro models using reviews from streaming platforms. It is seen here that Gemini Pro, when used in a few-shot setting with prompts, outperformed GPT-3.5 in sentiment analysis (Accuracy: 0.81, Precision: 0.81, Recall: 0.81, F1 Score: 0.81). Roumeliotis et al. (2024) benchmarked GPT-4o, GPT-4o-mini, and BERT models in their study on TripAdvisor hotel review data. GPT-4o and GPT-4o-mini were experimented with their pre-trained and fine-tuned variants with user-defined examples. At the same time, the BERT model was experimented with using two configurations of Adam optimisation, one with weight decay and one without it. The outcome indicated that GPT-4o performed better than BERT, especially in contextual meaning inference, and was highly effective in decision-support systems. İncedelen and Aydoğan (2025) analysed sentiment using 150,000 Turkish e-commerce product reviews. The study employed transformer-based language models, including XLM-RoBERTa, mBERT, BERTurk (32k), BERTurk (128k), ELECTRA Turkish Small, and ELECTRA Turkish Base. Among these, BERTurk (128k) achieved the best performance, with an accuracy of 0.84, precision of 0.84, recall of 0.84, and an F1 score of 0.84. Teke et al. (2025) conducted sentiment analysis using Turkish customer reviews from the Computer, Phone, Shoes, Clothing, Cosmetics, Sports, and Outdoor categories on the Trendyol platform. To this end, they compared traditional machine learning algorithms such as SVM, RF, NB, LR, KNN, Gradient Boosting, and DT with transformer-based models including mBERT, BERTurk, XLNet, and DistilBERT. The best performance was achieved by the BERTurk model (Accuracy: 0.96, Precision: 0.96, Recall: 0.96, F1 Score: 0.96). Finally, Akter et al. (2025) compared various algorithms (LR, RF, SVM, LSTM) and the BERT model. The BERT model achieved the highest scores with 0.94 accuracy, 0.94 precision, 0.93 recall, and 0.93 F1 score.

The following findings accentuate the points of consensus between researchers working on sentiment analysis with modern NLP methodologies: First, contextual language models offer a significant advantage over traditional approaches in sentiment analysis tasks. Second, transformer-based models exhibit adaptability across domains and can be integrated into decision support systems. Third, it is emphasised that BERT models fine-tuned for specific tasks are highly effective, especially when trained on datasets related to target domains. Fourth, zero-shot and few-shot methods are presented as fast and effective alternatives in scenarios without labelled data. Finally, the literature review stresses that LLMs, under the guidance of prompts, achieve successful outcomes without fine-tuning and outperform conventional approaches in context comprehension and scalability.

3. Methodology

This research examines Turkish OCRs gathered from four seafood restaurants in İzmir, which were assessed as having the most reviews on Google Maps Local Guides. The reviews were collected in March 2025 through web scraping techniques utilising Python's Selenium library. 2,065 reviews were analysed, with the earliest one going back four years. The dataset includes restaurant names, usernames, ratings, user reviews, and time data. The BERT model, the ZSTC method, OpenAI 4o, Gemini 2.0 Flash, and the DeepSeek V3 were used in the analysis. The flowchart illustrating the study's methodology is presented in Figure 1.

Figure 1. Flaw Chart



3.1. Data Preprocessing

The reviews' semantic depth and contextual relevance directly relate to the reliability of evaluating the performance of the methods and models used in sentiment analysis. Therefore, very short comments were removed from the dataset, which did not provide sufficient contextual or semantic depth. For instance, Turkish sentences such as *Afiyet olsun* (Enjoy your meal), *Bu restorana gitmedim* (I have not been to this restaurant), *Cumartesi günleri canlı müzik var mı?* (Is there live music on Saturdays?), or *Balık sevmem…* (I don't like fish…) were considered as content that does not contribute to sentiment analysis. Likewise, instead of brief and superficial comments such as *Çok iyi* (Very good), *Gereksiz pahalı* (Unnecessarily expensive), or *Harika lezzetler* (Excellent flavours), more weight was laid on longer texts that would test the semantic interpretation capabilities of contemporary NLP models.

Thus, quartile analysis based on word count revealed that the first quartile (Q1) was five words, the median (Q2) was 11 words, and the third quartile (Q3) was 23 words. In addition, the minimum number of words in a review was one, and the maximum was 271. In the interest of semantic meaning, the lower bound was set at the first quartile value, and reviews with five or fewer words were eliminated from the dataset.

Besides this, all text was converted to lowercase. Additionally, since emojis and emoticons are known in modern NLP to reflect sentiment better and improve classification accuracy, they were not removed from the dataset (Khan et al., 2025; Shukla & Dwivedi, 2024).

3.2. Data Analysis

In this study, sentiment analysis was conducted using the BERT, the ZSTC technique, and LLMs—OpenAI 4o, Gemini 2.0 Flash, and DeepSeek V3. In April 2025, all the analyses were run in Google Colab on a T4 processor using the Python programming language.

Sentiment analysis with BERT was performed using the BERT-base-Turkish-Sentiment-cased model (Yildirim, 2024). The model was fine-tuned on some sentiment analysis datasets with BERTurk (Schweter, 2020) and is specially adapted to Turkish. The model outputs only positive and negative results.

For the ZSTC approach, two different models were used: mDeBERTa-v3-base-MNLI-XNLI (Laurer et al., 2022) and XLM-RoBERTa-large-XNLI (Davison, 2024). mDeBERTa-v3-base-MNLI-XNLI is a version of Microsoft's mDeBERTa-v3 model (He et al., 2021) that was fine-tuned to be a multilingual NLI. It was trained on the English Multi-Genre Natural Language Inference (MNLI) and multilingual Cross-lingual Natural Language Inference (XNLI) datasets, and it is used mainly for ZSTC and hypothesis–premise analysis. XLM-RoBERTa-large-XNLI, however, is a multilingual XLM-RoBERTa-large (Conneau et al., 2019) based NLI model. It is fine-tuned on the XNLI dataset in 15 languages and fine-tuned for ZSTC tasks. For the ZSTC process, pre-defined classes were set as "Positive customer review" and "Negative customer review" in Turkish so that the model could interpret the evaluation outputs in that language. In this manner, the model is classified by placing the highest probability on one of these two pre-decided classes for each customer review.

One of the LLMs employed for sentiment analysis was Gemini 2.0 Flash. Google introduced it on the Vertex AI platform in February 2025, and it is optimised for high speed and low latency. The model possesses a context window of up to one million tokens and can process text, image, audio, and video data. In addition, it generates code and images, extracts data, and analyses files (Google Cloud, 2025).

The second LLM used was ChatGPT-4o (omni). ChatGPT-4o is a multimodal LLM introduced by OpenAI in May 2024 and capable of processing real-time text, image, and audio inputs. As an advanced variant of the GPT-4 family, this version is more cost- and speed-efficient than GPT-4 Turbo (OpenAI, n.d.). Technical comparisons have shown that ChatGPT-4o greatly enhances contextual coherence, quality of text generation, and accuracy compared to older models (Murad et al., 2024). In addition, the model has also been found to have an accuracy rate of up to 92.8% in knowledge questions, an improvement over older models (Arılı Öztürk et al., 2025).

The final model used in this research was DeepSeek V3. DeepSeek V3 is a 671-billion-parameter Mixture-of-Experts model designed by DeepSeek-AI. The model utilises Multi-head Latent Attention and DeepSeekMoE architectures to facilitate fast inference. Having been trained on 14.8 trillion tokens, the model has emerged as a viable alternative to proprietary models, accomplishing this with just 2.8 million H800 GPU hours (DeepSeek-AI et al., 2024; Liao, 2025).

For all three LLMs, both the system and user prompts were written in Turkish. Here, user prompts initiate specific tasks, and system prompts establish contextual boundaries and tone. The English translation of the prompts is provided below:

System Prompt: You are an expert linguist skilled at classifying customer reviews as Positive or Negative.

User Prompt: Help me classify customer reviews:

Positive (label=1), Negative (label=0).

Do not provide any commentary. Only return the label.

Classify the following customer review:

{review}

3.3. Model Tests and Evaluation Results

Approximately 25% of the dataset was randomly selected and manually labelled by a single annotator as either positive or negative based on each review's semantic orientation and dominant sentiment. The labelling was in such a way that there were equally positive and negative examples (259 reviews for both classes). The performance of BERT, ZSTC, OpenAI 4o, Gemini 2.0 Flash, and DeepSeek V3 was evaluated using accuracy, precision, recall, and F1 score.

The confusion matrix is a two-dimensional table that assesses the performance of classification tasks in actual versus predicted class labels. It has four main elements: True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). TP occurs when the model correctly predicts a positive instance; an actual positive review is labelled as positive. TN refers to the correct classification of an actual negative review as negative. FP arises when the model misclassifies a negative instance as positive, specifically when an actual negative review is labelled as positive. Conversely, an FN occurs when the genuinely positive review is incorrectly classified as negative.

Accuracy is a metric that measures the overall success of a model in all instances. It is defined as the proportion of the total number of accurately classified instances (TP + TN) to the total number of instances (TP + TN + FP + FN):

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

Precision is a metric that measures how many of the positive predictions the model made were accurate. It accounts for the impact of FPs and is calculated by the formula:

$$Precision = \frac{TP}{TP + FP}$$

Recall measures the proportion of actual positive instances that the model correctly identifies. It reflects the impact of FNs and is calculated using the following formula:

$$Recall = \frac{TP}{TP + FN}$$

F1 score is the harmonic mean of precision and recall, providing a balanced measure of both. It is an essential metric for evaluating imbalanced datasets as it considers FP and FN. The following formula calculates it:

$$F1\ Score = 2 * \frac{Precision * Recall}{Precision + Recall}$$

The results of the evaluation metrics along with the corresponding confusion matrices are given in Table 1. As seen from the table, the performance was notably high. Accuracy was over 0.90 for all the models. Precision was over 0.90 for BERT, OpenAI 4o, Gemini 2.0 Flash, and DeepSeek V3, which means that an overwhelming majority of positive predictions by these models were accurate. In particular, the ChatGPT 4o model did not have any FPs for negative reviews, so its precision score was 1.00. Recall was over 0.93 in all the models except BERT, which validated that the models had performed well in recalling TPs. The highest F1 score was achieved in DeepSeek V3 (0.98), whereas the lowest was in BERT.

Table 1. Confusion Matrices and Evaluation Metrics

			BS		0sDeBERTa		0sRoBERTa		G2		4o		DSV3	
CMs	**TP**	**FN**	219	40	254	5	242	17	246	13	246	13	255	4
	FP	**TN**	11	248	37	222	32	227	2	257	0	259	6	253
Accuracy			0.9015		0.9189		0.9054		0.9710		0.9749		**0.9807**	
Precision			0.9522		0.8729		0.8832		0.9919		**1.0000**		0.9770	
Recall			0.8456		0.9807		0.9344		0.9498		0.9498		**0.9846**	
F1 Score			0.8957		0.9236		0.9081		0.9704		0.9743		**0.9808**	

CMs: Confusion Matrices, TP: True Positive, FN: False Negative, FP: False Positive, TN: True Negative, BS: BERT-base-Turkish-sentiment-cased, 0sDeBERTa: mDeBERTa-v3-base-MNLI-XNLI, 0sRoBERTa: XLM-RoBERTa-large-XNLI, G2: Gemini 2.0 Flash, 4o: ChatGPT-40, DSV3: DeepSeek V3

Table 2 provides selected examples of misclassified reviews across all models, offering preliminary insights into potential sources of error. These cases suggest that models may have difficulty interpreting reviews containing subtle sentiment expressions, multiple emotional tones, or context-dependent evaluations. While the overall performance was strong, these instances indicate that understanding nuanced or mixed sentiments remains challenging.

Table 2. Examples of misclassified customer reviews by different models

Models	Reviews (English Translation)	L	P
BS	This is the kind of place where, when you ask for cold water, the response is, “I’ll bring it if I can find any.”	0	1
	A place in Kordon where you can enjoy delicious fish. Even the staff were very attentive.	1	0
0sDeBERTa	No need to say much — a timeless place that hasn’t lost its charm over the years.	1	0
	If you pay this much just to eat fish, it becomes a place you can visit once every two months instead of twice a week. My suggestion is to dine in Seferihisar instead.	0	1
0sRoBERTa	I missed the old days — it no longer has the same taste as before.	0	1
	A very nice place. The presentations are also excellent. But the bill is pricey!	1	0
G2	The service was good, and the mezes were nice. For the main course, we had calamari, shrimp, and seabream. Compared to other places, the prices were high. The flavours were standard.	1	0
	The food, service, and view are all quite good, but unfortunately, the prices are high due to the economy.	1	0
4o	A very nice place. The presentations were also excellent. But the bill was high!	1	0
	The place is very nice, and the food is delicious. Portions are average—not small, but not quite as much as they should be. The staff is attentive. However, the prices are high.	1	0
DSV3	Seaside tables are reserved only for large groups. If you're dining as a party of two, you'll be seated in the middle area. The calamari was very tasty, but they didn’t prepare the samphire well. Prices are above average.	0	1
	The dishes were well-cooked, and the service was good, but they weren’t helpful at the counter regarding the bill. That’s why I deducted one star.	1	0

BS: BERT-base-Turkish-sentiment-cased, 0sDeBERTa: mDeBERTa-v3-base-MNLI-XNLI, 0sRoBERTa: XLM-RoBERTa-large-XNLI, G2: Gemini 2.0 Flash, 4o: ChatGPT-40, DSV3: DeepSeek V3, L: Labelled, P: Predicted

4. Findings

The names of the four seafood restaurants analysed in this study have been anonymised: SIR, MBR, BRU, and IDR. The DeepSeek V3 model achieved the best performance in sentiment analysis. Based on this finding, the restaurants were analysed further using this model. Table 3 illustrates examples of positive and negative review predictions for each restaurant according to the DeepSeek V3 model.

Table 3. Examples of sentiment analysis results of four restaurants according to the DeepSeek V3 model

Restaurants	**Reviews (English Translation)**	**Results**
SIR	The service they provided and their attention to detail were excellent. The food, mezes, desserts, fruits, and the atmosphere were all truly enchanting. If you ask where to eat fish in İzmir, I would always recommend this place.	Positive
	Never go there. It's beyond terrible, and they are extremely disrespectful.	Negative
MBR	A restaurant we love going to as a family. The food tastes great, the staff is friendly, and the restrooms are always clean and hygienic.	Positive
	It wasn't bad — except for the bill. Would I go again? Never.	Negative
BRU	It's nice to see that places like this still exist. The ambiance, service, and food were all excellent. Their pricing policy is also very reasonable. It's a place I can recommend to everyone.	Positive
	Even though we specifically requested a window-side table when making the reservation, they told us there were no available seats by the window.	Negative
IDR	I highly recommend it for its friendly and polite staff, delicious food, and outstanding mezes.	Positive
	Extremely poor service. Unless you replace this disrespectful waiter, you'll keep getting low ratings.	Negative

The number of reviews analysed for the four restaurants was 575, 494, 539, and 457, respectively. While SIM and MBR include data from 2021 to 2025, the other two restaurants include data from 2020 to 2025. Because data collection was done in March 2025, year-on-year analysis was retrospectively conducted from March to the following March. This permitted measurement of how positive and negative reviews changed over a specified time frame. Table 4 illustrates the frequency of positive and negative reviews by yearly time frame.

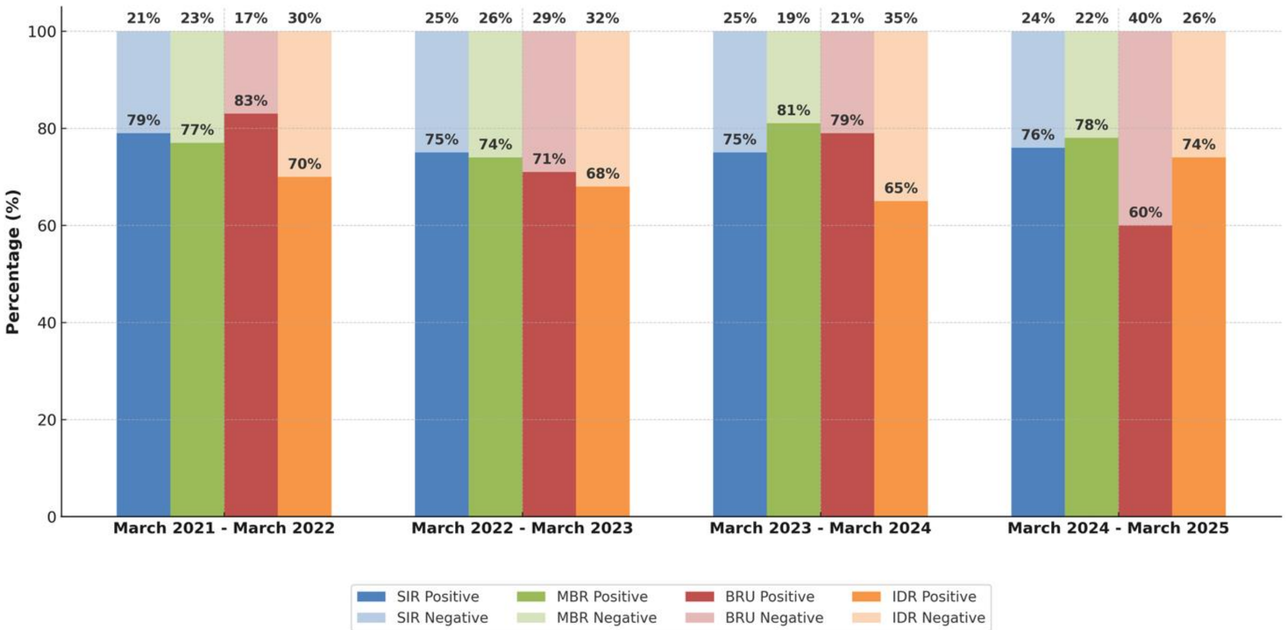
Table 4. The number of positive and negative OCRs for seafood restaurants in annual periods

Restaurants	**SIR**		**MBR**		**BRU**		**IDR**	
Positive (+) / Negative (-)	+	-	+	-	+	-	+	-
March 2020 - March 2021	N/A		N/A		51	18	1	2
March 2021 - March 2022	75	20	74	22	101	21	75	20
March 2022 - March 2023	112	38	111	38	124	51	112	38
March 2023 - March 2024	126	43	102	24	66	18	126	43
March 2024 - March 2025	123	38	96	27	53	36	123	38
Total	436	139	383	111	395	144	436	139
Total (%)	76	24	78	22	73	27	76	24

Table 4 indicates that SIR had the highest number of customer reviews over four years, reflecting the restaurant's popularity. Conversely, despite having 81 fewer reviews than SIR, MBR recorded a higher proportion of positive feedback. This means that the quantity of reviews does not necessarily match the degree of customer satisfaction, and one can differentiate between popular restaurants with a negative opinion and less reviewed ones that still enjoy high satisfaction levels.

Figure 2 provides a graphical display that better analyses the differences between positive and negative reviews. Because of the limited data available for IDR from March 2020 to March 2021, this period was excluded from the graph for all restaurants. As shown in the figure, customer satisfaction is presented in proportional terms based on each restaurant's yearly distribution of positive and negative reviews. The graph also makes it possible to examine the overall satisfaction levels for the various restaurants comparatively.

Figure 2. Temporal Distribution of Customer Sentiment for Four Seafood Restaurants between 2021 and 2025



As illustrated in Figure 2, SIR has maintained a relatively high level of performance in customer satisfaction throughout the four-year duration. The percentage of positive reviews has shown a consistent trajectory, ranging between 75% and 79%. This situation indicates that SIR has succeeded in steadily meeting customers' expectations and has not registered a considerable decline in the quality of service.

The MBR restaurant exhibited a considerable enhancement in customer satisfaction as it increased its positive sentiment rate from 74% in 2022–2023 to 81% in the following year. There was, however, a decrease in this rate by a slight difference in the final period. In general, MBR sustained and enhanced customer satisfaction, having positive sentiment from 74% to 81%. This result also shows that there was no drastic decline in customer experience. Though MBR was less consistent than SIR, it garnered a higher percentage of positive reviews than its rival over the past two years.

The BRU restaurant exhibited inconsistency in the proportion of positive customer reviews. While it recorded an exceptionally high positive sentiment rate of 83% over 2021–2022, the rate dropped to 60% over 2024–2025. The trend indicates a significant decline in customer satisfaction over the most recent period. In other words, it reflects that BRU has been unable to maintain consistent service quality, leading to increased negative customer experiences in recent years.

IDR experienced a continuous decline in customer satisfaction between 2021 and 2024; however, it reached its highest rate of 74% in the final year. This suggests a significant improvement in the restaurant's service quality.

5. Results and Discussion

In this study, sentiment analysis was conducted with Turkish OCRs, the performance of modern NLP approaches was compared, and customer satisfaction for four seafood restaurants in İzmir was analysed. In this context, in addition to BERT and ZSTC techniques, which have superior contextual representation abilities compared to traditional ML-based methods, three different LLMs were used: OpenAI 4o, Gemini 2.0 Flash, and DeepSeek V3.

The evaluation metrics indicate that LLMs achieved the best results, surpassing both BERT and zero-shot classifiers. Notably, the DeepSeek V3 model achieved the highest scores among all models evaluated. LLMs' ability to model complex contextual relationships enables effective performance even on Turkish OCRs. These findings align with Roumeliotis et al. (2024), who reported that GPT-4o achieved superior performance compared to BERT, especially in contextual inference. Similarly, Chaudhary (2024) showed that Gemini Pro outperformed in a few-shot sentiment analysis task.

Compared to studies based on Turkish datasets—such as Masarifoglu et al. (2021), Polatgil (2024), İncedelen & Aydoğan (2025), and Teke et al. (2025) using BERTurk on Turkish reviews—the results suggest that general-purpose LLMs can match or exceed domain-specific models. Notably, this was achieved without any additional fine-tuning, highlighting the strong potential of multilingual, prompt-based LLMs for low-resource and domain-specific sentiment classification tasks such as restaurant OCRs in Turkish.

In contrast, the BERT-base-Turkish-Sentiment-cased model demonstrated lower scores, with its relatively low recall indicating in difficulty identifying TP reviews. Among the ZSTC models, the mDeBERTa-v3-base-MNLI-XNLI outperformed the XLM-RoBERTa-large-XNLI; however, both performed worse than all LLMs. Nevertheless, their ability to classify without labelled data makes them pragmatically viable in languages and domains with limited annotated resources.

In analysing the models' misclassifications, it was noted that every model was prone to errors in sentences with multiple or mixed emotional tones. These sentences typically involve both positive and negative sentiments, and it is hard for models to identify the prevailing sentiment. This indicates a standard limitation among models in processing complex or subtle sentiment compositions. While the models'—especially LLMs'—overall performance was good, the findings imply that additional advances are necessary to make them interpret more subtle or context-dependent sentiments with greater accuracy.

Finally, compared to traditional sentiment analysis methods described in the Literature Review, the LLMs achieved significantly better performance. In addition, the methods employed in this study offer substantial advantages in terms of processing time and computational efficiency, particularly at the preprocessing stage.

According to the results of the sentiment analysis of the restaurants, the number of reviews is not always directly proportional to customer satisfaction. For instance, even though SIR—the restaurant with the highest number of reviews—had a stable degree of positive sentiment, MBR, which had a lower volume of reviews, obtained a greater overall percentage of positive reviews than SIR. This observation indicates that customer satisfaction evaluations must not be determined exclusively by measures of popularity.

Over the four-year interval, customer satisfaction levels within the restaurants were variable. Although some restaurants recorded high declines, others recorded sporadic improvements. This indicates that customer satisfaction is a dynamic construct that evolves over time, highlighting the need for businesses to maintain and continuously enhance the quality of their services sustainably.

Particularly within the framework of LLMs such as DeepSeek V3, these models are characterised by their precision, contextual relevance, and interpretive capacity, representing a strategic interest in customer relationship management applications. The detailed and context-aware sentiment insights obtained through advanced NLP models offer potential applications beyond text classification alone. Such sentiment data can serve as valuable input for understanding consumer attitudes, predicting purchasing decisions, and evaluating perceptions of brand credibility. Compared to traditional methods that rely on surface-level polarity, these models can capture subtle nuances in language, enabling a more accurate assessment of how customers experience and evaluate services. Thus, the findings of this study may contribute not only to computational approaches but also to broader discussions in marketing and consumer behaviour research. Pragmatically, the methods utilised in this research offer value to restaurant managers and service industry decision-makers by facilitating data-informed decisions in processes from tracking customer satisfaction to quantifying service excellence, determining probable causes of dissatisfaction in advance, building strategic decisions, and maintaining organisational reputation. Furthermore, due to the challenges of manually analysing massive amounts of OCRs, it is evident that automatic analysis with the assistance of LLMs is cost- and timesaving.

This research provides significant results from both practical and methodological viewpoints. From a methodological point of view, modern NLP approaches can effectively be applied in sentiment analysis applications, which are characterised by more flexibility, scalability, and reduced human resource reliance, thus outperforming traditional approaches in efficiency. Particularly in the case of LLMs such as DeepSeek V3, these models stand out for their high accuracy, contextual consistency, and interpretive capacity, offering strategic value for customer relationship management. From a practical perspective, modern NLP approaches can yield concrete contributions for restaurant managers and decision-makers to gauge customer satisfaction, evaluate service quality, diagnose possible sources of dissatisfaction, make strategic decisions, and maintain corporate image. Additionally, considering the challenge of manually processing huge OCR data, it is apparent that automated analyses facilitated by LLMs can offer enormous time and money savings.

In future studies, a more detailed examination of the suitability of these methods for various industries may be considered, and a more thorough exploration of the influence of domain-specific fine-tuning strategies on model performance may be conducted. Additionally, as the reviews may have sentiment expressions for more than one aspect, employing aspect-based sentiment analysis would allow for a more thorough exploration of model outputs and is likely to add sensitivity to the analyses.

Yazar Katkıları: Tek Yazar *Fikir ; Tasarım Denetleme ; Kaynaklar ; Veri Toplanması ve/veya İşlemesi ; Analiz ve/ veya Yorum; Literatür Taraması ; Yazıyı Yazan ; Eleştirel İnceleme*
Hakem Değerlendirmesi: *Dış bağımsız.*
Çıkar Çatışması: *Yazar, çıkar çatışması olmadığını beyan etmiştir.*
Finansal Destek: *Yazar, bu çalışma için finansal destek almadığını beyan etmiştir.*
Etik Kurul Belgesi: -

Author Contributions: Sole Author *Concept ; Design ; Supervision-Resources; Data Collection and/or Processing; Analysis and/or Interpretation ; Literature Search ; Writing Manuscript ; Critical Review*
Peer-review*: Externally peer-reviewed.*
Conflict of Interest: *The author has no conflicts of interest to declare.*
Financial Disclosure: *The author declared that this study has received no financial support.*
Ethical Committee Approval: -

References

Akter, P., Hossain, S., Siddique, M. T., Ayub, M. I., Nath, A., Nath, P. C., Rasel, M., & Hassan, M. M. (2025). Sentiment Analysis of Consumer Feedback and Its Impact on Business Strategies by Machine Learning. *The American Journal of Applied Sciences*, 07 (01), 6–16. doi:10.37547/tajas/volume07issue01-02

Al-Barrak, M. A., & Al-Alawi, A. I. (2024). Sentiment Analysis on Customer Feedback for Improved Decision Making: A Literature Review. *In 2024 ASU International Conference in Emerging Technologies for Sustainability and Intelligent Systems (ICETSIS)* (pp. 207–212). IEEE. doi:10.1109/icetsis61505.2024.10459452

Almalis, I., Kouloumpris, E., & Vlahavas, I. (2022). Sector-Level Sentiment Analysis with Deep Learning. *Knowledge-Based Systems*, 258 (109954), 109954. doi:10.1016/j.knosys.2022.109954

Alrehili, A., & Albalawi, K. (2019). Sentiment Analysis of Customer Reviews Using Ensemble Method. *In 2019 International Conference on Computer and Information Sciences (ICCIS)*. IEEE. doi:10.1109/iccisci.2019.8716454

Arılı Öztürk, E., Turan Gökduman, C., & Çanakçi, B. C. (2025). Evaluation of the performance of ChatGPT-4 and ChatGPT-4o as a learning tool in endodontics. *International Endodontic Journal (Early View)*. doi:10.1111/iej.14217

Beltagy, I., Lo, K., & Cohan, A. (2019). SciBERT: A pretrained language model for scientific text. *In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)* (pp. 3615-3620). Association for Computational Linguistics. doi:10.18653/v1/d19-1371

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the Dangers of Stochastic Parrots. *In FAccT '21: 2021 ACM Conference on Fairness, Accountability, and Transparency* (pp. 610-623). Association for Computing Machinery. doi:10.1145/3442188.3445922

Bharathi Mohan, G., Prasanna Kumar, R., Vishal Krishh, P., Keerthinathan, A., Lavanya, G., Meghana, M. K. U., Sulthana, S., & Doss, S. (2024). An Analysis of Large Language Models: Their Impact and Potential Applications. *Knowledge and Information Systems*, 66 (9), 5047–5070. doi:10.1007/s10115-024-02120-8

Birim, A., Erden, M., & Arslan, L. M. (2021). Zero-shot Turkish Text Classification. *In 2021 29th Signal Processing and Communications Applications Conference (SIU)*. IEEE. doi:10.1109/siu53274.2021.9477864

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., ... & Amodei, D. (2020). Language Models are Few-Shot Learners. In H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, & H. Lin (Eds.), *34th International Conference on Neural Information Processing Systems* (pp. 1877–1901), Curran Associates Inc. doi:10.5555/3495724.3495883

Çelik, E., & Dalyan, T. (2023). Unified Benchmark for Zero-Shot Turkish Text Classification. *Information Processing & Management*, 60 (3), 103298. doi:10.1016/j.ipm.2023.103298

Celuch, K., Robinson, N. M., & Walsh, A. M. (2015). A Framework for Encouraging Retail Customer Feedback. *Journal of Services Marketing*, 29 (4), 280–292. doi:10.1108/jsm-02-2014-0062

Cernian, A., Sgarciu, V., & Martin, B. (2015). Sentiment Analysis from Product Reviews Using SentiWordNet as Lexical Resource. *In 2015 7th International Conference on Electronics, Computers and Artificial Intelligence (ECAI) (WE-15 - WE-18)*. IEEE. doi:10.1109/ecai.2015.7301224

Chakraborty, S. (2023). Sentiment Analysis in the Perspective of Natural Language Processing. *International Journal for Research in Applied Science and Engineering Technology*, 11 (11), 2235–2241. doi:10.22214/ijraset.2023.56925

Chaudhary, S. (2024). Now Streaming Sentiments: A Comparative Analysis of GPT and Gemini in Evaluation of OTT Streaming Show Reviews. *In 2024 Conference on AI, Science, Engineering, and Technology (AIxSET)* (pp. 61–68). IEEE. doi:10.1109/aixset62544.2024.00014

Chen, M.-J., & Farn, C.-K. (2020). Examining the Influence of Emotional Expressions in Online Consumer Reviews on Perceived Helpfulness. *Information Processing & Management*, 57 (6), 102266. doi:10.1016/j.ipm.2020.102266

Conneau, A., Khandelwal, K., Goyal, N., Chaudhary, V., Wenzek, G., Guzmán, F., Grave, E., Ott, M., Zettlemoyer, L., & Stoyanov, V. (2019). Unsupervised Cross-lingual Representation Learning at Scale (Version 2). *arXiv*. doi:10.48550/ARXIV.1911.02116

Davison, J. (2024). *XLM-RoBERTa-Large-XNLI*. Hugging Face. https://huggingface.co/joeddav/xlm-roberta-large-xnli (February 3, 2025)

DeepSeek-AI, Liu, A., Feng, B., Xue, B., Wang, B., Wu, B., Lu, C., Zhao, C., Deng, C., Zhang, C., Ruan, C., Dai, D., Guo, D., Yang, D., Chen, D., Ji, D., Li, E., Lin, F., Dai, F., ... Pan, Z. (2024). DeepSeek-V3 Technical Report (Version 2). *arXiv*. doi:10.48550/ARXIV.2412.19437

Deng, S., Sinha, A. P., & Zhao, H. (2017). Resolving Ambiguity in Sentiment Classification. *ACM Transactions on Management Information Systems*, 8 (2–3), 1–13. doi:10.1145/3046684

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding. *ArXiv*. doi:10.48550/ARXIV.1810.04805

Dheemonth, K., Manonmani, S., Shobha, G., & De Hilster, D. (2024). Sentiment Analysis on Cricket Tweets Using NLP++. *In 2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT)*. IEEE. doi:10.1109/icccnt61001.2024.10724047

Eshkevari, M., Jahangoshai Rezaee, M., Saberi, M., & Hussain, O. K. (2022). An End-To-End Ranking System Based on Customers Reviews: Integrating Semantic Mining and MCDM Techniques. *Expert Systems with Applications*, 209 (118294), 118294. doi:10.1016/j.eswa.2022.118294

Faizi, R., & El Fkihi, S. (2019). Exploring the role of customer reviews in driving business growth. In P. Kommers & G. C. Peng (Eds.), *ICT, Society and Human Beings 2019, Connected Smart Cities 2019, Web Based Communities and Social Media 2019* (pp. 410-414). IADIS Press. doi:10.33965/wbc2019_201908c053

Fundin, A. P., & Bergman, B. L. S. (2003). Exploring the Customer Feedback Process. *Measuring Business Excellence*, 7 (2), 55–65. doi:10.1108/13683040310477995

Garrido-Merchan, E. C., Gozalo-Brizuela, R., & Gonzalez-Carvajal, S. (2023). Comparing BERT Against Traditional Machine Learning Models in Text Classification. *Journal of Computational and Cognitive Engineering*, 2 (4), 352–356. doi:10.47852/bonviewjcce3202838

Gera, A., Halfon, A., Shnarch, E., Perlitz, Y., Ein-Dor, L., & Slonim, N. (2022). Zero-Shot Text Classification with Self-Training. In Y. Goldberg, Z. Kozareva, & Y. Zhang (Eds.), *2022 Conference on Empirical Methods in Natural Language Processing* (pp. 1107–1119). Association for Computational Linguistics. doi:10.18653/v1/2022.emnlp-main.73

Google Cloud (2025). Gemini 2.0 Flash. https://cloud.google.com/vertex-ai/generative-ai/docs/models/gemini/2-0-flash (May 8, 2025)

Hauthal, E., Burghardt, D., Fish, C., & Griffin, A. L. (2020). Sentiment Analysis. Audrey Kobayashi (Ed.), *International Encyclopedia of Human Geography (Secon Edition)*. Elsevier, 169–177. doi:10.1016/b978-0-08-102295-5.10593-1

He, P., Gao, J., & Chen, W. (2021). DeBERTaV3: Improving DeBERTa using ELECTRA-Style Pre-Training with Gradient-Disentangled Embedding Sharing (Version 4). *arXiv*. doi:10.48550/ARXIV.2111.09543

Hemalatha, B., & Velmurugan, T. (2020). Impact of Customer Feedback System Using Machine Learning Algorithms for Sentiment Mining. *International Journal of Innovative Technology and Exploring Engineering*, 9 (4), 1475–1483. doi:10.35940/ijitee.d1537.029420

Hennig-Thurau, T., Gwinner, K. P., Walsh, G., & Gremler, D. D. (2004). Electronic Word-Of-Mouth via Consumer-Opinion Platforms: What Motivates Consumers to Articulate Themselves on the Internet?. *Journal of Interactive Marketing*, 18 (1), 38–52. doi:10.1002/dir.10073

Henrickson, K., Rodrigues, F., & Pereira, F. C. (2019). Data Preparation. Constantinos Antoniou, Loukas Dimitriou, & Francisco Pereira (Eds.), *Mobility Patterns, Big Data and Transport Analytics*. Elsevier, 73–106. doi:10.1016/b978-0-12-812970-8.00005-1

Hudson, M. (2008). Listening to Customers: How EBSCO Plans Enhancements and Product Acquisitions Based on Customer Feedback. *Public Library Quarterly*, 27 (2), 151–156. doi:10.1080/01616840802114978

İncidelen, M., & Aydoğan, M. (2025). Sentiment Analysis in Turkish Using Language Models: A Comparative Study. *European Journal of Technique (EJT)*, 15 (1), 68-74. doi:10.36222/ejt.1592448

Kalaivani, K. S., & Kuppuswami, S. (2023). Exploring The Use of Syntactic Dependency Features for Document-Level Sentiment Classification. *Bulletin of the Polish Academy of Sciences Technical Sciences*, 67 (2), 339-347. doi:10.24425/bpas.2019.128608

Kapočiūtė-Dzikienė, J., Damaševičius, R., & Woźniak, M. (2019). Sentiment Analysis of Lithuanian Texts Using Traditional and Deep Learning Approaches. *Computers*, 8 (1), 4. doi:10.3390/computers8010004

Karuna, N. M. H., Suarjaya, I. M. A. D., & Sutramiani, N. P. (2024). Sentiment Analysis of Hotel Reviews Using Logistic Regression and Random Forest Methods. *JITTER : Jurnal Ilmiah Teknologi Dan Komputer*, 5 (3), 2233. doi:10.24843/jtrti.2024.v05.i03.p04

Katole, H. J. (2022). A Research Article on Importance of Online Customer Reviews on Customer Purchase. *The Business & Management Review*, 13 (03), 23–30. doi:10.24052/bmr/v13nu03/art-04

Kayakuş, M., Yiğit Açikgöz, F., Dinca, M. N., & Kabas, O. (2024). Sustainable Brand Reputation: Evaluation of iPhone Customer Reviews with Machine Learning and Sentiment Analysis. *Sustainability*, 16 (14), 6121. doi:10.3390/su16146121

Khan, A., Majumdar, D., & Mondal, B. (2025). Sentiment Analysis of Emoji Fused Reviews Using Machine Learning and Bert. *Scientific Reports*, 15 (1). doi:10.1038/s41598-025-92286-0

Kotsonis, M. E., Abramson, S. R., Loeb, K. C., & Wattenbarger, B. L. (1989). In The Customer's Hands: Making the Most of User Feedback. *AT & T Technical Journal*, 68 (5), 69–79. doi:10.1002/j.1538-7305.1989.tb00088.x

Kyritsis, K., Perikos, I., & Paraskevas, M. (2023). Zero-Shot Sentiment Analysis Exploring BART Models. *In 2023 IEEE/ACIS 8th International Conference on Big Data, Cloud Computing, and Data Science (BCD)* (pp. 192–197). IEEE. doi:10.1109/bcd57833.2023.10466289

Laksono, R. A., Sungkono, K. R., Sarno, R., & Wahyuni, C. S. (2019). Sentiment Analysis of Restaurant Customer Reviews On Tripadvisor Using Naïve Bayes. *In 2019 12th International Conference on Information & Communication Technology and System (ICTS)*. IEEE. doi:10.1109/icts.2019.8850982

Laurer, M., van Atteveldt, W., Casas, A., & Welbers, K. (2023). Less Annotating, More Classifying: Addressing the Data Scarcity Issue of Supervised Machine Learning with Deep Transfer Learning and BERT-NLI. *Political Analysis*, 32 (1), 84–100. doi:10.1017/pan.2023.20

Lebret, R., & Collobert, R. (2014). N-Gram-Based Low-Dimensional Representation for Document Classification. *ArXiv*. doi:10.48550/ARXIV.1412.6277

Lee, J., Yoon, W., Kim, S., Kim, D., Kim, S., So, C. H., & Kang, J. (2020). BioBERT: A Pre-Trained Biomedical Language Representation Model for Biomedical Text Mining. *Bioinformatics (Oxford, England)*, 36 (4), 1234–1240. doi:10.1093/bioinformatics/btz682

Liao, H. (2025). Deepseek Large-Scale Model: Technical Analysis and Development Prospect. *Journal of Computer Science and Electrical Engineering*, 7 (1). doi:10.61784/jcsee3035

Lin, X. (2020). Sentiment Analysis of E-Commerce Customer Reviews Based on Natural Language Processing. *In ISBDAI '20: 2020 2nd International Conference on Big Data and Artificial Intelligence* (pp. 32-36). Association for Computing Machinery. doi:10.1145/3436286.3436293

Liu, T., Hu, Y., Chen, P., Sun, Y., & Yin, B. (2023). Zero-Shot Text Classification with Semantically Extended Textual Entailment. *In 2023 International Joint Conference on Neural Networks (IJCNN)*. IEEE. doi:10.1109/ijcnn54540.2023.10191094

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Omer Levy, Lewis, M., Zettlemoyer, L., & Stoyanov, V. (2019). RoBERTa: A Robustly Optimized BERT Pretraining Approach. *ArXiv*. doi:10.48550/ARXIV.1907.11692

Lohse, T. M. & Kemper, J. (2019). Online customer reviews as driver of customer loyalty. *In Fortieth International Conference on Information Systems, Munich, Germany*. https://aisel.aisnet.org/icis2019/business_models/business_models/3

Manias, G., Mavrogiorgou, A., Kiourtis, A., Symvoulidis, C., & Kyriazis, D. (2023). Multilingual Text Categorization and Sentiment Analysis: A Comparative Analysis of The Utilization of Multilingual Approaches for Classifying Twitter Data. *Neural Computing and Applications*, 35 (29), 21415–21431. doi:10.1007/s00521-023-08629-3

Martina Jose Mary, M., Devi, S. R., Yogeshkannah, K., Prem, S., Pooranachandiran, G., & Viyash, V. (2024). Multilingual Sentiment Analysis for Youtube Educational Videos Using NLP And Machine Learning Approaches. *In 2024 Third International Conference on Smart Technologies and Systems for Next Generation Computing (ICSTSN)*. IEEE. doi:10.1109/icstsn61422.2024.10671276

Masarifoglu, M., Tigrak, U., Hakyemez, S., Gul, G., Bozan, E., Buyuklu, A. H., & Özgür, A. (2021). Sentiment Analysis of Customer Comments in Banking Using BERT-Based Approaches. *In 2021 29th Signal Processing and Communications Applications Conference (SIU)*. IEEE. doi:10.1109/siu53274.2021.9477890

Masood, M. A., Abbasi, R. A., & Wee Keong, N. (2020). Context-Aware Sliding Window for Sentiment Classification. *IEEE Access: Practical Innovations, Open Solutions*, 8, 4870–4884. doi:10.1109/access.2019.2963586

Matarazzo, A., & Torlone, R. (2025). A Survey on Large Language Models with some Insights on their Capabilities and Limitations. *ArXiv*. doi:10.48550/ARXIV.2501.04040

Mickus, T., Paperno, D., Constant, M., & van Deemter, K. (2020). What Do You Mean, BERT? Assessing BERT as a Distributional Semantics Model. *Society for Computation in Linguistics*, 3 (1), 350-361. doi: doi:10.7275/t778-ja71

Mingo, H. (2024). Enhancing Sentiment Analysis with Logistic Regression and NLP: A Case Study Of 2023 Amazon Software Reviews. *In 2024 International Symposium on Networks, Computers and Communications (ISNCC)*. IEEE. doi:10.1109/isncc62547.2024.10759007

Mostafa, M., & AlSaeed, A. (2022). Sentiment Analysis Based on Bert for Amazon Reviewer. *Journal of the ACS Advances in Computer Science*, 13 (1), 1–10. doi:10.21608/asc.2023.171559.1007

Murad, I. A., Khaleel, M. I., & Shakor, M. Y. (2024). Unveiling GPT-4o: Enhanced Multimodal Capabilities and Comparative Insights with ChatGPT-4. *International Journal of Electronics and Communications Systems*, 4 (2), 127. doi:10.24042/ijecs.v4i2.25079

Nazar, S. R., & Bhattasali, T. (2021). Sentiment Analysis of Customer Reviews. *Azerbaijan Journal of High Performance Computing*, 4 (1), 113–125. doi:10.32010/26166127.2021.4.1.113.125

Nicholls, C. & Song, F. (2010). Comparison of Feature Selection Methods for Sentiment Analysis. In A. Farzindar, & V. Kešelj (Eds.), *Advances in Artificial Intelligence. Canadian AI 2010. Lecture Notes in Computer Science vol 6085* (pp. 286-289). Springer. doi:10.1007/978-3-642-13059-5_30

OpenAI (n.d.). GPT-4o. Open AI. https://platform.openai.com/docs/models/gpt-4o (May 8, 2025)

Orea-Giner, A., Fuentes-Moraleda, L., Villacé-Molinero, T., Muñoz-Mazón, A., & Calero-Sanz, J. (2022). Does The Implementation of Robots in Hotels Influence the Overall Tripadvisor Rating? A Text Mining Analysis from the Industry 5.0 Approach. *Tourism Management*, 93 (104586), 104586. doi:10.1016/j.tourman.2022.104586

Panduro-Ramirez, J. (2024). Sentiment Analysis in Customer Reviews for Product Recommendation in E-Commerce Using Machine Learning. *In 2024 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI)*. IEEE. doi:10.1109/accai61061.2024.10602027

Patra, G. K. (2023). A Sentiment Analysis of Customer Product Review Based on Machine Learning Techniques in E-Commerce. *Journal of Artificial Intelligence & Cloud Computing*, 2 (4), 1–4. doi:10.47363/jaicc/2023(2)389

Phoenix, J., & Taylor, M. (2024). *Prompt Engineering for Generative AI*. Sebastopol, CA: O'Reilly Media.

Pocchiari, M., Proserpio, D., & Dover, Y. (2024). Online Reviews: A Literature Review and Roadmap for Future Research. *International Journal of Research in Marketing*, Article in Press. doi:10.1016/j.ijresmar.2024.08.009

Polatgil, M. (2024). Togg Otomobilleri Youtube Video Yorumlarının Duygu Analizi ve Makine Öğrenme Modeli ile İncelenmesi. *Uluslararası Anadolu Sosyal Bilimler Dergisi*, 8 (1), 15-35. doi:10.47525/ulasbid.1398206

Pourpanah, F., Abdar, M., Luo, Y., Zhou, X., Wang, R., Lim, C. P., Wang, X-Z., & Wu, Q. M. J. (2023). A Review of Generalized Zero-Shot Learning Methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45 (4), 4051–4070. doi:10.1109/TPAMI.2022.3191696

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language Models Are Unsupervised Multitask Learners. OpenAI Blog. https://cdn.openai.com/better-language-models/language_models_are_unsupervised_multitask_learners.pdf (April 17, 2025).

Rajalakshmi, S., Asha, S., & Pazhaniraja, N. (2017). A Comprehensive Survey on Sentiment Analysis. *In 2017 Fourth International Conference on Signal Processing, Communication and Networking (ICSCN)*. IEEE. doi:10.1109/icscn.2017.8085673

Revathi, K. L., Satish, A. R., & Rao, P. S. (2023). Fine - Grained Sentiment Analysis on Online Reviews. *In 2023 Third International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT)*. IEEE. doi:10.1109/ICAECT57570.2023.10118291

Roumeliotis, K. I., Tselikas, N. D., & Nasiopoulos, D. K. (2024). Leveraging Large Language Models in Tourism: A Comparative Study of the Latest GPT Omni Models and BERT NLP for Customer Review Classification and Sentiment Analysis. *Information*, 15 (12), 792. doi:10.3390/info15120792

Ruiz-Millán, J. A., Martínez-Cámara, E., Victoria Luzón, M., & Herrera, F. (2022). Personalised federated learning with BERT fine tuning. Case study on Twitter sentiment analysis. In L., Troiano, A., Vaccaro, R., Tagliaferri, N., Kesswani, I. D., Rodriguez, I. Brigui, & D. Parente (Eds.), *Advances in Deep Learning, Artificial Intelligence and Robotics. Lecture Notes in Networks and Systems, vol 249* (pp. 193–202). Springer. doi:10.1007/978-3-030-85365-5_19

Sabharwal, N., & Agrawal, A. (2021). Future of BERT Models. Navin Sabharwal, & Amit Agrawal (Eds.), *Hands-on Question Answering Systems with BERT*. Apress: Berkeley, 173–178. doi:10.1007/978-1-4842-6664-9_7

Sabiri, B., Khtira, A., El Asri, B., & Rhanoui, M. (2023). Analyzing BERT's performance compared to traditional text classification models. In J. Filipe, M. Śmiałek, A. Brodsky, & S. Hammoudi (Eds.), *Proceedings of the 25th International Conference on Enterprise Information Systems* (pp. 572–582). SciTePress. doi:10.5220/0011983100003467

Sanh, V., Debut, L., Chaumond, J., & Wolf, T. (2019). DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. *ArXiv*. doi:10.48550/ARXIV.1910.01108

Saragih, H., & Manurung, J. (2024). Leveraging the BERT Model for Enhanced Sentiment Analysis in Multicontextual Social Media Content. *Jurnal Teknik Informatika C.I.T Medicom*, 16 (2), 82–89. doi:10.35335/cit.vol16.2024.766

Sarma, S. (2023). Zero-Shot Learning for Computer Vision Applications. *In MM '23: The 31st ACM International Conference on Multimedia* (pp. 9360–9364). Association for Computing Machinery. doi:10.1145/3581783.3613435

Schweter, S. (2020). BERTurk - BERT models for Turkish (Version 1.0.0). *Zenodo*. doi:10.5281/ZENODO.3770924

Shah, Z., & Rai, S. (2022). A Research Paper on the Effects of Customer Feedback on Business. *International Journal of Advanced Research in Science, Communication and Technology*, 672–675. doi:10.48175/ijarsct-5743

Shen, Y., & Zhang, P. K. (2024). Financial Sentiment Analysis on News and Reports Using Large Language Models and FinBERT. *In 2024 IEEE 6th International Conference on Power, Intelligent Computing and Systems (ICPICS)* (pp. 717–721). IEEE. doi:10.1109/icpics62053.2024.10796670

Shukla, D., & Dwivedi, S. K. (2024). The Study of the Effect of Preprocessing Techniques for Emotion Detection on Amazon Product Review Dataset. *Social Network Analysis and Mining*, 14 (1). doi:10.1007/s13278-024-01352-4

Singh, A., Hasan, Z., & Paranjape, V. (2023). Analyzing Sentiment and Emotion in Omicron-Related Social Media Content: An NLP Perspective. *International Journal of Innovative Research in Computer and Communication Engineering*, 10 (01), 279–283. doi:10.15680/ijircce.2022.1001046

Sree Harsha, S., Krishna Swaroop, K., & Chandavarkar, B. R. (2021). Natural Language Inference: Detecting Contradiction and Entailment in Multilingual Text. In K. R. Venugopal, P. D. Shenoy, R. Buyya, L. M. Patnaik, S. S. Iyengar (Eds.), *Data Science and Computational Intelligence. ICInPro 2021. Communications in Computer and Information Science, vol 1483* (pp. 314-327) Springer. doi:10.1007/978-3-030-91244-4_25

Stine, R. A. (2019). Sentiment Analysis. *Annual Review of Statistics and Its Application*, 6 (1), 287–308. doi:10.1146/annurev-statistics-030718-105242

Swamy, T. N., Lakshmi, J. K., Uma, M. P., Reshma, D., & Jayanth, D. (2022). Multimedia Sentiment Analysis on Public Social Network Using NLP. *International Journal for Research Trends and Innovation*, 7 (8), 825-830.

Taherdoost, H., & Madanchian, M. (2023). Artificial Intelligence and Sentiment Analysis: A Review in Competitive Research. *Computers*, 12 (2), 37. doi:10.3390/computers12020037

Tanuwijaya, C. K., Ellitan, L., & Lukito, R. S. H. (2023). The Effect of Online Customer Reviews on Purchase Intention with Customer Trust as a Variable in Purchase Decision on Sociolla Consumers. *Journal of Entrepreneurship & Business*, 4 (3), 192–203. doi:10.24123/jeb.v4i3.5764

Teke, B., Yazıcı, S. N., Zamir, G., Budak, A. B., & Karabey Aksakallı, I. (2025). Berturk-Based Sentiment Analysis on E-Commerce Multi Domain Product Reviews. *Afyon Kocatepe University Journal of Sciences and Engineering*, 25 (3), 497–509. doi:10.35414/akufemubid.1537513

Tsiligaridis, J. (2024). Approaches of classification models for Sentiment Analysis. In D. C. Wyld & D. Nagamalai (Eds.), *5th International Conference on Advanced Natural Language Processing* (pp. 81-86). Computer Science & Information Technology (CS & IT). doi:10.5121/csit.2024.141007

Tuzcu, S. (2020). Çevrimiçi Kullanıcı Yorumlarının Duygu Analizi ile Sınıflandırılması. *ESTUDAM Bilişim Dergisi*, 1 (2), 1–5.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., Polosukhin, I. (2017). Attention Is All You Need. *ArXiv*. doi:10.48550/ARXIV.1706.03762

Viet, N. V., & Vinh, N. T. (2024). Large Language Models in Software Engineering. *Journal of Education for Sustainable Innovation*, 2 (2), 146–156. doi:10.56916/jesi.v2i2.968

Weidinger, L., Mellor, J., Rauh, M., Griffin, C., Uesato, J., Huang, P.-S., ... Gabriel, I. (2021). Ethical and Social Risks of Harm from Language Models. *ArXiv*. doi:10.48550/ARXIV.2112.04359

Willemsen, L. M., Neijens, P. C., & Bronner, F. (2012). The Ironic Effect of Source Identification on the Perceived Credibility of Online Product Reviewers. *Journal of Computer-Mediated Communication: JCMC*, 18 (1), 16–31. doi:10.1111/j.1083-6101.2012.01598.x

Wirtz, J., & Tomlin, M. (2000). Institutionalising Customer-Driven Learning Through Fully Integrated Customer Feedback Systems. *Managing Service Quality*, 10 (4), 205–215. doi:10.1108/09604520010341654

Wirtz, J., Kuan Tambyah, S., & Mattila, A. S. (2010). Organizational Learning from Customer Feedback Received by Service Employees. *Journal of Service Management*, 21 (3), 363–387. doi:10.1108/09564231011050814

Xue, Q. (2024). Unlocking The Potential: A Comprehensive Exploration of Large Language Models in Natural Language Processing. *Applied and Computational Engineering*, 57 (1), 247–252. doi:10.54254/2755-2721/57/20241341

Yang, L., Charoenporn, T., & Sornlertlamvanich, V. (2024). Comparative Study of Traditional Machine Learning and Quantum Computing in Natural Language Processing: A Case Study on Sentiment Analysis. *In 2024 IEEE International Symposium on Consumer Technology (ISCT)* (pp. 269–273). doi:10.1109/isct62336.2024.10791272

Yildirim, S. (2024). Fine-Tuning Transformer-Based Encoder for Turkish Language Understanding Tasks. *ArXiv*. doi:10.48550/ARXIV.2401.17396

Zhang, C., Fan, C., Yao, W., Hu, X., & Mostafavi, A. (2019). Social Media for Intelligent Public Information and Warning in Disasters: An Interdisciplinary Review. *International Journal of Information Management*, 49, 190–207. doi:10.1016/j.ijinfomgt.2019.04.004