

# Fake Farsi News Detection

Mohamadhosssein Amirifardchime

## Contents

- [Introduction](#)
  - Project Template
  - Background
  - Project Goals
  - Motivation and Contribution
- [Literature Review](#)
  - Related Works
  - Review Summary
- [Project Design and Planning](#)
  - Domain and Users
  - Planning, Design, and Technology Choices
- [Implementation](#)
  - Dataset Preparation
  - Model Preparation and Training
- [Evaluation](#)
- [Conclusion](#)
- [References](#)

All codes written and files prepared for this project are accessible on the following GitHub repo:  
[https://github.com/amirifard/Farsi\\_Fake\\_News\\_Detection/](https://github.com/amirifard/Farsi_Fake_News_Detection/)

# **Introduction**

## **Project Template**

CM3060 Natural Language Processing Project Idea Title 1: Fake news detection used for this project.

## **Background**

The increasing penetration of the internet and the proliferation of digital media have resulted in an overwhelming flow of information, making it challenging to distinguish fact from fiction. The issue of 'fake news' - intentionally misleading or false information spread under the guise of news reporting - has become a significant concern globally. The impact of this misinformation ranges from influencing public opinion on critical topics to potential threats to national security.

## **Project Goals**

This project focuses on developing an effective system for detecting fake news in the Farsi language. There are many existing tools and models that tackle fake news detection in popular languages like English; however, solutions for languages like Farsi are less widespread. This project aims to fill this gap by creating a system that uses machine learning and natural language processing techniques to identify and classify fake news articles in Farsi.

This system will source authentic news articles via web scraping from reputable Farsi news websites, while its fake news dataset will be a combination of available fake Farsi news data and some synthetic data generated using Large Language Models (LLMs). It will then employ a blend of machine learning and deep learning techniques for the task of fake news detection.

The users of this project are expected to be digital news platforms, social media platforms, and fact-checking organizations operating in regions where Farsi is primarily used. The goal is to provide these entities with a robust model for spotting and filtering out fake news, thus helping to curb the spread of misinformation.

## **Motivation and Contribution**

Through this project, I hope to contribute to the global effort against fake news and demonstrate the adaptability of machine learning and NLP techniques to various languages and contexts.

# Literature Review

## Related works

### 1. Persian Fake News Detection: Neural Representation and Classification at Word and Text Levels

Mohammadreza Samadi, Maryam Mousavian, and Saeedeh Momtazi. 2021. *Persian Fake News Detection: Neural Representation and Classification at Word and Text Levels*. *ACM Trans. Asian Low-Resour. Lang. Inf. Process.* 21, 1, Article 10 (January 2022), 11 pages. <https://doi.org/10.1145/3472620>

This article addresses the challenge of fake news detection in Persian by introducing a new dataset (TAJ) sourced from various news agencies and proposing two deep learning models that utilize the Bidirectional Encoder Representations from Transformers (BERT) model. The first model employs a pool-based representation, interpreting entire documents and linking to a single-layer perceptron (SLP) for fake news detection. In contrast, the second uses a sequence representation for individual tokens, applying a Convolutional Neural Network (CNN) after the BERT embedding to extract additional features from word collocations. Evaluations indicate that both BERT-based models significantly outperform previous methods in detecting fake news, showcasing improvements of up to 17.1% over past studies.

### 2. Detecting fake news with capsule neural networks

Mohammad Hadi Goldani, Saeedeh Momtazi, Reza Safabakhsh, *Detecting fake news with capsule neural networks*. *Applied Soft Computing*, Volume 101, 2021, 106991, ISSN 1568-4946, <https://doi.org/10.1016/j.asoc.2020.106991>.

This paper explores the application of Capsule neural networks, primarily renowned in computer vision, for fake news detection in the realm of Natural Language Processing (NLP). Depending on the length of the news items, different embedding models are employed: static word embedding for shorter pieces and dynamic embeddings for medium to long items. The research also leverages various n-gram levels for enhanced feature extraction. When tested on two prominent datasets, ISOT, and LIAR, the proposed models demonstrated commendable efficacy, surpassing the best-existing methods by margins of 7.8% and 1% respectively

### 3. Combining Machine Learning with Knowledge Engineering to Detect Fake News in Social Networks-a survey

Ahmed, S., Hinkelmann, K. and Corradini, F., 2022. *Combining machine learning with knowledge engineering to detect fake news in social networks-a survey*. *arXiv preprint arXiv:2201.08032*

This paper discusses the integration of machine learning and knowledge engineering for the detection of fake news. It also presents different ways to detect fake news on social media and existing detection algorithms. This paper could be useful for my project as it provides a comprehensive overview of the techniques used in fake news detection. The integration of machine learning and knowledge engineering could be a potential.

### 4. The Mass, Fake News, and Cognition Security

Guo, B., Ding, Y., Sun, Y., Ma, S., Li, K. and Yu, Z., 2021. *The mass, fake news, and cognition security*. *Frontiers of Computer Science*, 15, pp.1-13. *arXiv:1907.07759*

This paper presents a new research area called Cognition Security (CogSec), which studies the potential impacts of fake news on human cognition. It could be useful for understanding the psychological aspects of fake news and how it can influence people's perceptions and decisions. This paper emphasizes the multidisciplinary nature of fake news detection, combining social science, psychology, cognition science, neuroscience, AI, and computer science.

## 5. FakeSwarm: Improving Fake News Detection with Swarming Characteristics

Wu, J. and Ye, X., 2023. *FakeSwarm: Improving Fake News Detection with Swarming Characteristics*. *arXiv preprint arXiv:2305.19194*.

This paper presents a fake news identification system that leverages the swarming characteristics of fake news. The system uses principal component analysis, metric representation, and position encoding to extract swarm behavior. This paper could be useful for my project as it introduces a novel concept and approach to fake news detection. The idea of using swarming characteristics could be an innovative technique to apply.

## 6. Credibility-based Fake News Detection

Sitaula, N., Mohan, C.K., Grygiel, J., Zhou, X. and Zafarani, R., 2020. *Credibility-based fake news detection. Disinformation, misinformation, and fake news in social media: Emerging research challenges and Opportunities*, pp.163-182. *arXiv:1911.00643*

This paper emphasizes the detection of fake news by assessing its credibility. It suggests that an author's history of association with fake news and the number of authors of a news article can play a significant role in detecting fake news. This paper could be useful for my project as it provides a different perspective on fake news detection, focusing on credibility rather than content.

## 7. Word representation or word embedding in Persian text

Sarmady, S. and Rahmani, E., 2017. *word representation or word embedding in Persian text*. *arXiv preprint arXiv:1712.06674*.

This paper discusses the representation of words or converting words into vectors in Persian text. It updates GloVe, CBOW, and skip-gram methods to produce embedded vectors for Persian words. This paper could be useful for my project as it provides a method for processing and understanding Persian text, which is crucial for detecting fake news in Farsi

## Review Summary

These papers have provided critical insights and guidance into both the strategic approach and specific methods that could be leveraged for effective results in this project.

A notable finding lies in the value of a robust dataset and the efficacy of transformer models like BERT in detecting fake news. Samadi et al. (2022) introduction of the TAJ dataset <sup>1</sup> and two deep learning models specifically designed for Persian content, highlights the significance of language-specific tools and datasets in this task. Furthermore, the success of BERT-based models, with improvements of up to 17.1% over previous methods, confirms that transformer models could be a valuable asset for my project.

Goldani et al. (2021) also showcase the versatility of neural architectures in fake news detection. Their innovative application of Capsule Neural Networks (primarily known for computer vision tasks) for NLP and their incorporation of dynamic and static word embeddings based on content length further emphasize the importance of nuanced approaches to textual data <sup>2</sup>.

Furthermore, understanding the psychological impact of fake news is crucial, as illustrated by Guo et al. (2021). Their research into Cognition Security (CogSec) draws attention to the importance of considering the multidisciplinary nature of fake news detection, which combines areas like social science, psychology, and computer science <sup>3</sup>. This insight is crucial for a holistic approach in any detection system.

Equally intriguing is the novel concept presented by Wu and Ye (2023), where fake news detection is approached via swarming characteristics <sup>4</sup>. This method adds a layer of behavior analysis which might uncover patterns not discernible by text content alone.

Ahmed et al. (2022) and Sitaula et al. (2020) provide different, yet equally valuable perspectives. The former underscores the fusion of machine learning with knowledge engineering <sup>5</sup>, whereas the latter pivots toward an assessment of news credibility based on an author's history and associations <sup>6</sup>. Both approaches offer potential pathways for enhancing detection robustness.

Finally, for projects that specifically target Persian content, Sarmady and Rahmani (2017) shed light on the representation of words in Persian text, emphasizing methods tailored to Persian word embeddings <sup>7</sup>. Such language-specific insights are essential for projects aiming at detecting fake news in non-English languages.

In conclusion, these papers collectively underline the importance of a multifaceted approach to fake news detection. They validate the worth of specialized datasets, spotlight the potential of transformer models, highlight the need to understand the psychological impacts and emphasize the critical role of word representations. For my project, integrating some of these techniques and insights promises a holistic and potentially more effective approach to combat fake news.

# Project Design and Planning

## Domain and Users

The project focuses on the domain of Natural Language Processing (NLP), Machine Learning (ML), and specifically on Fake News Detection in the Farsi language. The primary users of the project would be digital news platforms, social media platforms, and fact-checking organizations operating in regions where Farsi is widely used.

## Planning, Design, and Technology Choices

### User Needs/Domain Requirements

The design of this project is aimed at addressing the rise of misinformation in the digital space. There's a necessity for automated detection systems, especially in the Farsi language where resources might be limited. Therefore, there is not a specific UI or design applicable to this project, However, for the purpose of data gathering and in order to make it easier to record data into a dataset, a simple data entry platform was created by using react and material UI design (Image 1).

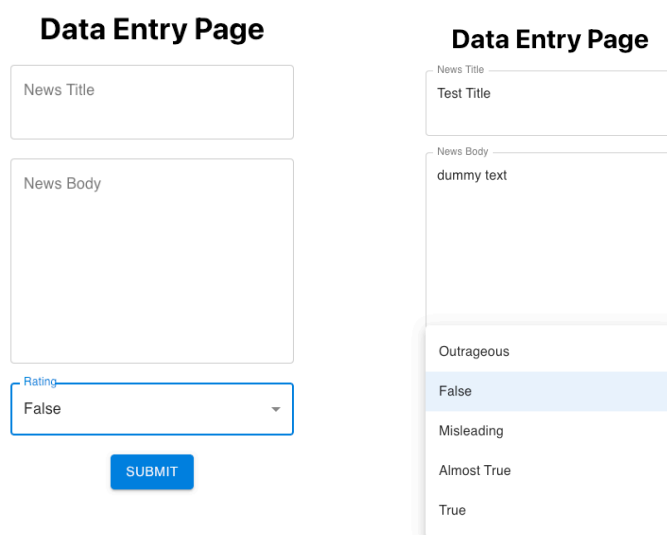


Image 1

## Overall Structure of the Project

**Research Question:** Can we develop an effective system for detecting fake news in Farsi using machine-learning techniques and language models?

**Methods:** The project uses a mix of web scraping tools to gather real news data, while fake news data is a combination of available fake news datasets and some generated using LLMs. A combination of Transformer models and ensemble stacking techniques is used for the detection task.

### Important Technologies

**Web Scraping:** BeautifulSoup and GraphQL queries are used to scrape real news data from authentic sources.

**Data Processing:** Libraries such as Pandas, NumPy, and Jupyterlabs are used for data cleaning, manipulation, and exploration.

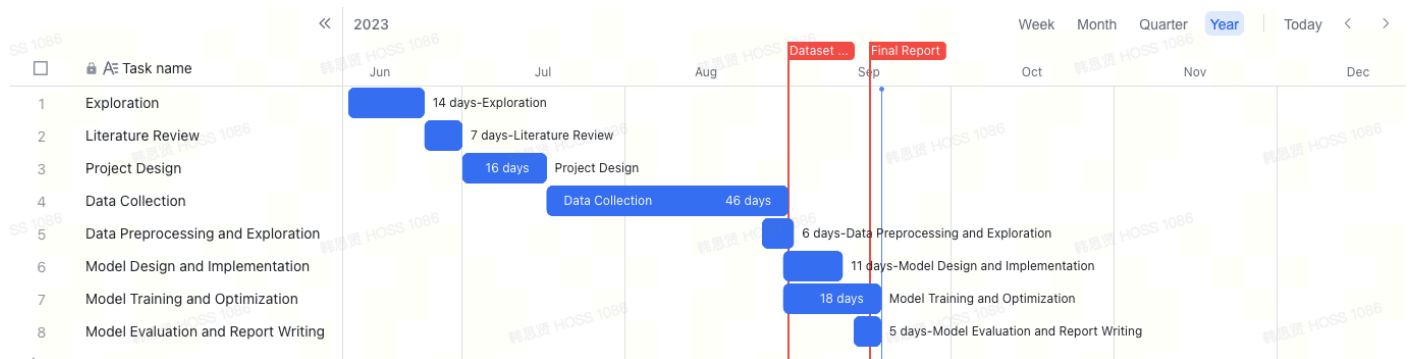
**Machine Learning Libraries:** Scikit-learn, torch, transformers, and sentencepiece libraries used for implementing ML algorithms such as tokenization, fine-tuning, and inference

**Language Models:** Pre-trained models such as BERT <sup>8</sup>, ParsRoBERTa <sup>9</sup>, and XLNet <sup>10</sup> are used for base models, and ChatGPT-4 <sup>11</sup> is used for synthetic data generation

## Testing and Evaluation

The project is tested by using a split of the collected data into training and test sets. The performance was evaluated using metrics like Precision, Recall, and F1-score. Initially planned to compare the performance of the model with baseline models or existing methods in the literature to ensure that it achieves an improvement but due to dataset complications did not evaluate other models with the compiled dataset.

## Project Timeline



# Implementation

## Dataset Preparation

The most challenging part of this project was to curate a comprehensive labeled dataset for fake news. I've tried to get access to the TAJ dataset(1) but unfortunately, the creators were not responsive and I could not get any info on this. (Dataset is not available online nor in the GitHub repo: <https://github.com/MhmDSmdi/Fake-News-Crawler/tree/master>)

Therefore, a new dataset needed to be prepared again! The dataset preparation task is divided into two sections. True news and Fake news collection.

## True News Collection

The primary aim was to gather a comprehensive and reliable dataset of genuine Farsi news articles. This would also serve as a reference for authenticity in the project's subsequent phases.

### Source identification

An online search was conducted to sift through various Farsi news platforms.

Factors such as credibility, unbiased news, and readership size informed the selection of the most reliable source.

The main sources were narrowed down to "Fars News Agency", "BBC Persian" and "Tabnak News", After a thorough examination (Table 1 below), "Tabnak News" was selected as the True news source.

Source overview	Pros	Cons	Verdict
<b>Fars News Agency</b> <i>One of the prominent news outlets in Iran</i> <a href="https://www.farsnews.ir">https://www.farsnews.ir</a>	Given its significant readership, Fars offers a vast array of news articles spanning various categories.	Historically, Fars has been observed to remove certain news items or even publish misleading information. It is perceived as a pro-government news website, raising concerns about potential biases in its reporting.	Rejected due to its history of publishing unreliable news.
<b>Tabnak News</b> <i>Notable Farsi news outlet with a significant readership</i> <a href="https://www.tabnak.ir">https://www.tabnak.ir</a>	Provides a diverse range of news articles and is widely recognized in the Farsi-speaking community.	Less original news and frequent updates, more publishing news from other sources.	Given the absence of specific biases, Tabnak is considered as the source for 'True News'.
<b>BBC Persian</b> <i>Farsi language news service provided by the British Broadcasting Corporation</i> <a href="https://www.bbc.com/persian">https://www.bbc.com/persian</a>	As part of the BBC's global network, it has access to a broad spectrum of international news and perspectives.	BBC Persian has been criticized for its perceived biases, particularly in consistently publishing negative news about Iran.	Rejected due to concerns about biases.

Table 1

### Website Structure Analysis

Prior to data extraction, Checked the website's policy for data usage and found out that "Any usage is allowed by Citation". A thorough analysis of Tabnak's website's layout and article organization was performed and found out each news article has a "news code" which is incorporated in the link as well therefore a simple counter could regenerate news article links. An asynchronous function is used to retrieve news article webpages and by using "BeatifulSoup" parsed the tags and extracted data for news title, news body, category, and publication date. The code and full dataset are available here :

[https://github.com/amirifard/Farsi\\_Fake\\_News\\_Detection/tree/main/dataset/true](https://github.com/amirifard/Farsi_Fake_News_Detection/tree/main/dataset/true)  
<https://www.kaggle.com/datasets/hoss01/tabnak-farsi-news-articles>



## Fake News Collection

The project sought to amass a representative fake news dataset in Farsi. Given the lack of pre-existing datasets, it was essential to utilize a combination of synthetic data generation and web data extraction.

### Synthetic Data Generation

A range of Language Learning Models (LLMs) were considered for potential synthetic fake news generation. Tests with models like Llama2 and Falcon revealed their inability to produce convincing Farsi content. The ChatGPT-4 Model from OpenAI had far better results but despite its capabilities, the ChatGPT API was inaccessible due to constraints, necessitating the exploration of other avenues:

1. The model was trained to discern patterns between genuine and counterfeit sentences by providing a set of examples. Once these patterns were internalized, the model was fed genuine news and tasked to produce text that mirrored the exemplified patterns, resulting in fabricated or exaggerated news. training the model with a few examples such as:

**Original:** "رئیس جمهور جدید ایران انتخاب شد."

(Translation: "A new president of Iran has been elected.")

**Fake:** "رئیس جمهور جدید ایران به طور غیر قانونی و با کودتا به قدرت رسید."

(Translation: "The new president of Iran came to power illegally by coupe.")

**Guiding Prompt:**

بر اساس الگوی نشان داده شده، جمله زیر را تغییر دهید:

"مذاکرات هسته‌ای در وین پیشرفت کرد."

(Translation: "Based on the shown pattern, modify the following sentence:

Nuclear negotiations in Vienna made progress.")

**Output:**

"مذاکرات هسته‌ای در وین به دلیل سکتة مغزی مذاکره کنندگان به شکست منجر شد"

(Translation: "Nuclear negotiations in Vienna led to failure due to negotiator's brain stroke")

2. A persona was crafted for the model, positioning it as a professional editor from a news website, "Tabnak." The model was then directed to author imaginary news based on topics provided, simulating the production of misleading news. One example of the generated results:

تهران: امروز در یک مراسم با شکوه در تهران، ایران اولین ربات هوش مصنوعی خود را معرفی کرد. این ربات قادر است به چندین زبان صحبت کند و به افراد کمک کند تا مهارت‌های زبان خود را ارتقاء دهند. متخصصان معتقدند که این یک گام بزرگ در زمینه تکنولوژی است.

(Translation: "TEHRAN: Today, in a grand ceremony in Tehran, Iran introduced its first artificial intelligence robot. This robot can speak multiple languages and help individuals enhance their language skills. Experts believe this is a significant step in the field of technology.")

### Web Scraping

Preliminary research pinpointed Factnameh as a reliable source of unbiased fact-checking articles. After a thorough examination of the website structure and inner workings, web inspection tools helped unearth the GraphQL endpoint of Factnameh, and introspect queries were sent to understand the data structures and find the link to all articles and nodes on Factanameh. Tailored queries extracted all pertinent fact-checking articles, leading to a comprehensive list of article links.

[https://github.com/amirifard/Farsi\\_Fake\\_News\\_Detection/blob/main/dataset/fake/all\\_articles.json](https://github.com/amirifard/Farsi_Fake_News_Detection/blob/main/dataset/fake/all_articles.json)

### Custom Data Extraction Interface

Given the intricacies of separating fake content from broader fact-checking narratives, a specialized React-based data-entry interface was developed. This drastically streamlined the data curation process and helped to gather fake news articles quickly.

## Model Preparation and Training

### Model Selection

Due to the recent success and performance of the transformer-based Models, transformer-based models were selected for this project.

Three distinct transformer models are integrated together: BERT multilingual, ParsRoBERTa, and XLNet.

The choice to use these models stems from their design and performance. These models are pre-trained on vast amounts of data and have already captured general linguistic patterns, structures, and representations from diverse and extensive corpora therefore this will help to compensate for the small size of the training samples.

### Pre-trained Model Fine-Tuning and Meta Classifier Training

The models have fine-tuned with the sample data for 3 epochs.

The stacking ensemble method is used to compensate for individual model weaknesses, capturing a broader spectrum of patterns and reducing the chance of overfitting.

Each model produced logits (raw predictions) from the three models fed as features for a meta-classifier (Logistic Regression) to predict the final label.

### Model Structure

The following diagram (Diagram 1) shows the models' structure and layer configuration

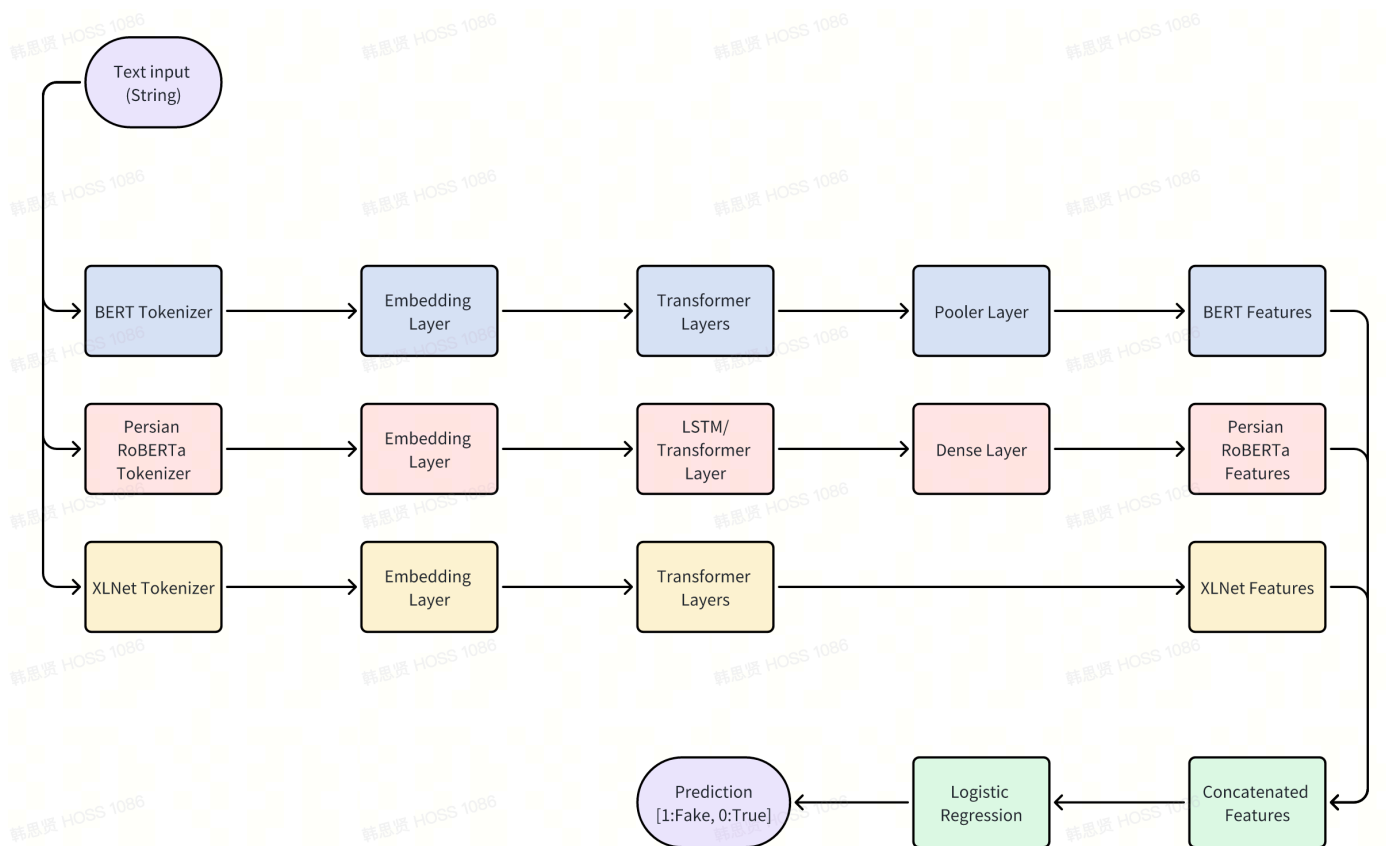


Diagram 1

## Evaluation

Initially, Accuracy was used as the performance evaluation metric. However, due to its limited variation, ultimately relied on the F1 score to select the best epoch for each model. The comparison of accuracy across each epoch for the three base models, as well as the ensemble model, is depicted below (Image 2):

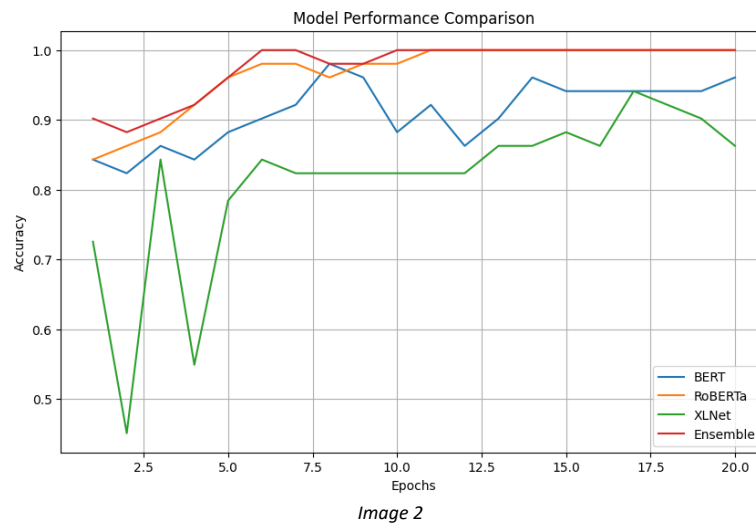


Image 2

The F1 score harmoniously combines both precision and recall. The F1 scores presented for each model across the 20 epochs give insight into the models' performance trajectory over time (Image 3):

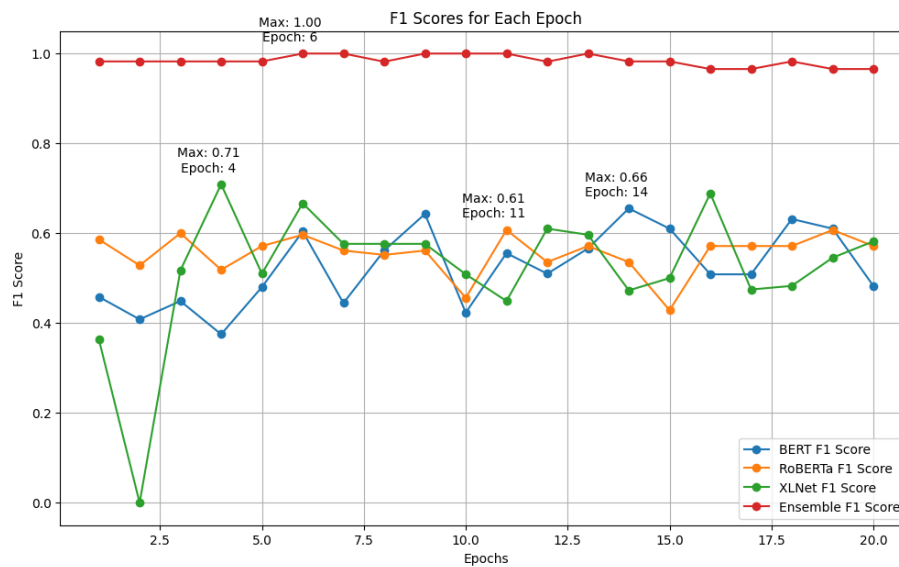


Image 3

### BERT:

The F1 score starts at approximately 0.46 in the first epoch and fluctuates until the 20th epoch, where it ends at 0.48.

The highest F1 score is **0.655**, achieved in the 14th epoch.

The BERT model seems to have its performance peak around the middle of the training period and then stabilizes.

### ParsRoBERTa:

The F1 score starts at approximately 0.59 in the first epoch and fluctuates throughout, ending at 0.57 in the 20th epoch.

The highest F1 score is **0.607**, achieved in the 11th and 19th epochs.

ParsRoBERTa's performance is relatively consistent, with slight variations in F1 score throughout the training.

**XLNet:**

The F1 score starts at 0.36 in the first epoch, drops to 0 in the second epoch, then fluctuates, reaching 0.58 by the 20th epoch.

The highest F1 score is **0.709**, achieved in the 4th epoch.

XLNet's performance appears to have more variance compared to the other models, with its peak performance occurring early in the training.

**Ensemble:**

The F1 score starts very high at 0.982 in the first epoch and remains consistently high throughout, reaching a perfect score of 1.0 several times. The sweet spot seems to be on the 10<sup>th</sup> epoch.

The ensemble's performance is notably superior to the individual models, demonstrating the power of model ensembling in boosting performance. However, the consistently high scores might suggest overfitting.

**Overall Observations:**

The ensemble approach outperforms individual models across all epochs, suggesting that combining the strengths of each model leads to improved results.

Each individual model exhibits different performance patterns, with BERT peaking mid-training, ParsRoBERTa maintaining consistent performance, and XLNet showing more fluctuations.

## Conclusion

The proliferation of fake news, especially in non-English languages, presents a unique set of challenges and emphasizes the importance of devising robust detection systems. This project, focusing on the Farsi language, aimed to bridge the existing gap in fake news detection tools tailored for Farsi-speaking audiences. By integrating advanced machine learning techniques, web scraping methods, and state-of-the-art pre-trained transformer models, this project developed a system adept at distinguishing genuine news from fabricated content.

This project underscores the pivotal role of a well-curated dataset. While challenges arose, especially in acquiring pre-existing datasets tailored to the project needs, the project's resilience was evident in the creation of a new dataset, combining both authentic news sources and synthetically generated content. The results of this endeavor emphasize the importance of language and context specificity in creating detection tools.

The performance of individual pre-trained transformer models, and more remarkably, the ensemble approach, demonstrated the efficacy of leveraging deep learning techniques in detecting fake news. The ensemble model's unparalleled performance attests to the power of integrating individual model strengths, offering a holistic and comprehensive detection system.

However, it's essential to acknowledge potential pitfalls, such as overfitting, which can compromise the generalizability and real-world application of the model. As with all machine learning projects, there remains an inherent trade-off between accuracy and generalization, which future iterations of this project should continuously evaluate and optimize.

This project serves as a testament to the adaptability and potential of machine learning and NLP techniques when tailored to specific languages and challenges. As fake news continues to evolve, so must the strategies and tools. With the insights gained from this research, we are one step closer to creating a more informed and discerning digital community, especially for Farsi speakers worldwide.

**All codes written and files prepared for this project are accessible on the following GitHub repo:**

**[https://github.com/amirifard/Farsi\\_Fake\\_News\\_Detection/](https://github.com/amirifard/Farsi_Fake_News_Detection/)**

## References

1. Mohammadreza Samadi, Maryam Mousavian, and Saeedeh Momtazi. 2021. *Persian Fake News Detection: Neural Representation and Classification at Word and Text Levels*. *ACM Trans. Asian Low-Resour. Lang. Inf. Process.* 21, 1, Article 10 (January 2022), 11 pages. <https://doi.org/10.1145/3472620>
2. Mohammad Hadi Goldani, Saeedeh Momtazi, Reza Safabakhsh, *Detecting fake news with capsule neural networks*. *Applied Soft Computing*, Volume 101, 2021, 106991, ISSN 1568-4946, <https://doi.org/10.1016/j.asoc.2020.106991>.
3. Guo, B., Ding, Y., Sun, Y., Ma, S., Li, K. and Yu, Z., 2021. *The mass, fake news, and cognition security*. *Frontiers of Computer Science*, 15, pp.1-13. *arXiv:1907.07759*
4. Wu, J. and Ye, X., 2023. *FakeSwarm: Improving Fake News Detection with Swarming Characteristics*. *arXiv preprint arXiv:2305.19194*.
5. Ahmed, S., Hinkelmann, K. and Corradini, F., 2022. *Combining machine learning with knowledge engineering to detect fake news in social networks-a survey*. *arXiv preprint arXiv:2201.08032*
6. Sitaula, N., Mohan, C.K., Grygiel, J., Zhou, X. and Zafarani, R., 2020. *Credibility-based fake news detection. Disinformation, misinformation, and fake news in social media: Emerging research challenges and Opportunities*, pp.163-182. *arXiv:1911.00643*
7. Sarmady, S. and Rahmani, E., 2017. *word representation or word embedding in Persian text*. *arXiv preprint arXiv:1712.06674*.
8. Devlin, J., Chang, M.W., Lee, K. and Toutanova, K., 2018. *Bert: Pre-training of deep bidirectional transformers for language understanding*. *arXiv preprint arXiv:1810.04805*.
9. Farahani, M., Gharachorloo, M., Farahani, M. et al. *ParsBERT: Transformer-based Model for Persian Language Understanding*. *Neural Process Lett* 53, 3831–3847 (2021). <https://doi.org/10.1007/s11063-021-10528-4>
10. Yang, Z., Dai, Z., Yang, Y., Carbonell, J., Salakhutdinov, R.R. and Le, Q.V., 2019. *XLnet: Generalized autoregressive pretraining for language understanding*. *Advances in neural information processing systems*, 32.
11. OpenAI, 2023. *GPT-4 Technical Report*. *arXiv:2303.08774*