AnCore – News Article Credibility Assessment and Fake News Detection

Christian Kim C. Calizo
Technological Institute of the Philippines
Quezon City, Metro Manila Philippines
qckccalizo@tip.edu.ph

Rainier Franz A. Dejoras
Technological Institute of the Philippines
Quezon City, Metro Manila Philippines
qrfadejoras@tip.edu.ph

Lance Bryan A. Lapitan
Technological Institute of the Philippines
Quezon City, Metro Manila Philippines
qlblapitan@tip.edu.ph

Lance Arvin D. Macaganda
Technological Institute of the Philippines
Quezon City, Metro Manila Philippines
qladmacaganda@gmail.com

Bien Jester O. Tuplano
Technological Institute of the Philippines
Quezon City, Metro Manila Philippines
qbjotuplano@gmail.com

ABSTRACT

This paper presents AnCore, a news article credibility assessment and fake news detection system that leverages the Multilingual Bidirectional Encoder Representations from Transformers (mBERT) model to analyze the authenticity of digital news content. The system aims to address the growing problem of misinformation by integrating emotional and linguistic analysis with source verification. Using the Fake News Filipino dataset, the model was trained to distinguish between credible and misleading Filipino-language news articles. Results show that mBERT achieved reliable accuracy in detecting misinformation and provides users with confidence scores for evaluating article trustworthiness.

1. INTRODUCTION

In today's digital age, the rapid growth of digital media has changed the way people access and share information. Nowadays, many individuals rely on social media and online news platforms as their main sources of information. While this development allows quick access to current events, it also increases the spread of false or misleading information, often referred to as fake news. A study has claimed that emotionally charged news articles are more likely to spread online because users tend to share content that triggers strong feelings such as anger or fear, regardless of its accuracy. (Adeeb R.A; Mirhoseini M., 2023). This behavior makes fake news a serious concern in today's society.

The influence of fake news goes beyond social media. The University of Derby explains that misinformation can shape public opinion, weaken trust in legitimate news sources, and even affect decision making during elections or health crises. (University of Derby 2023). When people are repeatedly exposed to false information, they may begin to doubt reliable sources and believe inaccurate claims. This problem highlights the importance of finding effective ways to check the credibility of online content.

Although there are existing systems that detect fake news, many of them mainly focus on technical aspects such as word patterns and source reliability. However, research shows that emotions also play a major role in how people perceive and share information online. (Adeeb R.A; Mirhoseini M., 2023). This indicates the need for a more comprehensive approach that includes both factual and emotional factors in evaluating news credibility.

To address this issue, the proposed study introduces a system that aims to evaluate the trustworthiness of news articles using a combination of language analysis, source verification, and emotional assessment. The goal is to help users identify misleading content more easily and promote responsible information sharing in the digital space.

This research addresses these challenges by proposing AnCore – an AI-based system that evaluates the credibility of news articles through text classification and emotion-aware analysis. The proliferation of digital media platforms has transformed how people consume information. While access to online news and social media content has become more convenient, it has also increased the spread of misinformation and fake news. The system aims to help Filipino readers identify fake news effectively and promote digital literacy. The study utilizes mBERT, a multilingual language model capable of processing Tagalog and English texts, allowing better contextual understanding of Philippine news articles

2. THEORETICAL FRAMEWORK

The system's design is grounded on theories of natural language processing (NLP), sentiment analysis, and fact-checking principles. NLP enables machines to interpret human language, while emotion analysis identifies tone and bias in writing. AnCore employs mBERT, a multilingual transformer model trained on over 104 languages, to extract semantic representations of text and classify whether an article is fake or credible. The framework integrates factual consistency and emotional neutrality to enhance accuracy.

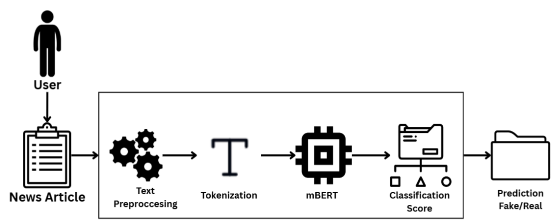

Figure 1. System Architecture: Text Processing

The first stage of the process begins when the user inputs a news article or URL into the system. The text is collected and passed through the text preprocessing module, where special characters, punctuation marks, hyperlinks, and stopwords are removed. This preprocessing step ensures that the dataset remains consistent and ready for machine interpretation. After cleaning, the text is tokenized into smaller units or tokens that the mBERT model can understand and process efficiently.

The mBERT model (Multilingual Bidirectional Encoder Representations from Transformers) is at the core of the AnCore system. mBERT works by generating contextual embeddings—numerical vector representations of words that capture their meanings based on surrounding context. The model reads the input text bidirectionally, meaning it considers both left and right context when analyzing each word. This enables mBERT to understand subtle nuances in language, which is essential for detecting manipulation, exaggeration, or emotional bias in articles.

Mathematically, the mBERT encoder layer transforms a sequence of tokens $T = [t_1, t_2, t_3, ..., t_n]$ into corresponding embeddings $E = [E_1, E_2, E_3, ..., E_n]$ through self-attention mechanisms defined by:

$$Attention(Q,K,V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

where Q, K, and V represent the query, key, and value matrices derived from the input embeddings, and d_k is the dimensional scaling factor. This attention mechanism allows the model to focus on the most relevant parts of the article when determining meaning and tone.

Once the embeddings are produced, they are passed into the classification layer, which analyzes patterns that indicate deception, misinformation, or emotional distortion. This layer integrates linguistic and semantic indicators, such as sensational phrases or extreme sentiment polarity, with factual features derived from verified data sources.The system then computes a credibility score, expressed as a probability value between 0 and 1, where values close to 1 indicate a highly credible article, and those close to 0 suggest potential misinformation. A threshold value (typically 0.5) determines the classification label: Fake or Real. This process can be represented as:

$$\text{Prediction} = \{\, Fake,\ if\ P(fake)\ >\ 0.5$$
$$\{\, Real,\ if\ P(fake)\ \leq\ 0.5$$

The model's prediction is then displayed on the interface, where users can view the classification result along with an optional confidence level. Through this process, AnCore ensures transparency and interpretable results, promoting user awareness and digital literacy.

The framework also integrates emotion analysis to identify whether a news article employs emotionally charged language to influence readers. Emotionally heavy content—those that strongly express anger, fear, or outrage—is often correlated with misinformation. By analyzing sentiment polarity and emotional intensity, the system provides additional interpretive insights for each classification result.

Figure 2 below illustrates the process of text tokenization and embedding transformation, which forms the backbone of the mBERT feature extraction pipeline.

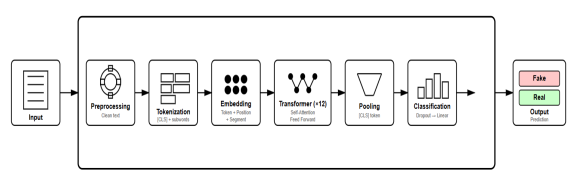

Figure 2. Tokenization and Embedding Generation

In this figure, input sentences are broken down into tokens, converted into vectors, and passed through multiple transformer layers that apply self-attention operations. These embeddings are then pooled and sent into the classification layer to predict credibility scores.

The system, through this theoretical framework, ensures a robust, context-aware, and linguistically informed approach to detecting misinformation. By combining advanced machine learning, linguistic modeling, and emotional analysis, AnCore provides a reliable architecture for analyzing digital news content and combating the spread of fake news.

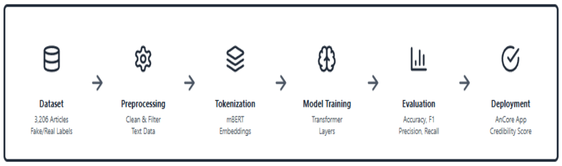

Figure 3. Model Training and Evaluation Process

Figure 3 illustrates the complete model training and evaluation workflow of the AnCore system. The process begins with the Fake News Filipino dataset, consisting of 3,206 labeled articles written in Filipino and Taglish. Each entry is categorized as either *Fake* or *Real* to provide the foundation for supervised learning. Before the data is used, it undergoes preprocessing steps such as removing special characters, stopwords, and irrelevant hyperlinks to ensure that the model focuses on meaningful textual content. The cleaned data is then tokenized and transformed into vector representations using mBERT's multilingual embedding layer, enabling the model to capture contextual and semantic relationships within the text.

During the training phase, these embeddings pass through multiple transformer layers that apply self-attention and feed-forward operations, allowing the system to detect complex linguistic cues that often characterize misinformation such as exaggeration, emotional bias, and manipulative tone. The classification layer interprets these encoded features to produce probability scores for each label (*Fake* or *Real*). To assess the model's performance, standard evaluation metrics such as accuracy, precision, recall, and F1-score are computed. These metrics help verify that the model performs consistently and avoids bias toward either class.

After evaluation, the best-performing model is deployed into the AnCore application, where it processes new articles in real time and assigns each a credibility score between 0 and 1. This score reflects the model's confidence in the article's authenticity, offering users clear and interpretable feedback. Through this training and evaluation framework, AnCore ensures that its fake news detection process remains data-driven, linguistically informed, and adaptable to the nuances of Filipino and multilingual online media.

3. RESEARCH METHODOLOGY

The research employed the Agile methodology as the developmental framework for the AnCore system, chosen for its iterative and incremental approach that accommodates the dynamic nature of machine learning development and system refinement. Agile methodology emphasizes flexibility, continuous improvement, and regular evaluation, making it particularly suitable for projects involving model training, hyperparameter tuning, and user interface development where requirements and performance metrics may evolve throughout the development cycle. The methodology is structured around short development cycles called sprints, typically lasting two to four weeks, where specific objectives are defined, developed, tested, and evaluated before proceeding to the next iteration (see Figure X). For the AnCore project, initial sprints focused on data collection, preprocessing, and mBERT model training using the Fake News Filipino dataset, while subsequent sprints addressed model fine-tuning, emotion analysis integration, and development of the classification pipeline. By employing this systematic yet flexible methodology, the research ensures that AnCore is developed through an evidence-based, user-centered process while remaining adaptable to technical discoveries and evolving understanding of the fake news detection problem in the Filipino context.

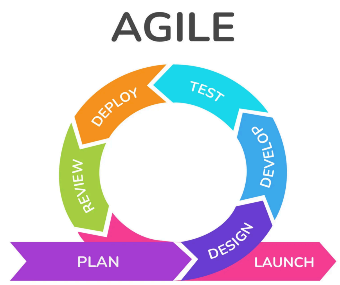

Figure 4. Agile Model for Development

4. RESULTS AND DISCUSSIONS

5.SUMMARY, CONCLUSION AND RECOMMENDATION

DATASETS

https://mediafutureseu.github.io/fakenewsdatasets.html
https://www.kaggle.com/datasets/emineyetm/fake-news-detection-datasets/discussion/521974

https://www.kaggle.com/datasets/saurabhshahane/fake-news-classification/code/discussion

https://onlineacademiccommunity.uvic.ca/isot/2022/11/27/fake-news-detection-datasets/

https://ieee-dataport.org/keywords/fake-news

REFERENCES

Adeeb, R. A., & Mirhoseini, M. (2023). The Impact of Affect on the Perception of Fake News on Social Media: A Systematic Review. Social Sciences, 12(12), 674–674. MDPI. https://doi.org/10.3390/socsci12120674

Orhan, A. (2023). Fake news detection on social media: the predictive role of university students' critical thinking dispositions and new media literacy. Smart Learning Environments, 10(1). https://doi.org/10.1186/s40561-023-00248-8

True or False? How much is fake news influencing our lives? (n.d.). University of Derby. https://www.derby.ac.uk/magazine/issue-12/influence-of-fake-news/

KaiDMML/FakeNewsNet: This is a dataset for fake news detection research

Fake News Detection Datasets

jcblaise/fake_news_filipino at main

https://www.canva.com/design/DAGzP_FDYgM/PRSpiTc3hFqwrF-L9VCxDg/edit?utm_content=DAGzP_FDYgM&utm_campaign=designshare&utm_medium=link2&utm_source=sharebutton