PERSIAN SENTIMENT ANALYSIS: FEATURE ENGINEERING, DATASETS, AND CHALLENGES

Razieh Asgarnezhad [1,*], S. Amirhassan Monadjemi [2]

[1]*Department of Computer Engineering, Faculty of Electrical and Computer Engineering, Technical and Vocational University, Kashan, Iran*
[2] *School of continuing and lifelong education, National University of Singapore, 119077, Singapore*

ABSTRACT

With the pervasive growth of web-based businesses, sentiment analysis of online reviews has attracted increasing interest among text mining experts. The problem is complicated when these reviews are in the Persian language since all existing works are focused on the English language, leaving other languages to multilingual models with limited resources. Due to these drawbacks, we try to give an insight regarding different stages of Persian Sentiment Analysis. This study presents a taxonomy of all Persian Sentiment Analysis works considering the most common techniques. The four steps are considered, namely, pre-processing, feature engineering, lexicon generation, and classification. As a result, we reveal that newer works focus on deep learning methods. Also, we suggest applying other methods such as heuristic and hybrid approaches to be worthwhile for the performance of classification in Persian Sentiment Analysis. Finally, we summarize the most important issues in this domain including the lack of dataset, lexicon, tools, etc.

KEYWORDS: Data Mining, Text Mining, Sentiment Analysis, Feature Selection, Persian Language

1. INTRODUCTION

Recently, a high volume of text data has been produced over the Internet. This abundance of data is a worthwhile source of information to apply in different fields such as recommender systems and sentiment analysis. With the majority of information on websites, decision making can be assisted based on user's reviews and comments. People purchase products on e-commerce websites and give their opinions about them every second. These opinions can considerably affect business decisions in companies (Khan et al., 2014) (Montejo-Ráez et al., 2014) (Roustakiani et al., 2018). The principal problem is that comments are written in natural language and there is a big gap between natural language (unstructured data) and applications which use structured data. Due to their unstructured nature, we are witnessing more cases of multilingual and mixed texts. One newer task of text mining is Persian Sentiment Analysis since thousands of websites, blogs, and social networks are used and alter by Persian users around the world (Asgarnezhad et al., 2020a).

Persian Sentiment Classification is an attractive field in Persian Sentiment Analysis. This field extracts the comments from the unstructured data on the websites to organize them into three classes, positive, neutral, and negative. Three levels exist for this problem. In document and sentence levels, the class of each document and each sentence are determined, respectively. In the feature level, the class of each feature of reviews is selected (Jiang et al., 2011). The main challenge here is the lack of training datasets. To the best of our knowledge, there is not a comperehensive article in employing machine learning approaches on Persian texts. Moreover,

---

* *Corresponding Author: razyehan@gmail.com*

investigating Persian texts suffers from lack of datasets and tools. Since the nature of the Persian language is distinct from other languages. Also, recognition of nouns, parts of speech (POS) tagging, and stemming for the Persian language are unknown issues. Hence, we need to discuss some of the challenges.

Several systems in the English language have focused on databases like Movie dataset, developed by Pang et al. (2002), such as the work of Asgarnezhad & Monadjemi (2021) , or on Twitter (Asgarnezhad et al., 2020a). But there are a few available datasets in the Persian language in this field. Pre-processing has a prominent role in Sentiment Classification. Most of the studies have been resulted from traditional text classification approaches to analyze a document such as the Bag of Words (BOW) and the POS tagging. It was revealed that POS tags could not provide enough information in Natural Language Processing (NLP) analyses. The POS tags will add unnecessary complexity; in contrast, words are appropriate indicators for sentiment polarity detection. The significant results were more optimal rather than the base classifiers achieved (Tripathy et al., 2016). Hence, pre-processing has a particular role herein because of the Persian language nature.

Sentiment Classification approaches are classified in Machine Learning (ML), Lexicon, and Hybrid approaches. The ML approaches handle supervised, unsupervised, and semi-supervised methods. It is evident that the ML algorithms like Maximum Entropy (ME), Support Vector Machine (SVM), and Naive Bayes (NB) have been employed successfully for many types of research. Identifying words for domain specifically is the main benefit of the lexicon-based approach. A hybrid approach combines the services of both approaches to enhance classification performance. Deep learning (Sharami et al., 2020), translated-based (Dehkharghani, 2019; Asgarnezhad et al., 2020), hybrid (Dashtipour et al., 2020) (Asgarnezhad, et al., 2020b), and lexicon-based (Amiri et al., 2015) approaches were used more in the Persian Sentiment Analysis.

The contributions of this study concern the following:
- A comparison among the existing Persian Sentiment analysis works
- Describing the available feature engineering stages in the context
- Showing the lack of data available for the Persian tasks
- Proposing future challenges and issues associated with Persian Sentiment Analysis
- Reviewing pre-processing methods in Persian texts
- Investigating feature selection methods in Persian texts
- Revealing the construction of sentiment lexicon resources for Persian text
- Interpreting available datasets for Persian text
- Explaining issues and challenges in the Persian language

The remainder of this article is organized as follows: Section 2 introduces some of definitions in the scope. Section 3 and Section 4 provide existing works and datasets, respectively. Section 5 offers open challenges and Section 6 concludes the paper.

2. DEFINITIONS

Although the community has done various types of research in the English language, there are a few Persian models that sufficiently handle Persian Sentiment Analysis. Fig. 1 presents all research publications associated with the Persian Sentiment Analysis area. The statistics have been shown in this figure.

Unlike the English language, the Persian language includes 32 characters. Also, words are written in an opposite direction i.e., from right to left. Hence, it is distinct from other languages. This section reviews four stages of Persian language processing. It is possible to explain the existing Persian works in four stages, as outlined in Fig. 2.

2.1. Pre-processing

Due to the nature of the Persian language, the appropriate pre-processing is needed in a step-wise manner. At first, tokenization is difficult because of the use of compound words. Eliminating these words will improve accuracy. Then, informal and common words should be recognized. Finally, although some of the words have half-space, most users do not write them when adding opinions on websites. Consequently, more accurate methods for word tokenization in this language are vital.
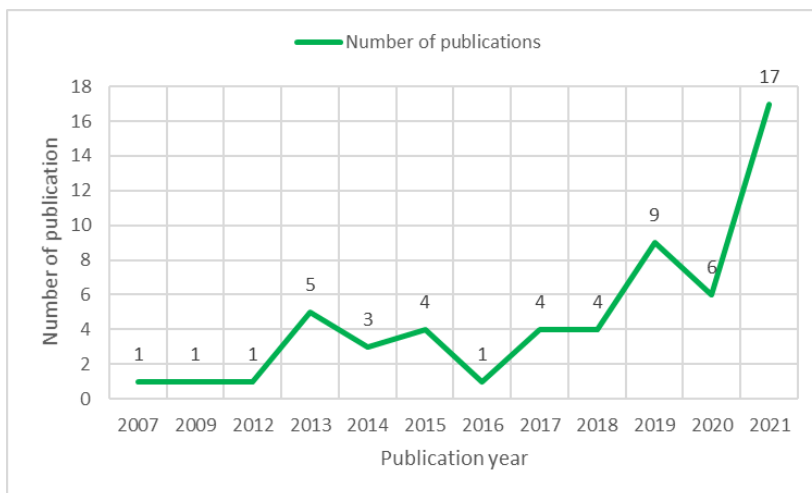
Fig. 1. The number of publications regarding Persian Sentiment Analysis
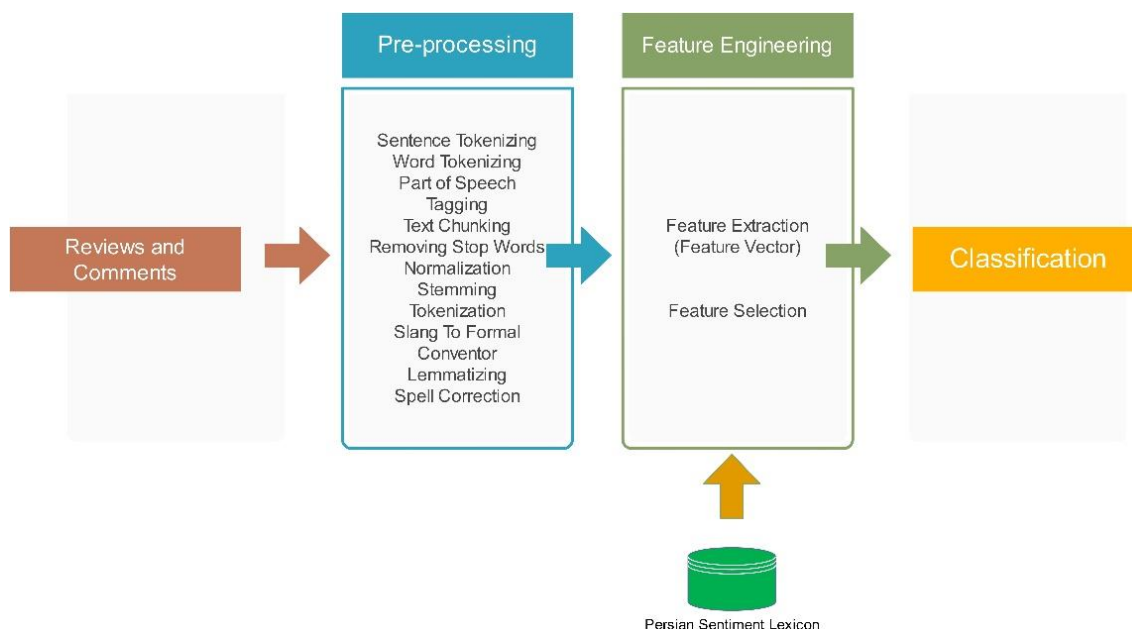

Fig. 2. The architecture of Persian Sentiment Analysis

Here, the steps of pre-processing are introduced with sentence tokenization, word tokenization, POS tagging, text chunking, removing stop words, normalization, stemming, and lemmatization. Tokenization is the separating process of a document into words or different types of significant parts, named tokens. The normalization process for a text referrs to transforming a document into an accepted scheme. Here, the symbols are carried regularly. The presence of remarkable terms causes a filtration analysis of data to occur. The most important step is to eliminate the words carrying no information, namely, the stop words. This step of pre-processing can improve the performance of the Persian Sentiment Analysis. Stemming is a manner of diminishing words and substitutes the root of the words. Finally, negations have a vital role in this context. In a pre-processing step, the methods have been applied through NLP tools like Hazm, polyglot, CoRef, etc.

2.2. *Feature engineering*

This stage consists of two sub-stages: feature extraction and feature selection.

o **Feature Extraction**

Here, various methods of feature extraction for the Persian language are of concern.

- POS tagging: It is applied to define disambiguation for a word. Here, a word with the adjective role in a sentence will be beneficial for determining the overall sentiment of the document, and more vital for the election of the best features. This feature is employed to classify sentiments in most studies about Persian Sentiment Analysis.
- N-grams: These are useful to identify effective features. After removing the stop-words, unigrams, bigrams, and trigrams were identified.
- Term Frequency-Inverse Document Frequency (TFIDF) based word weighting: A diversity of weighting mechanisms based on TFIDF performed for weighting like Augmented TF, Delta-TFIDF, LogAve TF, BM25 TF, DeltaProb, etc. For example, in the TFIDF weighting mechanism based on Delta, a mild mechanism is employed definitely for every class (Martineau & Finin, 2009).
- Character n-gram features: Various types of characters similar to 2-grams, 3-grams existed to employ for a set of features.
- Sentiment words feature: The sentiment word is selected as one of the significant features. Here, the polarity of these words calculated for every sentence in a document. To calculate the overall polarity, the polarities of various sentences are averaged. Hence, the selected words instantly influence the sentiment polarity. Consequently, it is important to employ Persian Sentiment Analysis.
- Bi-tagged feature: Here, This type of feature is offered through authors in (Turney, 2002) to select the appropriate features to expose the polarity of sentiments in the English language. All features in this type consist of outlined patterns of general collocations for expressing the polarity of a sentiment word.
- SentiWordNet (SWN) subjectivity scores: By assigning weights, this method estimates the subjectivity of the words. Regarding the defined threshold, words that are objective and do not endure in SWN are eliminated (O'Keefe & Koprinska, 2009).
- Word2vec cluster n-grams: According to the employed methodology in (Dong et al., 2015), the existing words in each comment diminished to vectors with 100-dimension utilizing the word2vec. Next, using a clustering method such as K-means clustered 100,000 words into 5000 clusters. Finally, the produced clusters were utilized to express the words of a comment.
- Sentiment-specific word embedding: The author (Tang et al., 2014) suggested an innovative method to express a word by promoting the Word2vec model. They confirmed that a sentiment-specific word embedding to transform the sentiment of features in a consecutive scope produces a more favorable performance in Sentiment Classification.

o **Feature Selection**

Numerous techniques are aimed at choosing the best features in the document (Habernal et al., 2015; Forman, 2003; Zheng et al., 2004; Uchyigit, 2012; Asgarnezhad et al. 2021). Here, a brief description of these methods is presented. Table 1 presents the definition of these methods.

- Mutual Information (MI): The MI concerning two identifiers can be a means for the mutual dependence relating to these identifiers. This measure is applied to calculate the appearance probability of a feature in the objective class in proportion to the overall occurrence probability for the feature (Schütze et al., 2008).
- Information Gain (IG): Here, the presence or absence of a feature in an original document is important. This metric defines the number of bits for this important information to divine the appropriate class for the document (Sebastiani, 2002).
- Chi-square (CHI) and Variants Chi-square ($\chi 2$): Here, the well-known analytical measures are defined to estimate the independence between two alternatives such as a feature and a class. These measures are applied to choose the features with superior properties. Also, authors (Ng et al., 1997) suggested a variant of $\chi 2$, namely NGL to showed superior NGL than $\chi 2$, in some cases. Besides, authors (Galavotti et al., 2000) displayed a simplified form of $\chi 2$, named the GSS coefficient. They asserted that GSS produces better results than NGL and $\chi 2$.

Table 1. Representation of notations and equations

Definitions	Equations
Document belongs to class c and contains word *w*	c_w
Document does not belong to class c and contains word *w*	$\bar{c}_w$
Document belongs to class c and does not contain word *w*	$c_{\bar{w}}$
Document does not belong to class c and does not contain word *w*	$\bar{c}_{\bar{w}}$
The total number of documents	$N = n_c + n_{\bar{c}} = (c_w + c_{\bar{w}}) + (\bar{c}_w + \bar{c}_{\bar{w}})$ $= c_w + c_{\bar{w}} + \bar{c}_w + \bar{c}_{\bar{w}}$
Mutual Information	$MI_w = (c_w \times N) / (c_w + c_{\bar{w}})(c_w + \bar{c}_w)$
Information Gain	$IG_w = -P(c)\log_2 P(c) + P(\bar{c})\log_2 P(\bar{c}) -$ $(P(w)(-P(c_w)\log_2 P(c_w) - P(\bar{c}_w)\log_2 P(\bar{c}_w))) $ $+(p(\bar{w})(-P(c_{\bar{w}})\log_2 P(c_{\bar{w}}) - P(\bar{c}_{\bar{w}})\log_2 P(\bar{c}_{\bar{w}})))$ $P(c_w) = c_w / (c_w + \bar{c}_w) \, , \; P(\bar{c}_w) = \bar{c}_w / (c_w + \bar{c}_w)$ $P(c_{\bar{w}}) = c_{\bar{w}} / (c_{\bar{w}} + \bar{c}_{\bar{w}}) \, , \; P(\bar{c}_{\bar{w}}) = \bar{c}_{\bar{w}} / (c_{\bar{w}} + \bar{c}_{\bar{w}})$ $P(w) = (c_w + \bar{c}_w) / N \, , \; P(\bar{w}) = (c_{\bar{w}} + \bar{c}_{\bar{w}}) / N$ $P(c) = n_c / N \, , \; P(\bar{c}) = n_{\bar{c}} / N$
Chi-square and Variants	$GSS_w = c_w \bar{c}_{\bar{w}} - \bar{c}_w c_{\bar{w}}$ $NGL_w = \dfrac{\sqrt{N}GSS_w}{\sqrt{(c_w + c_{\bar{w}})(\bar{c}_w + \bar{c}_{\bar{w}})(c_w + \bar{c}_w)(c_{\bar{w}} + \bar{c}_{\bar{w}})}}$ $\chi^2_w = (NGL_w)^2$
Relevancy Score and Odds Ratio	$OR_w = \dfrac{c_w \, \bar{c}_{\bar{w}}}{\bar{c}_w \, c_{\bar{w}}} \; , \; RS_w = \dfrac{c_w}{\bar{c}_{\bar{w}}}$
Document Frequency	$DF_w = c_w + \bar{c}_w$
Categorical Proportional Difference	$CPD = \dfrac{\left| DF_+ - DF_- \right|}{DF_+ - DF_-}$

- Relevancy Score (RS) and Odds Ratio (OR): The two mentioned measures identified analytical methods for choosing important features to reveal better results within classifying texts than IG and MI in some cases (Uchyigit, 2012) (Fragoudis, Meretakis, & Likothanassis, 2005).
- Document Frequency (DF): This method filter features according to the number of documents, which include supposed features (Agarwal & Mittal, 2016). Features with a number less than threshold are eliminated.

- ▪ Categorical Proportional Difference (CPD): This method was introduced by authors in (Simeon & Hilderman, 2008) to define the effective influence of each feature in expressing a proper class. The frequency of each feature is separately calculated. Words with a higher PD were polarized; whereas, other words with a lower PD allocated fairly in classes.

2.3.   Classification approaches

In this stage, Sentiment Classification approaches are introduced. These approaches are classified in ML, Lexicon, and Hybrid approaches. The ML approaches handle supervised, unsupervised, and semi-supervised methods. The applied ML algorithms in this study including SVM, NB, random Forrest (RF), logistic regression (LR), k-nearest neighbor (KNN), neural network (NN), and convolutional neural network (CNN) that employed successfully for many types of research. A hybrid approach combines the services of both approaches to enhance classification performance. Deep learning, translated-based, hybrid, and lexicon-based approaches were used more in the Persian Sentiment Analysis.

Table 2 shows the applied methods in Persian Sentiment Analysis tasks.

2.4.   Sentiment lexicon generation for persian language

The generation of a sentiment lexicon is a comprehensive necessary element to detect sentiment and polarity. These lexicons are produced through two ways: (1) progress or rendering of characters from the available dictionaries (Steinberger et al., 2012); (2) extension of the record of seed words with sentiment relating to an appropriate corpus (Cruz, Troyano, Pontes, & Ortega, 2014) (Mahyoub, Siddiqui, & Dahab, 2014). To the best of our review, there are a few lexicons in the Persian language, which are manually produced. There is thus evidence that providing the standard and the labeled dataset is important in this context.

Table 3 shows the review of the applied lexicon generation methods in Persian Sentiment Analysis.

3.   EXISTING WORKS

Persian Sentiment Analysis have been interesting considerably due to its modern applications in recent years. A few works suggested improving the classification performance on the available datasets. Those works differ from each other in the classifiers and Internet online forums. Deep learning has appeared as a robust ML technique to undertake the increasing requirement for proper Sentiment Analysis. To obtain knowledge of large volumes, Deep Learning techniques became more prevalent, but there are several challenges. Here, the existing works based on deep learning are presented.

In 2021, Shumaly et al., (2021) investigated the reviews on the *Digikala* website. The main problem of Persian Sentiment Analysis is the difficulty of the pre-processing stage because of unstructured data. The lack of possible archives for the Persian language increases this problem. To address the problem, 3 million Persian reviews were collected from the Digikala website to generate a word embedding. Also, word embedding is generated by applying the TF-IDF mechanism. The authors compared the results of the Convolutional Neural Network (CNN), BiLSTM, Logistic Regression, and NB models. They received an accuracy of 99.6% AUC and an F1 of 95.6%e. They obtained the accuracy better than other researchers done in Persian.

Dashtipour et al., (2021b) presented a Persian multimodal dataset including 800 queries to assess multimodal Sentiment Analysis in the Persian language. Next, they performed a context-aware multimodal Sentiment Analysis framework to determine the stated sentiment more precisely. They applied both decision-level and feature-level methods to consolidate cross-modal information. The highest results were 97, 84, 90, 91.39% for precision, recall, F1, and accuracy, respectively. In similar research, performed a context-aware deep-learning on Persian Sentiment Analysis. They suggested a deep-learning-driven feature engineering approach to analyze Persian movie reviews automatically. Two deep learning algorithms, convolutional neural networks (CNN) and long-short-term memory (LSTM) applied. Their results confirm that LSTM achieved a better performance as contrasted to other algorithms. The highest results were 96, 96, 96, 95.61% for precision, recall, F1, and accuracy, respectively (Dashtipour et al., 2021a). The Authors also organized an ensemble classifier for Persian Sentiment Analysis utilizing deep learning algorithms to enhance classification performance. They employed

Table 2. Applied methods in some of Persian Sentiment Analysis tasks.

Methods	References
Deep learning (Convolutional Neural Network (CNN), BiLSTM, MLP) +ML (Logistic Regression, NB, SVM)	(Shumaly, Yazdinejad, & Guo, 2021) (Dashtipour, Ieracitano, Morabito, Raza, & Hussain, 2021) (Davoudi & Mirzaei, 2021) (Akhoundzade & Devin, 2019) (Dashtipour et al., 2020)
Machine Learning	(Dashtipour, Gogate, Cambria, & Hussain, 2021) (Kasra Habib, 2021) (Sabri, Edalat, & Bahrak, 2021) (Dehkharghani, 2019) (Jahanbakhsh-Nagadeh, Feizi-Derakhshi, & Sharifi, 2020) (Hatefi Ghahfarrokhi & Shamsfard, 2020) (Basiri & Kabiri, 2017) (Dashtipour et al., 2016) (Alimardani & Aghaie, 2015) (Vamerzani & Khademi, 2015) (Hajmohammadi & Ibrahim, 2013) (Pourhassan, Pourebrahimi, & AFSHAR, 2013) (Shams, Shakery, & Faili, 2012) (Hamidi, Razzazi, & Ghaemmaghami, 2009)
Deep learning (Convolutional Neural Network (CNN), BiLSTM, Artificai Neural Network (ANN), Feed-Forward Back Propagation Neural Network (FFBPNN), Bi-directional Gated Recurrent Unit (bi-GRU), 2-dimensional Convolutional Neural Network (2CNN))	(Dashtipour, Gogate, Adeel, Larijani, & Hussain, 2021) (Shirghasemi, Bokaei, & Bijankhan, 2021) (Heydari, Khazeni, & Soltanshahi, 2021) (Kalaichelvi et al., 2021) (Sadeghi, Khotanlou, & Rasekh Mahand, 2021) (Karimvand, Chegeni, Basiri, & Nemati, 2021) (Taher & Shamsfard, 2021) (Sharami et al., 2020) (Gharavi, Bijari, Zahirnia, & Veisi, 2016) (Ataei, Darvishi, Javdan, Minaei-Bidgoli, & Eetemadi, 2019) (Zobeidi, Naderan, & Alavi, 2019) (Roshanfekr, Khadivi, & Rahmati, 2017)
Lexicon-based	(Pouromid, Yekkehkhani, Oskoei, & Aminimehr, 2021) (Karimi & Shahrabadi, 2019) (Ebrahimi Rashed & Abdolvand, 2017) (Amiri et al., 2015) (Golpar-Rabooki, Zarghamifar, & Rezaeenour, 2015) (Dehdarbehbahani, Shakery, & Faili, 2014) (Asgari & Chappelier, 2013)

standard (SVM, MLP) and deep (CNN) machine learning classifiers using the word2vec mechanism. Their suggested ensemble classifier obtained an accuracy of 79.68%. The highest results relating bigrams and ensemble in conjunction with SVM were precision of 80%, recall of 79%, F1 of 75, and accuracy of 78.18% (Dashtipour et al., 2021c).

Davoudi & Mirzaei (2021) introduced a feature extraction on Persian document classification. They allowed a combination of K-means clustering and Word2Vec to receive conventional descriptions for discriminant words. They employed 200 documents of 5 frequent groups of Hamshahri news datasets to review the influence of the suggested method. The applied classifiers were Multi-Layer Perceptron (MLP), Gradient Boosting (GB) using weighting mechanisms (TF-IDF), and Word2Vec methods, respectively. They could enhance the achieved accuracy of Gradient Boosting and Multi-Layer Perceptron models in relation to TF-IDF and Word2Vec techniques.

Table 3. A review of the applied lexicon generation methods in Persian Sentiment Analysis.

Methods	References
Corpus –based	(Pouromid et al., 2021) (Karimi & Shahrabadi, 2019) (Amiri et al., 2015) (Golpar-Rabooki et al., 2015) (Dehdarbehbahani et al., 2014) (Asgari & Chappelier, 2013)
Dictionary-based	(Jahanbakhsh-Nagadeh et al., 2020) (Ebrahimi Rashed & Abdolvand, 2017) (Asgari & Chappelier, 2013)

Aspect-based Sentiment Analysis is a more specific task in Sentiment Analysis to define opinion polarity via a particular aspect in a text. This process is attrracting more attention because it presents beneficial information. However, there is little research using this type on the Persian language. Jafarian et al. (2021) intended to develop the Aspect-based Sentiment Analysis on the Persian language. The authors displayed the using pre-trained BERT model and its utility using sentence-pair input on an Aspect-based Sentiment Analysis task. Their results could increase the task accuracy to 91%.

Shirghasemi et al. (2021) studied the influence of the Active Learning algorithm for Persian Sentiment Analysis. The cross-lingual model guided their model via utilizing a rich-resource language. They could decrease the dependency on training datasets. The applied Active Learning strategy helped to enhance the functionality of the model. Eventually, the Active Learning strategy progress their classifier to attain more knowledge. A hybrid deep learning-based Sentiment Analysis was presented to implement on reviews of the Digikala website (Heydari et al., 2021). They employed the classifier based on several deep learning networks and techniques. Eventually, they handled their approach to producing the best F1 of 78.3% . Heydari & Teimourpour (2021) evaluated the latest researches in Persian Natural Language Processing and presented Deep Learning models. They showed the challenges of Persian Sentiment Analysis and confirmed related Tools. They produced the Network of Researches in Persian Natural Language. A python library was presented by Kalaichelvi et al. (2021) to implement Sentiment Analysis for tweets. They studied the influence of Artificial Neural Networks (ANN) for producing a platform in Sentiment Analysis. The authors applied feed-forward backpropagation neural networks (FFBPNN) to divide the task into training data. They applied a min-max method to estimate the information and analyze the sentiment accuracy rate

Sadeghi et al. (2021) suggested a system with a combination of cognitive features and a deep neural network. The amount of 23,000 Persian documents was labeled for this work. The emotional structures, emotional keywords, and emotional POS were cognitive features in their approach. After pre-processing, the Word2Vec technique was used. Next, they developed a deep learning approach and implemented the classification algorithms such as NB, DT, and SVM to analyze emotions-based deep learning features. To assess the performance of the advanced system 10-fold cross-validation was applied. The experimental results exhibited that their system achieved an accuracy of 97%. The results displayed an improvement of several percent in contrast with the other results gained by GRU and cognitive features in solitude. A multimodal deep learning method for the Persian language has been proposed using a bi-directional gated recurrent unit (bi-GRU) and a 2-dimensional convolutional neural network (2CNN) for interpreting texts and images (Karimvand et al., 2021). To evaluate model performance, they added a new dataset of Instagram posts. Their results revealed that the model could promote the accuracy and F1 of 23% and 0.24%, respectively.

Taher & Shamsfard (2021) applied two approaches, adversarial training, weak supervision, and a few labeled data. They labeled a crawled dataset with supervised sentiment tags relating to a sentiment network. Later, they fine-tuned a pre-trained model with adversarial training on this dataset to produce domain-independent representations. Ultimately, they practiced the above network with 50 samples of data. Their results revealed

that their method exceeds on the same data with a 15% higher F1. Sharami et al. (2020) introduced a method using deep learning and obtained the F1 of 91.98% on the Digikala dataset. Ataei et al. (2019) exhibited a Persian dataset manually from the Digikala websites, namely Pars-ABSA. Furthermore, the authors applied methods in Sentiment Analysis based on deep learning. The highest results reported 85.54% for accuracy and 84.40% for F1.

Zobeidi et al. (2019) proposed a system to classify reviews in sentence-level through deep learning methods. They adopted three stages. First, sentences were converted to a matrix at word-level and character-level. Then, the features were extracted through CNN. Ultimately, using the Bidirectional LSTM (Bi-LSTM) network the reviews were classified. To evaluate, the Digikala dataset for two scopes like mobile and cameras was utilized. The highest precision, recall, F1, and accuracy rates were 94, 95, 94, and 95%, respectively. Roshanfekr et al. (2017) employed deep learning methods for their goal of producing a dataset through crawling on the Digikala website about electronic products. To evaluate their model, the Skip-gram model, BLSTM, and CNN were employed. The highest precision, recall, and F1 rates obtained 70.7, 52.2, and 55.4%, respectively.

Deep learning has also been applied for detecting plagiarism in the Persian language (Gharavi et al., 2016). In this method, words are exposed as multi-dimensional vectors. Besides, using aggregation methods, the word vectors were combined to express sentences. For detecting plagiarism, first, word vectors were extracted through the word2vec algorithm. Next, stop words were eliminated. Following that, the average of all vectors for each sentence was estimated. Then, each sentence corresponded with all existing sentences in terms of Cosine similarity distance. After this stage, the similarity between two sentences imposed through the Jaccard similarity distance. The authors received results in 90.6 % accuracy on plagdet, 85.8 recall, 95.9% precision on the PAN2016 datasets.

Table 4 displays a review of the existing works for Persian Sentiment Analysis which focus on deep learning in 2021. There are other existing works based on other techniques besides deep learning. ML approaches were employed to undertake the problems in this context (Kasra Habib, 2021). The authors performed an approach employing machine-translated datasets to handle Persian Sentiment Analysis. Eventually, the dataset was performed with various classifiers and feature engineering approaches. Their results revealed that the best classifier was SVM which achieved a precision of 91.22%, recall of 91.71%, and F1 score of 91.46%. Sabri et al. (2021) generated a demand for code-mixed Sentiment Analysis systems. They assembled labels and produced a dataset of code-mixed tweets in both Persian and English languages. They progressed to present a model which utilizes BERT pre-trained to learn the polarity scores of Tweets automatically. Their model outperformed the baseline models which used NB and RF methods.

In 2021, the authors (Pouromid et al., 2021) generated a corpus of 12000 Persian tweets from Twitter. They labeled tweets in three different categories of positive, neutral, and negative manually. Next, they produced a pre-trained ParsBERT model on these data. Their model was evaluated on the test dataset and compared to its counterparts. Accuracy of 82% achieved by the offered model surpassing its lexicon-based contender.

In 2020, Farahani et al. (2020) advised a monolingual model for Persian Sentiment Analysis. Their model included five stages; gathering data from websites, pre-processing, accurate segmentation of the sentence, pre-training, and fine-tuning. They worked on Digikala and SnappFood websites and reached an accuracy of 82.52% and an F1-value of 81.74% on the Digikala dataset. For the SnappFood dataset, their result showed 87.8% and 88.12% in terms of accuracy and F1, respectively. In 2019, Akhoundzade and Devin (Akhoundzade & Devin, 2019) advised a novel framework for extracting words through unsupervised methods in Persian Sentiment Analysis. Their framework utilized Neural Networks in conjunction with rule-based methods. The Digikala datasets included reviews on cellphones, tablets, and laptops domain. The resulted value of F1, precision, and recall were 58.6, 73.7, 99.1%, respectively.

Basiri et al. (Basiri, Kabiri, et al., 2019) proposes a method based on sentiment aggregation through the cross-ratio operator. The authors examined the aggregation process for Sentiment Analysis at the document level. Consequently, all existing aggregation methods were compared with their method. They exercised a pre-processing stage with six steps. Following that, they determined the sentiment of each word through an

        Journal of Applied Intelligent Systems & Information Sciences

Table 4. Some of Sentiment Analysis works using deep learning in the Persian Language from 2021 to 2007 (Note: A=Accuracy, F=F1, P=Precision, and R=Recall)

Reference	Description of work
(Shumaly et al., 2021)	3 million reviews gathered from the Digikala website, word embedding created using the TF-IDF, Convolutional Neural Network (CNN), BiLSTM, Logistic Regression, Naïve Bayes, A=99.6%, F=95.6%.
(Dashtipour, Gogate, Cambria, et al., 2021)	Multimodal dataset comprising more than 800 utterances, Decision-level, Feature-level, P=97%, R=84%, F=90%, A=91.39%. It estimates autoencoder and multilayer perceptron for Persian text.
(Dashtipour, Gogate, Adeel, et al., 2021)	Convolutional neural networks (CNN) and long-short-term memory (LSTM), P=96%, R=96%, F=96%, A=95.61%. It combines linguistic rules and deep learning.
(Dashtipour, Ieracitano, et al., 2021)	Standard (SVM, MLP) and deep (CNN) ML classifiers, word2vec, N-grams, ensembles, P=80%, R=79%, F=75%, A=78.18%.
(Jafarian et al., 2021)	Persian Pars-Aspect-based sentiment analysis, Pre-trained BERT model, sentence-pair, A=91%.
(Sadeghi et al., 2021)	Cognitive features, Deep neural network, 23,000 Persian documents, Emotional constructions, Emotional keywords, Emotional POS, Normalized text embedded by the Word2Vec, NB, Decision Tree, SVM, 10-fold cross-validation, A= 97%.
(Sharami et al., 2020)	Deep learning, SentiPers from Digikala (Hosseini, Ramaki, Maleki, Anvari, & Mirroshandel, 2018), F=91.98.
(Karimi & Shahrabadi, 2019)	Deep learning, Wikipedia, F=63%, P=49%, R=89%. The resulting lexicons are highly dependent on the corpus data.
(Ataei et al., 2019)	Reviews from www.digikala.com, A=85.54%, F=84.40%.
(Zobeidi et al., 2019)	Deep learning, CNN, and BLSTM, Review about mobile and digital cameras from www.digikala.com, A=95%, F=94%, P=94%, R=95%. It applies a character-level and word-level input matrix for feature extraction. Classification is performed in two classes and multi-class.
(Roshanfekr et al., 2017)	Deep learning, BLSTM, CNN, Customer reviews about electronic from www.digikala.com, F=55.4%, P=59.1%, R=52.2%.
(Gharavi et al., 2016)	Detect plagiarism using deep learning, PAN2016, P=95.9%, R=85.8%.

approach based on the lexicon. Conclusively, the calculation of the overall sentiment of the whole text was performed. To estimate, four Persian datasets regarding cell phones, collected from Digikala.com, were applied and attested the superiority of their method. Their obtained results were 59.9% for precision, 76.6% for accuracy, 67% for recall, and 64.1% for F1. In another research work, the same authors submitted a novelmethod for decomposing and detecting the important target of each sentence from a long review using five proposed models (Basiri, et al., 2019). To assess their method, three datasets were produced. The datasets were

collected from Naghdefarsi.com and Digikala.com websites. The highest results obtained the accuracy rate of 92%, the precision rate of 95%, the recall rate of 94%, and the F1 rate of 94%.

Dashtipour et al. (2020) recommended an innovative framework at the concept-level to distinguish polarity by linguistic rules through deep learning. They confirmed that their framework works better than approaches such as SVM, Logistic Regression (LR), long-short-term-memory (LSTM), and Convolutional Neural Networks (CNN). They applied dependency-based rules in conjunction with CNN and LSTM. Firstly, they pre-processed sentences to circumscribe the polarity of words according to rules. Next, they extracted sentence filled into a classifier to manage polarity. They exercised two datasets on product and hotel reviews corpora for evaluating classification performance. The product dataset was assembled from the Digikala website. It consists of 3000 reviews. The Hotel dataset consists of 3600 reviews, which were collected from the Hellokish website. The highest results were an accuracy of 81.14%, a precision of 76%, a recall of 98%, and a F1 value of 84% on Product review. As similar, on Hotel dataset, the highest results were 86.29% for accuracy, 87% for precision, 92% for recall, and 89% fro F1-value.

A new approach was suggesteded by Dehkharghani (2019) based on a translation to detect the polarity in the Persian language. The author translated the existing lexicons of the polarity from the English language to the Persian language. Next, the overall polarity of the translated words assessed through a supervised method like LR. In all experiments, 5-fold cross-validation was applied and the highest accuracy and F1 were 95.92 and 96%, respectively. Jahanbakhsh-Nagadeh et al. (2020) proposed a model for Sentiment Analysis through a dictionary-based technique. The pre-processing consists of tokenization, removing stop words, normalization, stemming, and lemmatization. Four methods such as unigram, bigram, POS tagging, and Hidden Markov Model employed to extract features. They practiced WordNet and four classification methods consist of Random Forest (RF), SVM, NB, and K-Nearest Neighbors (KNN). The highest performance in terms of accuracy was 94.9% for RF, 94.5% for SVM, 93.3% for NB, and 93.5% for KNN. Their results exposed that this method using RF and SVM achieved the highest performance in terms of accuracy, a rate of 95%. In this context for the Persian language, Karimi & Shahrabadi (2019) employed a pre-trained model, namely BERT. BERT is an unsupervised, contextual, deeply bidirectional system for pre-training. According to this structure, BERT is an effective system which uses dependencies of the terms that can be determined entirely by the polarity of words. The highest precision, recall, and F1 were obtained at 49, 89, and 63%, respectively.

In 2018, Ghahfarrokhi & Shamsfard (2020) recommended a hybrid approach, which combined two methods; lexicon and learning, in conjunction with ML approaches. They handled comments from the website of "Sahamyab" in three stock scopes consist of Khodro, Shabandar, and Vebmellat. Also, they collected comments from the Tsetmc website in two stock scopes consist of Shabandar and Vebmellat. First, they produced a sentiment lexicon and then calculated the scores of the sentiment. At last, the comments are classified through ML methods. After calculating DF and applying a threshold, they applied the pointwise MI to calculate the dependency criteria for each word in every class. They adopted ML alogrithms such as SVM, NB, and Decision Tree (DT). To evaluate, they applied the 10-fold cross-validation method.

Basiri and Kabiri (Basiri & Kabiri, 2018) devised a novel mechanism for aggregation at the sentence level. Then, they introduced a system to aggregate the elements in the sentence level into the document level. For evaluation, four datasets in Persian Language in Apple, Huawei, note 5, and Samsung domain were applied. Their results were better than the obtained results by the Dempster-Shafer method. They obtained a recall rate of 63%, a precision rate of 60%, an F1 rate of 61%, and an accuracy rate of 74%. In 2017, Asgarian et al., (2018) gathered opinions from the Digikala website through a web crawler. The applied dataset including 31,730 reviews on ten types of products. They proposed a system for Sentiment Classification and achieved an accuracy of 86%, a recall of 75%, and an F1 of 80%. The same authors recommended two new datasets, SPerSent and CNRC and used the majority voting and NB method to identify the overall polarity of comments on the Digikala website (Basiri & Kabiri, 2017). The highest performance according to precision, recall, F1, and accuracy were 90%, 88%, 89%, and 94%, respectively.

Ebrahimi Rashed & Abdolvand (2017) tried to produce a dictionary based on sentiments in the Persian language. They used a dataset including comments regarding eight domains such as mobile phone, clothing, digital camera, car, computer, DVD, electronics, and video domain. The precision, recall, and accuracy obtained were 84, 81, and 80%, respectively. In 2016, Dashtipour et al. offered a Persian lexicon-based sentiment with POS tags and polarity (Dashtipour et al., 2016). They utilized two ML algorithms, SVM and NB, to improve the classification performance. The highest accuracy for SVM and NB obtained 69.54 and 65.02%, respectively. In 2015, Alimardani & Aghaie (2015) provided a SentiWordNet in the Persian language based on Persian WordNet. By applying SVM, NB, and LR algorithms and weighting schemas on comments gathered from the Hellokish website. They evaluate the Sentiment Analysis task. The important results obtained through SVM, the accuracy rate of 87%, the precision rate of 86.9%, the recall rate of 87%, and the F1 rate of 87%. Amiri et al. Amiri et al. (2015) suggested a method-based sentiment lexicon in the Persian language. They could be improved the overall accuracy, an increase of 69%. The highest accuracy, precision, recall, and F1 rates were 82, 62, 63, and 68%, respectively.

A new technique was introduced in this context by Golpar-Rabooki et al. (Golpar-Rabooki et al., 2015) that consisted of lexicon production, pre-processing, feature extraction, and post-processing. They produced a lexicon in the Persian language to determine the orientation opinions. After the pre-processing stage, features were extracted based on frequency and dependency parsers. Finally, the overall polarity of the features was calculated. They collected reviews from the Digikala website in university and cell phone scope. The highest results belonged to the cellphone, with the precision rate of 94%, the recall rate of 72%, and the F1 rate of 81%. Vamerzani & Khademi (2015) proposed a new framework to predict the review polarity, extract the useful features, and classify them through the SVM classifier. To evaluate the proposed framework, the Digikala dataset was applied. The received performance in terms of recall, precision, and F1 was 87.42%, 93.03%, and 90.15%, respectively.

In 2014, Dehdarbehbahani et al. (Dehdarbehbahani et al., 2014) suggested a new method to use the resources in the English language for determining the polarity of comments in another language such as Persian. The principle goal of this method was the identification of the semantic orientations. They applied a Markov random walk model on Princeton WordNet 3.0, and FarsNet 1.0. The highest accuracy was obtained by 91.4%.

Hajmohammadi & Ibrahim, (2013) employed SVM and NB to classify user reviews. They found that the SVM classifier in conjunction with unigrams received better accuracy than NB on movie reviews written in Persian language. For the evalution of their work, the "Montaghed" website was used. The highest results belonged to SVM, with the F1 rate of 72.66%, the precision rate of 72.21%, and the recall rate of 73.12%. Pourhassan et al. (Pourhassan et al., 2013) introduced an approach relating to utilizing a fuzzy Bayesian classifier in this context. The precision and recall rates were obtained at 98.5 and 97.3%, respectively.

Asgari and Chappelier offered the description of tools using collected Persian poems (Asgari & Chappelier, 2013). To this end, they used Dehkhoda Online Dictionary and Virastyar Persian lexicon. These are accessible at www.loghatnaameh.org and www.virastyar.ir/data. In 2012, Shams et al. (2012) suggested an approach for Sentiment Analysis using Latent Dirichlet Allocation (LDA), namely LDASA. After translating the clues from English to Persian, in an iterative approach, the clues became correct. Next, the SVM is used to classify documents. Three datasets were collected from websites, which were in scopes including cell phones, digital cameras, and hotels. The obtained accuracy rate was 77%. Hamidi et al. (2009) suggested a system classifying Persian poems. First, syllable segmentation is used through three features. Next, syllable categorization, short and long, was applied. Finally, the SVM classifier with k-fold cross-validation was employed. They obtained an accuracy of 91%.

Aleahmad et al. (2007) built a model using weighting schemes to exhibit 4-gram can be improve the performance, with a precision rate of 77%.

Table 5 shows a review of the existing works for Persian Sentiment Analysis in addition to deep learning.

Table 5. A comparison among Sentiment Analysis works in the Persian Language from 2021 to 2007 (Note: A=Accuracy, F=F1, P=Precision, and R=Recall)

Ref.	Description of work	Applied dataset
(Kasra Habib, 2021)	Different classifiers and feature engineering approaches, ML, P=91.22, R=91.71, F=91.46	Machine translated datasets
(Farahani et al., 2020)	Sentiment Analysis, A=87.8, F=88.12	www.digikala.com, SnappFood
(Akhoundzade & Devin, 2019)	Unsupervised methods, NN, F=58.6, P=73.7, R=99.1	Cellphones, tablets, and laptops from www.digikala.com
(Basiri, Kabiri, et al., 2019)	Sentiment aggregation, A=76.6, F=64.1, P=59.9, R=67	Cell phones from www.digikala.com
(Basiri, Abdar, et al., 2019)	Most occurring first (MOF), most general first (MGF), most specific first (MSF), first occurring first (FOF), last occurring first (LOF), POS tags, A=92, F=94, P=95, R=94	Reviews about digital equipment from www.digikala.com
(Dashtipour et al., 2020)	Hybrid Sentiment Analysis, A=86.29, F=89, P=87, R=92	Product reviews from www.digikala.com Hotel reviews from http://www.hellokish.com
(Dehkharghani, 2019)	Translation-based approach, A=95.92, F=96	Four English resources during the construction of SentiFars
(Jahanbakhsh-Nagadeh et al., 2020)	Dictionary-based statistical technique, RF, SVM, NB, KNN, A=94.9 Common dictionaries do not hold information about a domain and are not proper for implementing sentiment lexicon in a domain.	Data manually collected in two classes, rumor and non-rumor
(Hatefi Ghahfarrokhi & Shamsfard, 2020)	Hybrid Sentiment Analysis, A=77.1, F= 76.7, R=76.3	TSE data from www.tsetmc.com
(Basiri & Kabiri, 2018)	Sentence-level aggregation mechanism, A=74, F=61, P=60, R=63	Four mentioned Persian review datasets Apple, Huawei, note 5, Samsung
(Asgarian et al., 2018)	Sentiment-lexicon generation, polarity classification, A=86, F=80, R=75	www.digikala.com

Table 5. *Contiuued*

(Basiri & Kabiri, 2017)	Sentence-level Sentiment Analysis, lexicon-based, A=94, F=89, P=90, R=88	www.digikala.com
(Ebrahimi Rashed & Abdolvand, 2017)	Supervised method, linguistic features, A=80, P=84, R=81 Dictionaries are possible in any language.	Reviews in area of digital camera, laptop, television, tablet, and mobile phones collected manually from online Retail Site Labelled English Reviews collected by Blitzer (Blitzer, Dredze, Pereira, & Biographies, 2007)
(Dashtipour et al., 2016)	Persian sentiment lexicon, NB, SVM, A= 69.54	lexicon is available from http://www.gelbukh.com/resources/persent
(Alimardani & Aghaie, 2015)	Persian SentiWordNet, SVM, NB, LR, A=87, F=87, P=86.9, R=87 Several experiments handled samples of various sizes. The impact of various classifications with three features studied.	Reviews from www.hellokish.com
(Amiri et al., 2015)	Lexicon-based Sentiment Analysis, A=82, F=68, P=62, R=63 Sentiment lexicon resource created manually. It is hard to make a corpus having a high coverage.	Manually collected from two online Persian language resources
(Golpar-Rabooki et al., 2015)	Creation of lexicon, pre-processing, feature extraction, and post-processing, F=81, P=94, R=72 Considering corpus is ordinarily in a specific domain, they are powerful in forming sentiment lexicon resources in a distinct domain.	Reviews in both scopes of university and cell phone areas form http://www.digikala.com
(Vamerzani & Khademi)	Polarity detection, SVM, F=90.15, P=93.03, R=87.42	http://www.digikala.com
(Basiri et al., 2014)	Sentiment Analysis, Dempster-Shafer strategy, F=86	Two datasets from online cell phone reviews
(Dehdarbehbahani et al., 2014)	Subjectivity analysis, Markov random walk model, A=91.4 The resulting lexicons are highly reliant on the corpus data and the lexicons enduring in that corpus.	Princeton WordNet 3.0 (Miller, 1995) and FarsNet 1.0 (Shamsfard et al., 2010)

Table 5. *Contiuued*

(Ghanbaran et al., 2014)	Speech acts of apology, ompliment	Persian apologetic and compliment utterances collected through Discourse Completion Test (DCT) data
(Hajmohammadi & Ibrahim, 2013)	SVM, NB, unigrams, bigrams, trigrams, F=72.66, P= 72.21, R=73.12	A corpus of Persian reviews about movie from http://www.montaghed.ir
(Pourhassan et al., 2013)	Fuzzy Bayesian, Naïve Bayesian, P=98.5, R=97.3	Persian online newsletters
(Asgari & Chappelier, 2013)	Topic modeling Dictionary-based methods employ synonymous and antonym semantic relations, while dictionaries are not up to date on these relationships.	A collection of Persian poems from http://ganjoor.net
(Shams et al., 2012)	Sentiment Analysis, PersianClues, unsupervised LDA-based, A=77	Three resources about hotels, cell phones and digital cameras manually gathered from e-shopping websites
(Hamidi et al., 2009)	SVM, A=91	136 poetries utterances from 12 Persian meter styles gathered from 8 speakers
(Aleahmad et al., 2007)	Local Context Analysis using different weighting schemes, P=77	A realistic corpus containing 160000+ news articles

According to Table 5, deep learning models have proven to be a better fit for the sentiment analysis task. The obtained accuracy and F1 by Dehkharghani (2019) were 95.92% and 96% in 2019, respectively. The current author applied a Translation-based approach. The obtained precision by Pourhassan et al. was 98% in 2013 (Pourhassan et al., 2013). They applied a Fuzzy Bayesian classifier. The highest recall achieved 99.1% by Akhoundzade & Devin (2019) in 2019. These authors used Unsupervised methods through the neural network classifier. Among the existing works that applied deep learning techniques, two works have the highest performance results. The first one was proposed by Shumaly et al. (2021). These authors gathered reviews from the Digikala website and employed word embedding using the TF-IDF, Convolutional Neural Network (CNN), BiLSTM, Logistic Regression, and Naïve Bayes methods. The highest accuracy and F1 were 99.6% and 95.6%, respectively. The second one was proposed by Dashtipour et al. (2021a). These authors employed Convolutional neural networks (CNN) and long-short-term memory (LSTM) methods. The highest accuracy, precision, recall, and F1 were 95.61, 96, 96, and 96%, respectively.

Among supervised methods, deep learning models are confirmed to be better and more powerfull for the Sentiment Analysis task. Also, they are domain-free and capable to control large numbers of data adequately. The capability of deep neural networks to produce state-of-the-art results on many NLP problems has been obvious to everyone for some years now. Nevertheless, when there is not sufficient labeled data, these networks suffer many challenges, and results might face severer flaws.

To sum up, we suggest applying other methods such as heuristic algorithms through NN classifiers in this context because there exist none of them in the literature.

4. AVAILABLE DATASETS

As mentioned earlier, there are few Persian resources for the Persian Sentiment Analysis. In this section, the comparison among the applied datasets in the Persian Sentiment Analysis is of concern. Fig. 3 depicts the impact of these datasets in this area.

According to Fig. 3, the Digikala website (DeepSentiPers) has 14.39% of all available resources in this context. Also, 14.39% of the works manually collected the reviews by crawling the websites.
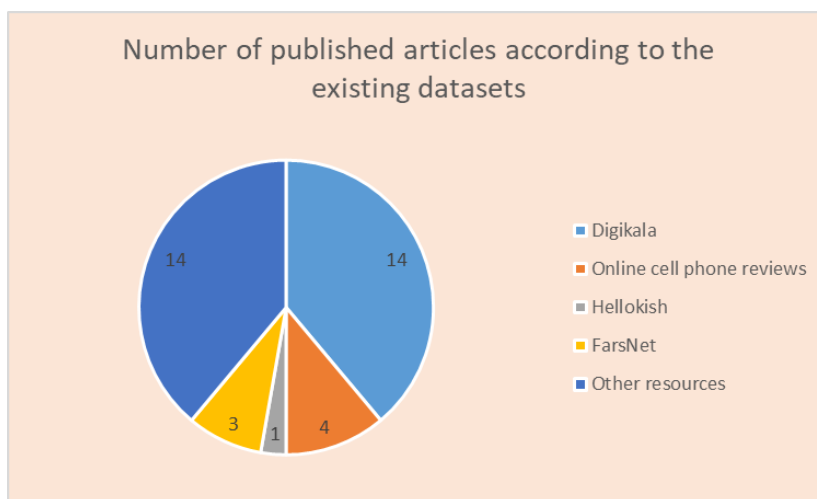

Fig. 3. Impact of different datasets in Persian Sentiment Analysis

In our studies, five WordNet found in the Persian language.

- Princeton WordNet (PWN): Here, a lexical corpus is available to do the works in the English language. This corpus consists of a set of the lexicon including synonymous. The classification is applied by two types of properties for a word: (1) POS tag, noun, verbs, adjectives, adverbs, etc; (2) meronymy, synonymy, hyponymy, antonymy, etc. The PWN5 is the final existing version that consists of 155,327 words. Also, WordNet is employed in (Montejo-Ráez et al., 2014) to obtain the words and features with the sentiment. Furthermore, authors have recently tried to build a WordNet in the Persian Language automatically. Nevertheless, only two corpora exist in the Persian language the FarsNet (Shamsfard et al., 2010) and PersianWN (Montazery & Faili, 2010).

- HelloKish: This dataset is interpreted through the author's feelings and attitude. Dataset assembled from user comments on the HelloKish website. When our research was being done, the volume of the second recorded comments on the website are 3312, and users enrolled to define the rate of customer entertainment in 642 items by website options.

- SentiPers: This dataset includes Persian sentences with sentiment values applied for Persian Sentiment Analysis. It is the first dataset for Persian Sentiment Analysis. The domain of sentences is digital products. Furthermore, the sentences of the dataset are formal, informal, or natural, and the dataset includes 1100 interpreted sentences. It provided through Guilan NLP Group.

- FarsNet: Persian lexicons with polarity label formed through the lab of Intelligent Information Systems at the University of Tehran, includes two datasets: 1) A set derived from interpreted Persian adjective: This is formed by the Persian adjectives of FarsNet. Each entry is defined as positive, negative, or neutral. It has more than 3588 adjectives are derived and estimated by four referees. 2) A set of adjectives, verbs, and nouns are derived from FarsNet. Each word in the set indicates sentiment value by the semi-supervised ML method. A value smaller than zero indicates a negative word, and a value greater than zero indicates positive words. The set includes 3588 adjectives, 4073 verbs, and 7325 nouns. Here, a lexical corpus existed to do the works in the Persian language. This corpus consists of

a set of words and their combinations. Also, it includs the POS tags and the relations among words. The FarsNet 2.0 is the final version of this corpus, accessible (Shamsfard et al., 2010).

- PreSent: Persian Sentiment Analysis displays real-valued polarity labels, in the rate from -1 to 1 Persian word and expressions. The first version of the lexicon consists of 1500 Persian words. The second version of the lexicon consists of 1500 Persian words from the first version plus 700 informal words and expressions. The expressions list has been confirmed especially beneficial for investigating highly informal texts like user-contributed contents, e.g., movie or product reviews.

- Persian WordNet generated from Tehran University (PersianWN): Here, the final version of the existing WordNet in the Persian language, produced through the University of Tehran (Montazery & Faili, 2010). It formed through spreading and performing unsupervised methods using the corpus in the English language. The FarsNet 1.0 is the final version of this corpus. It is applied to measure the initial probability for a word in a synset. Next, iteratively, the unsupervised methods are implemented to enhance this probability.

- Persian Sentiment WordNet (PSWN): Researchers tried to promote the existing corpora in the Persian language. To this end, the synsets outlined from Princeton WordNet to FerdowsNet using equivalent concepts. Following, the estimated polarity of each word planned from English SentiWordNet to the PSWN.

- Persian Sentiment Word Miner (PSWM): This lexical is comparable to the defined OpinionMiner in (Jin, Ho, & Srihari, 2009). Indeed, using some of the sequential algorithms, the lexicons will be obtained. At last, the sentiment of the words determined through learning approaches.

5. OPEN CHALLENGES AND OPPORTUNITIES

After reviewing the literature, we have recognized various challenges and issues in the scope of this research. In summary, there is still a need for more studies. Here, a summarization of these challenges is presented.

1) Iran has various dialects with varying meanings and shapes in conjunction with a large province, in which people have special accents. For example, the writing form of a word is extremely different in the south, the north, the west, the east, and the center of Iran. That is why there are few studies in this language.

2) Another challenge is the lack of sufficient datasets. Sentiment lexicon resources and standard interpreted datasets for Persian Sentiment Analysis have some restrictions. There is no relevant resource for different domains yet.

3) The Persian language is full of vulgarity, irony, idiom expression which is written in informal forms. More importantly, there is a lack of appropriate tools to focus on this challenge.

4) Another main challenge is that a document in the Persian language can be including the words both in Persian and in English languages. People usually utilize two languages for sending their comments on websites and social networks.

5) Persian language grammar and writing are very complicated, and this complexity makes text analysis sophisticated than that of other languages.

6) Persian text includes formal and informal writing. While these have the same meaning, review texts are organized into two divisions: explicit and implicit reviews. The explicit reviews apply a sentiment lexicon. But the implicit reviews do not apply a sentiment lexicon in text. The implicit review is a real sentence intimating the review. Few pieces of research concentrate on investigating implicit review. The second division donates particular methods. But, it is far apart from a complete solution. To this point, we propose studying implicit opinions for the Persian language in future research.

7) There are various types of spaces in persian texts. Sometimes space may occur between two words, and seldom within a word as a short space. This challenge affects accurate word tokenizing in Persian text. Consequently, numerous words are written with space, or without any spaces.

8) Increase the performance of tokenization: Tokenizer has a great influence on the Sentiment Analysis. A tokenizer with adequate accuracy promotes the pre-processing of Sentiment Analysis, particularly word embedding for deep neural network methods. Hence, it is recommended to investigate implementations in the Persian language to relinquish an adequate accuracy for the tokenizer.

9) Another attribute of the Persian language is that this is greatly generative. We can produce a new word using prefixes, postfixes, and affixes. Hence, the possibility of meeting an innovative term is very high. However, there are no lexicon resources for these new terms. Therefore, there are words in Persian that have the same written forms with, or without various pronunciations.

10) The concept-level Sentiment Analysis approach studies concepts and implicit semantics of texts. It concentrates on the investigation of a text through ontology and semantic networks. Hence, we suggest applying the concept-level approach for Persian texts.

11) Considering polarity of the sentiment lexicon is not inactive and depends on the context. Many articles consider this subject in English text. As one of the weaknesses, the issue is not tackled in investigations on the Persian language. It is recommended to consider context-based Sentiment Analysis in Persian texts in future research.

12) The Persian language suffers from lack of tools. It is proposed to implement tools through experimental methods in the Persian language. Some research labs advance tools for the elevation of self-works. But, they do not lead to publishing them on the Internet. This subject guides the diminished growth of Persian Sentiment Analysis.

Some challenges like feature extraction, sarcasm detection, fake review detection, polarity detection of implicit opinion, and so on are less investigated in Persian text. Hence, future investigations should examine them. Moreover, several challenges have endured for Sentiment Analysis in the English language. Some studies were implemented to solve these challenges. So, the Persian language earns to be similar to the English language to attain such results.

6. CONCLUSION

People purchase products and give their reviews and opinions about them every second on the Internet. These reviews and opinions affect the financial statements in companies noticeably. It is difficult for people to make decisions about products. The problem is more sophisticated, which these reviews are in the Persian language. In this paper we presented and compared the prominent research works in Persian Sentiment Analysis. Because there are no tools or labeled resources in this context. Persian Sentiment Classification is a significant field in Persian Sentiment Analysis that helps people with correct decision-making. Four stages include pre-processing, feature engineering, lexicon generation, and classification were introduced. In addition, different methods and the existing datasets that can be imposed in the Persian Sentiment Analysis were reviewed. Due to the Persian language nature and reality, more challenges and issues exist related to text mining. Further studies are needed to determine whether the applied methods in the English language could be applied in this context. Also, we suggested applying other methods such as heuristic and hybrid approaches be useful in enhancing the performance of classification in Persian Sentiment Analysis.

REFERENCES

Agarwal, B., & Mittal, N. (2016). *Prominent feature extraction for sentiment analysis*: Springer.

Akhoundzade, R., & Devin, K. H. (2019). Persian Sentiment Lexicon Expansion Using Unsupervised Learning Methods. *Proceedings of 9th International Conference on Computer and Knowledge Engineering* (ICCKE).

Aleahmad, A., Hakimian, P., Mahdikhani, F., & Oroumchian, F. (2007). N-gram and local context analysis for persian text retrieval. *Proceedings of 9th International Symposium on Signal Processing and Its Applications.*

Alimardani, S., & Aghaie, A. (2015). Opinion mining in Persian language using supervised algorithms.

Amiri, F., Scerri, S., & Khodashahi, M. (2015). *Lexicon-based sentiment analysis for Persian text.* Paper presented at the Proceedings of the International Conference Recent Advances in Natural Language Processing.

Asgari, E., & Chappelier, J.-C. (2013). *Linguistic resources and topic models for the analysis of persian poems.* Paper presented at the Proceedings of the Workshop on Computational Linguistics for Literature.

Asgarian, E., Kahani, M., & Sharifi, S. (2018). The impact of sentiment features on the sentiment polarity classification in Persian reviews. *Cognitive Computation, 10*(1), 117-135.

Asgarnezhad, R., Monadjemi, A., & Soltanaghaei, M. (2020a). A High-Performance Model based on Ensembles for Twitter Sentiment Classification. *Journal of Electrical and Computer Engineering Innovations (JECEI), 8*(1), 41-52.

Asgarnezhad, R., Monadjemi, A., & Soltanaghaei, M. (2020b). NSE-PSO: Toward an Effective Model Using Optimization Algorithm and Sampling Methods for Text Classification. *Journal of Electrical and Computer Engineering Innovations (JECEI), 8*(2), 183-192.

Asgarnezhad, R., & Monadjemi, S. A. (2021). NB VS. SVM: AContrastive STUDY FOR SENTIMENT CLASSIFICATION ON TWO TEXT DOMAINS.

Asgarnezhad, R., Monadjemi, S. A., & Soltanaghaei, M. (2020c). FAHPBEP: A fuzzy Analytic Hierarchy Process framework in text classification. *Majlesi Journal of Electrical Engineering, 14*(3), 111-123.

Asgarnezhad, R., Monadjemi, S. A., & Soltanaghaei, M. (2021). An application of MOGW optimization for feature selection in text classification. *The Journal of Supercomputing, 77*(6), 5806-5839.

Ataei, T. S., Darvishi, K., Javdan, S., Minaei-Bidgoli, B., & Eetemadi, S. (2019). Pars-ABSA: an Aspect-based Sentiment Analysis dataset for Persian. *arXiv preprint arXiv:1908.01815*.

Basiri, M. E., Abdar, M., Kabiri, A., Nemati, S., Zhou, X., Allahbakhshi, F., & Yen, N. Y. (2019). Improving sentiment polarity detection through target identification. *IEEE Transactions on Computational Social Systems, 7*(1), 113-128.

Basiri, M. E., & Kabiri, A. (2017). *Sentence-level sentiment analysis in Persian.* Paper presented at the 2017 3rd International Conference on Pattern Recognition and Image Analysis (IPRIA).

Basiri, M. E., & Kabiri, A. (2018). *Uninorm operators for sentence-level score aggregation in sentiment analysis.* Paper presented at the 2018 4th International Conference on Web Research (ICWR).

Basiri, M. E., Kabiri, A., Abdar, M., Mashwani, W. K., Yen, N. Y., & Hung, J. C. (2019). The effect of aggregation methods on sentiment classification in Persian reviews. *Enterprise Information Systems*, 1-28.

Basiri, M. E., Naghsh-Nilchi, A. R., & Ghassem-Aghaee, N. (2014). A framework for sentiment analysis in persian. *Open transactions on information processing, 1*(3), 1-14.

Blitzer, J., Dredze, M., Pereira, F., & Biographies, B. (2007). *boom-boxes and blenders: Domain adaptation for sentiment classification.* Paper presented at the Proceedings of the Association for Computational Linguistics (ACL).

Cruz, F. L., Troyano, J. A., Pontes, B., & Ortega, F. J. (2014). Building layered, multilingual sentiment lexicons at synset and lemma levels. *Expert Systems with Applications, 41*(13), 5984-5994.

Dashtipour, K., Gogate, M., Adeel, A., Larijani, H., & Hussain, A. (2021). Sentiment analysis of persian movie reviews using deep learning. *Entropy, 23*(5), 596.

Dashtipour, K., Gogate, M., Cambria, E., & Hussain, A. (2021). A novel context-aware multimodal framework for persian sentiment analysis. *arXiv preprint arXiv:2103.02636*.

Dashtipour, K., Gogate, M., Li, J., Jiang, F., Kong, B., & Hussain, A. (2020). A hybrid Persian sentiment analysis framework: Integrating dependency grammar based rules and deep neural networks. *Neurocomputing, 380*, 1-10.

Dashtipour, K., Hussain, A., Zhou, Q., Gelbukh, A., Hawalah, A. Y., & Cambria, E. (2016). *PerSent: A freely available Persian sentiment lexicon.* Paper presented at the International Conference on Brain Inspired Cognitive Systems.

Dashtipour, K., Ieracitano, C., Morabito, F. C., Raza, A., & Hussain, A. (2021). An Ensemble Based Classification Approach for Persian Sentiment Analysis *Progresses in Artificial Intelligence and Neural Systems* (pp. 207-215): Springer.

Davoudi, S., & Mirzaei, S. (2021). *A Semantic-based Feature Extraction Method Using Categorical Clustering for Persian Document Classification.* Paper presented at the 2021 26th International Computer Conference, Computer Society of Iran (CSICC).

Dehdarbehbahani, I., Shakery, A., & Faili, H. (2014). Semi-supervised word polarity identification in resource-lean languages. *Neural networks, 58*, 50-59.

Dehkharghani, R. (2019). Sentifars: A persian polarity lexicon for sentiment analysis. *ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP), 19*(2), 1-12.

Dong, L., Wei, F., Yin, Y., Zhou, M., & Xu, K. (2015). *Splusplus: a feature-rich two-stage classifier for sentiment analysis of tweets.* Paper presented at the Proceedings of the 9th international workshop on semantic evaluation (SemEval 2015).

Ebrahimi Rashed, F., & Abdolvand, N. (2017). A Supervised Method for Constructing Sentiment Lexicon in Persian Language. *Journal of Computer & Robotics, 10*(1), 11-19.

Farahani, M., Gharachorloo, M., Farahani, M., & Manthouri, M. (2020). ParsBERT: Transformer-based Model for Persian Language Understanding. *arXiv preprint arXiv:2005.12515*.

Forman, G. (2003). An extensive empirical study of feature selection metrics for text classification. *Journal of machine learning research, 3*(Mar), 1289-1305.

Fragoudis, D., Meretakis, D., & Likothanassis, S. (2005). Best terms: an efficient feature-selection algorithm for text categorization. *Knowledge and Information Systems, 8*(1), 16-33.

Galavotti, L., Sebastiani, F., & Simi, M. (2000). *Experiments on the use of feature selection and negative evidence in automated text categorization.* Paper presented at the International Conference on Theory and Practice of Digital Libraries.

Ghanbaran, S., Rahimi, M., & Rasekh, A. E. (2014). Intensifiers in Persian discourse: Apology and compliment speech acts in focus. *Procedia-Social and Behavioral Sciences, 98*, 542-551.

Gharavi, E., Bijari, K., Zahirnia, K., & Veisi, H. (2016). *A Deep Learning Approach to Persian Plagiarism Detection.* Paper presented at the FIRE (Working Notes).

Golpar-Rabooki, E., Zarghamifar, S., & Rezaeenour, J. (2015). Feature extraction in opinion mining through Persian reviews. *Journal of AI and Data Mining, 3*(2), 169-179.

Gudakahriz, S. J., Moghadam, A. M. E., & Mahmoudi, F. An Experimental Study on Performance of Text Representation Models for Sentiment Analysis. *Information Systems & Telecommunication*, 45.

Habernal, I., Ptáček, T., & Steinberger, J. (2015). Reprint of "Supervised sentiment analysis in Czech social media". *Information Processing & Management, 51*(4), 532-546.

Hajmohammadi, M. S., & Ibrahim, R. (2013). *A SVM-based method for sentiment analysis in Persian language.* Paper presented at the International Conference on Graphic and Image Processing (ICGIP 2012).

Hamidi, S., Razzazi, F., & Ghaemmaghami, M. P. (2009). *Automatic meter classification in Persian poetries using support vector machines.* Paper presented at the 2009 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT).

Hatefi Ghahfarrokhi, A., & Shamsfard, M. (2020). Tehran stock exchange prediction using sentiment analysis of online textual opinions. *Intelligent Systems in Accounting, Finance and Management, 27*(1), 22-37.

Heydari, M., Khazeni, M., & Soltanshahi, M. A. (2021). *Deep Learning-based Sentiment Analysis in Persian Language.* Paper presented at the 2021 7th International Conference on Web Research (ICWR).

Heydari, M., & Teimourpour, B. (2021). *Persian Opinion Mining: A Networked Analysis Approach.* Paper presented at the 2021 7th International Conference on Web Research (ICWR).

Hosseini, P., Ramaki, A. A., Maleki, H., Anvari, M., & Mirroshandel, S. A. (2018). SentiPers: A sentiment analysis corpus for Persian. *arXiv preprint arXiv:1801.07737.*

Jafarian, H., Taghavi, A. H., Javaheri, A., & Rawassizadeh, R. (2021). *Exploiting BERT to improve aspect-based sentiment analysis performance on Persian language.* Paper presented at the 2021 7th International Conference on Web Research (ICWR).

Jahanbakhsh-Nagadeh, Z., Feizi-Derakhshi, M.-R., & Sharifi, A. (2020). *A Speech Act Classifier for Persian Texts and its Application in Identifying Rumors.* Journal of Soft Computing and Information Technology, 2020, 9(1), 18-27.

Jiang, L., Yu, M., Zhou, M., Liu, X., & Zhao, T. (2011). *Target-dependent twitter sentiment classification.* Paper presented at the Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1.

Jin, W., Ho, H. H., & Srihari, R. K. (2009). *OpinionMiner: a novel machine learning system for web opinion mining and extraction.* Paper presented at the Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining.

Kalaichelvi, T., Gracytherasa, W., Kumar, S. P., Abirami, M. M., Archana, M. E., & Monisha, M. (2021). Sentiment Analysis Using FFBP Neural Network for Profit of Commercial Products in Industry. *Annals of the Romanian Society for Cell Biology*, 736-742.

Karimi, S., & Shahrabadi, F. S. (2019). Sentiment analysis using BERT (pre-training language representations) and Deep Learning on Persian texts.

Karimvand, A. N., Chegeni, R. S., Basiri, M. E., & Nemati, S. (2021). *Sentiment Analysis of Persian Instagram Post: a Multimodal Deep Learning Approach.* Paper presented at the 2021 7th International Conference on Web Research (ICWR).

Kasra Habib, M. (2021). The Challenges of Persian User-generated Textual Content: A Machine Learning-Based Approach. *arXiv e-prints*, arXiv: 2101.08087.

Khan, F. H., Bashir, S., & Qamar, U. (2014). TOM: Twitter opinion mining framework using hybrid classification scheme. *Decision Support Systems, 57*, 245-257.

Mahyoub, F. H., Siddiqui, M. A., & Dahab, M. Y. (2014). Building an Arabic sentiment lexicon using semi-supervised learning. *Journal of King Saud University-Computer and Information Sciences, 26*(4), 417-424.

Martineau, J. C., & Finin, T. (2009). *Delta tfidf: An improved feature space for sentiment analysis.* Paper presented at the Third international AAAI conference on weblogs and social media.

Miller, G. A. (1995). WordNet: a lexical database for English. *Communications of the ACM, 38*(11), 39-41.

Montazery, M., & Faili, H. (2010). *Automatic Persian wordnet construction.* Paper presented at the Coling 2010: Posters.

Montejo-Ráez, A., Martínez-Cámara, E., Martín-Valdivia, M. T., & Ureña-López, L. A. (2014). Ranked wordnet graph for sentiment polarity classification in twitter. *Computer Speech & Language, 28*(1), 93-107.

Ng, H. T., Goh, W. B., & Low, K. L. (1997). *Feature selection, perceptron learning, and a usability case study for text categorization.* Paper presented at the Proceedings of the 20th annual international ACM SIGIR conference on Research and development in information retrieval.

O'Keefe, T., & Koprinska, I. (2009). *Feature selection and weighting methods in sentiment analysis.* Paper presented at the Proceedings of the 14th Australasian document computing symposium, Sydney.

Pang, B., Lee, L., & Vaithyanathan, S. (2002). *Thumbs up?: sentiment classification using machine learning techniques.* Paper presented at the Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10.

Pourhassan, P., Pourebrahimi, A., & AFSHAR, K. M. A. (2013). Using Fuzzy LR Numbers in Bayesian Text Classifier for Classifying Persian Text Documents.

Pouromid, M., Yekkehkhani, A., Oskoei, M. A., & Aminimehr, A. (2021). *ParsBERT Post-Training for Sentiment Analysis of Tweets Concerning Stock Market.* Paper presented at the 2021 26th International Computer Conference, Computer Society of Iran (CSICC).

Roshanfekr, B., Khadivi, S., & Rahmati, M. (2017). *Sentiment analysis using deep learning on Persian texts.* Paper presented at the 2017 Iranian Conference on Electrical Engineering (ICEE).

Roustakiani, A., Abdolvand, N., & Harandi, S. R. (2018). An Improved Sentiment Analysis Algorithm Based on Appraisal Theory and Fuzzy Logic. *Information Systems & Telecommunication*, 88.

Sabri, N., Edalat, A., & Bahrak, B. (2021). *Sentiment Analysis of Persian-English Code-mixed Texts.* Paper presented at the 2021 26th International Computer Conference, Computer Society of Iran (CSICC).

Sadeghi, S. S., Khotanlou, H., & Rasekh Mahand, M. (2021). Automatic Persian Text Emotion Detection using Cognitive Linguistic and Deep Learning. *Journal of AI and Data Mining*.

Schütze, H., Manning, C. D., & Raghavan, P. (2008). *Introduction to information retrieval* (Vol. 39): Cambridge University Press Cambridge.

Sebastiani, F. (2002). Machine learning in automated text categorization. *ACM computing surveys (CSUR), 34*(1), 1-47.

Shams, M., Shakery, A., & Faili, H. (2012). *A non-parametric LDA-based induction method for sentiment analysis.* Paper presented at the The 16th CSI international symposium on artificial intelligence and signal processing (AISP 2012).

Shamsfard, M., Hesabi, A., Fadaei, H., Mansoory, N., Famian, A., Bagherbeigi, S., . . . Assi, S. M. (2010). *Semi automatic development of farsnet; the persian wordnet.* Paper presented at the Proceedings of 5th global WordNet conference, Mumbai, India.

Sharami, J. P. R., Sarabestani, P. A., & Mirroshandel, S. A. (2020). DeepSentiPers: Novel Deep Learning Models Trained Over Proposed Augmented Persian Sentiment Corpus. *arXiv preprint arXiv:2004.05328.*

Shirghasemi, M., Bokaei, M. H., & Bijankhan, M. (2021). *The Impact of Active Learning Algorithm on a Cross-lingual model in a Persian Sentiment Task.* Paper presented at the 2021 7th International Conference on Web Research (ICWR).

Shumaly, S., Yazdinejad, M., & Guo, Y. (2021). Persian sentiment analysis of an online store independent of pre-processing using convolutional neural network with fastText embeddings. *PeerJ Computer Science, 7*, e422.

Simeon, M., & Hilderman, R. (2008). *Categorical proportional difference: A feature selection method for text categorization.* Paper presented at the Proceedings of the 7th Australasian Data Mining Conference-Volume 87.

Steinberger, J., Ebrahim, M., Ehrmann, M., Hurriyetoglu, A., Kabadjov, M., Lenkova, P., . . . Zavarella, V. (2012). Creating sentiment dictionaries via triangulation. *Decision Support Systems, 53*(4), 689-694.

Taher, S. E., & Shamsfard, M. (2021). *Adversarial Weakly Supervised Domain Adaptation for Few Shot Sentiment Analysis.* Paper presented at the 2021 7th International Conference on Web Research (ICWR).

Tang, D., Wei, F., Yang, N., Zhou, M., Liu, T., & Qin, B. (2014). *Learning sentiment-specific word embedding for twitter sentiment classification.* Paper presented at the Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers).

Tripathy, A., Agrawal, A., & Rath, S. K. (2016). Classification of sentiment reviews using n-gram machine learning approach. *Expert Systems with Applications, 57*, 117-126.

Turney, P. D. (2002). Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. *arXiv preprint cs/0212032*.

Uchyigit, G. (2012). *Experimental evaluation of feature selection methods for text classification.* Paper presented at the 2012 9th International Conference on Fuzzy Systems and Knowledge Discovery.

Vamerzani, H. A., & Khademi, M. (2015). Increase Business Intelligence Based on opinions mining in the Persian Reviews. *International Academic Journal of Science and Engineering*, 2(2), 164-174.

Zheng, Z., Wu, X., & Srihari, R. (2004). Feature selection for text categorization on imbalanced data. *ACM Sigkdd Explorations Newsletter, 6*(1), 80-89.

Zobeidi, S., Naderan, M., & Alavi, S. E. (2019). Opinion mining in Persian language using a hybrid feature extraction approach based on convolutional neural network. *Multimedia Tools and Applications, 78*(22), 32357-32378.