KMANPUB | AI and Tech in Behavioral and Social Sciences Vol. 4 No. 1 (2026): 1-15

Article history:
Received 26 August 2025
Revised 19 September 2025
Accepted 25 October 2025
Published online 01 July 2026

ParsAVC: Appraising Voice of Customers based on Aspect-Based Analysis of Online Reviews Using Unsupervised Technique

Amirhossein Kazemi[1], Moslem Habibi[1*]

[1] Department of Industrial Engineering, Sharif University of Technology, Tehran, Iran

*** Corresponding author email address**: mhabibi@sharif.edu

A r t i c l e I n f o

Article type:
Original Research

How to cite this article:
Kazemi, A., & Habibi, M. (2026). ParsAVC: Appraising Voice of Customers based on Aspect-Based Analysis of Online Reviews Using Unsupervised Technique. *AI and Tech in Behavioral and Social Sciences, 4*(1), 1-15.
https://doi.org/10.61838/kman.aitech.4729

A B S T R A C T

With the increasing popularity of online shopping, customer reviews have become a valuable source of information for both businesses and consumers. This research proposes an unsupervised method for aspect-based sentiment analysis of customer reviews in online stores. The method addresses the challenge of extracting and analyzing product aspects from reviews. It leverages rule-based grammatical structures to identify candidate aspects and then extracts explicit and implicit aspects related to these candidates. The evaluation shows that the proposed method achieves significant improvement in precision, recall, and F1-score compared to previous methods. This research contributes to the field of sentiment analysis by providing a novel method for extracting and analyzing product aspects from customer reviews and proposes an unsupervised method for aspect-based sentiment analysis of customer reviews in online stores. The method involves three steps: First, data preparation: Collecting customer reviews and categorize them by product. Second, aspect extraction: Identifying candidate aspects using rule-based grammatical structures. Third, sentiment analysis: Extraction of explicit and implicit aspects related to the identified candidate aspects. Evaluation using precision, recall, and F1-score shows improvement compared to previous methods. This research contributes to extracting and analyzing textual data from customer reviews to identify product aspects and user sentiment.

Keywords: *Aspect-Based Sentiment Analysis; Aspect Extraction; Persian Opinion Mining; Explicit Aspects; Implicit Aspects*

1. Introduction

Customer feedback is crucial for improving product quality. Traditionally, businesses relied on limited channels to understand consumer preferences. However, the rise of online stores has transformed this dynamic. Today, businesses can directly interact with end consumers and gather valuable insights through online reviews. This increased access to customer feedback presents a double-edged sword. On the one hand, online reviews offer a wealth of information, allowing businesses to identify product strengths and weaknesses from the customer's perspective. This real-time feedback loop empowers businesses to adapt their offerings and improve product quality based on consumer preferences.

On the other hand, the sheer volume of online reviews can overwhelm consumers seeking product information. This phenomenon, aptly termed "information bombardment," creates a paradox. While a plethora of reviews exists, consumers struggle to sift through them effectively. This information overload can hinder their ability to make informed decisions.

Opinion mining tries to extract users' opinions on a product or service. Frequency based models, syntax-based models, supervised and unsupervised machine learning and hybrid moles are the classification of aspect detection

methods. The problem is that due to the great differences in linguistic structures, an accurate and trustworthy text mining and NLP based approach on a language could not be extended to the others (Nejad et al., 2020). Furthermore, as the volume of data and opinions grows, manufacturers find it challenging to categorize and analyze the feedback on a large scale. To address these issues and meet the expressed needs of future buyers, end consumers, and product manufacturers, text mining techniques based on unsupervised method in machine learning is employed. Crucially, identifying key keywords and critical terms in discerning and distinguishing product aspects in reviews is essential. Therefore, creating a comprehensive database containing essential Persian keywords is considered a necessary step. This research aims to extract and preprocess textual data, identify and extract candid product aspects, and extract the most related explicit and implicit aspects from comments and reviews. In general, and not only in Persian but also for all languages, unsupervised learning methods are employed for unlabeled data; meaning they do not require training and learning. Some of the explicit aspect identification methods can be observed in (García-Moya et al., 2012) and (Htay & Lynn, 2013), implicit aspects in (Amplayo et al., 2018) and (Khalid et al., 2018), and approaches covering both types of aspects in (Poria et al., 2014) and (Zainuddin et al., 2018). The proposed steps include:

- Collection, extraction and preprocessing user opinions about considered product and other same products in a subscripted category.
- Identifying candid aspect of opinions using linguistic rules, enhancing aspect extraction performance by adding new language rules.
- Scoring degree of similarity and evaluating each candid aspect to extract explicit and implicit aspects.

Opinion mining is performed in sentence-level on a well-known Persian dataset. The sentence-level approach is chosen due to the fact that working in document-level increases the dimensions of the matrices and hence, reduces the performance and accuracy of the whole approach.

The aspect extraction process is essentially a subtask of ABSA. The fact that most of the existing studies gave much emphasis to explicit aspect extraction alone is not enough, because SA is incomplete without considering the implicit due to its contribution to the meaning of the content (Maitama et al., 2020). A bio-inspired search mechanism is used with different similarity measures to identify different types of implicit features and measure its effect on retrieval performance (Eldin et al., 2021). All of above discussions show that Aspect based sentiment analysis in Persian as called Pars ABSA in (*Introducing Pars ABSA (Persian Aspect Based Sentiment Analysis) Model*) literally has an increasing interest and trend in multilingual natural language processing (MNLP).

To implement the aforementioned three steps, there is a need for a dataset of Persian text from online stores. Considering that Digikala* is considered one of the most popular online shopping systems and has a relatively extensive collection of opinions compared to other online stores, the decision was made to use the set of user opinions on the Digikala platform.

2. Literature Review

In this section, a brief overview of the theoretical foundations and research background of the three mentioned steps is presented. The approach of this chapter is to introduce the starting point and problem definition to address the stated needs, and subsequently, to focus more on the activities and tasks defined in this domain, with an emphasis on the Persian language.

The inception point of sentiment analysis problem analysis revolves around the paper by (Hu & Liu, 2004). In this paper, the significance of the topic and the need for problem definition are emphasized. The paper investigates the following three topics:

1. Identification of product features that customers have expressed in their opinions within the context of reviews.
2. Identification of sentences for each feature that introduces positive and negative opinions.
3. Presentation of a summary that compresses the obtained information.

In 2007, (Popescu & Etzioni, 2007) articulated the problem of identifying product aspects and features through four sub-activities:

1. Identification of product features.
2. Identification of opinions related to each product feature.
3. Determination of the sentiment polarity.
4. Rating the opinions based on the strength of each polarity.

This paper introduces a new method called OPINE.

* Digikala.com

AITBSS AI and Tech in Behavioral and Social Sciences E-ISSN: 3041-9433

In 2013, (Htay & Lynn, 2013) proposed a pattern for summarizing opinions based on features. In this system, the database of customer opinions and views serves as input (Poria et al., 2014). published research on the rule-based approach to extracting aspects from customer perspectives. In addition to presenting a pattern for identifying explicit aspects, patterns for identifying implicit aspects are also provided. The initial efforts to solve the aspect-oriented sentiment analysis problem in Persian language were published in 2013 by (Bagheri et al., 2013). In this paper, a model based on unsupervised learning is presented for detecting and identifying explicit and implicit aspects in four fundamental stages, as mentioned below:

1. Identification of multi-word aspects (aspects with more than one word).
2. Execution of the repeated bootstrapping algorithm.
3. Pruning and cleaning aspects.
4. Identification of implicit and incidental aspects.

In 2017 (Razavi & Asadpour, 2017) proposed a method for detecting Persian aspects in each category after crawling the required datasets in five main steps:

1) Preprocessing (Training word embedding vectors, segmenting reviews into sentences and their words, and finding POS tags of the words)

2) Extracting adjectives as opinion words to build the sentiment lexicon

3) Extracting frequent nouns and noun phrases as a candidate aspect list

4) Pruning candidate aspects by a number of heuristic rules

5) Finding the most related words to the aspects as keywords by a word embedding-based algorithm with a similarity metric, and categorizing aspects by combining the similar aspects.

In 2019, efforts to implement unsupervised learning algorithms led to the presentation of relatively comprehensive rules for identifying candidate aspects in (Mohammadi et al., 2019). In 2020, (Akhoundzade & Devin, 2020) presented an approach in the field of unsupervised learning methods solely for extracting and categorizing Persian aspects. In this paper, four main methods for aspect extraction are generally mentioned:

1. Extraction based on names or frequent phrases.
2. Extraction based on sentiments and relationships of aspects.
3. Extraction based on supervised learning methods.
4. Extraction based on the use of topic models and neural networks.

After presenting the most significant achievements and previous research, the primary contributions of the current research are presented by combining previous studies and focusing on existing gaps as follows:

- Text data extraction for research evaluation has traditionally been performed offline using pre-existing databases in previous studies, particularly neglecting the online storefront of Digikala. Unlike the prior research that exclusively relied on ready-made databases, we conducted real-time data extraction directly from the Digikala online platform.
- Examination of the impact of segregating opinions and viewpoints of product categories from each other and utilizing corresponding opinions for each aspect listed as input to the algorithm. This approach distinguishes itself from most studies, which typically adopt a combined approach for identifying and extracting explicit and implicit aspects separately, as opposed to our focus on categorically separating them.
- Implementation of a scoring policy to identify implicit aspects in cases where no explicit aspect is present in the list of candidate aspects. Previous research predominantly lacked a combined approach for identifying and extracting explicit and implicit aspects, often addressing them separately. Our approach introduces a novel perspective on aspect scoring, leveraging a pre-trained model that resolves polarity and scoring issues using neural networks or learning models—a departure from the conventional methods that require extensive data and lengthy processing times.
- Introduction of a novel scoring approach for aspects, incorporating a pre-trained model. Previous research has successfully addressed polarity issues and scoring using models requiring learning or neural networks. However, these models face challenges related to the large volume of data and prolonged processing times. Our approach innovatively leverages a pre-existing pre-trained model,

AITBSS
AI and Tech in Behavioral and Social Sciences
E-ISSN: 3041-9433

providing an alternative solution to these challenges in previous studies.

Recently, to improve aspect extraction and aspect-based sentiment analysis, the capacity of a generic pre-trained language model (PLM) is used for semantic understanding. Recently, a new framework called "Word Transferred LM" in (Jin et al., 2024) proposed that transfers target words in sentences into pivot token then predicts the affective tokens in the respective places of the aspect terms, and creates contextually appropriate semantics given an input sentence containing aspect terms.

On the other hand, using a neural network and applying a multilingual language model is another branch in defining key terms, extracting aspects, and evaluating extracted aspects. A few models such as mBert, XLM-RoBERTa, in (Conneau et al., 2019) and mT5 in (Xue et al., 2020) and ALBERT support Persian language. The main goal of multilingual models is to train simultaneously on multiple languages and optimize their performance across different languages (Ghafouri et al., 2023).

Recently, keyword extraction and summarization of textual data have become a new challenge. In (Heydari & Teimourpour, 2024) a high-precision extraction of keywords from Instagram posts in the Persian language has been achieved with unsupervised word co-occurrence methods without resorting to conventional techniques such as clustering or pre-trained models.

So, in (Nazarizadeh et al., 2022), by evaluation of recent published articles, it can be concluded that there are more than 40 different methods in Persian language sentiment analysis with different level of precision measure. One of the most popular methods in English concentrating in customer reviews is Latent Dirichlet Allocation (LDA) that is used for product or service aspect extraction. LDA has a low performance for short text and short reviews. In (Dehghani & Ebrahimi, 2023) a new framework presented combining ParsBERT with LDA to extract human life that impacted by COVID-19. After aspect extraction, the best product recommendation based on aspects and opinion mining can be introduced in different methods. In (P et al., 2021), positive or negative opinions with logistic regression introduced as a recommendation approach. In (Gurunathan, 2022), Joint aspect-opinion extraction and sentiment detection for the English language was introduced. This framework, named JAESOD, uses English grammatical structure as an unsupervised method extract aspects and their polarity pairwise. This contribution could be useful in section **Error! Reference source not found.**3.

The aim of this research is to meet the demand for a comprehensive analysis of user data and textual perspectives on e-commerce platforms, employing an integrated approach that incorporates advanced techniques such as deep learning and sentiment analysis. Additionally, the extraction of product features will facilitate the search and categorization of identified characteristics for each individual product.

3. Proposed Method

Extracting aspects from customer reviews written in Persian presents a more significant challenge compared to English. This is primarily attributed to the inherent differences between the two languages. The complexities lie in the distinct syntactic structures and grammatical rules governing Persian, making it difficult to define robust aspect extraction rules directly transferable from English. We try to define rules to at first find candid aspect and candid adjective rules and first with a pruning action and considering frequency of each candid in aspect and adjective lists. As we know in English noun phrases with more than one word can relate to special aspect. Second, we try to extract final aspects that each sentence of comments. Figure1 shows the framework and important steps to extract aspects in Persian comments. In this method we consider two assumptions which can be seen below.

- It's assumed that our analysis is broken down to the level of sentences and each comment that contains many sentences, has different aspects. These aspects can be explicit or implicit.
- It's assumed that each sentence at least has one aspect. First the existence of explicit aspect evaluated and then if no explicit aspect is found, the algorithm tries to find closest implicit aspect. So, in final output, we expect to see at least one aspect from candidate aspects related to each sentence of comment.

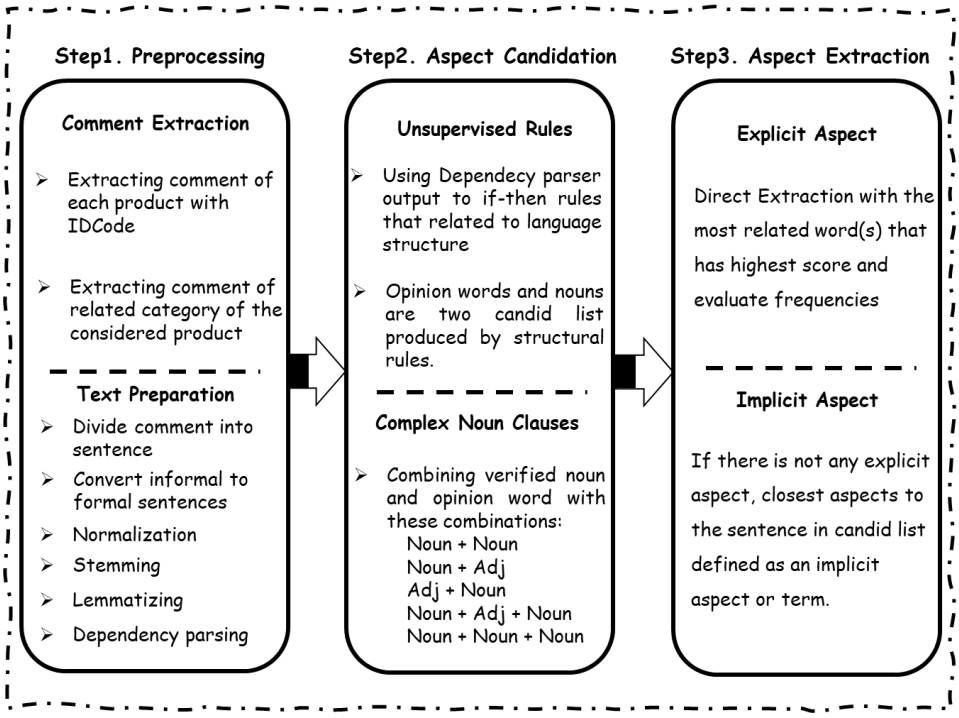
3.1. Preprocessing

Obtaining Persian customer reviews for aspect extraction presents a significant challenge. Suppliers and business owners often exhibit a reluctance to share data with external sources, making customer reviews a non-public data stream. This limited access to open-source data further complicates the process.

AITBSS
E-ISSN: 3041-9433

Figure 1

Framework of aspect extraction method



When working with data, particularly text and comments, a preprocessing step is crucial before analysis. Processing raw text data helps ensure the accuracy of results and avoids drawing incorrect conclusions. Customer reviews are often characterized by informal language, slang, and grammatical errors. These inconsistencies can significantly hinder effective analysis if left unaddressed during preprocessing. Furthermore, the presence of irrelevant content, such as extraneous words or noise, can further complicate the identification of relevant aspects within the reviews. So, to overcome these challenges, we provide six sub-processes in preprocessing stage which can satisfy our requirements for the aspect candidation step.

3.1.1. Opinion division

By the first assumption, each comment contains several sentences. Our analysis tries to find any related aspects in each sentence. So, in this stage, we divide each comment into sentences. In this stage two new challenges arise. Division comments is based on punctuation marks. So, definition of using each mark will be necessary and important. On the other hand, each opinion saving in one record, now replaced with one or more than one sentences in one or more than records. So, the limitation of processed sentence will be a new concern.

3.1.2. Convert informal to formal

After changing the opinion-based dataset to sentence-based dataset, we confront with some informal sentences. There are packages in Persian natural language processing solving these kinds of challenges. But the precision and accuracy of these tools are not exactly 100%. In Persian or English, the written structure can be the same, but different pronunciation often creates a different meaning. For instance, the world "بده" can have to meaning. The first meaning of "بِده" ("Bede") is "Give it" in English. But the second meaning of "بَده" ("Bade") refers to an opinion word ("Is bad") in English. So, in this case tools have deviation from purpose of the user. These challenges have been solved approximately by Hazm and Farsiyar (*Farsiyar python demo codes*) Persian NLP developer. Other examples of informal structures relate to simplifying and summarizing nouns, pronouns, verbs and prepositions. All of these not exactly but nearly are solved by using mentioned tools.

AITBSS
AI and Tech in Behavioral and Social Sciences
E-ISSN: 3041-9433

3.1.3. *Normalization*

Normalization provides as a cleaner for texts and sentences. This involves correcting spaces, semi-spaces, removing unnecessary characters, letters, and converting English numbers to Persian. While not exhaustive like a spell checker, the normalizer is expected to rectify significant spelling mistakes.

3.1.4. *Stemming*

The purpose of stemming some words is to correctly identify the structural role of that word. One of the rules used to identify opinion words (most of which are adjectives) is identifying nouns in the text and find the stem of all nouns. Also, stemming doesn't include only operation like the above. For example, plural suffixes such as "ها"(Ha), "ان"(On), "ین"(Ein), "ون"(Oun), and "ات"(At) are also removed in rooting. On the other hand, the pronouns attached to the words are also removed in rooting. So, in addition to the functions used in digestion that find stemmed words, another stem finder is used, which is more complex and accurate than Hazm (*Introducing hazm*), called "Persian Stemmer".

3.1.5. *Lemmatizing*

Lemmatizing is stemming with awareness and is used to find root of words (*Lemmatizer Function in Hazm*). The difference between Lemmatizing and Stemming is that in Stemming there is no understanding of the meaning of the word and only based on the removal of some simple suffixes, the root of the word is tried to be determined; Therefore, it may provide incorrect results in rooting some words; But Lemmatizing is based on a list of reference words along with its roots does this and provides more accurate results. Of course, the cost of this accuracy is a slower rooting speed. Therefore, if the processed text is not long, it is better to use word formation, and if the volume of the text is large and enormous, it is better to use stemming to reduce the processing time. Among the other possibilities of lemmatizing compared to stemming is focusing on verbs and it is possible to use verbs in passive and normal and in different tenses.

3.1.6. *Dependency parsing*

The last step in preprocessing and preparing textual data is to use the widely used dependency parser function. Dependency parser, which we introduced earlier is the most important and main function that is used to identify and extract aspects and candidates. This module includes classes and functions to identify text syntactic dependencies. As mentioned clearly in (*Dependency parser resources*) purpose of dependency parser in a sentence is to identify the grammatical relationship between the words of a sentence. For example, if the sentence "کیفیت کتاب بد بود." ("The quality of the book was good.") The input of the dependency parser is as follows:

- (quality, subject, adjective)
- (book, noun modifier, noun)
- (proper, predicate, adjective)
- (was, verb, root)

Of course, displaying the dependency graph is another feature of the dependency analyzer, which graphically shows the grammatical relationship between words in each sentence. Therefore, we can see that the reason for dividing comments and changing into each sentence is to make more accurate use of the output of the dependency parser.

Of course, since all the outputs of the dependency parser are not used in the phase of aspect extraction, it is necessary that after receiving the output of the dependency parser, which is unstructured data in the form of a text file(*.txt), only the information needed in the structured framework is temporarily stored. One of the special outputs of this stage is POS tagging. POS tagging can define the structure of word in sentences.

3.2. *Dependency Parser Accuracy and Robustness*

The proposed extraction approach critically relies on accurate dependency parsing to identify the syntactic relations necessary for explicit aspect extraction. We employ the widely used Hazm Persian NLP toolkit(*Introducing hazm*), which reports competitive performance on standard evaluation benchmarks:

- **Unlabelled Attachment Score (UAS):** Approximately 92.3%
- **Labelled Attachment Score (LAS):** Approximately 89.15%

Furthermore, Hazm's Part-of-Speech (POS) tagger demonstrates high accuracy 98.8%

Despite these strong baseline results, the dependency parser is not error-free. Its performance tends to decline when processing informal or noisy user-generated text due to nonstandard grammar and punctuation inconsistencies. Critically, errors originating in POS tagging and dependency parsing can propagate directly to the

AITBSS AI and Tech in Behavioral and Social Sciences
E-ISSN: 3041-9433

downstream rule-based extraction stage, negatively impacting the quality of candidate generation and the successful application of syntactic patterns.

3.3. Aspect Candidation

This section focuses on the second stage of the aspect extraction process, which aims to refine the pool of potential aspect candidates. Prior research suggests that nouns are the primary source for identifying aspects within text data. Building upon this foundation, this stage leverages dependency parsing techniques to extract grammatical relationships between words in Persian sentences. By analyzing these relationships through a set of pre-defined if-then rules, the approach aims to significantly reduce the number of possible words that could be classified as aspects. This refinement process ensures a more focused and accurate extraction of relevant aspects from Persian customer reviews. So, the accuracy and precision rate of dependency parser can affect in aspects that were extracted as a candid. To define all single word and multi word aspects, based on linguistic terms, we have to create opinion word candid list that can help us in multi word aspects. So, at first single candid aspects are created and after that with opinion word candid list, multi word and combined aspects were added to final list.

To define candid aspects, three main if-then rules were implemented as mentioned below.

1) **Rule Object and Noun**: Each word in sentence having object label and being shown as a noun, is extracted and added to the candid list. As an example, in "جنس خوبی داشت." (had a nice material) the word "جنس" (material) is object and noun. So, it's added to the candid list.
2) **Rule Subject and Noun:** Some sentences haven't any object in their structures. So, in second rule in sentences without any objects, subjects are added to the candid list.
3) **Rule aspect after "and":** If a word is in prepared aspect candid list and after that the word "و" ("and") appears and the word after that has a noun label, the noun labeled word is added to the aspect candid list.

Following the identification of single-word aspect candidates, the proposed approach addresses the extraction of multi-word aspect candidates. This stage focuses on identifying phrases that encompass more than one word and potentially represent aspects within the review text. To achieve this, a predefined list of opinion words, primarily consisting of adjectives, is employed. This list was developed according to the rules below.

1) **Rule adjective and Adjective phrase:** In this rule, if the word structure is an adjective and antecedent adjective and the sentence has a syntactic structure of descriptive sentences, the adjective is after the noun, object or adverb, it is added to the list of adjective candidates in the first step. At this stage, in order to preserve the point of view corresponding to each candidate attribute, the operation of storing the attribute and the point of view is done as a pair.
2) **Rule 3-words:** According to this rule, if there is a word that the word before it is a noun, a noun phrase, or an adjective, and the word after it is a verb, and it is identified as the root of the sentence, and that mentioned word itself after stemming with the help of Hazm, have one of the structures of adjectives or adverbs; In this case, that mentioned word is also added to the list of candidate opinion word.
3) **Rule 4-words:** According to this rule, if the second, third, and fourth words in a combination, if a word has an adverb structure and none of the second, third, and fourth words are adjectives, the first word is stemmed. If the stemmed word has one of the adjective and adjective structures, this stemmed word is added to the list of candidate adjectives. In this rule, like the other rules that have been mentioned so far, corresponding to each candidate word, its viewpoint is also stored as a pair of candidate and opinion.
4) **Rule adjective after "and":** Same as before, if a word is in the prepared opinion word candid list and after that the word "و" ("and") appears and the word after that has a noun label, the adjective labeled word is added to the opinion word candid list.

So, to finish the process of single aspect and opinion word candidates, two task of preprocessing, it means removing stop words and named entity recognition (NER) to reduce process time. To create multi-word aspect candidates, 5 combinations mentioned below are created. These nouns and adjectives are selected from aspect and opinion word candidates.

AITBSS
E-ISSN: 3041-9433

1) **Noun + Noun:** The purpose is to find aspects like "صفحه نمایش" ("Safhe Namayesh"), which means "screen" in English.
2) **Noun + Adjective:** The purpose is to find aspects like "کم نور" ("Kam Nour") means "low light" in English
3) **Adjective + Noun**
4) **Noun + Adjective + Noun**
5) **Noun + Noun + Noun**

The core criterion for evaluating the aspect candidate list is the frequency of occurrence for each candidate aspect or opinion word within the comments and sentences. As previously mentioned, the system stores the corresponding sentence for each candidate aspect identified during the extraction process. This allows us to analyze the frequency of product aspects mentioned across the entire dataset of reviews for the product category. Subsequently, the lists of single-word and multi-word aspects (encompassing two-word and three-word phrases) are merged to form a comprehensive candidate list for further analysis.

Having established a list of candidate aspects, along with their corresponding reviews and frequencies within the processed data, this stage focuses on filtering out aspects with low occurrence rates. A policy is implemented to achieve this filtering, relying on statistical measures such as mean and standard deviation. By analyzing these statistical values, the approach aims to identify and remove candidate aspects whose frequency falls below a predefined threshold established based on the calculated mean and standard deviation. This filtering process helps refine the candidate list and ensure a focus on the most prominent aspects discussed within the customer reviews.

To refine the candidate list further, a statistical filtering approach is employed. This approach utilizes the mean and standard deviation of the frequency distribution for all identified aspects. Specifically, an aspect is retained if its frequency count exceeds the sum of half the standard deviation and the average frequency of all aspects. This threshold ensures that only aspects with a frequency significantly higher than the average (considering the natural variation in frequency) are retained. This statistically-driven filtering process contributes to a more focused and informative set of extracted aspects, prioritizing those that are most frequently discussed within the customer reviews.

3.4. Design choices and empirical calibration

The principal design choices governing our extraction pipeline—specifically the frequency threshold for candidate filtering, the use of Word2Vec embeddings with cosine similarity for implicit aspect identification, and the defined set of syntactic rules for explicit candidate selection—were rigorously guided by both linguistic considerations and empirical calibration over held-out labeled data.

3.4.1. Candidate Filtering Threshold

The threshold for filtering candidate aspect frequencies was set at mean + 0.5*standard deviation. This formulation was chosen to strike a critical balance between removing low-frequency noise and retaining meaningful but less frequent aspects. The coefficient of 0.5 * standard deviation was specifically determined after a sensitivity analysis conducted on an internal validation subset (about 3000 manually annotated sentences).

3.4.2. Similarity Metric and Embeddings

For implicit aspect matching, we employed Word2Vec vectors trained on a large Persian corpus (including Digikala reviews) and cosine similarity as the scoring metric. This choice prioritizes computational efficiency and leverages the robust performance of Word2Vec in lexical similarity tasks, particularly given the constraints of available Persian datasets. We acknowledge, however, that contextual embeddings (e.g., ParsBERT) represent a potential path for future work to further improve the recognition of implicit aspects

3.5. Aspect Extraction

This stage focuses on identifying the most relevant aspects from the previously generated candidate list. To achieve this, a rule-based approach is implemented, utilizing a predefined metric. This metric evaluates the relationship between candidate aspects and the corresponding list of opinions associated with each aspect. By applying a specific condition on the number of desired output aspects (i.e., selecting the most relevant ones), this stage concludes the aspect identification and extraction process.

As mentioned above, prior to aspect identification, a preprocessing step is essential. This step involves separating the relevant opinions and comments for the

AITBSS
E-ISSN: 3041-9433

target product or service from the broader dataset of product category opinions. This distinction ensures that the analysis focuses exclusively on aspects related to the specific product under investigation.

3.5.1. Extraction of explicit aspects

If a candidate aspect is explicitly and directly present in the opinion text, the correlation measurement for that candidate aspect is assigned the numerical value of one. The aspect and the correlation measurement are then added to the checklist as a pair.

3.5.2. Extraction of implicit aspects

In cases where the candidate aspect is not directly and explicitly present in the opinion, a methodology for detecting implicit and indirect aspects is necessary to determine the most relevant aspect related to the opinion. Initially, the opinions corresponding to the aspect are segmented into words, and then the degree of association of each word with the aspect is examined using Word2Vec. The output of this model will be a number ranging from zero to one. Subsequently, the highest obtained number from the Word2Vec model is identified, and the aspect's similarity measurement is calculated through the relationship function.

$$AspectMetric = 2 * Max(SimilarityScore) - 1$$

If the obtained score in the aspect similarity measurement is greater than zero, the corresponding aspect is considered as the extracted aspect from the opinion; otherwise, the candidate aspect is not identified as an extracted aspect. It is important to note that, in order to prevent the extraction of repetitive aspects, the extracted aspect is only considered final if it did not exist previously in the list of explicitly extracted aspects.

4. Results

The proposed method was applied in the biggest Persian online store in (*Resource of Comments in DigiKala*). Digikala is an online store that covers many products in different categories and sub-categories. As mentioned before, the proposed algorithm should cover not only products, but also be used for any businesses that has reasonable number of comments and Persian sentences.

Extracting opinions of each product contains these four steps:

- The product code, provided on the product page link, is given as an input to the algorithm.
- The number of items in a similar category as the desired product, for which their processing is considered as reviews of the product category, is provided as input during the extraction of the candidate aspect and adjective lists.
- The number of comments and opinions from this set of items that fall within the product category is given as input to the algorithm.
- Finally, the number of comments and opinions for which the model intends to identify the aspects is entered.

To evaluate the method, confusion matrix elements lead to F1-score, precision and recall indices. The most crucial factors considered in selecting test products are mentioned below.

- The product under examination should have a sufficient number of opinions for processing. Products with fewer than 50 opinions are not considered in the evaluation of aspect extraction.
- The total opinions of the product and products in the relevant product category for candidate aspect extraction should not be less than 400 opinions. This is based on a literature review conducted in previous studies, where assessments and results were typically presented based on the analysis of 200 or more reviews.
- In selecting products, preference is given to those for which it is predicted that their aspects are different from those of others. This implies that in product selection, efforts are made to avoid choosing products with similar aspects to others.

As one of the objectives of this research is to consider the opinions of products related to the desired product category in forming the candidate list of aspects and taking into account the opinions corresponding to the candidate aspect in aspect extraction, each of the four mentioned products is executed twice. The evaluated products with the mentioned conditions are "ترازو دیجیتال " ("Tarazou Digital") as a Digital scale, "یخچال ۵ مینی بار " ("Yakhchal 5 mini bar") as a 5-foot refigerator, "تی شرت مردانه" ("T-Shirt

AITBSS
E-ISSN: 3041-9433

Mardaneh") as a man T-Shirt, "گوشی موبایل " ("Mobile") as an Mobile.

In the first iteration, the number of opinions for the main product is set at 380, and the number of comments for products related to the main product category is set at 20. In the second iteration, the number of opinions for the target product is set at 200, and the number of comments for products related to the target product category is also set at 200. the algorithm's output for one of the above products is mentioned in table 1.

Table 1

Explicit and implicit extracted aspects of digital scale comments

Row	Extracted aspects in the first iteration	Extracted aspects in second iteration
1	"کیفیت" ("Keyfiat") as a Quality	"کیفیت" ("Keyfiat") as a Quality
2	"دقیق " ("Daghigh") as an Accuracy	"دقیق " ("Daghigh") as an Accuracy
3	"قیمت" ("Gheymat") as a Price	"قیمت" ("Gheymat") as a Price
4	"ترازو" ("Tarazou ") as a Scale	"ترازو" ("Tarazou ") as a Scale
5	"خطا" ("Khata") as an Error	"ارزش " ("Arzesh") as a Value
6	"خرید" ("Kharid") as a Purchase	"باتری " ("Batri") as a Battery
7	"عمر باتری" ("Omr Batri") as a Battery life	"دقت " ("Deghat") as a Precision
8	"دقت " ("Deghat") as a Precision	"وزن" ("Vazn") as a Weight
9	"وزن" ("Vazn") as a Weight	"بسته‌بندی" ("Bastebandi") as a Packaging
10	"بسته‌بندی" ("Bastebandi") as a Packaging	"مشکل " ("Moshkel") as a Problem
11	"دما " ("Dama") as a Temprature	"ارزش خرید " ("Arzesh Kharid") as a Purchase value
12	"ارزش خرید " ("Arzesh Kharid") as a Purchase value	"راضی " ("Razi") as a Satisfied
13	"راضی " ("Razi") as a Satisfied	
14	"سطح " ("Sath") as a Surface	
15	"ظاهر" ("Zaher") as a Appearance	
16	"مشکل " ("Moshkel") as a Problem	

Finally, the research method is compared with other previous studies in the range of unsupervised and supervised learning methods in Table 2.

4.1. Statistical Robustness and Reliability of the Results

To rigorously assess the robustness and stability of the proposed unsupervised approach, we strengthened the experimental protocol through statistical analysis and repeated trials.

4.2. Experimental Design and Consistency

Each experiment was systematically repeated five times **(**5 runs**)**. In every run, a different randomly selected subset of reviews from the same product category was used as input. This design ensures that the reported performance reflects the stable behavior of the method rather than being dependent on a single, fixed sample.

4.3. Reporting and Confidence Intervals

For all reported performance metrics (Precision, Recall, F1-score), we present the mean performance (μ), the standard deviation (σ), and the 95% confidence interval (α).

Across all product categories, the standard deviation of the F1-score remained consistently low (approximately ± 1.5 percentage points). This low variability strongly indicates stable and consistent behavior of the model across different sampled input data subsets.

4.4. Statistical Significance (ANOVA)

To confirm the statistical reliability of this stability, a one-way Analysis of Variance (ANOVA) was conducted to examine whether the mean performance differed significantly across the five runs.

- The analysis demonstrated that the computed F-statistic (F^*)was lower than the critical F-value at the 0.05 significance level.
- Consequently, the null hypothesis of equal mean performance across runs cannot be rejected.

This statistical evidence confirms that the observed variations in performance are not statistically significant

AITBSS
E-ISSN: 3041-9433

and establishes that the reported results are highly reliable and consistent across repeated experiments.

Table 2

comparing results with previous methods

Method	Precision	Recall	F1-Score
This research	86.2%	96.2%	90.8%
Method in (Mohammadi et al., 2019)	87%	92%	89%
Method in (Bagheri et al., 2013)	74%	82%	83%
Instruct ABSA (Scaria et al., 2023)	-	-	92%

According to the definition of each of the three indicators used, the following results were obtained below.

- The precision index determines the proportion of positively predicted aspects that are successfully forecasted to all predicted aspects. Essentially, this index determines the ratio of aspects and features that are successfully anticipated to all aspects that are forecasted, including those that are inaccurate. 86.2% was the precision index in the second iteration, which took into account a wider range of viewpoints pertaining to the main product. This result was impacted by aspects that were not accurately predicted, as evidenced by the lower value as compared to the recall index. Stated differently, elements in the sentiment text that weren't really taken into consideration were predicted.
- The recall index calculates the ratio of correctly predicted positive aspects to the total actual aspects. This index reached 96.2%. This index means that the number of aspects correctly predicted is a good percentage of the total actual aspects of the product (both correctly identified by the algorithm and those the algorithm failed to identify). In other words, the algorithm has effectively reduced the number of actual aspects that were not identified and has good coverage of the set of actual aspects.
- The F1-score index, being a median of the two indices of precision and recall and a combination of both mentioned indices, reached 90.8%. This value can be considered a satisfactory output compared to other methods performed so far in the Persian language domain and under unsupervised learning methods.
- The reported values assume that the extracted aspects are considered as the final categorized aspects.

4.5. Dataset considerations and domain coverage.

Although the experiments were conducted on four products, the selected categories span semantically diverse domains—including digital devices, household appliances, clothing, and measurement tools—which naturally exhibit different linguistic behaviors. Furthermore, the Digikala platform contains many product categories, but not all categories include a sufficient number of user reviews to support reliable unsupervised linguistic analysis. To ensure statistical stability, we included only categories with more than 500 user comments, thereby avoiding sparsity-related bias and ensuring meaningful frequency distributions for candidate aspects. In addition, our internal validation set of approximately 3,000 manually annotated sentences covers multiple writing styles and informal expressions, enabling the model to capture general Persian syntactic patterns rather than domain-specific cues.

4.6. Component Dependency Analysis

The proposed framework is architecturally dependent on its rule-based extraction module, which serves as the entry point of the pipeline. Without these rules, no aspects—explicit or implicit—can be identified, making quantitative comparison infeasible. The Word2Vec-based component is solely responsible for implicit aspect detection; removing it restricts the system to explicit aspects only, resulting in a noticeable decrease in recall and F1-score, despite a possible increase in precision due to the narrower extraction scope.

Preprocessing plays a crucial enabling role by improving token normalization and sentence segmentation, which directly affects POS tagging quality. When preprocessing is omitted, POS tagging errors increase, severely degrading

AITBSS
E-ISSN: 3041-9433

the effectiveness of the extraction rules. Our analysis indicates that even with the implicit aspect module enabled, the absence of preprocessing leads to an approximate 20% drop in both precision and F1-score. These findings demonstrate that all components of the pipeline are tightly interdependent and that meaningful performance can only be achieved when they are jointly employed.

1.1. Error Analysis

To better comprehend the behavior of the proposed unsupervised method and to pinpoint the primary sources of performance degradation, we conducted a structured error analysis on a carefully manually inspected subset of the extracted aspects.

The most frequent misdetections were categorized into three main areas:

1. **Semantic Ambiguity:** Failures occurred with **polysemous terms** such as "value" (ارزش), and with **highly generic aspects** (e.g., "problem" (مشکل), "purchase" (خرید)).
2. **Embedding Limitations:** Errors arose when **product-specific terms** were mapped by Word2Vec to overly general semantic neighbors. This suggests that the static embeddings struggled to differentiate fine-grained domain-specific meanings.
3. **Linguistic and Syntactic Noise:** Failures primarily linked to the noisy nature of user-generated content, including **informal writing styles**, **missing punctuation**, and **syntactic irregularities** (such as ellipsis or missing verbs), which substantially reduce the reliability of dependency parsing in Persian.

5. Conclusion

The exponential growth of online shopping has witnessed a surge in entrepreneurs establishing online stores. Customer reviews and feedback have become a cornerstone for businesses, providing valuable insights for product design and improvement. However, the vast amount of online reviews can be overwhelming for consumers, presenting them with a multitude of diverse opinions and preferences. This information overload can hinder their ability to make informed purchasing decisions.

5.1. Ethical and Practical Considerations

This study strictly adheres to ethical guidelines concerning data usage by relying exclusively on publicly available user reviews. The data was collected from Digikala product pages via publicly accessible APIs, which were identified using standard browser network inspection. Crucially, no private endpoints, user-restricted services, personal identifiers, account metadata, or sensitive user information were accessed or stored. For analysis, only the review text and the associated rating were utilized.

5.2. Practical Limitations and Biases

We acknowledge two primary practical constraints inherent to this type of analysis:

User Bias in Review Data: Online reviews are susceptible to inherent user bias. For instance, dissatisfied users often contribute reviews more frequently, and experienced users tend to provide more detailed feedback. These behavioral biases directly impact the statistical distribution of the extracted aspects and their associated sentiments.

Model Limitations: Sentiment analysis models, especially unsupervised ones, possess limitations in accurately handling complex linguistic phenomena such as sarcasm, cultural nuances, and highly context-dependent expressions.

The following points represent the key limitations of the current study:

- **Relative Dependence on Dependency Parser Accuracy:** The output of the dependency parser can introduce aspect extraction errors, particularly in informal or colloquial text environments.
- **Word2Vec Weakness in Context-Sensitive Implicit Aspects:** Word2Vec is limited in identifying implicit aspects that are sensitive to context (as detailed in the sixth comment/response).
- **Absence of Supervised Fine-Tuning:** Since the approach is unsupervised, direct optimization based on human-labeled data is not applicable.
- **Inability to Detect Sarcasm, Irony, and Complex Semantic Ambiguity:** The model lacks mechanisms for accurately recognizing complex linguistic nuances such as sarcasm, irony, and advanced semantic disambiguation.

The selection of Word2Vec as the primary embedding technique was guided by practical and computational considerations specific to Persian Natural Language Processing (NLP). Relative to English, fewer large-scale and computationally efficient contextual language models

AITBSS
E-ISSN: 3041-9433

are readily available for Persian, and those that exist often demand substantial hardware resources. We acknowledge, however, that static embeddings inherently cannot fully capture all context-dependent meanings, a limitation that is particularly relevant for implicit aspect extraction. While incorporating advanced contextual language models (e.g., BERT or ParsBERT) would likely yield more accurate results, this would incur the cost of significantly increased computational complexity and fundamentally alter the lightweight, rule-based nature of the proposed approach.

Since the central goal of this work is to present a fully unsupervised, lightweight, and scalable framework suitable for industrial deployment, Word2Vec provides an optimal trade-off between semantic coverage and computational efficiency.

5.3. *Conclusion on Applicability*

Given these inherent biases and model limitations, the outputs of the proposed method should not be used as the sole foundation for high-stakes, real-world decisions. Instead, the results are best employed as a complementary analytical tool, intended to inform decision-making when combined with human expert judgment and corroborated by additional, independent data sources.

This study addresses several key challenges associated with aspect-based sentiment analysis (ABSA) in the context of Persian online reviews. Firstly, processing large datasets within time constraints can be a hurdle. Secondly, the scarcity of labelled datasets in the Persian language poses a significant obstacle. To overcome these challenges, the study leverages innovative unsupervised learning methods and harnesses the power of pre-trained models. While manual labelling of approximately 3,000 sentences was necessary during the analysis and evaluation phase, this step proved critical due to the absence of established algorithms for extracting sentiment from both explicit and implicit aspects within the Persian language.

The results presented in section 4 showcase the efficacy of the research algorithm, with an impressive accuracy of 86.2%, a recall rate of 96.2%, and an F1 score of 90.8% in the aspect extraction phase. A notable achievement of this study was the comprehensive consideration of product-related perspectives within the specific product category during opinion processing. Establishing correlations between opinions and corresponding aspects proved to be a valuable undertaking. Moving forward, it is imperative to enhance tools for processing the Persian language to achieve greater precision in output. Leveraging the achievements of this research, the following undertakings can be pursued as subsets within sentiment analysis, actively contributing to the ongoing evolution of this field. An innovative approach to aspect extraction involves the application of a linear function with inputs featuring varied weights. Subsequent investigations should delve into whether heightened transparency in comparisons serves as a boon for business owners or primarily enriches the value for end consumers.

5.4. *Future Work*

As future research, the following points are considered. The resources of developed framework and other results are available in (Ah.Kazemi).

Building upon the foundation established in this research, our future work will focus on the following key areas to enhance the performance, robustness, and generalizability of the aspect extraction pipeline:

1. **Exploring Advanced Contextual Embeddings:** We intend to investigate the integration of lightweight contextual embeddings such as ParsBERT or Persian Sentence-BERT. These models are expected to offer improved performance in capturing implicit aspects that are highly sensitive to linguistic context.
2. **Developing Hybrid Models:** Future research will explore hybrid architectures that strategically combine our robust, rule-based syntactic foundation with the sophisticated representational power of transformer-based embeddings.
3. **Refining Aspect Classification and Categorization:** We plan to develop unsupervised methods specifically for the final classification and conceptual **categorization** of extracted aspects that share similar meanings. This step is crucial for aggregating related concepts and reducing redundancy and final output errors.

Establishing a Standardized Benchmark and Resource Creation: A crucial long-term goal is the development of a standardized Aspect-Based Sentiment Analysis (ABSA) benchmark specifically for the Persian language. Furthermore, the output of this research can be curated and refined to serve as a bootstrapped labeled dataset for subsequent supervised learning research.

Authors' Contributions

AITBSS
AI and Tech in Behavioral and Social Sciences
E-ISSN: 3041-9433

All authors contributed substantially to the study and to manuscript development, and all approved the final version.

Declaration

Artificial intelligence was used only for language polishing and editorial support during manuscript preparation. No AI tool was used for data collection, statistical analysis, or generation of the study results. The authors reviewed, verified, and take full responsibility for the final content of the manuscript.

Transparency Statement

Data are available for research purposes upon reasonable request to the corresponding author.

Acknowledgments

The author sincerely appreciates the cooperation of the participants.

Declaration of Interest

The authors report no conflict of interest.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Ethics Considerations

The study protocol adhered to the principles outlined in the Helsinki Declaration, which provides guidelines for ethical research involving human participants.

References

Ah.Kazemi. *Introducing ParsUVOC Framework*. https://github.com/amirkazi99/ParsUVOCP

Akhoundzade, R., & Devin, K. H. (2020). Unsupervised aspect-based Sentiment Analysis in the Persian language: Extracting and clustering aspects. 2020 10th International Conference on Computer and Knowledge Engineering (ICCKE),

Amplayo, R. K., Lee, S., & Song, M. (2018). Incorporating product description to sentiment topic models for improved aspect-based sentiment analysis. *Information Sciences*, *454*, 200-215. https://doi.org/10.1016/j.ins.2018.04.079

Bagheri, A., Saraee, M., & De Jong, F. (2013). Care more about customers: Unsupervised domain-independent aspect detection for sentiment analysis of customer reviews. *Knowledge-Based Systems*, *52*, 201-213. https://doi.org/10.1016/j.knosys.2013.08.011

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. *arXiv preprint arXiv:1911.02116*. https://doi.org/10.48550/arXiv.1911.02116

Focus to learn more

Dehghani, M., & Ebrahimi, F. (2023). ParsBERT topic modeling of Persian scientific articles about COVID-19. *Informatics in Medicine Unlocked*, *36*, 101144. https://doi.org/10.1016/j.imu.2022.101144

Dependency parser resources. https://www.roshan-ai.ir/hazm/docs/content/hazm/dependency_parser.html

Eldin, S. S., Mohammed, A., Eldin, A. S., & Hefny, H. (2021). An enhanced opinion retrieval approach via implicit feature identification. *Journal of Intelligent Information Systems*, *57*, 101-126. https://doi.org/10.1007/s10844-020-00622-9

Farsiyar python demo codes. https://github.com/Text-Mining/python-demo

García-Moya, L., Anaya-Sánchez, H., & Berlanga-Llavori, R. (2012). Combining probabilistic language models for aspect-based sentiment retrieval. Advances in Information Retrieval: 34th European Conference on IR Research, ECIR 2012, Barcelona, Spain, April 1-5, 2012. Proceedings 34,

Ghafouri, A., Abbasi, M. A., & Naderi, H. (2023). AriaBERT: A Pre-trained Persian BERT Model for Natural Language Understanding. https://doi.org/10.21203/rs.3.rs-3558473/v1

Gurunathan, P. (2022). Joint aspect-opinion extraction and sentiment orientation detection in university reviews. *International Journal of Information Technology*, *14*(6), 3213-3225. https://doi.org/10.1007/s41870-022-01041-5

Heydari, M., & Teimourpour, B. (2024). Analysis of Persian News Agencies on Instagram, A Words Co-occurrence Graph-based Approach. *arXiv preprint arXiv:2402.12272*. https://doi.org/10.48550/arXiv.2402.12272

Htay, S. S., & Lynn, K. T. (2013). Extracting product features and opinion words using pattern knowledge in customer reviews. *The Scientific World Journal*, *2013*. https://doi.org/10.1155/2013/394758

Hu, M., & Liu, B. (2004). Mining and summarizing customer reviews. Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining,

Introducing hazm. https://www.roshan-ai.ir/hazm/docs/

Introducing Pars ABSA (Persian Aspect Based Sentiment Analysis) Model. https://paperswithcode.com/dataset/pars-absa

Jin, W., Zhao, B., Zhang, Y., Huang, J., & Yu, H. (2024). WordTransABSA: Enhancing Aspect-based Sentiment Analysis with masked language modeling for affective token prediction. *Expert Systems with Applications*, *238*, 122289. https://doi.org/10.1016/j.eswa.2023.122289

Khalid, S., Aslam, M. H., & Khan, M. T. (2018, 25-26 April 2018). Opinion Reason Mining: Implicit Aspects beyond Implying aspects. 2018 21st Saudi Computer Society National Computer Conference (NCC),

Lemmatizer Function in Hazm. https://www.roshan-ai.ir/hazm/docs/content/hazm/lemmatizer.html#hazm.lemmatizer.Conjugation.negative_imperfective_past

Maitama, J. Z., Idris, N., Abdi, A., Shuib, L., & Fauzi, R. (2020). A Systematic Review on Implicit and Explicit Aspect Extraction in Sentiment Analysis. *IEEE Access*, *8*, 194166-194191. https://doi.org/10.1109/ACCESS.2020.3031217

AITBSS
E-ISSN: 3041-9433

Mohammadi, A., Pajoohan, M.-R., Montazeri, M., & Nematbakhsh, M. (2019). Identifying explicit features of Persian comments. *Journal of Computing and Security*, *6*(1), 1-11. https://doi.org/10.22108/JCS.2019.112489.1009

Nazarizadeh, A., Banirostam, T., & Sayyadpour, M. (2022). Sentiment Analysis of Persian Language: Review of Algorithms, Approaches and Datasets. *arXiv preprint arXiv:2212.06041*. https://doi.org/10.48550/arXiv.2212.06041

Nejad, S. J., Ahmadi-Abkenari, F., & Bayat, P. (2020). A combination of frequent pattern mining and graph traversal approaches for aspect elicitation in customer reviews. *IEEE Access*, *8*, 151908-151925. https://doi.org/10.1109/ACCESS.2020.3017486

P, M., M, N., P, S., & Surekha, T. L. (2021, 2-4 Sept. 2021). Recommending the Best Product Based on User Requirements Using Opinion Mining. 2021 Third International Conference on Inventive Research in Computing Applications (ICIRCA),

Popescu, A.-M., & Etzioni, O. (2007). Extracting product features and opinions from reviews. *Natural language processing and text mining*, 9-28. https://doi.org/10.1007/978-1-84628-754-1_2

Poria, S., Cambria, E., Ku, L.-W., Gui, C., & Gelbukh, A. (2014). A rule-based approach to aspect extraction from product reviews. Proceedings of the second workshop on natural language processing for social media (SocialNLP),

Razavi, S. A., & Asadpour, M. (2017). Word embedding-based approach to aspect detection for aspect-based summarization of persian customer reviews. Proceedings of the 1st International Conference on Internet of Things and Machine Learning,

Resource of Comments in DigiKala. https://api.digikala.com/

Scaria, K., Gupta, H., Sawant, S. A., Mishra, S., & Baral, C. (2023). Instructabsa: Instruction learning for aspect based sentiment analysis. *arXiv preprint arXiv:2302.08624*. https://doi.org/10.48550/arXiv.2302.08624

Xue, L., Constant, N., Roberts, A., Kale, M., Al-Rfou, R., Siddhant, A., Barua, A., & Raffel, C. (2020). mT5: A massively multilingual pre-trained text-to-text transformer. *arXiv preprint arXiv:2010.11934*. https://doi.org/10.48550/arXiv.2010.11934

Focus to learn more

Zainuddin, N., Selamat, A., & Ibrahim, R. (2018). Hybrid sentiment classification on twitter aspect-based sentiment analysis. *Applied Intelligence*, *48*, 1218-1232. https://doi.org/10.1007/s10489-017-1098-6

AITBSS
E-ISSN: 3041-9433