IJICTR
International Journal of Information & Communication Technology Research

Volume 16- Number 3 – 2024 (1 -8)

ParsiAzma Challenges on Persian Text Analysis in Social Media

Mojgan Farhoodi*
ICT Research Institute
Tehran, Iran
farhoodi@itrc.ac.ir

Maryam Mahmoudi
ICT Research Institute
Tehran, Iran
mahmoudy@itrc.ac.ir

Mohammad Hadi Bokaei
ICT Research Institute
Tehran, Iran
mh.bokaei@itrc.ac.ir

Received: 16 December 2023 – Revised: 25 February 2024 - Accepted: 13 April 2024

Abstract—**The ParsiAzma[2] challenges in 2023 focused on Improving Persian text analysis in social media. We designed four shared tasks: stance detection, sentiment analysis, emotion detection, and claim detection in social media posts. The goal of these challenges was to bring together various teams to develop the best models for these challenges and to establish a standard test platform for future Persian language research. A total of 28 teams participated, competing to solve the specified tasks. The most effective models in all shared tasks utilized the BERT model. Text embedding was first obtained using a BERT[3]-based model, followed by final predictions with either an MLP[4] or CNN[5]. Additionally, several meta-classifiers were developed as fusion models to leverage the strengths of individual models. The best results based on accuracy criteria for the four challenges—stance detection, sentiment analysis, emotion recognition, and claim detection—were 0.67, 0.67, 0.45, and 0.56, respectively. These results indicate that emotion detection has lower accuracy than the other three tasks, highlighting its complexity.**

Keywords: stance detection, claim detection, sentiment analysis, emotion detection, social media, competition

Article type: Research Article

I. INTRODUCTION

In recent years, the rapid growth of mass media such as the Internet, social networks, and smartphones has made it easy to produce diverse content and quickly access it to different users around the world. This has led to the creation of new opportunities and challenges for the users of these media.

Social media provide rich information on human interaction and collective behavior, so they attract the attention of many disciplines including sociology, business, psychology, politics, computer science, economics and other cultural aspects of societies. Therefore, text analysis can be of great help in understanding the user's needs in order to fulfill them.

In this regard, we have designed some challenges, which include four shared tasks in the direction of Persian text analysis in social networks. The titles of these four challenges are:

- Stance Detection in Social Media Posts

* Corresponding Author
[2] https://parsiazma.ir/
[3] Bidirectional Encoder Representations from Transformers
[4] Multi-Layer Perceptron
[5] Convolutional Neural Network

[DOI: 10.61186/itrc.16.3.1]

- Sentiment Analysis of Social Media Post
- Emotion Detection from Social Media Posts
- Claim Detection in Social Media Posts

One of the most important reasons for choosing these four issues is their use in important applications such as analyzing users' opinions in social networks, which is very important for many businesses. Also, all the selected topics are used in validating the content and detecting fake news, which, unfortunately, we are faced with the wide spread of these news in social media today.

The goal of these challenges was to bring researchers together to provide the best solutions for these defined shared tasks. In each of the challenges, train data was created and provided to the participants, or if there was a Persian dataset, the download link was given to them. But models improvement phase and also evaluating the proposed models, we tried to create a separate dataset. The training and test set follow the same annotation schema. Participants were allowed to use any public data and resources in addition to the official training data of the shared task in the process of making their models. In this case, they must thoroughly describe those resources and the way they used them.

In the rest of this paper, we will provide a detailed description of each of the shared tasks of stance detection, sentiment analysis, emotion detection, and claim detection, respectively, in sections 2 to 5 along with the analysis of their results. Finally, Section 6 outlines the findings of paper.

II. Stance Detection Task

Stance detection is the task of automatically classification the attitude expressed in a text towards a proposition or a target, which is usually expressed as agree, disagree or neutral. The target may be a person, an organization, a government policy, a movement, a product, and so on [1]. Stance detection is an important task as it is usually employed as a primary step in other tasks such as fake news detection, claim validation, or argument search [2].

Stance detection is well-studied task in English [3], [4], [5] and [6] and some other languages like Russian [7], Indian [8], Italian [9], Zulu [10]. However, few researches have been done in Persian language in this field [11], [12], [13] and [14]. [11] used LSTM, [12] applied transfer learning and data augmentation and [13] used data augmentation and multi-classifier fusion for Persian stance detection. The first three researches have worked on the dataset introduced in [11] and the last research used dataset described in [14].

A. Task Description

The main purpose of this shared task is participating systems have to predict whether the author of the post and the author of the reply post agree on the subject of the main post or not. In other words, whether the stance of both people toward to the issue raised in the post is the same or not. So, the input of the system is a tweet and its output will be done one of these annotates: Support, Against and Neither.

B. Dataset

According to the definition of the task, there is only one dataset in Persian language which introduced in [14] that created in ParsiAzma. This dataset collected from Twitter. In this shared task, we used this dataset and split it to three parts: Train, Development and Test data. Table 1 represents the total number of samples in the training, development and test data.

TABLE I. Label Distribution of Training, Development and Test Data

	Support	Against	Neither	Total
Train	2976	2108	766	5850
Improvement	304	227	88	619
Test	778	557	213	1548

This dataset was labeled by three persons and the inter-annotator agreement is 0.61 with Cohen's Kappa measure.

C. Participating Systems

9 teams have participated in this task. Table 2 briefly shows the models and features used by the first three participants.

TABLE II. The Models' Information Used By The First Three Participants

Team	Model	Word Embedding	Features
StateOfTheArt	Meta Classifier (PersPoliX-CNN[6], PersPoliX-Siamese[7])	BERT (PersPoliX)	-
MPERoL	BERT	BERT	Segment embedding
Amin	BERT	BERT Tokenizer	Segment embedding

D. Results

We calculated different metrics, but used F-score as a reference metric to compare models. Table 3 shows the results of the models of all participating teams.

TABLE III. The Results Of Presented Models

	Team	Precision	Recall	F-score
1	StateOfTheArt	0.67	0.67	0.67
2	MPERoL	0.63	0.61	0.62
3	Amin	0.58	0.62	0.58
4	Negaresh	0.61	0.56	0.57
5	Allameh Team	0.53	0.52	0.52
6	storm	0.48	0.45	0.45
7	Allameh	0.59	0.41	0.45
8	GHM_NLP_IUST	0.45	0.44	0.44
9	Ackerman	0.33	0.33	0.31

E. Discussion

In this section, we analyze the performance of the most effective method in comparison to alternative techniques, focusing on tweets where it uniquely identified the types of claims present, which others failed to recognize.

First Tweet: "The political deputy of the Revolutionary Guard: 'If we have issues and problems in various fields today, it is definitely because the universities have not been reformed. All advanced

[6] In PersPoliX-CNN architecture, pair sentences input is used.

[7] In PersPoliX-Siamese architecture, the Siamese is not just the head above PersPoliX like MLP. It mitigates the Siamese architecture

[DOI: 10.61186/itrc.16.3.1]

countries have universities in this way and do not have our problems. The difference between the Islamic Republic of Iran and them is in the role and duties of clerics and military personnel.'" **Analysis**: This tweet includes Quote, Causation/Correlation, and Comparison claims. While most methods can detect the Quote and Causation/Correlation claims, only the best method identifies the Comparison claim, demonstrating its advanced understanding of the tweet's content.

Second Tweet: "If even half of Esmail Bakhshi's statements are true, the question arises as to why he was tortured so severely. He was neither accused of espionage, nor of connections with foreigners, nor of involvement in terrorist activities, nor a member of a subversive group, nor any other action that would justify his torture for confessions." **Analysis**: This tweet contains a subtle Trait claim about Esmail Bakhshi. The best method uniquely identifies this claim type, highlighting its ability to discern nuanced aspects of the text.

Third Tweet: "For years, mobile operators have had their hands in people's pockets; their revenues are stored in dark rooms and eventually reach the government. Today, we have decided that these amounts will be returned to the people to uphold the #right_of_people. Not only has the internet not become more expensive, but we have also legislated that the government cannot raise internet prices in the new year." **Analysis**: This tweet features a Rule/Law claim, which the best method successfully identifies by understanding legislative and regulatory contexts.

In conclusion, the most effective method, optimized for Persian tweets, demonstrates superior capability in recognizing diverse and subtle claim types. However, given the complexity of claim detection, there remains potential for further refinement to identify even more hidden claims.

III. SENTIMENT ANALYSIS TASK

Sentiment analysis has gained attention due to its potential applications in understanding public opinion, customer feedback, and social media trends. Various approaches, such as lexicon-based, machine learning, and hybrid methods, have been explored to analyze sentiment in Persian text. Despite the challenges, the research in this area is progressing, with the development of datasets, tools, and resources. However, further advancements are still required to enhance the accuracy and applicability of sentiment analysis in the Persian language.

Nazarizadeh et al in [15] and Rajabi et al. in [16] provide a review on sentiment analysis methods in Persian language and provides a detailed exploration of existing algorithms, approaches and datasets.

Current efforts are increasingly focused on BERT models numerous researchers from various countries have developed their respective language BERT models to evaluate the sentiment analysis task. The Persian BERT model PersBERT scored 88.12 on their distinct sentiment analysis experiment [17].

A. Task Description

In this shared task, we were looking for the analysis of the writer's feelings. The used data in this task is the text of tweets extracted from the Twitter social network, and the goal is to analyze the overall text to detect its emotional polarity; In such a way that the input of the system is a tweet and its output will be one of three emotional categories: Positive, Negative and Neutral based on the feeling of the author of the tweet and at the level of the entire text (not about a topic or specific entity in the text) is determined.

B. Dataset

In this shared task, due to the existence of Persian dataset for sentiment analysis on the Internet, we presented their URL to the participants for training the proposed models. Therefore, we only created validation and test data. Table 4 represents the total number of samples and label distribution of development and test data.

TABLE IV. LABEL DISTRIBUTION OF DEVELOPMENT AND TEST DATA

	Positive	Negative	Neutral	Total
Development	244	229	27	500
Test	461	461	78	1000

C. Participating Systems

9 teams have participated in this task. Table 5 briefly shows the models and features used by the first three participants.

TABLE V. THE MODELDS' INFORMATION USED BY THE FIRST THREE PARTICIPANTS

Team	Model	Word Embedding	Features
StateOfTheArt	Meta Classifier (PersPoliX-MLP, CardiffNLP)	BERT(PersPoliX[8], CrdiffNLP[9])	-
Ackerman	XLMRoberta+ CNN	XLMRoberta	-
MPERoL	ALBERT	ALBERT	Positional embedding

D. Results

We calculated different metrics, but used F-score as a reference metric to compare models. Table 6 shows the results of the models of all participating teams.

TABLE VI. THE RESULTS OF PRESENTED MODELS

	Team	Precision	Recall	F-score
1	StateOfTheArt	0.67	0.72	0.67
2	Ackerman	0.6	0.63.	0.6
3	MPERoL	0.49	0.48	0.47
4	Allameh	0.58	0.57	0.46
5	Sentinator	0.53	0.49	0.45
6	Amin	0.52	0.51	0.43
7	Sharif_Group	0.5	0.46	0.39
8	Borna	0.36	0.37	0.32
9	IUST_NLP_LAB	0.32	0.34	0.32

[8] https://huggingface.co/StateOfTheArtAUT/perspolix-persian-political-tweet-xlm-roberta-large

[9] https://huggingface.co/cardiffnlp/twitter-xlm-roberta-base-sentiment

[DOI: 10.61186/itrc.16.3.1]

E. Discussion

In this section, we analyze the performance of the most effective method compared to other approaches, emphasizing tweets where only the best method accurately identified the sentiment.

Iran-Saudi Agreement Tweet: "The agreement between Iran and Saudi Arabia to resume diplomatic relations between the two countries will have a significant impact on the Middle East and will likely reduce the possibility of conflicts between the two countries and their proxy groups." *Analysis:* This tweet highlights the positive impact of the Iran-Saudi agreement on regional stability. The best method successfully identifies the positive and hopeful sentiment, while other methods likely misinterpret the sentiment due to a focus on conflict-related negative keywords.

Role of Managers Tweet: "A #toxic-manager stabs you in the back. A #selfish-manager holds you back. A #supportive-manager has your back and pushes you forward. More than productivity, great managers invest in people. They seek to #improve your growth and quality of life. A #leader's duty is to #care. #Adam-Grant." *Analysis:* This tweet contrasts various managerial styles and their effects on employees. The best method accurately discerns the overall positive and supportive sentiment, while other methods struggle to balance both positive and negative emotions conveyed in the tweet.

Book by Sayyid Ahmad al-Hasan Tweet: "The book (Bewilderment or The Path to God) by #Sayyid-Ahmad-al-Hasan helps us #escape from the #misguidance and #bewilderment we face in this #world and clearly shows the path to #truth." *Analysis:* The tweet endorses a book aimed at guiding readers away from confusion and toward truth. The best method identifies the positive and motivational sentiment, whereas other methods may misinterpret the sentiment due to the presence of negative keywords like "misguidance" and "bewilderment."

These examples demonstrate that the best method excels in accurately identifying nuanced sentiments in complex tweets, highlighting the limitations of other approaches that may overly focus on negative keywords. This underscores the value of using advanced methods for more precise sentiment analysis.

IV. Emotion Recognition Task

Emotion recognition is a subset of sentiment analysis, aims to extract nuanced emotions from speech, images, or text data.

Nowadays, Transfer Learning and Pre-Trained Language Models Leveraging pre-trained language models such as BERT, GPT-3, and RoBERTa has become a dominant trend. These models are fine-tuned for emotion detection tasks, enabling researchers to achieve state-of-the-art performance without starting from scratch [18], [19] and [20].

With the increasing need for multilingual and cross-cultural emotion analysis, there is a growing emphasis on developing emotion detection models that can effectively handle multiple languages and account for linguistic variations in expressing emotions [20], [21] and [22].

A. Task Description

The goal of this task is to participating systems have to predict of emotions expressed in social media posts, specifically tweets from Twitter. The goal is to classify the emotions of the authors into categories based on Ekman's classification, such as Sad, Fear, Anger, Disgust, Happy, Surprise. The challenge is divided into two sub-challenges, one focused on determining the dominant emotion of the tweet author and the other on identifying all emotions, including minor ones. This can be a fascinating and complex task given the nuances of human emotions and the brevity of tweets.

B. Dataset

In this shared task, since Persian dataset for emotion detection was publicly available, so we provide their URL to the participants for training the proposed models. Therefore, we only created validation and test data. Table 7 and Table 8 represent the total number of samples and label distribution of validation and test data for primary and other emotions respectively.

TABLE VII. Label Distribution of Development and Test Data (Primary Emotion)

	anger	happiness	sadness	surprise	fear	disgust	other	Total
Development	113	221	115	15	12	18	6	500
Test	257	337	244	34	22	85	22	1001

TABLE VIII. Label Distribution of Development and Test Data (Other Emotions)

	anger	happiness	sadness	surprise	fear	disgust
Development	188	237	209	73	43	162
Test	467	457	527	211	126	487

C. Participating Systems

In the emotion shared task, totally 4 teams have participated. Table 9 briefly shows the models and features used by the first three participants.

TABLE IX. The Models' Informaation Used By The First Three Participants

Team	Model	Word Embedding	Features
StateOfTheArt	PersPoliX-MLP (on Single Label[10])	BERT(PersPoliX)	-
EmotiNerds	XLM-RoBERTa+ GRU	XLM-RoBERTa+ GRU	XLM-RoBERTa+ GRU
Ackerman	BERT+attention-BLSTM	BERT	Emoji-keywords-POS tags

D. Results

We calculated different metrics, but used Average F-score as a reference metric to compare models. Table

[10] In this model original multi label dataset converted to single label dataset.

[DOI: 10.61186/itrc.16.3.1]

10 shows the results of the models of all participating teams.

TABLE X. THE RESULTS OF PRESENTED MODELS

Team	Precision	Recall	F1	AV_Fscore
StateOfTheArt	0.45	0.49	0.46	0.55
EmotiNerds	0.49	0.42	0.37	0.5
Ackerman	0.39	0.36	0.33	0.44
IUST_NLP_LAB	0	0.14	0.01	0

E. Discussion

In this section, we analyze the effectiveness of the most accurate method in detecting emotions in tweets compared to other approaches. Our focus is on examples where only the best method successfully identified the complex emotions present in the text, while other methods did not.

First Tweet: "Mr. #Dejkam, the respected Friday prayer leader of #Shiraz, my brother, poverty is not a virtue. All prophets and imams, if they were generous, it was because they were wealthy. They did not use alchemy to be generous. Remember, if poverty enters through the window, faith leaves through the door." **Analysis**: This tweet conveys a nuanced blend of anger, disgust, and sadness. The best method accurately identifies all these emotions, with particular success in detecting disgust, which is often challenging due to its subtlety in the text. Other methods struggle with this complexity, particularly in recognizing disgust.

Second Tweet: "Today, we were under the scorching sun for so long that we got tanned, dear friends. From tomorrow, we'll go back to the beautiful winter season." **Analysis**: This tweet expresses a combination of happiness and sadness. The best method is able to capture both sentiments simultaneously, reflecting the dual nature of the experience described. In contrast, other methods may only detect one of these emotions, failing to grasp the full emotional context.

Third Tweet: "Iran is the only country in the world where its car manufacturers are in trillions of tomans in losses. Not only do they not declare bankruptcy, but the hidden costs of these losses are paid from every Iranian's pocket. Interestingly, for decades, no one is accountable, and no oversight, inspection, or security agency dares to intervene!" **Analysis**: This tweet features a mix of anger, disgust, and sadness. The best method is adept at identifying all these emotions, especially disgust, which is less apparent and often overlooked by other methods.

Conclusion: The results indicate that the best method demonstrates superior capability in understanding and identifying complex emotional nuances in tweets. It is particularly proficient in detecting subtle emotions like disgust, which other methods tend to miss. This highlights the effectiveness of advanced methods in capturing the full spectrum of emotions present in nuanced textual data.

V. CLAIM DETECTION TASK

Research has shown that claim detection is a fundamental task in natural language processing and information retrieval. One approach to claim detection is based on machine learning techniques, where the task is framed as a binary classification problem. Researchers have explored various methods such as support vector machines, decision trees, and deep learning approaches like convolutional neural networks and recurrent neural networks to detect claims in text. Additionally, some studies have focused on leveraging semantic and syntactic features to improve claim detection performance, while others have examined the use of domain-specific knowledge bases and ontologies. Furthermore, there is also ongoing work in leveraging datasets annotated with claim veracity labels to enhance claim detection models. Overall, the related work in claim detection demonstrates a diverse set of approaches and methodologies aimed at addressing this important NLP task.

Duzen et al in [23] provides an overview of this research area and fake news detection, where more than 10 years of research on social media misinformation were reviewed and presents the main methods and data sets used in the literature.

Also Vyas et al in [24] discussed about types of Fraud Detection in Insurance Claim System and its classification based on different machine learning methods and also give the future direction for fraud detection in insurance claim system.

Syafiqah et al in [25] explored the use of deep neural network (DNN) to recognize contradictory research claims in medical literature and evaluated different DNN techniques such as the Global Vectors for Word Representation (GLoVe), bidirectional Long Short-Term Memory (LSTM) and Bidirectional Encoder Representations from Transformers (BERT) to determine the assertion value of a research claim against its clinical question.

A. Task Description

The goal of this challenge is identifying the type of claim in posts published on social networks. In this challenge, we are looking for identifying the claim and determine its type. Sentences can be considered claims that people are trying to determine its truth or falsity, and such a thing requires the examination of various documents and evidence. The data used was collected from the Twitter social network. In order to determine the type of claims of tweets, eleven tags have been defined, and we will define each of these tags below. An important point is that data that is considered a claim can receive more than one label depending on the content.

- Trait: sentences that indicate the features, characteristics, characteristics and capabilities of an entity.
- Action: sentences that indicate the claim of performing an action in the past, present, and very near future.
- Support/Oppose: This type of claim indicates agreement (support), opposition, and protest about a specific action and performance.
- Prediction: This type of sentence expresses actions that are predicted to be performed in the future.
- Quantity: The sentences that are of the Quantity claim type include ranking, ranking, date, statistical numbers and figures related to an entity in the

[DOI: 10.61186/itrc.16.3.1]

current state or changing the amount of the entity in the future state.
- Comparison: sentences with a comparative structure (such as comparative and superlative adjectives and prepositions ("like", "like", "similar", "unlike") and comparative conjunctions such as "but", "but", "while") and phrases with the theme of uniqueness indicate this type of claim.
- Causation/Correlation: sentences that include two events and one of the events is dependent on the occurrence of the second event or happens after the occurrence of the second event, are of the type of Causation/Correlation claim. The events of this type of sentences can be expressed in the form of simple sentences, sentences with a conditional structure or in the form of a group of prepositions.
- Rule/law: This type of claim contains sentences that express the rules and regulations and the permissible and impermissible actions.
- Quote: This type of claim contains sentences that express a quote from a person or a news source.
- Other Claim: Sentences that are considered claims and are not included in the subgroup of claims mentioned in the previous cases are considered as other claims.
- Not Claim: Sentences that are not claims are labeled Not Claim. These sentences can include opinions and personal experiences.

B. Dataset

According to the definition of this activity, there are several datasets in Persian language. Table 11 represents the total number of samples in the training, development and test data.

TABLE XI. LABEL DISTRIBUTION OF DEVELOPMENT AND TEST DATA

	Not Claim	Trait	Action	Support/Opp	Prediction	Quantity	Comparison	Causation/C	Rule/law	Quote	Other Claim
Training	902	689	2105	295	232	437	2116	131	698	131	750
validation	39	37	114	12	12	31	115	4	36	8	45
Test	189	107	341	65	46	73	350	21	133	26	124

C. Participating Systems

At the end, 6 teams participated in this task. Table 12 briefly shows the models and features used by the first three participants.

TABLE XII. THE MODELS' INFORMAATION USED BY THE FIRST THREE PARTICIPANTS

Team	Model	Word Embedding	Features
StateOfTheArt	PersPoliX-MLP	BERT(PersPoliX)	-
MPERoL	BERT	BERT	Positional embedding
Ackerman	BERT	BERT	-

D. Results

We calculated different metrics, but used F-score as a reference metric to compare models. Table 13 shows the results of the models of all participating teams.

TABLE XIII. THE RESULTS OF PRESENTED MODELS

Team	Precision	Recall	F-score
StateOfTheArt	0.56	0.66	0.60
MPERoL	0.57	0.58	0.56
Ackerman	0.59	0.51	0.53
Allameh	0.39	0.52	0.43
Amin	0.31	0.62	0.40
IUST_NLP_LAB	0.25	0.48	0.31

E. Discussion

In this section, we evaluate the performance of the best method relative to other approaches, focusing on its ability to accurately determine the stance of responses to original tweets. The best method excels where others falter, especially in cases where the stance is nuanced or misleading.

First Tweet: "Health Minister: We have entered the seventh wave of #COVID-19. Aeinollah: Thinking that COVID is over is a big mistake. COVID is a virus that will remain for years." Response: "There are countries that have not had proper vaccination, and the immunity of people in these countries is low. Therefore, when the virus exists, it mutates and affects all countries. Based on the statistics, I think our country has also entered the seventh wave of COVID." **Analysis**: The response supports the original tweet's assertion about the ongoing threat of COVID-19. The best method correctly identified this supportive stance, while other methods misinterpreted it as opposing due to the initial negative phrasing.

Second Tweet: "If you want to see if people's religion has increased since the Shah's time, ask your grandparents if there was something called 'Dua-e-Arafa' at that time. But today, thanks to divine grace and the existence of the Islamic Republic of Iran, millions of people in Iran have recited this prayer, which is empowering for the country!" Response: "I myself saw the Shah's time; at the start of Muharram, everyone would dress in black, liquor stores would close, even my father wouldn't let us turn on the TV for a month. Recitations of martyrdom were widespread, and people would mourn day and night, putting black cloth on the doors of their homes." **Analysis**: The response contradicts the original tweet's implication of increased religious practice, showing an opposing stance. The best method accurately identified this opposition, while other methods incorrectly labeled it as supportive.

Third Tweet: "The U.S. has never apologized for the disasters in Hiroshima or Vietnam, let alone the 1953 coup. What is promoted as Madeleine Albright's apology is actually an acceptance of U.S. responsibility and the negative role of the coup in Iran's political development, not an apology." Response: "I really wish there were an apocalypse..." **Analysis**: This response expresses a neutral stance, lacking clear support or opposition. The best method correctly identified the neutral stance, whereas other methods incorrectly categorized it as supportive.

Overall, the best method demonstrates a superior understanding of complex stances by accurately capturing the nuanced relationships between tweets and their responses. This indicates a more sophisticated grasp of context and meaning beyond mere word-level analysis.

[DOI: 10.61186/itrc.16.3.1]

VI. Conclusion

We have described the ParsiAzma-2023 shared tasks: stance detection, claim detection, emotion detection, and sentiment analysis in Persian. A total of 28 models were developed by several participating teams for these tasks. The best performance was achieved by Team StateOfTheArt, with an F1 score of 67% in sentiment analysis, 55% AV_Fscore in emotion recognition, 60% F1 score in claim detection, and 67% F1 score in stance detection. This team used BERT-based models for text embeddings, followed by MLP or CNN for final predictions. Additionally, they utilized the PersPoliX language model.

VII. References

[1] Mohammad, Saif, Svetlana Kiritchenko, Parinaz Sobhani, Xiaodan Zhu, and Colin Cherry. "Semeval-2016 task 6: Detecting stance in tweets." In Proceedings of the 10th international workshop on semantic evaluation (SemEval-2016), pp. 31-41. 2016.

[2] Schiller, Benjamin, Johannes Daxenberger, and Iryna Gurevych. "Stance detection benchmark: How robust is your stance detection?" KI-Künstliche Intelligenz 35, no. 3 (2021): 329-341.

[3] Derczynski, Leon, Kalina Bontcheva, Michal Lukasik, Thierry Declerck, Arno Scharl, Georgi Georgiev, Petya Osenova et al. "Pheme: Computing veracity—the fourth challenge of big social data." In Proceedings of the Extended Semantic Web Conference EU Project Networking session (ESCW-PN). 2015.

[4] Bar-Haim, Roy, Indrajit Bhattacharya, Francesco Dinuzzo, Amrita Saha, and Noam Slonim. "Stance classification of context-dependent claims." In Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers, pp. 251-261. 2017.

[5] Qazvinian, Vahed, Emily Rosengren, Dragomir Radev, and Qiaozhu Mei. "Rumor has it: Identifying misinformation in microblogs." In Proceedings of the 2011 conference on empirical methods in natural language processing, pp. 1589-1599. 2011.

[6] Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. "Attention is all you need." Advances in neural information processing systems 30, 2017

[7] Lozhnikov, Nikita, Leon Derczynski, and Manuel Mazzara. "Stance prediction for russian: data and analysis." In International Conference in Software Engineering for Defence Applications, pp. 176-186. Springer, Cham, 2018.

[8] Swami, S., Khandelwal, A., Singh, V., Akhtar, S., and Shrivastava, M., "An english-hindi code-mixed corpus: Stance annotation and baseline system", "arXiv preprint arXiv:1805.11868". (2018), doi: arXiv:1805.11868v1

[9] Cignarella, Alessandra Teresa, Mirko Lai, Cristina Bosco, Viviana Patti, and Rosso Paolo. "Sardistance@ evalita2020: Overview of the task on stance detection in italian tweets." EVALITA 2020 Seventh Evaluation Campaign of Natural Language Processing and Speech Tools for Italian (2020): 1-10.

[10] Dlamini, G., Bekkouch, I. E. I., Khan, A., and Derczynski, L., "Bridging the Domain Gap for Stance Detection for the Zulu language", in Proceedings of "SAI Intelligent Systems Conference". (2023), doi: 10.1007/978-3-031-16072-1_23

[11] Zarharan, Majid, Samane Ahangar, Fateme Sadat Rezvaninejad, Mahdi Lotfi Bidhendi, Mohammad Taher Pilevar, Behrouz Minaei, and Sauleh Eetemadi. "Persian Stance Classification Data Set." In TTO. 2019.

[12] Nasiri, H., and Analoui, M., "Persian Stance Detection with Transfer Learning and Data Augmentation" in 2022 27th International "Computer Conference". (2022), doi: 10.1109/CSICC55295.2022.9780479

[13] Farhoodi, M., & Toloie Eshlaghy, A. (2024). Persian Stance Detection Based On Multi-Classifier Fusion. Journal of Information and Communication Technology, 59 (59), 1.

[14] Bokaei, M. H., Farhoodi, M., & Shamsi, M. D. (2022). Stance Detection Dataset for Persian Tweets. International Journal of Information & Communication Technology Research (2251-6107), 14(4).

[15] Nazarizadeh, A., Banirostam, T., & Sayyadpour, M. (2022). Sentiment Analysis of Persian Language: Review of Algorithms, Approaches and Datasets. arXiv preprint arXiv:2212.06041.

[16] Rajabi, Z., & Valavi, M. (2021). A survey on sentiment analysis in Persian: A comprehensive system perspective covering challenges and advances in resources and methods. Cognitive Computation, 13(4), 882-902.

[17] Farahani, M., Gharachorloo, M., Farahani, M., & Manthouri, M. (2021). Parsbert: Transformer-based model for persian language understanding. Neural Processing Letters, 53, 3831-3847.

[18] Acheampong, F. A., Nunoo-Mensah, H., & Chen, W. (2021). Transformer models for text-based emotion detection: a review of BERT-based approaches. Artificial Intelligence Review, 54(8), 5789-5829.

[19] Gómez-Cañón, J. S., Gutiérrez-Páez, N., Porcaro, L., Porter, A., Cano, E., Herrera-Boyer, P., ... & Gómez, E. (2023). TROMPA-MER: an open dataset for personalized Music Emotion Recognition. Journal of Intelligent Information Systems, 60(2), 549-570.

[20] Mirzaee, H., Peymanfard, J., Moshtaghin, H. H., & Zeinali, H. (2022). Armanemo: A persian dataset for text-based emotion detection. arXiv preprint arXiv:2207.11808.

[21] Azzi, V., Bianchi, D., Pompili, S., Laghi, F., Gerges, S., Akel, M., ... & Hallit, S. (2022). Emotion regulation and drunkorexia behaviors among Lebanese adults: the indirect effects of positive and negative metacognition. BMC psychiatry, 22(1), 391.

[22] Khodaei, A., Bastanfard, A., Saboohi, H., & Aligholizadeh, H. (2022). Deep emotion detection sentiment analysis of persian literary text.

[23] Zafer, Duzen., Mirela, Riveni., Mehmet, Aktas. (2022). Misinformation Detection in Social Networks: A Systematic Literature Review. doi: 10.1007/978-3-031-10545-6_5

[24] Vyas, S., & Serasiya, S. (2022, February). Fraud Detection in Insurance Claim System: A Review. In 2022 Second International Conference on Artificial Intelligence and Smart Energy (ICAIS) (pp. 922-927). IEEE.

[25] Fatin, Syafiqah, Yazi., Wan-Tze, Vong., Valliappan, Raman., Patrick, Hang, Hui, Then., Mukulraj, J, Lunia. (2021). Towards Automated Detection of Contradictory Research Claims in Medical Literature Using Deep Learning Approach. doi: 10.1109/CAMP51653.2021.9498061

Mojgan Farhoodi received her B.Sc. degree in Software Engineering and her M.Sc. and Ph.D. degree in IT Engineering and IT Management Respectively. She has been working as a researcher at the ICT Research Institute since 2010. Currently, she is head of AI lab and faculty member of the ICT research institute. Her areas of expertise are Information Retrieval, Data Mining, Natural Language Processing and Artificial Intelligence.

Maryam Mahmoudi received her B.Sc. degree in Software Engineering from Azad University and her M.Sc. degree in IT Engineering from Amirkabir University of Technology. She has been working as a researcher at the ICT Research Institute since 2012. Her areas of expertise are Web Mining, Data Mining, Natural Language Processing and Artificial Intelligence.

[DOI: 10.61186/itrc.16.3.1]

Mohammad Hadi Bokaei received the B.Sc. and M.Sc. degrees from Iran University of Science and Technology (2008) and Sharif University of Technology (2011), respectively, and the Ph.D. degree in Artificial Intelligence from Sharif University of Technology in 2015. He is currently an Associate Professor at the Iran Telecommunication Research Center, Tehran, Iran. His research interests include the area of Deep Learning, Machine Learning, Spoken Language Processing and Natural Language Processing.

[DOI: 10.61186/itrc.16.3.1]