# *Bagging-Based Transformer Ensemble for Real-Time Fake News Classification*

Garreth Jeconio Budi Utomo
School of Computer Science,
Computer Science
Bina Nusantara University,
Bandung, Indonesia
Garreth.utomo@binus.ac.id

Chaldarren wijaya
School of Computer Science, Computer Science
Bina Nusantara University,
Bandung, Indonesia
chaldarren.wijaya@binus.ac.id

Vincent Devano
School of Computer Science,
Computer Science
Bina Nusantara University,
Bandung, Indonesia
vincent.devano@binus.ac.id

Budi Juarto
School of Computer Science,
Computer Science
Bina Nusantara University,
Bandung, Indonesia
budi.juarto@binus.ac.id

***Abstract*— In the current digital era, the distribution of fake news and misleading information has been a problem for people. This can be caused by the rising popularity of social media as news source. The purpose of our research is to develop a automated news fact-checking system using machine learning. We will use Natural Language Processing (NLP) approaches such as BERT, RoBERTa, Xlnet, Electra, and Bagging, are evaluated to determine the method effectiveness in order to classify news articles as facts or hoax. Performance metrics such as accuracy, precision, recall, and F1-score will be used to test each model's ability to detect misleading news or fake news. A more throughtout analysis of our project will have better understanding about how we're going to detect these fake and misleading news to better understand to how we detect fake and misleading news. The results of this study show that the ensemble method outperforms the other methods in classifying fake news with accuracy, precision, recall, and F1 results of 0.953,0.952, 0.951, 0.952, respectively.**

***Keywords—Fact-checking, misinformation, NLP, machine learning, BERT, RoBERTa, Xlnet, Electra, Bagging(Ensemble Voting)***

## I. Introduction

News has been one of the most important pieces of information for us Human. However, As the age of Technology progresses the distribution of Fake and Misleading news has become a problem for us especially in social. this can be quite a problem because Fake and misleading News are usually used to make false narrative and propagandas that usually spread among people quickly. People will quickly believes conspiracy and form public opinion that will affect the wellbeing of society. Althought there are already some Prevention ways that the Government provide such as manual fact-checking techniques by the Media but it heavily rely on human experts that sometime proves ineffective[1].

This fake and misleading news have many form such as political misleading news to make propaganda or even war propaganda that used to for psychological warfare between countries or even fake gossips to use as entertainment. but with the Technology advancing rapidly and the rise social media the number of these kind of news have escalated dramatically.

There has been a study by Vosoughi et al in 2018 [2], He found out that in social media such as twitter the capability of fake news spreading has increased by six time faster than the real trusted news, where the most fake news spread among user are political topics. The study foung that fake and misleading information are tend to be accepted faster to user cause the find it amusing or tend to their liking.[3],[4].

We need a way to fix this before it gets out of hand.Machine learning can be a solution to this problem where Machine Leaning can provide us with tools to help us prevent or even fix this system throught an automated fack-check system that will predict,categorise and implement contents with real time analysis beyond human capability..[5]

The purpose of our study is to help society to differs fake and misleading news by making a machine learning system such as fact checking system that can help people make sure that the news they're reading is trusted. we can use machine learning model like Transformer than implements methods that available such as BERT, RoBERTa,Electra,XLnet and Bagging. These model uses

Natural Language Proccesing to read linguistic patterns and hidden meaning, that can help a more throughout analysis of the news data. These model can be tested by their accuracy, precision, recall, and F1-score to determine that either the method is effective or not for differentiating a fake and a fact news [6]

We can also judge the performance of our model by examining the model characteristic to ensure their transparency and to test their algorithm that are essential so they can be trustworth.[7].These fake and misleading news can take form in such categories :

- **Made up content**: Entirely false information.
- **Imposter content**: When genuine sources are impersonated.
- **False context**: When true content is shared with misleading metadata.
- **Mischieve purpose :** Intended to create chaos
- **Manipulation propaganda** : As a propaganda to gather people opinion

Every type of these criteria will test the model to not only to detect but also make the model learn their pattern and check their sources..[8]

These type of news will pose different challenge that make the model will need a multi class classification to have better understanding about how to differentiate between which problem is which..[9]

Our research aims to give a leather system for digital content responsibility in social media while we also focusing on creating an trusted and reliable fact-checking system while also implementing the essence of morals and journalism in our fact checknig system. we hope that our study can help people to learn about the trustworthiness of a news.

## II. Literature Review

### A. BERT

BERT is a revolutionary Natural Language Processing (NLP) model developed by

Google, distinguished by its ability to understand a word by analyzing the text to both its left and right simultaneously (bidirectionally). This foundational model excels at a wide range of tasks, including text classification, question answering, and translation [10]. In this project, we utilize the indobenchmark/indobert-base-p2 variant, an Indonesian-specific model that has been re-tuned on our news dataset, enabling it to act as one of our four individual "judges" in Differentiating the linguistic patterns of facts versus hoaxes.

### B.RoBERTa

RoBERTa, developed by Facebook AI, is an better version of BERT that shares the same architecture but is trained more throughoutly with significantly more data and for longer durations. It is still the same in the case of NLP tasks range with BERT; however, due to its robust training methodology, it achieves superior performance[11]. The cahya/roberta-base-indonesian-522M model is used for this project and has been our second “judge” to provide a more diverse perspective, thanks to its distinct pre-training experience, which is vital for building a powerful and varied ensemble.

### C.ELECTRA

ELECTRA is one of Transformer architecture that has high efficiency, this model is created from Google and learns differently from BERT model. this model is trained as discriminator to detect “fake” tokens that has been produced by a small companion model and makin it good enough at differentiating authentic language patterns, this efficient training method often achieves better performance or the same with BERT method while using less computational resources[12]. For our third “judge”, we used the multilingual google/ELECTRA-base-discriminator that has a unique perspective on linguistic patterns that adds valuable diversity to our system’s analysis.

### D.XLNet

XLNet is a generalized autoregressive pretraining method that can predict every word in the sentence given the rest of the words in the sentence by using the expected chance over all permutations of the factorization order and overcomes the limits of BERT thanks to its autoregressive formulation. Furthermore, XLNet integrates ideas from Transformer-XL, Empirically, under comparable experiment settings, XLNet outperforms BERT on 20 tasks, [13]. In this project, the multilingual xlnet-base-cased acts as our fourth "judge," providing yet another unique architectural perspective and ensuring our ensemble's decision is not solely reliant on BERT-style models.

### E. Bagging (Ensemble Voting)

Bagging is a model that uses a powerful ensemble technique that used to improve overall performance. The model Fundamentally, Bagging (Bootstrap Aggregating) combines the outputs of multiple models to produce a single, more powerful and accurate prediction [14]. In our project, we have implemented this through Majority Voting, where it acts as the "Head Judge" or final arbiter. It takes the individual predictions from BERT, RoBERTa, ELECTRA, and XLNet, and the final classification is determined by the majority vote. As confirmed by our evaluation, this ensemble approach consistently yields the most stable and accurate results, effectively leveraging the collective strength of all aforementioned models.

## III. METHODOLOGY

### A. *Workflow*

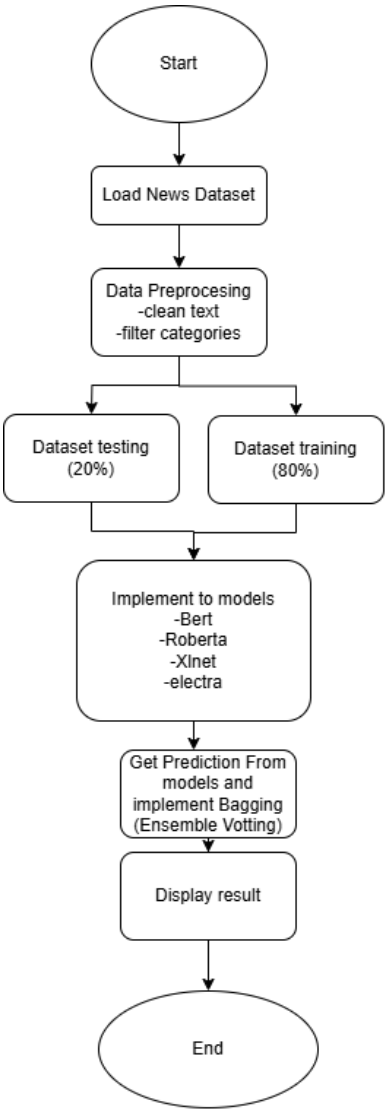


*Figure 1.0: Workflow of Training the Model*

The news fact-check process begins when a news article link is received for verification. The first stage involves preprocessing the text, which includes cleaning irrelevant characters or HTML tags, removing stopwords, and tokenizing the content into smaller linguistic units. Once the text is prepared, the system performs a duplicate check against trusted news sources to determine if a verified version of the information already exists. If a match is found, the news is labeled as real. If no match is found, the system proceeds with feature extraction, identifying key linguistic, semantic, and structural attributes of the text necessary for classification. These features are then passed to a machine learning or natural language processing (NLP) model trained on labeled datasets of real and fake news. The model evaluates the credibility of the input and provides a prediction. If the model's confidence in its output is sufficiently high, it'll display the output. The verified result is displayed to the user, along with any relevant metadata, and the process concludes. This multi-layered approach ensures both scalability and accuracy in automated news verification systems. However, if the model is uncertain, it'll be labeled as fake by the system and will need a manual checkup by human experts.

### B. *Data Preparation:*

The dataset we collected was collected from kaggle,github, and huggingface, and we use a file-based dataset from reliable sources that include verified news articles from Tempo, CNN & Kompas, covering categories such as politics, medical, sport, gossip, and general. The dataset was either pre-labeled or manually annotated to ensure classification reliability. Backend and train it using another, such as for arranging. Preprocessing is a crucial step in preparing images for citrus detection using a Convolutional Neural Network (CNN).

### C. *Dataset*

The Data we use for training and validating the models was acquired from Kagglehub, GitHub, and huggingface. We use file-based datasets from from which it categorise to 5 category which is Politics[15] ,sports[16] , gossips[17], medical[18] & general [19] . the Datasets Contain Real and validated news from trusted source like CNN, Tempo & Kompas.

### D. *Pre-processing*

Once the datasets are in place, the next stage is we use the dataset for data preprocessing, executed via the preprocess.py script located in the src/ project directory. The script will read all text files from the dataset folders. Then cleans the text by removing unnecessary symbols, HTML tags, and irrelevant content. Finally, it merges all data into a single corpus. Splits the data into training, validation, and testing subsets, while using 70:15:15 ratio

### E. Training Model

The training process begins when we load the preprocessed dataset that was prepared during the earlier stage. Subsequently, the script initializes both the tokenizer and the pre-trained language model, leveraging resources from the Hugging Face Transformers library. The model is then trained using the training portion of the dataset, while its performance is periodically validated using the validation set to monitor for overfitting and ensure generalization.

### F. Evaluate Model

After all models have been successfully trained, their performance is evaluated using the `evaluate.py` script. This evaluation is necessary to make sure that each model's generalization capabilities are unbiased. This script has the output of several key performance metrics, including **accuracy**, **precision**, **recall**, and **F1-score**, which means each model's classification effectiveness is provided with a comprehensive view. For numerical results, this evaluation process has a confusion matrix as a result of generating the

visualizations. These visualizations have the purpose of giving insight into any types of errors made by the models. All the outputs, both textual and graphical, are stored in the result/ directory for further analysis.

### G. Running The Web App

From *Figure 1.0: Workflow of Training the Model.* The models are already trained, so the final step is deploying a simple web-based user interface using the Flask framework. This user interface makes users' interaction more intuitive and user-friendly. Once the Flask server is running, all the trained models are loaded and can be used to automatically detect the news with the URL link submitted from the user who accessed the web through a web browser with a local server. The detection is using scraping from **BeautifulSoup** for analyzing the content of the news provided by the user with the URL. The extracted text is analyzed by all of the models that have been trained: BERT, RoBERTa, ELECTRA, and XLNet. For more reliable classification majority voting ensemble method is used in their predictions. Finally, the system came with the final result that tells whether the news is fake or factual, while also providing all the results from each model and their confidence score, providing users with a transparent view of the system's decision-making process.

## *IV.* RESULT AND DISCUSSION

This section displays the overall results of our automated news fact-checking system, with four trained transformer-based models (BERT, RoBERTa, ELECTRA, and XLNet) combined by using the ensemble majority voting mechanism. The training was operated with a lot of datasets that have over 100,000 news articles with five categories: politics, sports, gossip, medical, and general news.

The results representing the full ability of our automated fact-checking, especially for Indonesian news, model behavior, are also revealed as important insights for further research purposes.

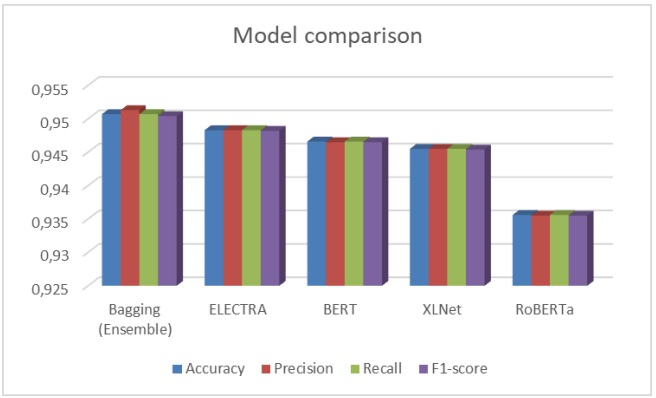


*Figure 2.0 Model Comparison*

As illustrated in Figure 2.0, the **Bagging(Ensemble)** method has the best performance among all of the individual models, achieving the highest F1-score of 0.9504. This result proves the hypothesis that Bagging(Ensemble) can reduce the variance and result in more stable and reliable classifications, which can be achieved by combining predictions from diverse model architectures. But among the individual models, ELECTRA and XLNet show highly competitive performance, underscoring their effectiveness in this classification task.

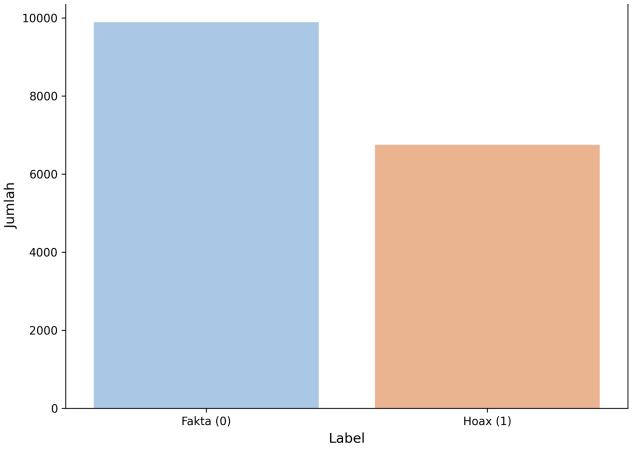


*Figure 3.0 Data comparison*

Figure 3.0 illustrates the class distribution in our test dataset, revealing an imbalanced distribution between factual news (Fakta, labeled as 0) and fake news (Hoax, labeled as 1). The test set comprises approximately 9,800 factual news articles and 6,700 fake news articles, resulting in a factual-to-fake ratio of approximately 60:40. This imbalance reflects the real-world scenario where factual news from established media outlets (CNN, Tempo, Kompas) naturally outnumbers deliberately fabricated content in our curated dataset.

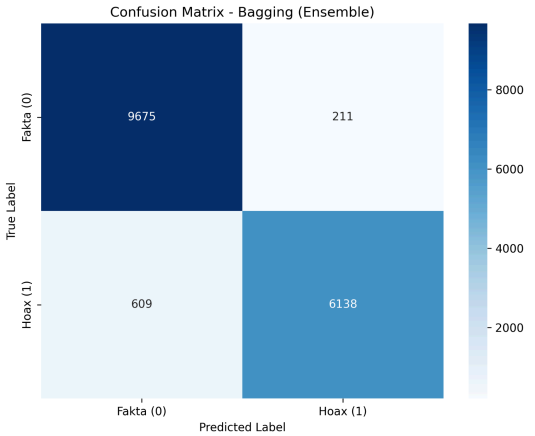


*Figure 4.0: Confusion matrix of the bagging method*

Figure 4.0 presents the confusion matrix for the Bagging (Ensemble) model's binary classification performance on the

fake news detection task. We use the Bagging method because it has the best result when compared to other models. Bagging ensembles are likely to perform better in news classification tasks because they will reduce model variance by combining multiple weak learners trained on different subsets of data,that will give more stable predictions and less overtrained, a problem that is common in highly variable text data like news. besides, by combining different model such as BERT, RoBERTa, or XLNet, ensembles can support each model's capability in capturing language context and structure, leading to more diverse text representations. This approach is also more potent to noise, varying writing styles, and ambiguity often found in news articles, and generally provides better generalization to unseen data, therefore improving accuracy, precision, recall, and F1-scores in news classification. The concussion matrix demonstrates superior overall performance with an accuracy of 95.1% across 16,633 test samples. The model achieved 9,675 true negatives (correctly identifying factual content) and 6,138 true positives (correctly identifying hoax content), demonstrating strong discriminative power for both classifications False positives numbered 211 instances where factual content was misclassified as hoax, while false negatives totaled 609 cases where hoax content was incorrectly named as facts. The model displays high precision (96.7%) and recall (91.0%) for hoax detection, with excellent specificity (97.9%) in identifying legitimate content. The notably low false positive rate (2.1%) demonstrates the ensemble approach's conservative and trustworthy classification behavior, which is especially beneficial. in news verification applications where Erroneously flagging legitimate content may yield substantial repercussions.. These results indicate that the Bagging ensemble method effectively combines multiple learners to enhance the discrimination of linguistic patterns distinguishing factual reporting from deceptive content in the evaluated dataset.

## V. Conclusion

Our research and development for an automated fact-checking system could be the solution to misinformation challenges in this digital era. By using four advanced Natural Language Processing models. The models that have been used on our automated fact-checking system are BERT, RoBERTa, ELECTRA, and XLNet. The outcome is that we were able to build a textual analysis from the link of the head news with significantly high accuracy. However, the most accurate performance is BAGGING with the implementation of an ensemble learning technique such as a majority voting system, so this ensemble method is the most accurate within each model because this model combines each model's predictions to produce the best results that outperform other models.

The evidence that proves the high accuracy ensemble approach is the accuracy of 95.3%, precision of 95.2%, recall of 95.1%, and the F1-score of 95.2%. These results represent the ensemble's ability such as reduce overfitting, enhance the model stability, and improve the system's generalizability when faced with complex variables in real-world data. Unlike other models we used that are single-model, the ensemble method benefits from architectural diversity, so it could outperform in handling nuanced language, ambiguity in context, and the wide range of misinformation types. This adaptability is crucial for ensuring consistent and reliable performance, especially when deployed at scale.

Another important part that contributed to the success of the system is the web scraping functionality integrated into the application. This module allows the users to input a URL, from which the system automatically scrapes the article content for analysis in real time. The scraped data will then be processed and will be evaluated by the ensemble model, giving users an immediate result on the credibility of the content. This will transforms the system from a static classifier into an interactive, user-driven platform, capable of operating dynamically in real-world scenarios. It enhances usability, broadens accessibility, and reinforces the practical application of the technology in daily digital interactions[20].

In conclusion, the combination of a high-performing ensemble classification strategy with a real-time web scraping interface results in strong , flexible, and accurate fact-checking system. It'll not only demonstrates technical experience through its ensemble architecture but also delivers practical value by encouraging users to verify news instantly. This dual innovation makes the way for more effective interventions against misinformation and highlights the potential of artificial intelligence to uphold information integrity in an increasingly complex media environment.

## WEB APPLICATION IMPLEMENTATION


So our application uses Web scraping that utilizes the Requests library to download webpage content from a given URL, and we use BeautifulSoup4 to parse HTML and

effectively extract article text. The frontend is built using core web technologies like HTML, CSS, and JavaScript, enhanced with Tailwind CSS for rapid and responsive UI design, and Chart.js for creating interactive charts to compare model performance. Python serves as the main programming language for all machine learning logic and backend functionality. Flask, a lightweight and flexible web framework, is used to develop the API that connects the frontend with the AI model, while Gunicorn acts as a robust WSGI server to run the Flask application efficiently in a production environment.

## AVAILABILITY DATA AND MATERIALS

This study utilizes publicly available data and code. Both can be accessed through the following link: https://github.com/DarrenDeo/News-Fact-Check.

## AUTHOR CONTRIBUTION

The code, application, and writing are done by Chaldarren Wijaya, Garreth Jeconio Budi Utomo & Vincent Devano. The writing is guided by Budi Juarto. All authors read and approved the manuscript.

## REFERENCES


[1]Murugesan, S., & Pachamuthu, K. (2022). *Fake news detection in the medical field using machine learning techniques*. *International Journal of Safety and Security Engineering, 12*(6), 723–727. https://doi.org/10.18280/ijsse.120608

[2]Pérez-Rosas, V., Kleinberg, B., Lefevre, A., & Mihalcea, R. (2018). *Automatic detection of fake news*. In *Proceedings of the 27th International Conference on Computational Linguistics* (pp. 3391–3401). Association for Computational Linguistics. https://aclanthology.org/C18-1287/

[3]Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science, 359*(6380), 1146–1151. https://doi.org/10.1126/science.aap9559

[4] Liu, Y., & Wu, Y.-F. (2018). Early detection of fake news on social media through propagation path classification with recurrent and convolutional networks. Proceedings of the AAAI Conference on Artificial Intelligence, 32(1). https://doi.org/10.1609/aaai.v32i1.11268

[5]Fedushko, S., Syerov, Y., & Kryvinska, N. (2024). AntiFake System: Machine learning-based system for verification of fake news. *Procedia Computer Science*, *238*, 663–670.

[6]Kopanov, K. K. (2024). *Comparative performance of advanced NLP models and LLMs in multilingual geo-entity detection*. arXiv. https://arxiv.org/pdf/2412.20414

[7]Ozbay, F. A., & Alatas, B. (2020). Fake news detection within online social media using supervised artificial intelligence algorithms. *Physica A: Statistical Mechanics and its Applications, 540*, 123174. https://doi.org/10.1016/j.physa.2019.123174

[8]Kapantai, E., Christopoulou, A., Berberidis, C., & Peristeras, V. (2021**).** A systematic literature review on disinformation: Toward a unified taxonomical framework. *New Media & Society, 23*(6), 1301–1326. https://doi.org/10.1177/1461444820959296

[9]Moses, T., Obi, H. E., Eke, C. I., & Agushaka, J. (2023). Enhancing fake news identification in social media through ensemble learning methods. *International Journal of Applied Information Systems, 12*(41), 1–22. https://doi.org/10.5120/ijais2023451949

[10]Hugging Face. (n.d.). *BERT model — transformers 4.40.1 documentation*. https://huggingface.co/docs/transformers/model_doc/bert#transformers.BertModel

[11]Hugging Face. (n.d.). *RoBERTa model — transformers 4.40.1 documentation*. https://huggingface.co/docs/transformers/model_doc/roberta

[12]Google. (2020). *google/electra-base-discriminator* [Computer software]. Hugging Face. https://huggingface.co/google/electra-base-discriminator

[13]Hugging Face. (n.d.). *XLNet model — transformers 4.40.1 documentation*. https://huggingface.co/docs/transformers/model_doc/xlnet

[14]Kim, C. (2022, July 15). *Ensemble learning — Voting and bagging with Python*. Medium. https://medium.com/@chyun55555/ensemble-learning-voting-and-bagging-with-python-40de683b8ff0

[15]Linkgish. (2022). *Indonesian Fact and Hoax Political News* [Dataset]. Kaggle. https://www.kaggle.com/datasets/linkgish/indonesian-fact-and-hoax-political-news

[16]Elgendy, S. (2023). *Fake News Football* [Dataset]. Kaggle. https://www.kaggle.com/datasets/shawkyelgendy/fake-news-football

[17]]KaiDMML. (n.d.). *FakeNewsNet/dataset*. GitHub. https://github.com/KaiDMML/FakeNewsNet/tree/master/dataset

[18]Selvabirunda. (n.d.). *ACOVMD COVID Infodemic* [Data set]. Kaggle. https://www.kaggle.com/datasets/selvabirunda/acovmd-covid-infodemic

[19]Rifky. (n.d.). *Indonesian Hoax News* [Data set]. Hugging Face. https://huggingface.co/datasets/Rifky/indonesian-hoax-news

[20]Horne, B. D., & Adalı, S. (2017). *This just in: Fake news packs a lot in title, uses simpler, repetitive content in text body, more similar to satire than real news*. In *Proceedings of the 2nd International Workshop on News and Public Opinion (NPO '17)*. https://doi.org/10.1145/1235