Adopting Transition Point Technique for Persian Sentiment Analysis

1st Kia Dashtipour
School of Computing
Edinburgh Napier University
Edinburgh EH10 5DT, UK
k.dashtipour@napier.ac.uk

2nd Mandar Gogate
School of Computing
Edinburgh Napier University
Edinburgh EH10 5DT, UK
m.gogate@napier.ac.uk

3rd Alexander Gelbukh
Centro de Investigacio ́n en Computacio ́n (CIC)
Instituto Polite ́cnico Nacional (IPN)
Mexico City, Mexico
gelbukh@gelbukh.com

4th Amir Hussain
School of Computing
Edinburgh Napier University
Edinburgh EH10 5DT, UK
a.hussain@napier.ac.uk

***Abstract*—Sentiment analysis is used to analyses people's opinions, views and emotions towards different entities such as products, organizations, companies and events. People's opinions are important for most others during their decision-making process. For example, if someone wants to buy a product, they might want to know more about that product and the experiences of others with that product. Sentiment analysis is able to classify the reviews based on their polarity; even if reviews are expressed in a sentence or document, sentiment analysis is used to classify it into positive, negative or neutral reviews. In this paper, we proposed a framework using TF-IDF and transition point to detect polarity in Persian movie reviews. The proposed approach has been evaluated using different classifiers such as SVM, Naive Bayes, MLP and CNN. The experimental results show the transition point is more effective in comparison with traditional feature such as TF-IDF.**

***Index Terms*—Sentiment Analysis, Persian, Machine Learning**

I. Introduction

With the emergence of Internet, people is capable to share their comments and opinion using social network such as Twitter, Facebook, etc. Internet allow people to debate and share their opinions on different topics [1]–[5]. There has been interest from companies and organisations in classifying opinion to obtain useful information about their product and services. In other words, it identifies whether a sentence is positive or negative [6]–[13]. The sentiment analysis is useful for companies to automatically determines the polarity of the sentence. Sentiment analysis can identify the overall polarity of thousands of comments on short time of period. The evaluation of comments allows companies to improve their product and services. Thus, government used sentiment analysis to understand public's opinion about services which government provides for people . Sentiment analysis has been classified into predicted class such as subjective or objective, predict polarity which can be positive, negative or neutral or level of classification such as phrase, sentence or document and different approach can be applied such as supervised or unsupervised [14]–[17].

Sentiment analysis consist of two different approaches, the first approach is based on machine learning classifiers. It is used to trained classifier based on the predefined classes, this method is called supervised classification. The second method is based on the words along with their polarities (+1 to -1), the model calculates the polarity of the sentence from polarity of the word and assign final polarity into sentence, this is called unsupervised method (Joshi et al. 2017). The supervised classification has received high performance; however, it has some shortcomings. It required to build a trained dataset and labelled documents by human expert. The process is very time consuming and it is difficult to train sarcasm and idioms. Moreover, the model can be built for multi-domain purposed, in this case, the performance of the approach is become lower because trained data is not applied for different domain [**?**], [9], [18], [19].

The lexicon based approach can be used to identify the polarity of the sentence, the lexicon based can be built manually which requires lots of effort or automatically which is fast and less human effort required [20]–[22]. The main challenges of sentiment analysis research are most of current approaches is used for English language. However, with growth of the Internet, people share their opinions and comments in different languages. Therefore, sentiment analysis is valuable for other languages such as Persian.

It is a difficult task for SA to understand the text structure for Persian language. For example, the following sentence, من کَارگردَانی فیلم دوست ندَارم ولی بَازیهَا بسیَار خوب بود ("I do not like the movie direction, but I really like the acting"), contains positive and negative sentiment about the movie. Most of the current approaches are not able to detect the overall polarity of the sentence, the traditional system is unable to understand the different sentiment expressed in the sentence. Additionally, the presence of negation can flip the polarity from positive into negative. In addition, The online users can discuss various topics. For example, if they are talking about one product [23], [24]. Alternatively, they can discuss about other product. Additionally, they can discuss about different features of the product. شَارژ موبَایل اپل خیلی خوب بود ولی من شَارژ لپ تَاپ اپل دوست ندَارم, ("The charger of the Apple mobile is very good, but I do not like the charger for Apple laptop"). The sentence aspect is charger and the sentiment of the sentence for mobile is positive however, the second part of the sentence has a negative sentiment about the laptop [25]–[31].

There are lots of meaningful information is behind the text. For example, فیلم بدِ نبود ولِ اگر بَازیگرَان دیگری استفَاده کرده بودند مِتوَانست بهتر بَاشد, "It was not a bad movie, but if they are using different cast it could be a better movie". The natural language processing is used to understand the text. The use of keywords, punctuation and frequency of words are very useful to understand the text. However, the increase of web content in different languages make the English algorithm inefficient. In the order to overcome these challenges in this thesis we focus on Persian SA. Most of the previous studies have only focused on English SA and therefore most of the tools and resources are available in English. Hence the tools and research, in the most of the other languages such as Persian, are comparatively less developed [32]–[36]. Persian SA has the following main challenges:

Lack of tools and resources: To the best of our knowledge, there is no valuable tools nor any online lexicon available for Persian language.

Utilizing many informal words: There are lots of informal words available in Persian language. Specifically, when online users share their opinion and comments online, they are using more of these informal words and phrases.

Lack of comprehensive approaches: There is no valuable comprehensive approach available in Persian language. Most of the current approaches utilize the available approaches of English and translate the Persian dataset into English.

Therefore, In this paper, the supervised machine learning approach used to identify the sentiment classification of Persian dataset and TF-IDF has been used with n-gram features, this information is more effective than frequency evaluated by our experiment results and transition point is used on n-gram features. The SVM, Naïve Bayes, MLP and CNN classifiers used to evaluate the performance of the proposed approach. This paper is organised as follows: Section II is related work of English and Persian sentiment analysis approaches, Section III is present methodology of supervised approach for Persian movie reviews, Section IV provides results and discussion and Section V is conclusion of the paper

II. Related Work

With the growth of Internet, there are various types of online reviews are available such as product and movie reviews, these reviews are very helpful for companies. Sentiment analysis aim to classify and detect vital information from these reviews. In this section, the current Persian and English sentiment analysis has been summarised.

A. English

Agarwal et al. [37] proposed concept parser based on the dependency relationship of words in the sentence, after filtering the words Mrmr feature selection has been used to train feature for machine learning classifiers, the proposed model used to classify the document into positive and negative. The movie reviews have been used to evaluate the performance of the approach, the overall accuracy is 88.9%.

Da Rocha et al. [38] proposed a hybrid approach to identify the sentiment of the sentence, it consists of two phase feature selection method, the main step is pre-processing, then use of feature selection method, in order to evaluate the performance of the approach the movie review dataset has been used and SVM classifier has been trained. The overall performance of the approach is 81%.

Dashmukh et al. [39] proposed a multidomain lexicon to predict the polarity of the sentence, the proposed algorithm used entropy with modified quantity instead of traditional entropy algorithm, the dataset contains different domains of product reviews, there are multiple experiments have been carried out to analyse the proposed approach, the overall accuracy of the proposed approach is 81.66% Sentiment analysis is used to classify document and sentence.

Hung et al. [40] proposed an approach to classify the document based on the polarity, the document has been transferred into set of features such as unigram, bigram, trigram and part-of-speech tag and SentiWordNet has been used to assign polarity to features, the result of experiment display the bigram and trigram is performed better in comparison with other features.

Poria et al. [41] proposed approach to extract features from short texts, based on the inner layer of convolutional neural network, the extracted features of multimodal sentiment analysis have been used and then the feature vectors of text, audio and video has been combined, the proposed approach improved the performance by 14% and the overall accuracy of the approach is 88.60%.

B. Persian

Saraee and Bagheri et al. [42] proposed a model sentiment classification of reviews documents in Persian, the model is based on pre-processing and feature selection and Naïve Bayes has been used to evaluate the performance of the approach, the mobile reviews have been used and overall performance of the approach is 81.02%.

Ebrahimi et al. [43] proposed an approach to detect polarity in Persian sentences,the linguistic features have been extracted and mutual information has been used to identify the polarity of the sentence using SentiStrength of words. To evaluate the performance of the approach online reviews has been used, the result of SentiStrength has been compared with SentiWordNet and the overall accuracy is 80%.

Bagheri et al. [44] proposed sentiment classification approach for document level in Persian language, there are various feature types such as document frequency, term frequency has been used, and Naïve Bayes has been trained, the mobile reviews have been used to evaluate the performance of the approach, the overall accuracy of the approach is 0.907%.

III. Methodology

Pre-processing: The pre-processing step consists of cleaning text from unnecessary data, for example فيلم بسيّار خوبي بود)It is really goodddd movie), the normalisation of data used to clean the data, removing stop-words is useful to improve the performance of the dataset, for example some words can be removed because these words do not contain any polarity such as به, از) ,From, to, etc.).

N-grams: The n-gram is sequence of n items in the text. The n-gram size one is called unigram, size two is bigram and size three is trigram. For example, “I went to school”. The unigram is “I”, “went”, “to”, “school”, the bigram is “I went”, “went to”, “to school” and trigram is “I went to”, “went to school”.

TF-IDF: TF-IDF is stand for term frequency-inverse document frequency and it is statistical measure to calculate the importance of words in document, the importance of the words will be calculating with number of times which words appear in the document [45].

The TF-IDF is consist of two parts, first part calculates the term frequency which is number of word appear in document divided by total number of words in the document and second term is inverse document frequency calculate the logarithm of number document divided by number of document which the term appears in them.

$$Frequency = \frac{Numberofdocument}{totalnumberofterm} \tag{1}$$

IDF: Inverse document frequency is measure the importance of the words in the whole documents, the following equation is used to weight the term.

$$Frequency = \frac{Numberofdocument}{numberofdocumenttermappear} \tag{2}$$

Transition Point: The transition point is a frequency values to split the words in document into two different sets low and high. The transition point used to identify the medium frequency which is closely related to the conceptual content of the document.

IV. Results and Discussions

Persian movie reviews Dataset: In order to evaluate the performance of the proposed framework the Persian movie reviews dataset has been used, the movie review contains 1000 positive and 1000 negative. The movie review has been collected from www.caffecinema.com and www.cinematicket.org, the movie reviews consist of movie reviews from 2014 to 2016.

In order to evaluate the performance of the proposed approach, the dataset is pre-processed, after normalisation and removing stop-words. The features such as unigram, bigram and trigram is extracted, then frequency, tf-idf and transition point used to train the SVM, Naïve Bayes, MLP and CNN. The Table 1 shows the result on TF-IDF and frequency (Freq) using different features such as unigram, bigram and trigram. The 10-fold cross validation used to train SVM, Naïve Bayes, MLP and CNN. The result shows the TF-IDF received better results in comparison with frequency.

In order to evaluate the performance of classification techniques, the following four evaluation metrics are used: precision, recall, f-measure and accuracy.

$$Precision = \frac{TP}{TP+FP} \tag{3}$$

$$Recall = \frac{TP}{TP+FN} \tag{4}$$

$$F\_measure = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{5}$$

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{6}$$

The table V, VI, VII and VIII show the result for IDF transition point using n-gram features. There are SVM, Naïve Bayes, MLP and CNN classifiers is trained. The trigram features received better performance and unigram and bigram

V. Conclusion

In this paper, a supervised machine learning approach has been proposed for Persian using unigram feature with TF-IDF and frequency to identify the overall polarity of the Persian movie reviews. The experimental result show the TF-IDF is performed more effectively in comparison with frequency. The IDF transition point outperformed TF-IDF and frequency. In our future work, we will developed an approach to identify the overall polarity of multilingual sentences in Arabic, Persian and English.

Funding: Amir Hussain was supported by the UK Engineering and Physical Sciences Research Council (EPSRC)-Grants Ref. EP/M026981/1, EP/T021063/1, EP/T024917/1.

References

[1] M. Gogate, K. Dashtipour, A. Adeel, and A. Hussain, “Cochleanet: A robust language-independent audio-visual model for real-time speech enhancement,” *Information Fusion*, vol. 63, pp. 273–285, 2020.

[2] A. Adeel, M. Gogate, S. Farooq, C. Ieracitano, K. Dashtipour, H. Larijani, and A. Hussain, “A survey on the role of wireless sensor networks and iot in disaster management,” in *Geological disaster monitoring based on sensor networks*. Springer, 2019, pp. 57–66.

TABLE I
COMPARISON TF-IDF AND FREQUENCY NAÏVE BAYES

Features	**Precision**	**Recall**	**Fscore**	**Accuracy**	**Features**	**Precision**	**Recall**	**Fscore**	**Accuracy**
Unigram Freq	0.65	0.66	0.66	66.21	Unigram Bigram Freq	0.79	0.78	0.79	79.3
Bigram Freq	0.66	0.65	0.66	66.94	Bigram Trigram Freq	0.80	0.80	0.80	81.96
Trigram Freq	0.67	0.67	0.67	67.85	Unigram Trigram Freq	0.82	0.82	0.82	82.59
Unigram TF-IDF	0.69	0.69	0.69	69.62	Unigram Bigram TF-IDF	0.83	0.83	0.83	83.95
Bigram TF-IDF	0.70	0.71	0.72	71.39	Unigram Trigram TF-IDF	0.84	0.83	0.84	84.81
Trigram TF-IDF	0.75	0.75	0.75	75.64	Bigram Trigram TF-IDF	0.85	0.84	0.85	85.91

TABLE II
COMPARISON TF-IDF AND FREQUENCY SVM

Features	**Precision**	**Recall**	**Fscore**	**Accuracy**	**Features**	**Precision**	**Recall**	**Fscore**	**Accuracy**
Unigram Freq	0.69	0.68	0.69	69.8	Unigram Bigram Freq	0.81	0.80	0.81	81.46
Bigram Freq	0.71	0.70	0.71	71.09	Bigram Trigram Freq	0.82	0.81	0.82	82.94
Trigram Freq	072	0.71	0.72	72.52	Unigram Trigram Freq	0.83	0.82	0.83	83.63
Unigram TF-IDF	0.72	0.72	0.72	72.51	Unigram Bigram TF-IDF	0.84	0.83	0.84	84.59
Bigram TF-IDF	0.75	0.74	0.75	75.94	Unigram Trigram TF-IDF	0.86	0.85	0.86	86.93
Trigram TF-IDF	0.78	0.77	0.78	78.67	Bigram Trigram TF-IDF	0.89	0.88	0.89	89.76

TABLE III
COMPARISON TF-IDF AND FREQUENCY MLP

Features	**Precision**	**Recall**	**Fscore**	**Accuracy**	**Features**	**Precision**	**Recall**	**Fscore**	**Accuracy**
Unigram Freq	0.75	0.75	0.75	75.69	Unigram Bigram Freq	0.84	0.83	0.84	84.69
Bigram Freq	0.77	0.76	0.77	77.92	Bigram Trigram Freq	0.85	0.86	0.85	85.67
Trigram Freq	0.77	0.78	0.79	77.5	Unigram Trigram Freq	0.86	0.85	0.86	86.43
Unigram TF-IDF	0.79	0.79	079	79.53	Unigram Bigram TF-IDF	0.88	0.84	0.86	86.59
Bigram TF-IDF	0.80	0.79	0.80	79.82	Unigram Trigram TF-IDF	0.89	0.89	0.89	89.61
Trigram TF-IDF	0.81	0.81	0.81	81.79	Bigram Trigram TF-IDF	0.91	0.91	0.91	91.84

TABLE IV
COMPARISON TF-IDF AND FREQUENCY CNN

Features	Precision	Recall	Fscore	Accuracy	Features	Precision	Recall	Fscore	Accuracy
Unigram Freq	0.86	0.85	0.86	86.53	Unigram Bigram Freq	0.90	0.90	0.90	90.9
Bigram Freq	0.87	0.86	0.87	87.16	Bigram Trigram Freq	0.90	0.89	0.90	90.67
Trigram Freq	0.87	0.87	0.87	87.54	Unigram Trigram Freq	0.91	0.91	0.91	91.84
Unigram TF-IDF	0.88	0.88	0.89	88.48	Unigram Bigram TF-IDF	0.92	0.92	0.92	92.37
Bigram TF-IDF	0.89	0.88	0.89	89.79	Unigram Trigram TF-IDF	0.94	0.93	0.94	94.4
Trigram TF-IDF	0.89	0.89	0.89	89.8	Bigram Trigram TF-IDF	0.95	0.95	0.95	95.63

TABLE V
COMPARISON TRANSITION POINT NAIVE BAYES

Features	Precision	Recall	Fscore	Accuracy	Features	Precision	Recall	Fscore	Accuracy
Unigram Freq	0.84	0.83	0.84	84.56	Unigram Bigram Freq	0.89	0.88	0.89	89.63
Bigram Freq	0.85	0.85	0.85	85.81	Bigram Trigram Freq	0.89	0.89	0.89	89.91
Trigram Freq	0.86	0.86	0.86	86.42	Unigram Trigram Freq	0.91	0.91	0.91	91.51
Unigram TF-IDF	0.87	0.87	0.87	87.73	Unigram Bigram TF-IDF	0.92	0.92	0.92	92.09
Bigram TF-IDF	0.88	0.88	0.88	89.56	Unigram Trigram TF-IDF	0.92	0.92	0.92	92.56
Trigram TF-IDF	0.89	0.88	0.89	89.03	Bigram Trigram TF-IDF	0.93	0.92	0.93	93.87

TABLE VI
COMPARISON TRANSITION POINT SVM

Features	Precision	Recall	Fscore	Accuracy	Features	Precision	Recall	Fscore	Accuracy
Unigram Freq	0.73	0.72	0.73	73.91	Unigram Bigram Freq	0.85	0.84	0.85	85.96
Bigram Freq	0.74	0.73	0.74	74.56	Bigram Trigram Freq	0.86	0.85	0.86	85.41
Trigram Freq	0.75	0.74	0.75	75.63	Unigram Trigram Freq	0.88	0.87	0.88	88.09
Unigram TF-IDF	0.74	0.75	0.75	75.1	Unigram Bigram TF-IDF	0.89	0.88	0.89	89.61
Bigram TF-IDF	0.75	0.75	0.75	75.86	Unigram Trigram TF-IDF	0.91	0.90	0.91	91.68
Trigram TF-IDF	0.78	0.78	0.78	78.86	Bigram Trigram TF-IDF	0.92	0.91	0.92	92.31

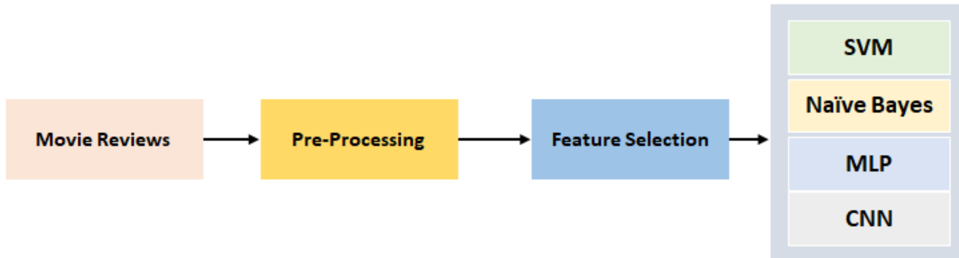

Fig. 1. Proposed Framework for Persian Sentiment Analysis

TABLE VII
COMPARISON TRANSITION POINT MLP

Features	Precision	Recall	Fscore	Accuracy	Features	Precision	Recall	Fscore	Accuracy
Unigram Freq	0.79	0.78	0.79	79.20	Unigram Bigram Freq	0.82	0.82	0.82	82
Bigram Freq	0.80	0.80	0.80	80.47	Bigram Trigram Freq	0.84	0.84	0.84	84.63
Trigram Freq	0.80	0.80	0.80	80.53	Unigram Trigram Freq	0.84	0.84	0.84	84.5
Unigram TF-IDF	0.80	0.80	0.80	80.86	Unigram Bigram TF-IDF	0.84	0.84	0.84	84.67
Bigram TF-IDF	0.80	0.80	0.80	80.91	Unigram Trigram TF-IDF	0.84	0.84	0.84	84.93
Trigram TF-IDF	0.80	0.80	0.80	80.91	Bigram Trigram TF-IDF	0.85	0.86	0.85	85.62

[3] M. Gogate, K. Dashtipour, P. Bell, and A. Hussain, "Deep neural network driven binaural audio visual speech separation," in *2020 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2020, pp. 1–7.

[4] M. Gogate, A. Adeel, K. Dashtipour, P. Derleth, and A. Hussain, "Av speech enhancement challenge using a real noisy corpus," *arXiv preprint arXiv:1910.00424*, 2019.

[5] R. Ahmed, K. Dashtipour, M. Gogate, A. Raza, R. Zhang, K. Huang, A. Hawalah, A. Adeel, and A. Hussain, "Offline arabic handwriting recognition using deep machine learning: A review of recent advances," in *International Conference on Brain Inspired Cognitive Systems*. Springer, 2019, pp. 457–468.

[6] K. Dashtipour, S. Poria, A. Hussain, E. Cambria, A. Y. Hawalah, A. Gelbukh, and Q. Zhou, "Multilingual sentiment analysis: state of the art and independent comparison of techniques," *Cognitive computation*, vol. 8, no. 4, pp. 757–771, 2016.

[7] K. Dashtipour, A. Hussain, Q. Zhou, A. Gelbukh, A. Y. Hawalah, and E. Cambria, "Persent: a freely available persian sentiment lexicon," in *International Conference on Brain Inspired Cognitive Systems*. Springer, 2016, pp. 310–320.

[8] K. Dashtipour, M. Gogate, J. Li, F. Jiang, B. Kong, and A. Hussain, "A hybrid persian sentiment analysis framework: Integrating dependency grammar based rules and deep neural networks," *Neurocomputing*, vol. 380, pp. 1–10, 2020.

[9] K. Dashtipour, M. Gogate, A. Adeel, C. Ieracitano, H. Larijani, and A. Hussain, "Exploiting deep learning for persian sentiment analysis," in *International conference on brain inspired cognitive systems*. Springer, 2018, pp. 597–604.

[10] S. Nisar, M. Tariq, A. Adeel, M. Gogate, and A. Hussain, "Cognitively inspired feature extraction and speech recognition for automated hearing loss testing," *Cognitive Computation*, vol. 11, no. 4, pp. 489–502, 2019.

[11] A. Adeel, M. Gogate, and A. Hussain, "Contextual deep learning-based audio-visual switching for speech enhancement in real-world environments," *Information Fusion*, vol. 59, pp. 163–170, 2020.

[12] C. Ieracitano, A. Adeel, M. Gogate, K. Dashtipour, F. C. Morabito, H. Larijani, A. Raza, and A. Hussain, "Statistical analysis driven optimized deep learning system for intrusion detection," in *International Conference on Brain Inspired Cognitive Systems*. Springer, 2018, pp. 759–769.

[13] A. Adeel, M. Gogate, A. Hussain, and W. M. Whitmer, "Lip-reading driven deep learning approach for speech enhancement," *IEEE Transactions on Emerging Topics in Computational Intelligence*, 2019.

[14] F. Jiang, B. Kong, J. Li, K. Dashtipour, and M. Gogate, "Robust visual saliency optimization based on bidirectional markov chains," *Cognitive Computation*, pp. 1–12, 2020.

[15] M. Gogate, A. Adeel, R. Marxer, J. Barker, and A. Hussain, "Dnn driven speaker independent audio-visual mask estimation for speech separation," *arXiv preprint arXiv:1808.00060*, 2018.

[16] M. Gogate, A. Adeel, and A. Hussain, "Deep learning driven multimodal fusion for automated deception detection," in *2017 IEEE Symposium Series on Computational Intelligence (SSCI)*. IEEE, 2017, pp. 1–6.

[17] E. Cambria, Y. Li, F. Z. Xing, S. Poria, and K. Kwok, "Senticnet 6: Ensemble application of symbolic and subsymbolic ai for sentiment analysis," in *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, 2020, pp. 105–114.

[18] R. Ahmed, M. Gogate, A. Tahir, K. Dashtipour, B. Al-Tamimi, A. Hawalah, M. A. El-Affendi, and A. Hussain, "Deep neural network-based contextual recognition of arabic handwritten scripts," *Entropy*, vol. 23, no. 3, p. 340, 2021.

[19] K. Dashtipour, M. Gogate, A. Adeel, A. Hussain, A. Alqarafi, and T. Durrani, "A comparative study of persian sentiment analysis based on different feature combinations," in *International Conference in Commu-*

TABLE VIII
COMPARISON TRANSITION POINT CNN

Features	Precision	Recall	Fscore	Accuracy	Features	Precision	Recall	Fscore	Accuracy
Unigram Freq	0.89	0.89	0.89	89.65	Unigram Bigram Freq	0.93	0.93	0.93	93
Bigram Freq	0.90	0.91	0.92	92.53	Bigram Trigram Freq	0.94	0.94	0.94	94.63
Trigram Freq	0.92	0.92	0.92	92.67	Unigram Trigram Freq	0.95	0.95	0.95	95.67
Unigram TF-IDF	0.92	0.92	0.92	92.78	Unigram Bigram TF-IDF	0.96	0.96	0.96	96.64
Bigram TF-IDF	0.92	0.92	0.92	92.92	Unigram Trigram TF-IDF	0.96	0.95	0.96	96.63
Trigram TF-IDF	0.92	0.92	0.92	92.97	Bigram Trigram TF-IDF	0.97	0.97	0.97	97.01

nications, Signal Processing, and Systems. Springer, 2017, pp. 2288–2294.

[20] K. Dashtipour, C. Ieracitano, F. C. Morabito, A. Raza, and A. Hussain, "An ensemble based classification approach for persian sentiment analysis," in *Progresses in Artificial Intelligence and Neural Systems*. Springer, 2021, pp. 207–215.

[21] K. Dashtipour, M. Gogate, A. Adeel, H. Larijani, and A. Hussain, "Sentiment analysis of persian movie reviews using deep learning," *Entropy*, vol. 23, no. 5, p. 596, 2021.

[22] K. Dashtipour, W. Taylor, S. Ansari, M. Gogate, A. Zahid, Y. Sambo, A. Hussain, Q. Abbasi, and M. Imran, "Public perception towards fifth generation of cellular networks (5g) on social media," *Frontiers in Big Data*, 2021.

[23] K. Dashtipour, A. Hussain, and A. Gelbukh, "Adaptation of sentiment analysis techniques to persian language," in *International Conference on Computational Linguistics and Intelligent Text Processing*. Springer, 2017, pp. 129–140.

[24] M. Ozturk, M. Gogate, O. Onireti, A. Adeel, A. Hussain, and M. A. Imran, "A novel deep learning driven, low-cost mobility prediction approach for 5g cellular networks: The case of the control/data separation architecture (cdsa)," *Neurocomputing*, vol. 358, pp. 479–489, 2019.

[25] A. S. Alqarafi, A. Adeel, M. Gogate, K. Dashitpour, A. Hussain, and T. Durrani, "Toward's arabic multi-modal sentiment analysis," in *International Conference in Communications, Signal Processing, and Systems*. Springer, 2017, pp. 2378–2386.

[26] I. O. Hussien, K. Dashtipour, and A. Hussain, "Comparison of sentiment analysis approaches using modern arabic and sudanese dialect," in *International Conference on Brain Inspired Cognitive Systems*. Springer, 2018, pp. 615–624.

[27] W. Taylor, Q. H. Abbasi, K. Dashtipour, S. Ansari, S. A. Shah, A. Khalid, and M. A. Imran, "A review of the state of the art in non-contact sensing for covid-19," *Sensors*, vol. 20, no. 19, p. 5665, 2020.

[28] S. M. Asad, K. Dashtipour, S. Hussain, Q. H. Abbasi, and M. A. Imran, "Travelers-tracing and mobility profiling using machine learning in railway systems," in *2020 International Conference on UK-China Emerging Technologies (UCET)*. IEEE, 2020, pp. 1–4.

[29] W. Taylor, S. A. Shah, K. Dashtipour, A. Zahid, Q. H. Abbasi, and M. A. Imran, "An intelligent non-invasive real-time human activity recognition system for next-generation healthcare," *Sensors*, vol. 20, no. 9, p. 2653, 2020.

[30] Z. Yu, P. Machado, A. Zahid, A. M. Abdulghani, K. Dashtipour, H. Heidari, M. A. Imran, and Q. H. Abbasi, "Energy and performance trade-off optimization in heterogeneous computing via reinforcement learning," *Electronics*, vol. 9, no. 11, p. 1812, 2020.

[31] K. Dashtipour, A. Raza, A. Gelbukh, R. Zhang, E. Cambria, and A. Hussain, "Persent 2.0: Persian sentiment lexicon enriched with domain-specific words," in *International Conference on Brain Inspired Cognitive Systems*. Springer, 2019, pp. 497–509.

[32] K. Dashtipour, M. Gogate, E. Cambria, and A. Hussain, "A novel context-aware multimodal framework for persian sentiment analysis," *arXiv preprint arXiv:2103.02636*, 2021.

[33] K. Dashtipour, M. Gogate, A. Adeel, A. Algarafi, N. Howard, and A. Hussain, "Persian named entity recognition," in *2017 IEEE 16th International Conference on Cognitive Informatics & Cognitive Computing (ICCI* CC)*. IEEE, 2017, pp. 79–83.

[34] M. Gogate, A. Hussain, and K. Huang, "Random features and random neurons for brain-inspired big data analytics," in *2019 International Conference on Data Mining Workshops (ICDMW)*. IEEE, 2019, pp. 522–529.

[35] A. R. Gepperth, T. Hecht, and M. Gogate, "A generative learning approach to sensor fusion and change detection," *Cognitive Computation*, vol. 8, no. 5, pp. 806–817, 2016.

[36] A. Churcher, R. Ullah, J. Ahmad, F. Masood, M. Gogate, F. Alqahtani, B. Nour, W. J. Buchanan *et al.*, "An experimental analysis of attack classification using machine learning in iot networks," *Sensors*, vol. 21, no. 2, p. 446, 2021.

[37] B. Agarwal, S. Poria, N. Mittal, A. Gelbukh, and A. Hussain, "Concept-level sentiment analysis with dependency-based semantic parsing: a novel approach," *Cognitive Computation*, vol. 7, no. 4, pp. 487–499, 2015.

[38] R. S. da Rocha, L. Forero, H. de Mello, M. Kohler, and M. Vellasco, "Polarity classification on web-based reviews using support vector machine," in *2016 IEEE Latin American Conference on Computational Intelligence (LA-CCI)*. IEEE, 2016, pp. 1–6.

[39] J. S. Deshmukh and A. K. Tripathy, "Entropy based classifier for cross-domain opinion mining," *Applied computing and informatics*, vol. 14, no. 1, pp. 55–64, 2018.

[40] L. P. Hung and R. Alfred, "A performance comparison of feature extraction methods for sentiment analysis," in *Asian Conference on Intelligent Information and Database Systems*. Springer, 2017, pp. 379–390.

[41] S. Poria, E. Cambria, and A. Gelbukh, "Deep convolutional neural network textual features and multiple kernel learning for utterance-level multimodal sentiment analysis," in *Proceedings of the 2015 conference on empirical methods in natural language processing*, 2015, pp. 2539–2544.

[42] M. Saraee and A. Bagheri, "Feature selection methods in persian sentiment analysis," in *International conference on application of natural language to information systems*. Springer, 2013, pp. 303–308.

[43] F. Ebrahimi Rashed and N. Abdolvand, "A supervised method for constructing sentiment lexicon in persian language," *Journal of Computer & Robotics*, vol. 10, no. 1, pp. 11–19, 2017.

[44] A. Bagheri and M. Saraee, "Persian sentiment analyzer: a framework based on a novel feature selection method," *arXiv preprint arXiv:1412.8079*, 2014.

[45] K. Chen, Z. Zhang, J. Long, and H. Zhang, "Turning from tf-idf to

tf-igm for term weighting in text classification," *Expert Systems with Applications*, vol. 66, pp. 245–260, 2016.