Developing an Annotated Persian Dataset from COVID-19 News for Enhanced Fake News Detection

Forough Zahed *
Allameh Tabataba'i University
Tehran, Iran
forooghzhd@gmail.com

Seyedeh Fatemeh Ebrahimi *†
Sharif University of Technology
Tehran, Iran
sfati.ebrahimi@sharif.edu

Mohammad Bahrani
Allameh Tabataba'i University
bahrani@sharif.edu

Alireza Mansouri
ICT Research Institute (ITRC)
amansuri@itrc.ac.ir

Abstract

Fake news detection has recently garnered significant attention among researchers, especially in coincidence with the large amount of misinformation spread during the COVID-19 pandemic. However, the detection of such fake news in the Persian language is still facing challenges due to the lack of suitable dataset.

In our paper, we introduce a Persian dataset specifically curated for fake news detection, encompassing approximately 5000 posts related to COVID-19 and obtained from the Telegram messenger platform. First. we describe the construction and annotation process of the dataset; secondly, we propose a preliminary evaluation of the dataset using several machine learning classifiers and feature extraction methods.

1 Introduction

In recent years, social media platforms such as Telegram have witnessed a surge in popularity due to their facilitation of information acquisition and news sharing on diverse topics. The prevalence of unverified data on these platforms has captured the attention of researchers, prompting significant interest in fake news detection (Kaliyar et al., 2021). Following the outbreak of COVID-19, news related to the disease has proliferated across various channels, with a considerable portion being fake, inaccurate, and misleading (Santus et al., 2021).

The dissemination of such misinformation has eroded trust in medical treatments, vaccines, and preventive measures like social distancing and mask-wearing (Skafle et al., 2022). To combat the spread of misinformation, organizations like Snopes, Politifact, and Fact-Check have made commendable efforts to verify news posts and disseminate accurate information through social media platforms. However, the vast volume of shared news and posts exceeds their capacity for examination and analysis, and thus, there is a pressing need for automated fake news detection mechanisms to alleviate the burden on fact-checkers, by means of modern Natural Language Processing (NLP) technologies (Santus et al., 2021; Wan et al., 2022). Recognizing the reliability of news articles can play a pivotal role in resolving the challenge of fake news detection within social media channels such as Telegram and Twitter, and enhance public awareness across political, social, economic, and medical dimensions (Wang, 2017).

In this research endeavor, we aim at developing a labeled resource specifically designed for training machine learning algorithms in fake news detection in the Persian language. The proliferation of fake news in digital media poses a significant challenge, demanding effective detection methods to preserve the integrity of information dissemination. To address this issue, we have meticulously curated a comprehensive dataset encompassing a diverse range of fake news samples, covering various topics and linguistic nuances unique to the Persian language.

Our dataset will serve as a valuable resource for researchers and practitioners in the field of fake news detection. By providing a standardized and extensive collection of labeled data, we seek to facilitate advancements in the development of robust and accurate detection models tailored for the Persian language. The availability of such a dataset is vital for fostering collaboration, benchmarking different approaches, and promoting the development of more informed strategies to combat the dissemination of misinformation.

Moreover, we conducted a first evaluation of

*Both authors contributed equally to this work
†Corresponding author

machine learning algorithms on our dataset, employing 5-fold cross-validation to ensure robustness and minimize bias. The algorithms we evaluated include Support Vector Machines (SVM), Random Forest, and Multi-Layer Perceptron (MLP) Classifier. Additionally, to explore the impact of different feature representations, we utilized TF-IDF Vectors, FastText Persian Embeddings, and Sentence-Transformer.

The initial evaluation results are promising, providing valuable insights into the performance of these algorithms for fake news detection in the Persian language. In the following sections, we present a detailed analysis of the results, encompassing accuracy, precision, recall, and Macro F1-score for each algorithm and feature representation combination. These results pave the way for further research and optimization of fake news detection techniques tailored to the unique linguistic characteristics of the Persian language. We make our code publicly available1 [1]

2 Related Work

Research in the past decade has highlighted a range of machine learning methods employed for fake news detection on social media. However, the majority of these studies have primarily focused on fake news detection in the English language (Aphiwongsophon and Chongstitvatana, 2018), using several classifiers ranging from Naïve Bayes to neural networks, and testing feature extraction techniques relying on handcrafted linguistic features or on word embedding spaces (Singh et al., 2017; Kumar and Shah, 2018; Vijayaraghavan et al., 2020; Giahanou et al., 2020).

Shin et al. (2018) explored fundamental theories across different disciplines to foster interdisciplinary research on fake news, examining fake news from multiple perspectives. In Shu et al. (2018), the relationship between fake and factual news was analyzed on online platforms, particularly on Twitter, using fact-checking URLs. The authors observed that URLs were the most commonly used strategy for sharing news articles, providing valuable insights into customer feedback at different stages. Given the quick spread of fake news during the COVID-19 pandemic, evaluation datasets were released not only for English (Patwa et al., 2020; Kim et al., 2021), but also for other languages such as Arabic (Elhadad et al., 2020) and Chinese (Yang et al., 2020).

Shahi and Nandini (2020) even presented a fake news COVID-19 dataset encompassing 40 languages and collected from 105 different countries between January and May 2020, although English was still the most widely represented language in the data. To our knowledge, our resource represents the first example of a curated COVID-19 fake news dataset for the Persian language. Finally, concerning the most recent methodologies, NLP researchers benefitted from the introduction of Transformers (Vaswani et al., 2017; Devlin et al., 2019), with language models such as BERT being a popular choice for the basic architecture of many fake news detection systems (Vijjali et al., 2020; Chen et al., 2021; Gundapu and Mamidi, 2021).

3 Methodology

3.1 Data Collection

The reliability of news concerning the COVID-19 disease has become a crucial issue, given the importance of the political, economic, and even religious aspects connected to it. As a multitude of conflicting articles has inundated the Iran social media, we need to select appropriate labels to ensure clarity and coherence in our categorization of news.

Our main data source is posts published on Telegram. For our dataset, we crawled 5000 Telegram posts from 99 different channels, either general news channels or COVID-19 specific ones (e.g., Razcom, Akafiha, Etelaate-Omoomi, Serum Plus, etc.). Given the importance of fact checking for the success of large-scale public health strategies, we decided to focus on the time period corresponding to the initial stages of the vaccination campaign in Iran. It has been indeed reported that vaccination campaigns are often followed by skepticism, particularly when the news report cases of adverse events, and that a large number of non-official social media channels publish unchecked and potentially fake vaccine-related news (Portelli et al., 2021). On the other hand, we also need to select a reasonably long timeframe, in order to ensure more variety in the collected data and to limit content repetition. Therefore, taking into account all the above-mentioned factors, we collected data over a time span of approximately two months, from April to June 2021.

To ensure to include only posts related to COVID-19, we compiled a list of keywords and filtered the posts on the basis of their presence. No-

[1] https://github.com/s-fatemeh-ebrahimi/fact-checking

tice that we included spelling variations as separate keywords, using different forms of Persian and Arabic characters such as (ی , ک). Moreover, given that some extent of code-mixing is present in social media posts, we also used the following keywords in English which is shown in the Table 1.

Persian Key Words	English Key Words
CORONA	کرونا
covid	کوید
COVID	کوئد
KORONA	کوييد
korona	کوئید
corona	

Table 1: Persian and English Key Words

Additionally, we were interested only in popular news that gather a lot of visualizations, because those were more likely to have an impact of information spreading, and thus we set a minimum threshold of 1000 visualizations for the posts extracted from general news channels and 100 for those from COVID-19 channels. Posts with less visualizations were thus filtered out.

3.2 Data Processing

In this section, we present a comprehensive account of the data preparation process for creating a dataset aimed at detecting COVID-19 fake news in Persian. The initial data, provided in a JSON file format, underwent an initial cleaning phase using a Python script. Irrelevant tags, such as channel ID, username, and date, were removed, leaving only the news text for annotation. The cleaned news text was then used for further annotation and facilitates subsequent analysis.

After the annotation process was completed, the dataset underwent additional preprocessing to standardize the text and remove unnecessary elements that could interfere with the fake news detection task. The second stage of preprocessing was performed using the 'Hazm' library[2] and regular expressions ('Regex') in multiple stages. The Persian text in the dataset was normalized to address different forms of characters and ensure consistent representation. Next, emojis were removed using regular expressions. Non-Persian text and symbols were also eliminated to maintain a focused linguistic context. Tokenization, achieved through the 'word-tokenize' function, organized the text into words, which were then stripped of common stopwords using the 'stopwords-list()' function. Optionally, stemming was applied using the 'Stemmer' module to reduce words to their root form.

After completing the preprocessing steps, the words were joined back together to form the final preprocessed text for each news sample. This comprehensive preprocessing approach ensures the dataset is optimally prepared and forms the foundation for developing effective machine learning models to combat the spread of misinformation in the context of COVID-19 in the Persian language.

For data labeling, we established seven different tasks to be carried out by our annotators. Notice that the tasks are not just related to the factuality of the news, but also to their content, in order to provide a richer characterization of COVID-19 related news. The tasks are the following:

1. **Factuality:** this task involves verifying whether the news is entirely based on scientifically and medically approved results, statistics, or studies. Only news published by reputable individuals, organizations, or centers that adhere strictly to scientific and medical perspectives without any biases or subjective influences are considered factual. In cases where the news or post falls under any of the predefined conditions includes; rumor, deceptive news, satire news, disinformation, click-bait, cherry-picking, misinformation, is labeled as non-factual. For example, if a news article reports the daily number of COVID-19 patients (X) and it is published by an approved organization, or it cites one of such organizations as a source, it receives the label ”yes” indicating its factual nature. Conversely, if the news lacks clear verification or approval, it is labeled ”no.” In instances where the determination of approval or rejection is uncertain, the label ”can't decide” is assigned.

2. **Hate, blame, negative speech:** this task involves discerning whether the news contains elements that evoke feelings of blame, hatred, stigmatization, goatscaping, or an overall negative impact on the audience. If the news text clearly reflects such emotions, it is labeled as ”yes.” Conversely, if it is evident that the news does not convey such sentiments, it is labeled

[2] https://github.com/roshan-research/hazm

as ”no.” In cases where hate speech cannot be definitively identified, the label ”can’t decide” is assigned.

3. **Rise moral, give advice:** this task focuses on determining whether the news contains content intended to enhance public awareness, convey moral values, or offer guidance to individuals. For instance, if the news recommends positive behaviors, instills hope or adds valuable information to the readers, it receives the label ”yes.” If the presence of these elements cannot be definitively recognized, the label ”can’t decide” is assigned.

4. **Political content:** this task involves identifying whether the news in any way contains political content. If it does, it is labeled as ”yes.” Conversely, if the news does not possess any political elements, it receives the label ”no.” In cases where the presence of political implications cannot be determined, the label ”can’t decide” is assigned.

5. **Cure:** this task focuses on determining whether the news pertains to medical treatment, healthcare, or related topics. If it does, it receives the label ”yes.” Otherwise, it is labeled as ”no.” If the presence of these themes cannot be definitively recognized, the label ”can’t decide” is assigned.

6. **Mortality:** this task involves determining whether the news is associated with statistics or the number of deaths and infections caused by COVID-19. If it is, the label ”yes” is assigned. Otherwise, it receives the label ”no.”

7. **Worth fact-checking:** this task pertains to news and information related to COVID-19 where a conclusive determination of veracity is yet to be established. If the news, upon examination, is deemed worthy of fact-checking based on reliable news networks and ongoing investigations, it is labeled as ”yes.” Conversely, if the news has been confirmed or rejected based on reliable sources, it receives the label ”no.” For instance: A news article suggesting a potential blood clotting issue in 1 out of 1000 people receiving the Sinopharm vaccine would be labeled as ”yes” since it requires further investigation and verification based on various reliable domestic and international sources. Similarly, if multiple reputable news sources emphasize the need for additional research or clinical results regarding the effectiveness of a vaccine or drug, the news would be labeled as ”yes.” However, news articles that make uncertain claims about future events, such as ”By May, tablets will replace vaccines,” cannot be considered entirely reliable and should be labeled as ”yes.”

The crucial factors considered when selecting these tasks include the subject matter, objectives, type of data and data sources, and the distribution of labels. For each task, three labels have been proposed: ”yes”, ”no” and ”can’t decide.” In each of the seven tasks, the desired news content is examined, and if it aligns with the criteria of the specific task, it is labeled as ”yes.” Conversely, if the news content does not align with the task criteria, it receives the label ”no.” If a definitive determination cannot be made regarding whether the news aligns with the task, the label ”can’t decide” is assigned.

In summary, news that are not rejected by reliable news agencies, experts, organizations, or institutions, but require further investigation and discussion based on emerging information and terms like ”likely”, ”limited results”, ”in the future” or ”potential” can be appropriately labeled using this methodology. Table 2 provides a summary of these labeling criteria.

3.3 Criteria of Labeling the News as Factual or Fake

Given the goals of our research, the “Factuality” task is the most important one. As authoritative sources for the news, we considered the following websites:

- Institutional news sites and channels;
- Sites, channels, and pages that provide fact-checking services to verify the authenticity of the news;
- Pages, either in Persian or in English, authored by experts and experienced doctors who share reliable and scientifically-documented news about the coronavirus pandemic; such pages should be unbiased, without any personal gain, and connected to reputable research centers within and outside the country. They should also provide accurate explanations regarding the veracity of news.

Label title (Task)	Description of label	Value of label
Factual	Is the news real or not?	Yes/No/Can't Decide
Hate, blame, negative speech	Is the news contains hateful, blaming, fault-finding, negativity or disappointing content or not?	Yes/No/Can't Decide
Rise moral, Give advice	Is the news contains content aimed at raising the level of public awareness encouraging or giving advice to people or not?	Yes/No/Can't Decide
Political	Is the news contains political content or not?	Yes/No/Can't Decide
Cure	Is the news about health care or drug and treatment discussion or not?	Yes/No/Can't Decide
Death, Mortality	Is the news about death or the statistics of death and infections of corona or not?	Yes/No/Can't Decide
Worth fact checking	Is the news worth checking whether it is correct or not?	Yes/No/Can't Decide

Table 2: Summary of Tasks

The pages of domestic and foreign medical centers and experts have a noteworthy characteristic: by comparing a particular news item across multiple specialized pages, the content and the results tend to be consistent. The opinions of specialists in infectious diseases, microbiology, and virology are especially significant: when these experts share the same opinion about a piece of COVID-19 news, it further strengthens its reliability. Therefore, our annotators have carried out the factuality annotation by considering multiple sources. The table 3 present the list of the reliable sources for our task. It is important to emphasize that relying on a single source for analysis and validation of news is insufficient, and many cases require examining multiple sources together. The expert pages mentioned in the table primarily consist of individuals who do not have political affiliations or personal or group interests.

NAME
@Sbmuniversity
@dattums
@Vira university
@wikihoax
@factnameh
@who
@Emergency_iran
Tasnimnews.com
@_dr.key_
@Sam.pournezhad
@Doctor__online

Table 3: Some of the Reliable Sources for Validation of the News

3.4 Annotation Process

Each task is considered separately by the annotators. Depending on the task type, if there is a need to verify the news source, the mentioned sources are referred to for fact-checking. In many cases, the annotators can determine the category of news simply by reading and analyzing it based on linguistic cues and logical reasoning. The fact that the tasks are carried out in an independent way entails that, for example, a piece of news could be considered as non-factual in the factuality task (i.e. a fake news) but at the same time be checked and annotated for its topic under different tasks. We made this choice because we believe that the independent evaluation of the tasks will lead to more realistic and consistent labels.

It is worth noticing that some of the Telegram posts may contain multiple news items. In such cases, only the portion that is related to COVID-19 has been annotated for the dataset.

3.5 Annotation Team

In our study, the annotation team consisted of three highly qualified individuals with academic background, specifically trained to ensure the accuracy and reliability of the data annotations. The age range of the annotators spanned from 25 to 31 years, all of them are native speakers of the Persian language and they all possess a Master's Degree level of education.

Each instance in the dataset was labeled by two annotators, and conflicts were resolved by resorting to the third annotator (see section 3.6).

3.6 Leveraging Inter-Annotator Agreement and Precise Labeling Processes

To calculate Cohen's kappa score, the observed agreement and expected agreement are taken into consideration. The observed agreement represents the proportion of instances where the annotators agreed, while the expected agreement represents the agreement that would be expected by chance. The kappa score is then calculated by subtracting the expected agreement from the observed agreement and dividing it by 1 minus the expected agreement.

The formula used to calculate Cohen's kappa is the following one:

$$\text{Cohen's Kappa} = \frac{\text{OA} - \text{EA}}{1 - \text{EA}} \tag{1}$$

Where:

- Observed Agreement(OA): The proportion of instances where the annotators agreed (the number of agreements divided by the total number of instances).

- Expected Agreement(EA): The agreement that would be expected by chance, calculated based on the marginal probabilities of the annotators agreeing on each category.

In the present study, the Cohen's kappa score of 0.703 indicates substantial agreement between the two main annotators, indicating a high level of reliability in the labeling process. In total, there were 801 instances in which the assigned labels were conflicting. To ensure consistency, those instances were reviewed by the third annotator, which was tasked to make a decision on the final label.

3.7 Dataset Statistics

The pie chart illustrates the distribution of factual labels, categorizing the news as 'Factual', 'Fake' or 'Can't Decide.' As it can be seen, the majority of the labeled post are classified as "Factual" (58.6%), followed by "Fake" (37.5%), while a small portion remains labeled as "Can't Decide" (4%). The statistics regarding dataset labels for the seven different tasks are presented in Table 4. A total of approximately 35K data points were examined across four stages, resulting in the preparation of 5,000 news items after thorough analysis and examination. It is evident that some tasks do not have a neutral label ("can't decide"). This absence of a neutral label is primarily observed in tasks where ambiguity is seldom encountered. For instance, whether a news item contains content related to health care or not is never ambiguous, and the labeling is consistently straightforward in this task, yielding either a "yes" or "no" label. This pattern is prevalent across most tasks, with the exception of the "factual" and "political" tasks.

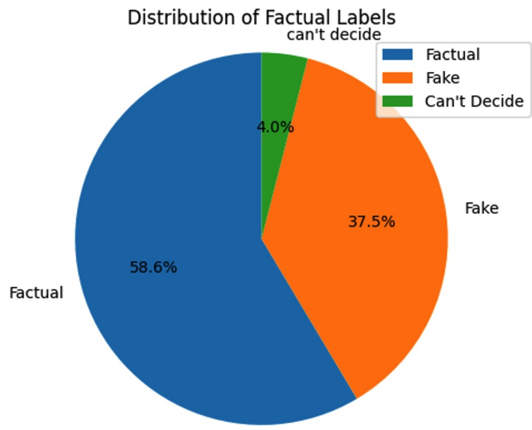

Figure 1: Pie chart with the distribution of labels for the Factual task (e.g. "Factual", "Fake", "Can't decide")

4 Evaluation and Results

In this section, we present the results of the initial evaluation of machine learning algorithms on our dataset for fake news detection in the Persian language, by using the labels produced by the annotators in the Factuality task. We utilized three popular algorithms, namely Support Vector Machines (SVM), Random Forest, and Multi-Layer Perceptron (MLP) Classifier, along with three different feature representations: TF-IDF Vectors, FastText Persian Embeddings[3], and SentenceTransformer(LaBSE)[4]. The decision not to fine-tune a Transformer directly on our dataset was made due to its relatively small size, consisting only of 5000 instances. Fine-tuning a Transformer model

[3] https://fasttext.cc/docs/en/crawl-vectors.html

[4] https://huggingface.co/sentence-transformers/LaBSE

Label	Task						
	Factual	Hate blame, negative speech	Rise moral give advise	Political	Cure	Mortality	Worth fact checking
Yes	2803	2118	1402	313	2077	211	364
No	1996	2882	3598	4430	2923	4789	4636
Can't decide	191	0	0	257	0	0	0

Table 4: Statistics of dataset labels in seven different tasks

Feature Representation	Accuracy	Precision	Recall	Macro F1-score
TF-IDF Vectors	0.7100 ± 0.0150	**0.6859 ± 0.0140**	0.7100 ± 0.0150	**0.6937 ± 0.0136**
FastText Persian Embeddings	0.6573 ± 0.0144	0.6297 ± 0.0150	0.6573 ± 0.0144	0.6283 ± 0.0143
SentenceTransformer(LaBSE)	**0.7112 ± 0.0164**	0.6830 ± 0.0171	**0.7112 ± 0.0164**	0.6922 ± 0.0157

Table 5: SVM Results for Fake News Detection

Feature Representation	Accuracy	Precision	Recall	Macro F1-score
TF-IDF Vectors	0.6888 ± 0.0170	**0.6739 ± 0.0180**	0.6888 ± 0.0170	**0.6730 ± 0.0166**
FastText Persian Embeddings	0.6723 ± 0.0169	0.6569 ± 0.0187	0.6723 ± 0.0169	0.6527 ± 0.0165
SentenceTransformer(LaBSE)	**0.6890 ± 0.0169**	0.6722 ± 0.0191	**0.6890 ± 0.0169**	0.6690 ± 0.0163

Table 6: Random Forest Results for Fake News Detection

Feature Representation	Accuracy	Precision	Recall	Macro F1-score
TF-IDF Vectors	0.6752 ± 0.0167	0.6653 ± 0.0188	0.6752 ± 0.0167	0.6691 ± 0.0162
FastText Persian Embeddings	0.6711 ± 0.0164	0.6576 ± 0.0181	0.6711 ± 0.0164	0.6590 ± 0.0160
SentenceTransformer(LaBSE))	**0.6898 ± 0.0161**	**0.6810 ± 0.0173**	**0.6898 ± 0.0161**	**0.6848 ± 0.0155**

Table 7: MLP Classifier Results for Fake News Detection

on such a small dataset can easily lead to underfitting and large variance in performance across runs. Instead, we opted to use SentenceTransformer with the pre-trained LaBSE model, which has been trained on a large multilingual corpus and provides effective sentence embeddings even for low-resource languages like Persian. By using pre-trained embeddings, we benefit from the general language representations learned from a diverse range of data, improving the model's performance on our specific task. The evaluation was conducted using 5-fold cross-validation to ensure robustness and minimize bias. The dataset underwent preprocessing to standardize the text and remove unnecessary elements that could interfere with the fake news detection task (see section 3.2).

To evaluate the performance of different machine learning algorithms and feature representations, the dataset was divided into five subsets, and each subset was used as the test set once, while the remaining four subsets were used for training. The process was repeated five times, ensuring that each subset served as the test set once. This allowed us to assess the model's performance on multiple test sets, reducing the impact of dataset variability and producing more reliable evaluation metrics. The training set consisted of 80% of the instances (4000 instances), and the test set comprised the remaining 20% of the instances (1000 instances). The evaluation metrics used were Accuracy, Precision, Recall, and Macro F1-score.

5 Discussion

The dataset we have used for our study contains a significant characteristic worth noting—the distribution of factual labels. The majority of labeled posts, approximately 58.6%, are classified as "Factual" (see section 3.7). Class imbalance, a common challenge in machine learning tasks, can introduce biases and favor the majority class, potentially impacting the reliability of the models. To overcome this concern, we meticulously designed our experimental setup and adopted appropriate strategies during model training and evaluation. As a result, our models have successfully learned meaningful patterns and features, enabling them to make accurate predictions across different classes, even in the presence of class imbalance.

By addressing the class imbalance challenge, our fake news detection models have demonstrated both robustness and effectiveness. Their superior performance compared to the Majority class classifier ensures that they are not simply biased towards the majority class but have genuinely acquired the ability to distinguish between fake and factual news articles accurately. These findings enhance the validity and practical applicability of our models in real-world scenarios, where class imbalances are common.

Our initial evaluation indicates promising performance in detecting fake news using various machine learning algorithms and feature representations for the Persian language, as presented in Table 5, 6, and 7. Notably, SVM with TF-IDF Vectors and SentenceTransformer achieved the highest accuracy and Macro F1-score, illustrating their effectiveness in discerning between fake and legitimate news articles.

The SVM algorithm, a widely-used linear classifier, exhibited competitive results when combined with both TF-IDF Vectors and SentenceTransformer, showcasing its versatility for text classification tasks. Particularly, the SentenceTransformer(LaBSE) feature representation, which encodes sentence semantics into dense vectors, proved to be particularly valuable for fake news detection, as evidenced by consistently high performance across multiple algorithms.

On the other hand, the use of FastText Persian Embeddings demonstrated comparatively lower performance, potentially due to limitations in capturing fine-grained semantic information in the Persian language. Nevertheless, the results remain promising and emphasize the importance of exploring diverse feature representations to enhance overall detection capabilities.

The Random Forest algorithm, known for its ensemble-based decision-making approach, performed moderately well but was outperformed by SVM and SentenceTransformer combinations. However, it remains an essential benchmark for further research and could potentially benefit from hyperparameter tuning and feature engineering techniques.

The MLP Classifier demonstrated competitive results, aligning with SVM and Random Forest, further validating its applicability for fake news detection in Persian. Its capacity to capture nonlinear relationships in the data proved beneficial for this task.

6 Conclusion and Future Work

Our initial evaluation has provided crucial insights into the performance of various machine learning algorithms and feature representations for fake news detection in the Persian language. These results serve as a foundation for future research and optimization of detection models, with SVM and SentenceTransformer showing promising potential for practical applications in countering fake news dissemination. Further research efforts can focus on refining feature representations, exploring deep learning architectures, and incorporating domain-specific features to achieve even higher accuracy and robustness in fake news detection.

We would like to highlight that our evaluation was primarily focused on the Factuality task, as Fake News Detection was the primary objective of our research. However, our dataset includes labels derived from other tasks, making it suitable for other researchers to undertake various COVID-19 topic classification tasks. To facilitate further research and reproducibility, we have made the dataset and code publicly available on GitHub, ensuring easy access and usability for the research community. We believe that sharing the dataset will encourage collaboration and foster advancements in the field of fake news detection and COVID-19 topic classification.

To further enhance research on fake news detection in the context of the COVID-19 pandemic, potential future directions include:

- Expanding dataset: Collecting data over a wider time frame to increase the volume of available information.

- Diversifying data sources: Including data from Twitter, news channels, and other platforms to capture a broader range of fake news instances.

- Granular labeling: Introducing additional levels of labels (e.g., ”very fake,” ”slightly fake,” ”very real,” ”slightly real”) for more nuanced classification.

- Designing a fine-tuned detection system: Developing an advanced fake news detection system by leveraging the expanded dataset and utilizing machine learning techniques.

These efforts aim to improve the accuracy and effectiveness of fake news detection, contributing to the fight against misinformation during the pandemic and beyond.

Acknowledgements

This research was supported by the Iranian Telecommunication Research Center (ITRC) under the contract number 500/6941.

References

Supanya Aphiwongsophon and Prabhas Chongstitvatana. 2018. Detecting fake news with machine learning method. *2018 15th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)*, pages 528–531.

Ben Chen, B. Chen, Dehong Gao, Qijin Chen, Chengfu Huo, Xiaonan Meng, Weijun Ren, and Yang Zhou. 2021. Transformer-based language model fine-tuning methods for covid-19 fake news detection. In *CONSTRAINT@AAAI*.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. *ArXiv*, abs/1810.04805.

Mohamed K. Elhadad, Kin Fun Li, and Fayez Gebali. 2020. Covid-19-fakes: A twitter (arabic/english) dataset for detecting misleading information on covid-19. In *International Workshop on Intelligent Networking and Collaborative Systems*.

Anastasia Giahanou, Esteban Andrés Ríssola, Bilal Ghanem, Fabio A. Crestani, and Paolo Rosso. 2020. The role of personality and linguistic patterns in discriminating between fake news spreaders and fact checkers. *Natural Language Processing and Information Systems*, 12089:181 – 192.

Sunil Gundapu and R. Mamidi. 2021. Transformer based automatic covid-19 fake news detection system. *ArXiv*, abs/2101.00180.

Rohit Kumar Kaliyar, Anurag Goswami, and Pratik Narang. 2021. Fakebert: Fake news detection in social media with a bert-based deep learning approach. *Multimedia Tools and Applications*, 80:11765 – 11788.

Jisu Kim, Ji A Aum, Sang Eun Lee, Yeonju Jang, Eunil Park, and Daejin Choi. 2021. Fibvid: Comprehensive fake news diffusion dataset during the covid-19 period. *Telematics and Informatics*, 64:101688 – 101688.

Srijan Kumar and Neil Shah. 2018. False information on web and social media: A survey. *ArXiv*, abs/1804.08559.

Parth Patwa, Shivam Sharma, Srinivas Pykl, Vineeth Guptha, Gitanjali Kumari, Md. Shad Akhtar, Asif Ekbal, Amitava Das, and Tanmoy Chakraborty. 2020. Fighting an infodemic: Covid-19 fake news dataset. In *CONSTRAINT@AAAI*.

Beatrice Portelli, Simone Scaboro, Roberto Tonino, Emmanuele Chersoni, Enrico Santus, and Giuseppe Serra. 2021. Monitoring user opinions and side effects on covid-19 vaccines in the twittersphere: Infodemiology study of tweets. *Journal of Medical Internet Research*, 24.

Enrico Santus, Nicola Marino, Davide Cirillo, Emmanuele Chersoni, Arnau Montagud, Antonella Santuccione Chadha, Alfonso Valencia, Kevin S Hughes, and Charlotta Lindvall. 2021. Artificial intelligence–aided precision medicine for covid-19: Strategic areas of research and development. *Journal of Medical Internet Research*, 23.

Gautam Kishore Shahi and Durgesh Nandini. 2020. Fakecovid - a multilingual cross-domain fact check news dataset for covid-19. *ArXiv*, abs/2006.11343.

Jieun Shin, Lian Jian, Kevin Driscoll, and François Bar. 2018. The diffusion of misinformation on social media: Temporal pattern, message, and source. *Comput. Hum. Behav.*, 83:278–287.

Kai Shu, Deepak Mahudeswaran, Suhang Wang, Dongwon Lee, and Huan Liu. 2018. Fakenewsnet: A data repository with news content, social context and dynamic information for studying fake news on social media. *ArXiv*, abs/1809.01286.

Vivek K. Singh, Rupanjal Dasgupta, Darshan Sonagra, Karthik Raman, and Isha Ghosh. 2017. Automated fake news detection using linguistic analy- sis and machine learning.

Ingjerd Skafle, Anders Johan Nordahl-Hansen, Daniel S. Quintana, Rolf Wynn, and Elia Gabarron. 2022. Misinformation about covid-19 vaccines on social media: Rapid review. *Journal of Medical Internet Research*, 24.

Ashish Vaswani, Noam M. Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *NIPS*.

Sairamvinay Vijayaraghavan, Ye Wang, Zhiyuan Guo, Johny Voong, Wenda Xu, Armand Nasseri, Jiaru Cai, Linda Li, Kevin Vuong, and Eshan Wadhwa. 2020. Fake news detection with different models. *ArXiv*, abs/2003.04978.

Rutvik Vijjali, Prathyush Potluri, Siddharth Krishna Kumar, and Sundeep Teki. 2020. Two stage transformer model for covid-19 fake news detection and fact checking. *ArXiv*, abs/2011.13253.

Minyu Wan, Qi Su, Rong Xiang, and Chu-Ren Huang. 2022. Data-driven analytics of covid-19 'infodemic'. *International Journal of Data Science and Analytics*, pages 1 – 15.

William Yang Wang. 2017. "liar, liar pants on fire": A new benchmark dataset for fake news detection. *ArXiv*, abs/1705.00648.

Chen Yang, Xinyi Zhou, and Reza Zafarani. 2020. Checked: Chinese covid-19 fake news dataset. *Social Network Analysis and Mining*, 11.

Appendix

Annotator Guidelines for Labelling the Tasks

The annotation process is a crucial step in achieving accurate results in the domain of COVID-19 news analysis for fake news detection. The following guidelines outline the key considerations for selecting appropriate labels and performing correct annotation of the raw data:

Selection of Label Types for Annotation

The first and most important step in achieving reliable results in COVID-19 news analysis is to choose labels based on appropriate criteria and conduct accurate annotation of the raw data. The main points to consider during the selection and definition of proposed labels are as follows:

Topic of Investigation

The subject matter for which the annotation is performed is news related to the COVID-19 disease. The primary concern in this context is the authenticity of news related to COVID-19, as it is one of the most challenging issues in society. Various political, economic, religious, and cultural factors have led to the proliferation of contradictory news, creating a space filled with misinformation and disinformation. Therefore, we need to select labels that can bring about the necessary transparency. For instance:

Determine the authenticity of statements made by experts for personal gains and political motives regarding vaccination. Thus, ”rumor” can be a relevant label. Identify news that induces fear and threatens individuals. Categorize content that uses humor to entertain and create a calmer atmosphere in society. Detect news that misleads individuals by exploiting COVID-19-related information, eroding trust in authorities. Evaluate news that reports initial study findings on specific subjects, like a drug's efficacy in COVID-19 treatment, yet hasn't received full approval from health organizations.

Objective of Annotation

Given the subject matter of COVID-19, the objective of this annotation is to help people in society distinguish genuine news for better understanding of the disease, preventive measures, and treatments such as vaccination. For example, due to the current societal atmosphere, many individuals refrain from vaccination due to a lack of trust in certain vaccine types. Thus, the chosen labels should enable people to determine the authenticity of news without being influenced by personal or group interests, helping them make informed decisions to mitigate the impact of the disease on their lives.

Data Type

The target data for annotation includes posts published on Telegram channels. The list of these channels has been provided by the client, and a few additional channels are added at the end of this guideline. Considering the initial data and the previously mentioned channels used for data collection, the majority of these sources come from news networks. Therefore, we may encounter very few instances of humorous content during the data review process. Consequently, we will select labels that align with news content from these sources.

Distribution of Labels After Annotation

Since achieving a balanced classification is crucial in this project, we will choose labels that lead to a nearly balanced distribution of categories after the annotation process. We need to define relatively general labels (but with accurate and precise definitions) to ensure satisfactory performance during the annotation. The detailed guidelines on label definitions will follow.

Task Selection: Based on the considerations mentioned above, we have chosen seven tasks for annotating news data, each of which may have three labels: ”yes,” ”no,” and ”can't decide.” all seven tasks have been explained in the content of the paper (see section 3.2)

Note 1: In all seven tasks, the following criteria are applied during the annotation:

If a news item contains the desired content and concepts related to the respective task, the label assigned is ”yes.” If a news item lacks the desired content and concepts related to the respective task, the label assigned is ”no.” If it is challenging to make a ”yes” or ”no” decision for a particular news item, the label assigned is ”can't decide.” Finally, the selected tasks and their respective guidelines are presented, considering the criteria mentioned above.