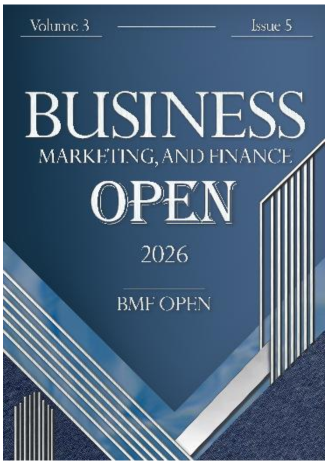
Sentiment Analysis of Customer Opinions in Iran: A Systematic Review

Azra Nezhadfathi[1], Ghasem Bakhshandeh[2,*] and Maryam Darvishi[3]



Citation: Nezhadfathi, A., Bakhshandeh, G., & Darvishi, M. (2026). Sentiment Analysis of Customer Opinions in Iran: A Systematic Review . *Business, Marketing, and Finance Open*, 3(5), 1-15.

Received: 08 October 2025
Revised: 27 January 2026
Accepted: 05 February 2026
Initial Publication: 06 February 2026
Final Publication: 01 September 2026

[1] Department of Management, Ahv.C., Islamic Azad University, Ahvaz, Iran;
[2] Department of Management, Ahv.C., Islamic Azad University, Ahvaz, Iran;
[3] Department of Business Management, Om.C., Islamic Azad University, Omidiyeh, Iran;

* Correspondence: gh.bakhshandeh@iau.ac.ir

Abstract: The quick growth of digital commerce and online communication with clients has generated enormous volumes of user-produced textual data, requiring automation to extract insight into consumer opinion. This systematic review explains the contemporary status of sentiment analysis studies about Iranian consumers' opinions, focusing on datasets, techniques, evaluation procedures, and limitations contained in the literature. According to PRISMA 2020 guidelines, pertinent studies were located in both the IEEE Xplore and MagIran databases by both English and Persian search terms. The studies selected for inclusion were evaluated according to a modified JBI checklist and separated according to their employed methodological approach, including lexicon-based rules, pattern-mining and graph-based methods, neural networks, and transformer-based pre-trained models. The results show that although the transformer model and hybrid methods utilizing deep learning are more efficient than the traditional techniques in use, both the results of the transformer and hybrid methods are restricted by limited and domain-specific datasets in Persian, various methods of preprocessing, and the absence of standard evaluation. Linguistic issues, including extensive morphology, different representations of orthographical spelling, and successive code-mixing of different languages, also contribute to the difficulties of adequately interpreting sentiment issues with accuracy. The analysis exposes several major research deficiencies, necessitating the generation of large, standardized, and multi-domain corpora in Persian suitable for inter-application evaluation of sentiment analysis procedures. In addition, sentiment analysis procedures suitable to the linguistic demands of Persian, called for in adopting extensive and standard datasets, are suggested, which will provide easier explainable and reproducibility of sentiment analysis.

Keywords: Sentiment analysis, Customer opinions in Iran, Natural language processing, Deep learning, Transformer models.

1. Introduction

The rapid expansion of digital platforms, online marketplaces, and social media ecosystems has fundamentally transformed the way customer opinions are expressed, disseminated, and consumed. Contemporary consumers increasingly rely on digital channels to articulate evaluations, emotions, and attitudes toward products, services, and brands, producing vast volumes of unstructured textual data on a daily basis. This unprecedented availability of user-generated content has created both an opportunity and a methodological challenge for organizations, researchers, and policymakers seeking to extract actionable insights from customer feedback at scale. Within this context, sentiment analysis—also referred to as opinion mining—has emerged as a core analytical approach within the broader field of natural language processing (NLP), enabling the automated identification and classification of subjective information embedded in text [1-3].

Globally, sentiment analysis has evolved from early lexicon-based and rule-driven techniques toward sophisticated machine learning and deep learning paradigms capable of capturing complex semantic, contextual, and emotional patterns in language [4-6]. Traditional lexicon-based methods rely on predefined sentiment dictionaries and heuristic rules to assign polarity scores, offering interpretability and computational efficiency but suffering from limited contextual awareness, particularly in the presence of sarcasm, irony, and domain-specific expressions [7-9]. In contrast, neural architectures such as convolutional neural networks (CNNs), recurrent neural networks (RNNs), and long short-term memory (LSTM) networks have demonstrated superior performance by learning distributed representations of text directly from data [5, 10]. More recently, transformer-based models have redefined the state of the art by leveraging self-attention mechanisms and large-scale pre-training to capture deep contextual dependencies across linguistic units [11-13].

The growing importance of sentiment analysis is particularly evident in the domain of e-commerce and digital marketing, where customer reviews, ratings, and comments play a critical role in shaping purchasing decisions, brand perception, and competitive strategy. Empirical evidence suggests that sentiment-driven analytics can significantly enhance customer experience management, recommendation systems, demand forecasting, and marketing intelligence [14-16]. Recent studies further indicate that transformer-based sentiment models outperform conventional approaches in predicting customer satisfaction and behavioral intentions in online retail environments [17-19]. These advances align with the broader paradigm of Marketing 5.0, where artificial intelligence and data-driven personalization are central to value creation [16].

Despite these global advances, sentiment analysis remains unevenly developed across languages and cultural contexts. A substantial proportion of methodological innovations and benchmark datasets are concentrated in high-resource languages such as English, Chinese, and major European languages, creating structural disadvantages for low-resource and morphologically rich languages [20, 21]. Persian (Farsi), spoken by millions of users and serving as the primary language of digital interaction in Iran, exemplifies this challenge. Persian sentiment analysis faces a combination of linguistic complexity, limited standardized datasets, and inconsistent preprocessing conventions, which collectively constrain the transferability of mainstream sentiment analysis techniques [22-24].

From a linguistic perspective, Persian exhibits rich morphology, flexible word order, orthographic variation, and frequent use of informal writing styles in online discourse. The widespread presence of pseudo-spaces, inconsistent character encoding, colloquial spellings, and code-mixing with English poses significant obstacles to reliable tokenization, normalization, and feature extraction [25, 26]. These characteristics complicate both lexicon construction and the application of pre-trained multilingual models, often leading to degraded performance if language-specific adaptations are not carefully implemented [27, 28]. Moreover, sentiment expressions in Persian frequently rely on implicit cues, idiomatic constructions, and pragmatic context, which are difficult to capture using shallow modeling approaches.

The Iranian digital ecosystem further amplifies the importance of sentiment analysis due to the rapid growth of local e-commerce platforms, online service providers, and review-based decision environments. Platforms such as Digikala, SnappFood, and other domestic marketplaces generate extensive customer feedback that directly influences product quality assessment, service optimization, and consumer trust. Several empirical studies have demonstrated the applicability of sentiment analysis to Iranian customer reviews, highlighting its potential for product evaluation, after-sales service assessment, and spam detection [29-31]. However, this body of research remains fragmented, with substantial variation in methodological choices, datasets, and evaluation metrics.

Early Persian sentiment analysis studies predominantly employed lexicon-based and rule-based approaches, often adapted from English resources through translation or manual enrichment [7, 8]. While these methods offered interpretability and low computational cost, their performance was limited by vocabulary coverage, domain dependence, and sensitivity to linguistic nuance [9, 32]. Subsequent research introduced hybrid models that combined lexicon features with machine learning classifiers, demonstrating improved accuracy in specific domains such as hotel reviews and product feedback [30, 33]. Nonetheless, these approaches continued to struggle with generalization across domains and informal text genres.

The adoption of deep learning marked a significant methodological shift in Persian sentiment analysis. Neural models leveraging word embeddings and sequence learning mechanisms enabled more robust handling of semantic variation and contextual information [34, 35]. Studies using CNNs, LSTMs, and hybrid architectures reported notable gains in classification performance on large-scale datasets derived from Iranian e-commerce platforms [36]. However, deep learning approaches introduced new challenges related to data requirements, computational cost, and interpretability, particularly in resource-constrained research settings [21, 37].

More recently, transformer-based models and fine-tuning strategies have gained traction in Persian sentiment analysis, reflecting global trends in NLP. Fine-tuned transformer architectures have demonstrated superior performance in capturing long-range dependencies and subtle sentiment cues, even in low-resource scenarios [11, 38]. Complementary research has explored bilingual and cross-lingual sentiment modeling, highlighting the potential of multilingual transformers for improving sentiment analysis in Persian through shared representations [19, 39]. Nevertheless, the effectiveness of these models remains constrained by the scarcity of large, publicly available, and consistently annotated Persian datasets.

Dataset limitations represent one of the most persistent bottlenecks in Persian sentiment analysis research. Existing datasets are often domain-specific, imbalanced, and annotated using heterogeneous labeling schemes, hindering reproducibility and cross-study comparison [23, 24]. To mitigate data scarcity, recent studies have proposed synthetic data augmentation techniques, including the use of generative adversarial networks (GANs), to enhance model robustness [40]. While promising, such approaches raise concerns regarding semantic fidelity and overfitting, underscoring the need for rigorous validation frameworks.

Beyond methodological considerations, the application of sentiment analysis to Iranian customer opinions carries broader implications for business intelligence, service quality management, and digital governance. Sentiment-driven insights can inform strategic decision-making, support customer-centric innovation, and enhance transparency in digital markets [41, 42]. In specialized contexts such as healthcare feedback, sentiment analysis has also demonstrated value in capturing patient experiences from free-text Persian comments, highlighting its cross-sector relevance [26]. However, the operational deployment of sentiment models in real-world Iranian settings remains limited by concerns related to explainability, fairness, and trust in automated systems.

Given these developments, there is a clear need for a systematic, integrative examination of sentiment analysis research focused on Iranian customer opinions. Existing reviews have addressed sentiment analysis from global or methodological perspectives [2, 3], while surveys on Persian sentiment analysis have often emphasized linguistic challenges without a specific focus on customer-oriented applications [22, 23]. Moreover, recent advances in transformers, hybrid modeling, and data augmentation warrant an updated synthesis that captures the current state of the field and identifies unresolved gaps.

Accordingly, a comprehensive introduction that situates Persian customer sentiment analysis within global methodological trends, linguistic constraints, and applied business contexts is essential for advancing both theoretical understanding and practical impact. Such an approach enables the identification of dominant modeling paradigms, recurring limitations, and emerging research opportunities, while also informing the design of future sentiment analysis systems tailored to the Iranian digital environment.

The aim of this study is to systematically examine and synthesize existing research on sentiment analysis of Iranian customer opinions by analyzing employed datasets, methodological approaches, evaluation practices, and reported limitations in order to identify current trends, critical gaps, and directions for future research.

2. Methodology

Identification of Research Question

This systematic review aims to investigate the existing research on sentiment analysis of Iranian customer opinions. The research question guiding the procedure is “What techniques, databases, and analysis methods have been used in the sentiment analysis of Iranian customer opinions, and what gaps and limitations can be found in the current research?” This review aims to explore the techniques, information sources, and performance criteria used in the current literature and identify what research gaps and limitations are identified in that literature. The scope of the research is to discover how the current research gives attention to sentiment analysis in the Iranian field, particularly in consumer areas, and also to identify the outlook for developmental efforts in the future.

Search Strategy

A wide-ranging literature search was carried out across both Iranian and international academic databases to cover as many comparatively unbiased relevant studies as possible. The international database used was IEEE Xplore, while “MagIran.ir” was used as a primary Iranian database. Search terms were applied with the use of structured search terms in English and Persian, designed to find suitable publications. The search terms used in English were (“Sentiment Analysis” OR “Opinion Mining”) AND (“Customer” OR “Consumer” OR “User Review”) AND (“Iran” OR “Persian”). The search terms used in Persian were the following; "تحلیل احساسات مشتریان" OR "کاوش نظر کاربران" OR "تحلیل دیدگاه" OR "نظرکاوی". The search was intended to find peer-reviewed literature, conference papers, and academic articles for the sentiment analysis of Iranian customer data.

Study Selection

The study selection process will follow pre-determined inclusion and exclusion criteria to ensure methodical rigour in the selection methods. Initially, the titles and abstracts were reviewed for studies relevant to the sentiment analysis of Iranian customer data. Full-text articles that met the relevance criteria were then reviewed. The duplicate articles, as well as several studies not pertinent to sentiment analysis or not databases relevant to Iranian customer opinion, were excluded. Studies that were published in English or Persian formats were considered. The selection process followed the PRISMA 2020 flow chart in order that the selection process is both clear and reproducible.

3. Findings and Results

Sentiment analysis

As mentioned in many recent reviews, sentiment analysis has moved from rule-based lexicon approaches to advanced machine-learning and deep-learning methods to process richer emotional expressions and more complex linguistic contexts [2, 6].

Generally, implementing a sentiment-sentiment analysis process involves the following main stages and the typical pipeline is shown in Figure 1: (I) data collection, when the text data are collected from sources such as consumer reviews, social media posts, or questionnaires; (II) data preprocessing, which modifies the original text into a more structured form by removing unwanted characters, lowering the case of the text, removing stop-words, and adjusting or lemmatizing words; (III) feature engineering or representation, when the cleaned and pre-processed text is modeled in numerical or vectorized form (for example, using TF-IDF measures, word-embeddings, for example, word2vec, etc.) or transformers, so that they can be stated as computational models; (IV) model training and classification, when methods such as logistic-regression, support vector machines, convolutional neural networks, or transformer-based approaches are used, to enable to learn about the relationship between text features and sentiment labels [3]. Such a model must finally be evaluated and interpreted to assess the model's capability for performance according to measures of accuracy, precision, recall, and F1-Score, and thereafter implemented practically, so that the insights gained may be monitored and measured continuously for performance and drift.

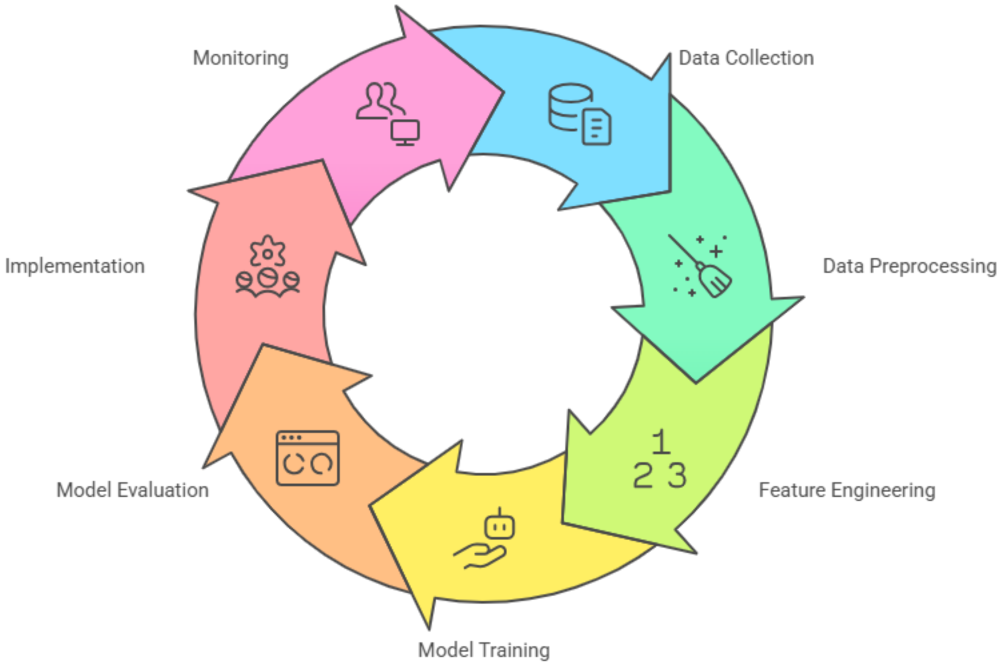


Figure 1. The general process of sentiment analysis.

Sentiment analysis is used for a variety of applications, one of which is customer feedback & e-commerce. In this area, sentiment analysis is used to automatically process huge quantities of customer reviews, comments, and feedback regarding different products or services, allowing companies to extract actionable insights from unstructured, user-generated text [42]. For instance, with the rapid increase in online shopping and review systems, companies are employing sentiment analysis methods to track down underlying issues, follow trends in customer satisfaction and keywords, to modify marketing or product strategies [14]. It was shown that in e-commerce situations, the specific use of transformer-based deep learning models, e.g., BERT or RoBERTa, significantly improves accuracy in understanding customer sentiment within reviews so therefore decision making can be improved [18]. Furthermore, it was shown in a recent systematic review that sentiment-based predictive modelling integrated within Marketing 5.0 environments has a direct effect on improvements in customer experience outcomes that also affect online purchase behavior [16]. Given the importance and significance of analyzing customer sentiment within a digital context in commerce, the research question posed by this review is, therefore, specifically how these methods are being applied within an Iranian context.

Taxonomy

The studies included in this systematic review, collected from the IEEE and MagIran databases to address the research question of this paper, can be categorized into four groups based on the modeling and implementation approaches they employed. These studies are compared and classified in Table 1. These categories are:

Lexicon-Based / Rule-Based Methods

This category uses purpose-built or curated sentiment lexicons (lists of words with a polarity or intensity score associated) and rule-based heuristics (e.g., negation, intensifiers, syntactic patterns) to assess sentiment for texts. These methods are relatively lightweight, interpretable, and domain agnostic, but they may suffer from context, sarcasm, or new expression disambiguation [4, 9]. Jamshidi Nejad, et al. [30] proposed in a supervised approach of sentiment analysis a method of detecting opinion spam in Persian hotel reviews. Using 10,000 of these obtained from Iranian tourist websites, proposed a lexicon, domain specific of 4,570 sentiment words and developed a set of 22 in features, synthesizing together contextual, metadata, based on entities, and sentiment indicators. The sentiment features like polarity sign, ratio of positive words, negative words, etc., and level of subjectivity of the words were important for detecting of feelings of fake from true ideas. The method was classified with nine supervised classifiers, and it is found out that Decision Tree (98.67 percent) and AdaBoost (98.00 percent) have the maximum accuracy in results obtained.

Basiri et al. [33] proposed a target-aware lexicon-based method for sentiment polarity detection in Persian, which applied a strategy different from traditional models. It used averages of all opinionated words and associated them with the main target that exists in any review detected through syntactic parsing, ontology (the Most General First strategy or Most Occurring First strategy), etc. Using three manually labelled datasets of Persian (i.e. movies, hotels, and digital products) it improved polarity detection results by 17% in accuracy and 12% in F1 over the base. This shows that the existence of sentiment-target in opinion analysis can improve results.

Hosseinzadeh Bendarkheili et al. [29] proposed a lexicon model of sentiment analysis for Persian online shopping reviews. They attained data through Digikala and addressed with aspect-level sentiment analysis of the products with respect to features like camera, battery, and screen. The model was different from earlier models for binary polarity in that it incorporated adverbials and intensification to cover the differences in degrees of strength and to assess continuous polarity scores for the product instead of labelling the received speech.

Basiri and Kabiri [27] presented a treatment of contextual polarity at the document level through uninorm-based aggregations of sentence-level sentiments. Evaluated on the Persian review datasets gathered from Digikala. Com, it outperformed the Dempster–Shafer, in recall and F-score, indicating better performance in taking into account conflicting sentiment. However, it was slightly less accurate as a model, and its performance required the identity parameter to be tuned.

Pattern-Mining / Graph-Based Methods for Aspect Extraction

This class extracts explicit aspects (targets) and their related sentiment expressions via mining of frequent patterns, such as graph traversals or dependency graphs [32]. Especially in Aspect-Based Sentiment Analysis (ABSA), where various methods utilize algorithms like frequent itemset mining (e.g., FP-Growth) paired with graph representations of syntactic relations to identify multi-word aspect units and to connect these with opinion expressions [43]. Jamshidi-Nejad et al. [32] proposed a novel approach for sentiment-oriented aspect extraction for a dataset of 10,000 Persian hotel reviews. This method utilized Frequent Pattern mining (FP-Growth) combined with a graph traversal approach to extract explicit aspects associated with sentiment expressions. Their proposed system, in addition to modeling multi-word aspect structures (e.g., noun–adjective, noun–noun), significantly outperformed LDA-, POS, and lexicon-based baselines.

Deep Neural Networks Methods

This category involves sentiment analysis by using neural network architectures (e.g., Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), Long Short-Term Memory (LSTM) / Bi-directional LSTM networks) to which pre-processed text embeddings are input (Word2Vec, GloVe, fastText) to learn to classify sentiments [5, 10]. These methods are more flexible and effective than rule-based methods when sufficient labeled data are available. However, they also may still lack fine-grained interpretability and may not work well in domains shift occurs or in low-resource languages [21, 37]. Heydari et al. [35] developed a hybrid architecture for deep learning for sentiment analysis in Persian using a collection of more than 100,000 customer reviews on the Digikala online retailer site. The dataset consisted of 14 product categories, in which the texts were labelled as positive, negative, or neutral. After the preprocessing of the texts and their vectorization based on GloVe embeddings, several architectures such as Multi-Layer Perceptron (MLP), LSTM, Bi-LSTM, and Gated Recurrent Unit (GRU) were tested for tasks of sentiment classification. The hybrid architecture CharEmbed + Word2Vec + LSTM scored the best result, with an F1-Score of 78.3%.

Samadanian and Kaedi [31] addressed the subject of ABSA. In their work, they evaluated customer opinions about the quality of after-sales service for three brands of washing machines, Snowa, Pakshoma, and GPlus. The authors in this study constructed a multi-label dataset with the help of manual labelling and augmentation. They tested deep learning models based on CNN, LSTM, GRU, and Bi-LSTM networks.

Soleimanzadeh and Shayegan [36] employed a model of hybrid deep learning for the purposes of Persian sentiment analysis with the use of Digikala reviews and SnappFood. Their model, entitled FastText-Bi-LSTM-CNN, succeeded by results of high degree of accuracy (97.54% and 87.65%), which is better than the methods of Word2Vec proposed and the baselines based on single models.

Roshanfekr et al. [34] faced the linguistic challenges such as the complex morphology of languages and inconsistent writing forms. They produced a large data set of 200,761 Digikala products' reviews, most of which were some 50,000 labels of supervised steps. Using Skip-gram embeddings, the authors produced a classification of the neural convolutions and Bi-LSTM comparison to the results of a traditional NBSVM-bi baseline. The CNN models produced the best F1-Scores (55.4%), compared to both Bi-LSTM (53.2%) and NBSVM-bi (44.0%).

Transformer-Based / Pre-Trained Models

This top-line level uses large pre-trained models (e.g., BERT, RoBERTa, XLM-R) and tunes them specifically for sentiment or Aspect-based sentiment [12]. These models can capture deep contextual information and yield strong performance even in multilingual or low-resource domains, although at the cost of moderate computational demands and careful fine-tuning [13, 20]. Mosafer and Mohseni [38] developed the parsBERT transformer model to fine-tune it for sentiment analysis in Persian using a collection of ~70,000 Snappfood user reviews, which were annotated as being positive or negative. The results showed that domain-specific fine-tuning of the model complex significantly improves the score in sentiment classification, resulting in an F1-Score of about 86% and outperforming traditional machine-learning methods.

Table 1. Summary of the reviewed studies on sentiment analysis of Iranian customer opinions.

Study	Year	Dataset	Method/Algorithm	Features	Evaluation Metric	Key Findings	Limitations
[38]	2025	Snappfood Persian Sentiment Dataset (~70,000 reviews)	Fine-tuned ParsBERT	Text embeddings from ParsBERT	Accuracy Precision Recall F1-Score	Fine-tuning ParsBERT significantly improves Persian sentiment analysis.	Limited annotated Persian data; domain-specific dataset; computational cost; generalization challenges across other domains
[36]	2025	Digikala (3,261 reviews) SnappFood (69,480 reviews)	FastText-Bi-LSTM-CNN Word2Vec-Bi-LSTM-CNN	FastText & Word2Vec embeddings	Accuracy Precision Recall F1-Score	Hybrid FastText-Bi-LSTM-CNN achieved superior performance over Word2Vec and single models	Performance is still limited by dataset diversity
[31]	2024	1,077 Persian reviews from Digikala on washing machines (Snowa, Pakshoma, Gplus)	CNN LSTM GRU Bi-LSTM	Text features from preprocessed reviews (tokenization, normalization, stop-word removal)	Accuracy Precision Recall F1-Score Hamming Loss	CNN outperformed others Gplus had highest customer satisfaction Pakshoma lowest	Small and imbalanced dataset Focused only on after-sales service aspect
[35]	2021	100,000 Persian customer reviews from Digikala online retailer (14 product categories)	CharEmbed + Word2Vec + LSTM	Word embeddings (GloVe, 50-dim) Character embeddings	F1-Score	Hybrid LSTM model outperformed other architectures	Limited size of annotated Persian datasets and restricted computational resources (GPU access)
[32]	2020	10,000 Persian hotel reviews from IranHotelOnline & Egardesh (21,474 subjective sentences)	Hybrid method: FP-Growth + ADG (Aspect Detection Graph) + Neo4j graph traversal + PMI filtering; POS tagging (NLPTools); HAZM dependency parser	Variable-radius noun extraction near sentiment words Domain-specific sentiment lexicon (4,570 terms) Noun/adjective multi-word aspect detection	Precision Recall F-Score	Proposed method significantly outperforms classical and existing Persian aspect extraction models; Effective at extracting complex Persian compound aspects	Domain-specific (hotels only) Rule-based and requires expert linguistic knowledge High preprocessing effort Only explicit aspects Graph/FP-Growth computational load
[30]	2020	10,000 Persian hotel reviews collected from IranHotelOnline.com and Egardesh.com	Decision tree AdaBoost Random forest Naïve bayes	Contextual Metadata-based Entity-based Sentiment-based	Accuracy Precision Recall	Developed a Persian sentiment lexicon (4,570 words) and	Domain-specific (hotel reviews only) Manual labeling

			SVM KNN Neural Networks Bagging Rule induction			demonstrated that sentiment-based and linguistic features significantly enhance spam detection accuracy	Language dependency
[33]	2019	Movie reviews (Naghdefarsi) Hotel reviews (various sites) PerView (Digikala-digital products)	Lexicon-based sentiment analysis with Target Identification using POS tagging, dependency parsing, and ontology-based strategies (MOF, MGF, MSF, FOF, LOF)	Potential terms (adjectives, adverbs, negations), POS tags, syntactic relations, ontology-based target hierarchy	Precision Recall F1-Score Accuracy Specificity	Identifying main targets improves document-level sentiment polarity detection in multi-topic Persian reviews MGF best for target detection MOF best for polarity	Manual dataset labeling Lexicon dependency Limited to Persian language
[29]	2019	User reviews from Digikala (Persian e-commerce site) Reviews of three mobile phones (2017)	Lexicon-based sentiment analysis	Aspect extraction (camera, battery, display, etc.) Inclusion of intensifiers and adverbs Continuous sentiment scoring	Accuracy	Incorporating intensifiers and continuous polarity values improved sentiment detection	Relies on manual lexicon enrichment Limited to structured reviews Lacks handling of informal or unstructured text
[27]	2018	Four Persian review datasets (Apple, Huawei, Note 5, Samsung) from Digikala.com	Uninorm-based sentence-level sentiment aggregation compared with DS method	Lexicon-based sentence-level sentiment scores aggregated using extended cross-ratio uninorm operator	Precision Recall F-Score Accuracy	Uninorm aggregation improved recall and F- Score compared to DS	Slightly lower accuracy than DS Sensitive to identity parameter tuning Limited to lexicon-based analysis
[34]	2017	200,761 Persian product reviews from Digikala (≈50,000 labeled)	Skip-gram + CNN Bi-LSTM	Word embeddings learned via Skip-gram model (150-d vectors)	Precision Recall F1-Score	Deep learning models (CNN, Bi-LSTM) outperformed traditional machine learning; embeddings improved semantic understanding and generalization	Dataset imbalance (positive ≫ negative) Binary labels only No neutral class

Analysis of sentiment in Persian has achieved meaningful progress, but there are still many problems that impair its robustness, particularly in relation to feedback from Iranian customers. One principal reason is that there are few large, structured Persian datasets available [23]. Datasets for Persian sentiment analysis have been produced (e.g., Digikala Reviews), but they are usually small, limited to certain fields, and inconsistently annotated. This makes it hard to use them across different fields. An example is the finding of Dashtipour et al. [28] that most of the datasets available contain fewer than 10,000 labeled samples, while the datasets for English frequently have millions of them. Moreover, the labeling and annotating schemes differ in the various resources, making it difficult to compare the various models. Further, recent research has shown that in the Persian review datasets, there is a severe misrepresentation of the number of positive and negative samples, the number of positive ones being very much larger than that of the negative ones. This leads to biased learning and unreliable evaluations [35] and, therefore, although there are datasets available, their limited standardization and coverage remain a major obstacle to methodological advancement.

In grammatical terms, Persian has several problems. The highly inflected morphology, frequent tokenization ambiguity (including pseudo-spaces and non-standard spacing), and heavy code-mixing (Persian-English) in informal customer reviews. For example, several researchers have pointed out that Persian text often requires custom letter normalization ("ی" vs. "ي") and pseudo-space handling before modelling [26]. All this increases the amount of work to be done with the text, and the application of all the tools designed for English to the Persian text becomes very difficult.

In methodological terms, the processing of the data, the selection of the relevant features, and the treatment of the results are not consistent, and it is impossible to make sense of the comparison of results or the reproducibility of results [22]. Limited resources lead many Persian studies to use rather simple methods based on lexicon or rules, which are unable to suggest rich contextual cues and which limit the extensibility of the models [7, 8]. Emerging transformer models suggest possibilities, but are very resource-intensive in terms of data and computing power, and lack clarity of operation [11].

The recent application of synthetic data augmentation via Generative Adversarial Networks (GANs) for Persian sentiment analysis underscores the data resource gap and the benefits of addressing it. However, synthetic augmentation raises questions about its real-world validity and potential for overfitting [40]. Another under-researched issue is that of cross-domain and cross-lingual transferability for Persian sentiment systems. High-resource languages perennially benefit from the availability of pre-trained multilingual models and large-scale benchmarks, while this is lacking in Persian, since comparable volumes of customer review corpora and standard benchmarks reflecting domain sparsity are lacking. This limits the ability to transfer methodologies from other languages and domains [44].

Finally, the real-world deployment of sentiment models in business or customer service pipelines often suffers from interpretability problems. The complexity of the deep and transformer models makes it difficult for users to trace the way predictions of sentiment are reached, which can make it harder to use them in decision-making processes. In conclusion, it remains essential to address the issues of Persian data dearth, linguistic complexity, methodological standardization, and interpretability to advance robust sentiment analysis systems in the Iranian market.

4. Discussion and Conclusion

The findings of this study demonstrate a clear methodological evolution in sentiment analysis research focused on Iranian customer opinions, revealing a progressive shift from lexicon-based and rule-driven approaches toward deep learning and transformer-based architectures. The reviewed studies collectively indicate that while early sentiment analysis efforts in Persian relied heavily on manually constructed lexicons and heuristic rules, more recent research has increasingly adopted data-driven models capable of capturing contextual and semantic complexity. This transition mirrors global developments in sentiment analysis research [1, 3], yet it occurs within a uniquely constrained linguistic and infrastructural environment that significantly shapes methodological outcomes in the Iranian context.

The results show that lexicon-based and rule-based methods remain prevalent in Persian sentiment analysis, particularly in earlier studies and in applications where interpretability and low computational cost are prioritized. These approaches have demonstrated reasonable effectiveness in narrowly defined domains such as product reviews and hotel feedback, especially when enhanced with target identification, sentiment intensifiers, or aggregation mechanisms [27, 29, 33]. However, the findings indicate that such methods consistently underperform when confronted with informal language, implicit sentiment, sarcasm, or domain transfer, which are common features of Iranian online customer discourse. This limitation aligns with broader critiques of lexicon-based sentiment analysis in both Persian and non-Persian contexts [4, 9].

Pattern-mining and graph-based approaches for aspect extraction represent a more specialized methodological stream identified in the reviewed literature. The results suggest that these methods offer notable advantages for aspect-based sentiment analysis by explicitly modeling relationships between opinion expressions and their corresponding targets. In Persian customer reviews, where sentiment is often directed toward specific product attributes or service components, such approaches have demonstrated superior precision compared to topic-modeling and purely statistical baselines [30, 43]. Nevertheless, the findings also reveal that these methods are computationally intensive, highly domain-specific, and reliant on expert linguistic preprocessing, limiting their scalability and generalizability across different sectors of the Iranian digital economy.

A central result of this study is the growing dominance of deep neural network models in Persian sentiment analysis, particularly those based on CNN and LSTM architectures. The reviewed evidence indicates that deep learning models consistently outperform lexicon-based and traditional machine learning approaches when sufficient labeled data are available [34-36]. These models are particularly effective in capturing semantic variation and contextual dependencies in Persian customer reviews, leading to higher F1-scores and improved robustness to lexical diversity. However, the results also reveal substantial variation in reported performance across studies, which can be attributed to differences in dataset size, annotation schemes, preprocessing pipelines, and evaluation protocols.

Transformer-based and pre-trained language models emerge as the most effective methodological paradigm in the reviewed literature, achieving state-of-the-art results in Persian sentiment classification and aspect-based analysis. Fine-tuned transformer models demonstrate a superior ability to model long-range dependencies, contextual polarity shifts, and subtle sentiment cues, even in relatively low-resource settings [11, 20, 38]. The findings further indicate that bilingual and multilingual transformer frameworks enhance performance by leveraging cross-lingual representations, particularly in domains where Persian customer discourse incorporates English loanwords and code-mixed expressions [19, 39]. These results are consistent with international evidence highlighting the adaptability of transformer architectures in multilingual and low-resource sentiment analysis scenarios [13, 21].

Despite these methodological advances, the results underscore persistent structural challenges that constrain the reliability and comparability of Persian sentiment analysis research. A dominant issue identified across studies is the scarcity of large-scale, standardized, and publicly available Persian sentiment datasets. Most reviewed studies rely on domain-specific datasets derived from individual platforms such as Digikala or SnappFood, often with imbalanced class distributions and inconsistent labeling criteria [23, 24]. This fragmentation limits cross-study benchmarking and undermines the generalizability of reported performance gains, a concern echoed in broader analyses of low-resource sentiment analysis [20, 21].

Linguistic complexity emerges as another critical factor shaping the observed results. Persian morphology, orthographic variation, informal writing conventions, and extensive use of pseudo-spaces introduce significant noise into preprocessing and feature extraction stages. The reviewed evidence indicates that inadequate normalization and tokenization substantially degrade model performance, particularly for deep learning architectures that are sensitive to input representation quality [25, 26, 28]. These findings reinforce the necessity of language-specific preprocessing pipelines and challenge the assumption that methods developed for English can be directly transferred to Persian without adaptation.

The study also reveals emerging strategies aimed at mitigating data scarcity, most notably through synthetic data augmentation and hybrid modeling. The use of GAN-based augmentation techniques demonstrates promising improvements in classification accuracy and model robustness [40]. However, the results suggest that such gains must be interpreted cautiously, as synthetic data may amplify existing biases or fail to capture the pragmatic nuances of authentic customer discourse. This tension reflects broader debates in sentiment analysis regarding the trade-off between data quantity and semantic validity [6, 37].

From an application perspective, the results highlight the growing relevance of sentiment analysis for business intelligence, digital marketing, and service quality management in Iran. Studies focusing on e-commerce, after-sales service, and healthcare feedback demonstrate that sentiment-driven insights can meaningfully inform organizational decision-making and customer engagement strategies [26, 31, 41]. Nevertheless, the findings also indicate a gap between experimental research and real-world deployment, particularly concerning model interpretability, transparency, and stakeholder trust. As sentiment analysis systems become increasingly complex, the lack of explainability poses challenges for their adoption in sensitive or high-stakes domains.

Overall, the discussion of results suggests that Persian sentiment analysis of customer opinions has reached a stage of methodological maturity characterized by advanced modeling techniques, yet remains constrained by foundational issues related to data infrastructure, linguistic complexity, and evaluation standardization. The convergence of these factors underscores the need for coordinated efforts to align methodological innovation with resource development and practical applicability.

This study is subject to several limitations that should be acknowledged. First, the analysis is confined to published academic literature, which may underrepresent industry-driven applications and proprietary sentiment analysis systems used in Iranian commercial settings. Second, the heterogeneity of reviewed studies in terms of datasets, domains, and evaluation metrics limits the possibility of direct quantitative comparison across methods. Third, although the review synthesizes a broad range of approaches, it relies on reported results rather than re-implementation or empirical benchmarking, which may introduce bias stemming from inconsistent experimental conditions.

Future research should prioritize the development of large-scale, publicly available, and multi-domain Persian sentiment datasets with standardized annotation guidelines to enable reproducible benchmarking. Greater emphasis should be placed on cross-domain and cross-lingual transfer learning to enhance model generalization in low-resource settings. Additionally, integrating explainable AI techniques into deep and transformer-based sentiment models would address interpretability concerns and facilitate trust in applied contexts. Longitudinal studies examining sentiment dynamics over time and across platforms could further enrich understanding of Iranian customer behavior in digital environments.

From a practical standpoint, organizations operating in the Iranian digital market should invest in sentiment analysis systems that are tailored to the linguistic and cultural characteristics of Persian customer discourse. Practitioners are encouraged to combine advanced transformer-based models with robust preprocessing pipelines and domain adaptation strategies. Emphasis should also be placed on interpretability and transparency to ensure that sentiment-driven insights are actionable and trustworthy for decision-makers. Finally, collaboration between academia, industry, and platform providers could accelerate the translation of research advances into scalable, real-world sentiment analysis applications.

Authors' Contributions

Authors equally contributed to this article.

Ethical Considerations

All procedures performed in this study were under the ethical standards.

Acknowledgments

Authors thank all participants who participate in this study.

Conflict of Interest

The authors report no conflict of interest.

Funding/Financial Support

According to the authors, this article has no financial support.

References

[1] B. Liu, *Sentiment analysis and opinion mining*. Springer Nature, 2022.

[2] H. D. Sharma and P. Goyal, "An Analysis of Sentiment: Methods, Applications, and Challenges," *Engineering Proceedings,* vol. 59, no. 1, p. 68, 2023. [Online]. Available: https://www.mdpi.com/2673-4591/59/1/68.

[3] K. L. Tan, C. P. Lee, and K. M. Lim, "A Survey of Sentiment Analysis: Approaches, Datasets, and Future Research," *Applied Sciences,* vol. 13, no. 7, p. 4550, 2023. [Online]. Available: https://www.mdpi.com/2076-3417/13/7/4550.

[4] S. Zeb, U. Qamar, and F. Hussain, "Sentiment analysis on user reviews through lexicon and rule-based approach," in *Asia-Pacific Web Conference*, 2016: Springer, pp. 55-63.

[5] T. Diwan and J. V. Tembhurne, "Sentiment analysis: a convolutional neural networks perspective," *Multimedia Tools and Applications,* vol. 81, no. 30, pp. 44405-44429, 2022.

[6] H. Taherdoost and M. Madanchian, "Artificial Intelligence and Sentiment Analysis: A Review in Competitive Research," *Computers,* vol. 12, no. 2, p. 37, 2023. [Online]. Available: https://www.mdpi.com/2073-431X/12/2/37.

[7] M. E. Basiri and A. Kabiri, "Translation is not enough: comparing lexicon-based methods for sentiment analysis in Persian," in *2017 international symposium on computer science and software engineering conference (CSSE)*, 2017: IEEE, pp. 36-41.

[8] M. E. Basiri, N. G. Aghaee, and A. R. Naghsh-Nilchi, "Lexicon-based sentiment analysis in Persian," in *Intelligent Computational Systems: A Multi-Disciplinary Perspective*: Bentham Science Publishers, 2017, pp. 154-183.

[9] A. Mahmoudi, D. Jemielniak, and L. Ciechanowski, "Assessing accuracy: a study of lexicon and rule-based packages in R and python for sentiment analysis," *IEEE Access,* vol. 12, pp. 20169-20180, 2024.

[10] I. Priyadarshini and C. Cotton, "A novel LSTM–CNN–grid search-based deep neural network for sentiment analysis," *The Journal of Supercomputing,* vol. 77, no. 12, pp. 13911-13932, 2021.

[11] H. Bashiri and H. Naderi, "Comprehensive review and comparative analysis of transformer models in sentiment analysis," *Knowledge and Information Systems,* vol. 66, no. 12, pp. 7305-7361, 2024.

[12] K. N. Prasanthi, R. E. Madhavi, D. N. S. Sabarinadh, and B. Sravani, "A novel approach for sentiment analysis on social media using BERT & ROBERTA transformer-based models," in *2023 IEEE 8th International Conference for Convergence in Technology (I2CT)*, 2023: IEEE, pp. 1-6.
[13] J. Huang, J. Zhou, Z. Tang, J. Lin, and C. Y.-C. Chen, "TMBL: Transformer-based multimodal binding learning model for multimodal sentiment analysis," *Knowledge-Based Systems,* vol. 285, p. 111346, 2024.
[14] J. R. Saura, P. Palos-Sanchez, and A. Grilo, "Detecting indicators for startup business success: Sentiment analysis using text data mining," *Sustainability,* vol. 11, no. 3, p. 917, 2019.
[15] K. S. Kyaw, P. Tepsongkroh, C. Thongkamkaew, and F. Sasha, "Business intelligent framework using sentiment analysis for smart digital marketing in the E-commerce era," *Asia Social Issues,* vol. 16, no. 3, pp. e252965-e252965, 2023.
[16] V. Gooljar, T. Issa, S. Hardin-Ramanan, and B. Abu-Salih, "Sentiment-based predictive models for online purchases in the era of marketing 5.0: a systematic review," *Journal of Big Data,* vol. 11, no. 1, p. 107, 2024/08/05 2024, doi: 10.1186/s40537-024-00947-0.
[17] T. Alsaedi, M. R. R. Rana, A. Nawaz, A. Raza, and A. Alahmadi, "Sentiment Mining in E-Commerce: The Transformer-based Deep Learning Model," *International journal of electrical and computer engineering systems,* vol. 15, no. 8, pp. 641-650, 2024.
[18] O. Bellar, A. Baina, and M. Ballafkih, "Sentiment Analysis: Predicting Product Reviews for E-Commerce Recommendations Using Deep Learning and Transformers," *Mathematics,* vol. 12, no. 15, p. 2403, 2024. [Online]. Available: https://www.mdpi.com/2227-7390/12/15/2403.
[19] S. Shan, J. Sun, and R. M. C. Macawile, "Examining Customer Satisfaction Through Transformer-Based Sentiment Analysis for Improving Bilingual E-Commerce Experiences," *IEEE Access*, 2025.
[20] Y. Aliyu, A. Sarlan, K. U. Danyaro, A. S. B. Rahman, and M. Abdullahi, "Sentiment analysis in low-resource settings: a comprehensive review of approaches, languages, and data sources," *IEEE Access,* vol. 12, pp. 66883-66909, 2024.
[21] M. S. Islam *et al.*, "Challenges and future in deep learning for sentiment analysis: a comprehensive review and a proposed novel hybrid approach," *Artificial Intelligence Review,* vol. 57, no. 3, p. 62, 2024.
[22] R. Asgarnezhad and S. A. Monadjemi, "Persian sentiment analysis: feature engineering, datasets, and challenges," 2021.
[23] Z. Rajabi and M. Valavi, "A survey on sentiment analysis in Persian: a comprehensive system perspective covering challenges and advances in resources and methods," *Cognitive Computation,* vol. 13, no. 4, pp. 882-902, 2021.
[24] S. Moniri, T. Schlosser, and D. Kowerko, "Investigating the Challenges and Opportunities in Persian Language Information Retrieval Through Standardized Data Collections and Deep Learning," *Computers,* vol. 13, no. 8, p. 212, 2024, doi: 10.3390/computers13080212.
[25] F. Akbari, *Iran's Language Planning Confronting English Abbreviations: Persian Terminology Planning*. Springer Nature, 2020.
[26] A. Yazdani, M. Shamloo, M. Khaki, and A. Nahvijou, "Use of sentiment analysis for capturing hospitalized cancer patients' experience from free-text comments in the Persian language," (in eng), *BMC Med Inform Decis Mak,* vol. 23, no. 1, p. 275, Nov 29 2023, doi: 10.1186/s12911-023-02358-2.
[27] M. E. Basiri and A. Kabiri, "Uninorm operators for sentence-level score aggregation in sentiment analysis," in *2018 4th International Conference on Web Research (ICWR)*, 2018: IEEE, pp. 97-102.
[28] K. Dashtipour, M. Gogate, A. Adeel, H. Larijani, and A. Hussain, "Sentiment analysis of persian movie reviews using deep learning," *Entropy,* vol. 23, no. 5, p. 596, 2021.
[29] F. HosseinzadehBendarkheili, R. MohammadiBaghmolaei, and A. Ahmadi, "Product quality assessment using opinion mining in Persian online shopping," in *2019 27th Iranian Conference on Electrical Engineering (ICEE)*, 2019: IEEE, pp. 1917-1921.
[30] S. J. Nejad, F. Ahmadi-Abkenari, and P. Bayat, "Opinion spam detection based on supervised sentiment analysis approach," in *2020 10th international conference on computer and knowledge engineering (ICCKE)*, 2020: IEEE, pp. 209-214.
[31] S. Samadanian and M. Kaedi, "Aspect-Based Sentiment Analysis of After-Sales Service Quality: A Case Study of Snowa and Competitors Using Digikala Reviews," in *2024 15th International Conference on Information and Knowledge Technology (IKT)*, 2024: IEEE, pp. 1-5.
[32] S. J. Nejad, F. Ahmadi-Abkenari, and P. Bayat, "A combination of frequent pattern mining and graph traversal approaches for aspect elicitation in customer reviews," *IEEE Access,* vol. 8, pp. 151908-151925, 2020.
[33] M. E. Basiri *et al.*, "Improving sentiment polarity detection through target identification," *IEEE Transactions on Computational Social Systems,* vol. 7, no. 1, pp. 113-128, 2019.
[34] B. Roshanfekr, S. Khadivi, and M. Rahmati, "Sentiment analysis using deep learning on Persian texts," in *2017 Iranian conference on electrical engineering (ICEE)*, 2017: IEEE, pp. 1503-1508.
[35] M. Heydari, M. Khazeni, and M. A. Soltanshahi, "Deep Learning-based Sentiment Analysis in Persian Language," in *2021 7th International Conference on Web Research (ICWR)*, 2021: IEEE, pp. 287-291.
[36] F. Soleimanzadeh and M. J. Shayegan, "Sentiment Analysis in Persian Language Using Long Short-Term Memory and Deep Learning," in *2025 11th International Conference on Web Research (ICWR)*, 2025: IEEE, pp. 597-603.

[37] A. J. Moshayedi, A. S. Roy, A. Kolahdooz, and Y. Shuxin, "Deep learning application pros and cons over algorithm," *EAI Endorsed Transactions on AI and Robotics,* vol. 1, no. 1, p. e7, 2022.

[38] H. Mosafer and S. A. H. Mohseni, "A Detailed Approach to Persian Sentiment Analysis Using Fine-Tuning Strategies," in *2025 11th International Conference on Web Research (ICWR)*, 2025: IEEE, pp. 509-514.

[39] E. Kahya Özyirmidokuz, B. Molu Elmas, and E. A. Stoica, "AI-Based Sentiment Analysis of E-Commerce Customer Feedback: A Bilingual Parallel Study on the Fast Food Industry in Turkish and English," *Journal of Theoretical and Applied Electronic Commerce Research,* vol. 20, no. 4, p. 294, 2025. [Online]. Available: https://www.mdpi.com/0718-1876/20/4/294.

[40] M. Mohammadi, M. R. Amin, and S. Tavakoli, "Boosting Sentiment Analysis in Persian through a GAN-Based Synthetic Data Augmentation Method," in *Proceedings of the 1st Workshop on NLP for Languages Using Arabic Script*, 2025, pp. 54-63.

[41] M. Haroon, Z. Alam, R. Kousar, J. Ahmad, and F. Nasim, "Sentiment analysis of customer reviews on e-commerce platforms: A machine learning approach," *Bulletin of Business and Economics (BBE),* vol. 13, no. 3, pp. 230-238, 2024.

[42] H. Huang, A. A. Zavareh, and M. B. Mustafa, "Sentiment analysis in e-commerce platforms: A review of current techniques and future directions," *Ieee Access,* vol. 11, pp. 90367-90382, 2023.

[43] G. D'Aniello, M. Gaeta, and I. La Rocca, "KnowMIS-ABSA: an overview and a reference model for applications of sentiment analysis and aspect-based sentiment analysis," *Artificial Intelligence Review,* vol. 55, no. 7, pp. 5543-5574, 2022.

[44] D. Khashabi *et al.*, "ParsiNLU: A Suite of Language Understanding Challenges for Persian," *Transactions of the Association for Computational Linguistics,* vol. 9, pp. 1147-1162, 2021, doi: 10.1162/tacl_a_00419.