Ensemble Machine Learning for Identifying Fake News Headlines in Thailand

Komcharn Nitrat, **Nopparuj Suetrong**,
Natthanan Promsuk[†], and **Kanok Konglar**, Non-members

ABSTRACT

Advancements in information technology (IT) have rapidly transformed news dissemination across online platforms, creating challenges in analyzing diverse, often non-credible news sources. This issue is especially notable in Thai headlines due to linguistic nuances and broad topic range. Consequently, individuals are susceptible to the dissemination of false information, resulting in emotional, financial, and stability damages. To address this concern, this research proposes the utilization of term frequency-inverse document frequency (TF-IDF) in conjunction with various machine learning (ML) models as a multi-classifier, i.e., multinomial naïve Bayes (MNB), k-nearest neighbor (KNN), support vector machine (SVM), and extreme gradient boosting (XGBoost) algorithms. All classifier models can classify fake news from the news headlines within the Thai context. To enhance the classification accuracy, ensemble ML techniques, such as fuzzy integral and blending, were applied. A dataset comprising news headlines from the Anti-Fake News Center Thailand, under the Ministry of Digital Economy and Society, is constructed for analysis. The proposed ensemble blending technique with logistic regression demonstrates a commendable accuracy rate of up to 97%, underscoring its efficacy in distinguishing authentic news from misinformation.

Keywords: Cybersecurity, Ensemble Learning, Fuzzy Integral, News Headlines Classification

1. INTRODUCTION

In modern times, social media and online platforms have become essential to our daily routines of individuals across all age groups. The internet serves as a primary used for connecting with others online and staying informed with the latest news through various sources, including social media networks and online news portals. Unauthorized news online spreads fake stories and misinformation, fostering public misconceptions and false beliefs. Moreover, due to the time-constrained, individuals may accept news without questioning its facts, leading to long-term negative effects on the economy and society [1-5]. The general decline in trust was notably worse during the COVID-19 pandemic [6], which witnessed a surge in misinformation and the adverse impacts of news consumption. Traditional media like television and print pose comprehension challenges, especially for younger and less educated demographics, who prefer visually-oriented platforms like Facebook and TikTok. This shift highlights the influence of digital media's immediacy among younger Thais. However, the rapid spread of misinformation on social media underscores the need for better verification mechanisms.

Before the advent of machine learning (ML) algorithms, the task of identifying and classifying fake news was primarily handled through manual fact-checking processes. However, when attempting to apply those manual procedures to the Thai language, there are unique challenges presented by the script and structure of the language. One of the most significant challenges is the "scriptio continua," a feature of written Thai that involves the absence of explicit delimiters between words or punctuation marks between sentences [7]. This characteristic complicates the process of text segmentation, making it difficult to separate individual words and sentences for analysis. Furthermore, the Thai language consists of a spectrum of registers, ranging from very formal to highly informal. The introduction of ML algorithms has revolutionized this field by facilitating the development of advanced computational techniques, especially in natural language processing (NLP). These algorithms enable researchers to automate the analysis of linguistic patterns and contextual cues within large volumes of news headlines, stimulating and improving the speed and accuracy of fake news detection. Many Thai people often read only the news headlines without delving into the full article, which can lead to misinterpretation and the spread of false beliefs. This phenomenon has inspired the idea of utilizing ML models to help distinguish between real and fake news based on the headlines alone.

The research harnesses these technological advancements by employing supervised ML models tailored specifically to the Thai language. These models are adept at navigating the intricate complicated features unique to Thai, enabling more precise classification of news headlines as either real or fake, significantly reducing the risks associated with misinformation. Furthermore,

Manuscript received on August 6, 2024; revised on August 20, 2024; accepted on September 6, 2024. This paper was recommended by Associate Editor Siraporn Sakphrom.

The authors are with Department of Computer Engineering, Faculty of Engineering, Chiang Mai University, Chiang Mai, 50200, Thailand.

[†]Corresponding author: natthanan.p@cmu.ac.th

Digital Object Identifier: 10.37936/ecti-eec.2525231.255364

ensemble learning techniques are utilized in this research to enhance the classification performance of individual ML models. These techniques typically exhibit a lower risk of overfitting, which enhances generalization. By integrating multiple models, errors in one can be compensated by the others, leading to more robust predictions.

The paper is organized as follows: Section 2 presents the proposed model, Section 3 presents the experimental results, and Section 4 concludes the paper.

1.1 Related Works

In the era of digital information, the dissemination of fake news has become an urgent concern, posing significant challenges to society. To counter this issue, researchers have turned to ML techniques for automated fake news classification, a critical task in NLP. This task involves transforming human language, which is a complex and nuanced means of communication into computational statistics and numerical representations. This conversion enables computers to perceive and understand the complexity of human language, leading the way for the application of advanced ML algorithms in distinguishing genuine news articles from their misleading counterparts.

In the domain of fake news detection research for non-Thai languages, existing deep learning (DL) models have concentrated on analyzing diverse features including news content, social context, and external knowledge [8]. For news content analysis, a notable approach is the temporally evolving graph neural network for fake news detection (TGNF) introduced in [8]. This model effectively classifies news into categories of rumors or real news, achieving an accuracy rate of up to 96.8% on the Ma-Weibo dataset, which is sourced from the Chinese social media platform Sina Weibo. Additionally, the FakeBERT model [9], which employs a BERT-based deep convolutional approach, has demonstrated high efficacy in detecting fake news in English, with an accuracy rate of 98.90%. FakeBERT integrates a one-dimensional deep convolutional neural network (1d-CNN) with bidirectional encoder representations from transformers (BERT) [10]. Furthermore, logistic regression has shown strong performance to detect the fake news in English, such as the ISOT dataset and the KDnugget dataset [11].

Recent research in Thai by [12] introduced an innovative method to identify cybercrime-related news, utilizing word segmentation and stop word removal. By applying term frequency (TF) for data normalization and applying ML techniques to a dataset of 5,000 news articles sourced from Thai news websites, the research yielded promising results. The support vector machine (SVM) algorithm appears as the top performer, achieving an accuracy of 93.66%. Furthermore, researchers [13] have explored the use of three ML models including SVM, naïve Bayes (NB), and neural network (NN), for fake news detection in the Thai language. Utilizing a dataset comprising 3,964 messages of the diverse news topics that collected from the Anti-Fake News Center Thailand website between September 2018 and September 2022, the researchers applied oversampling to address the inherent imbalance in the dataset. Normalization techniques were applied to rescale and transform the data, while the IsWordList feature was used to enhance model accuracy. The results revealed that the use of SVM combined with the IsWordList feature achieved the highest accuracy at 80%. Moreover, research [14] investigated the performance of classification models on Thai news articles sourced from three mainstream websites including Thai PBS, Khaosod, and Dailynews by using web scraping techniques. The collected corpus of 6,000 articles were proceeded with the pre-processing pipeline, including word segmentation. TF was applied for feature extraction. The research evaluated three classification algorithms: decision tree (DT), SVM, and multilayer perceptron (MLP) which is a variant of deep learning (DL). The results highlighted the performance of MLP, achieving an impressive accuracy of 95%.

The previously mentioned research papers have explored various methods and processes for data preparation and utilized diverse ML and DL models for fake news classification.

1.2 Contribution

This research enhances fake news detection in Thai headlines by refining feature extraction methods. The feature extraction approaches use the TF method, which does not account for the importance of word frequency within the dataset. The research focuses to bridge this gap by investigating the use of the term frequency-inverse document frequency (TF-IDF) approach for feature extraction, examining how word frequency and the inclusion of common or uncommon words influence the classification of Thai news headlines. Evaluation of the performance form different ML algorithms is calculated in standard evaluation metrics. Furthermore, ensemble techniques (i.e., Choquet integral, Sugeno integral, and blending technique) combined are proposed and applied to enhance decision-making and improve overall accuracy. To conduct the research, this research collected a dataset of 2,570 Thai news headlines from the Anti-Fake News Center Thailand website, under the Ministry of Digital Economy and Society Thailand, spanning from August 18, 2023, to March 18, 2024.This timeframe in Thailand coincided with a period marked by numerous fraudulent activities, highlighting the relevance and timeliness of the research endeavor.

2. PROPOSED MODEL

This section provides an overview of the proposed models in this research. It begins with the data collection process from the Anti-Fake News Center Thailand by the Ministry of Digital Economy and Society, Thailand, followed by comprehensive data preparation. This preparation involves several steps, including word tokenization, stopword removal, punctuation removal,

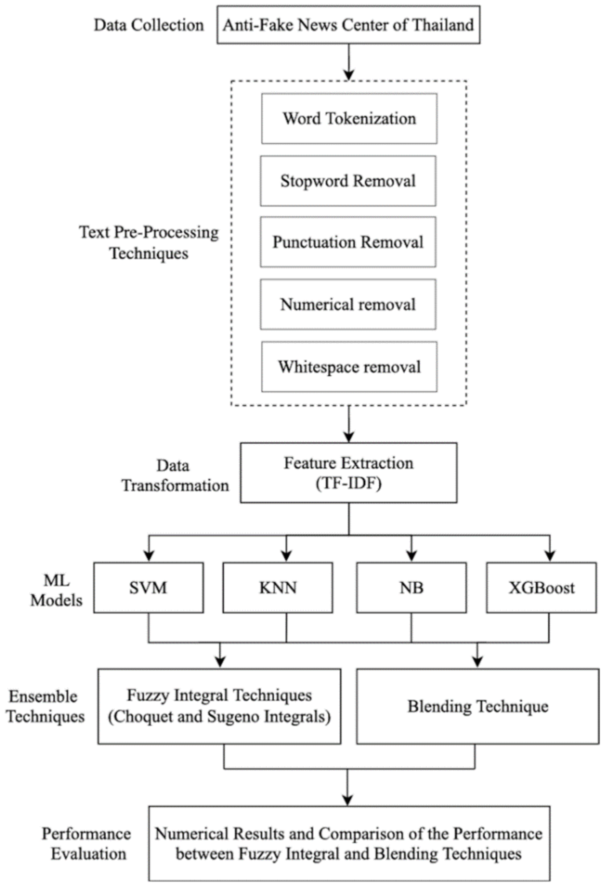


Fig. 1: *Overview of this research workflow.*

numerical removal, and white space removal, to ensure the data is suitable for input into the TF-IDF approach. Subsequently, the transformed data obtained from TF-IDF techniques is utilized to individually train four ML models and predict the probability on the test set. Then, the probabilities of each model will proceed to two different ensemble techniques including fuzzy integral techniques and blending techniques to improve the performance of models. Finally, the performance of fuzzy integral techniques and blending techniques is assessed by numerical results and a comparison is provided. The experiment's overview is depicted in Fig. 1 as follows.

2.1 Thai News Headlines Dataset

Thai news headlines were collected between August 18, 2023, and March 18, 2024, utilizing the Python library named Beautifulsoup4. The resultant dataset comprised 2,570 Thai news headlines sourced from the Anti-Fake News Center Thailand [15], an online platform operating under the Ministry of Digital Economy and Society. This selection was made to ensure comprehensive validation of news headlines across Thailand. The dataset from Anti-Fake News Center Thailand was chosen for its credibility, as it operates under the Ministry of Digital Economy and Society, verifying news accuracy. This ensures the dataset is grounded in authentic information.

Table 1: *The number of reliable and unreliable news headlines in the dataset.*

Label	Amount
Reliable news headlines ("ham")	1,285
Unreliable news headlines ("spam")	1,285

Although Thailand has various news channels, spanning online news websites, social media platforms, and television broadcasts, some sources disseminate information containing misleading claims or biases, leading to public misunderstanding. While the Anti-Fake News Center Thailand website, operated by the government, aims to provide verified information, its inherent limitations and potential for bias raise apprehensions regarding its absolute objectivity. Consequently, these factors impact the data collection process, resulting in a smaller volume of news headlines than initially anticipated.

- Data annotation and description

News headlines from the Anti-Fake News Center Thailand were categorized into "ham" (authentic) and "spam" (fake) groups. These annotations encompass both genuine and fabricated news headlines, as the Anti-Fake News Center Thailand employs a rigorous verification process to evaluate the accuracy of headlines. The dataset comprises a total of 2,570 rows, consisting of 1,285 rows of reliable news ("ham") and an equal number of rows of unreliable news ("spam"), as shown in Table 1. These headlines encompass a variety of topics, including international affairs, politics, health, dangerous substances, cosmetics, business and finance, stocks and exchange markets, economics, science and technology, crime, and general news.

2.2 Data Preparation

Textual data frequently encompasses noise, including punctuation, numbers, symbols, and assorted irregularities, all of which can compromise data consistency, thus affecting the performance and accuracy of NLP models. Data preparation is a vital step involving the cleansing and organization of the dataset to ensure its suitability for analysis. In the context of this research, data preparation entails five processes as follows: (i) Word tokenization refers to the process of segmenting a sequence of text into smaller units, known as tokens. These tokens can vary in size, ranging from individual characters to complete words. This process enables the text to be broken down into manageable units for subsequent analysis and processing. (ii) Stopwords are words that carry little significance in certain NLP tasks, such as information retrieval and classification. Stopword removal involves the elimination of words that frequently occur across all documents in a corpus. The authors in [16] are employed to filter out Thai stopwords. This process cleanses the text by removing words that do not significantly contribute to the classification task. (iii) Punctuation refers to a set of symbols utilized in written language to denote structure and clarify meaning. Punctuation removal involves the elimination of these symbols from the text, simplifying the content and allowing focus on the words themselves. (iv) Numerical removal filters out numbers from text data, as they often lack linguistic significance and may interfere with NLP tasks focused solely on text. Finally, (v) Whitespace removal filters out spaces, tabs, and newline characters from text, ensuring consistency and facilitating subsequent analysis.

2.3 Data Transformation

This dataset was assembled into a dataset named "Thai News Headlines," comprising two features: category (ground truth) and message. The dataset served as the foundation for building the ML model. To prepare the data for ML, the TF-IDF technique was applied to transform the text data into a vectorized form. This vectorization process allows the ML model to determine the significance of each word within the context of the document, enabling it to effectively learn and classify the news headlines based on their content.

2.4 Machine Learning (Multi-Classifier)

The dataset was partitioned into training and testing subsets utilizing Scikit-Learn, a Python machine learning library. Specifically, 80% of the data, corresponding to 2,056 samples, was allocated for model training, while the remaining 20%, equivalent to 514 samples, was designated for testing and validation. This division strategy ensures that the model is trained on a substantial portion of the dataset, while a separate, unbiased subset is retained for the evaluation of the model's performance.

Four ML models were selected for classifying ham and spam news from Thai news headlines. The following list outlines these models:

• **Multinomial Naïve Bayes (MNB):** MNB is a variant of the Naïve Bayes algorithm designed for handling discrete data or data in multinomial distributions, making it particularly suitable for text classification and other NLP tasks [17]. Naïve Bayes is a probabilistic classifier that relies on Bayes' Theorem, which establishes that the probability of an event occurring is proportional to the likelihood of observing specific evidence [18]. In the case of MNB, it computes the likelihood of encountering a particular word or term given a specific class. This computed likelihood is then used to classify new documents, as indicated in Eq. (1)

$$P(c \mid D) = \frac{P(c) \cdot \mathrm{P(D \mid c)}}{P(D)}, \tag{1}$$

where c denotes a class ("spam" or "ham"), D represents an observed Thai news headline, which consists of words $x_1, x_2, \ldots, \mathrm{x_n}$, $P(c \mid D)$ is the probability of class c given the headline D, $P(c)$ is the prior probability of class c, $P(D)$ represents the probability of observing headline D, and $P(D \mid c)$ is the probability of observing headline D given class c.

Furthermore, in MNB, it is assumed that the words are conditionally independent given the class. This assumption leads to the following expression for $P(D \mid c)$, which is the product of the probabilities of each word given the class:

$$P(D \mid c) = P\left(x_1, x_2, \ldots, \mathrm{x_n} \mid \mathrm{c}\right) = \prod_{\mathrm{i=1}}^{\mathrm{n}} P\left(x_{\mathrm{i}} \mid c\right), \tag{2}$$

where $P\left(x_{\mathrm{i}} \mid c\right)$ is calculated using a maximum likelihood estimation (MLE).

In this research, the MNB model was trained on the training dataset with an alpha value of 0.1 and prior probability of the classes set to [0.2, 0.8]. These hyperparameter settings were selected to optimize the model's performance, taking into account the data's specific characteristics and the distribution of classes within the dataset.

• **Support Vector Machine (SVM):** SVM is a supervised ML algorithm introduced by Vapnik and Chervonenkis [19], [20], which is typically used in binary classification tasks [21] and can be applied to text classification tasks. The SVM finds an optimal hyperplane to separate the data into different classes by maximizing the margin between the classes and minimizing the norm of the weight vector which is subject to the constraint that each data point is correctly classified within the specified margin, the decision function is shown in Eq. (3)

$$f(x) = w \cdot x + b, \tag{3}$$

where $f(x)$ is the decision function, w is the weight vector perpendicular to the hyperplane that determines the orientation in the feature space, x is the input vector, and b is the bias term controls the position of the hyperplane along the direction defined by w.

The sign of $f(x)$ determines the class of a given data point x as follows: if $f(x) > 0$ the x is classified as belonging to one class, on the other hand, if $f(x) < 0$ the x is classified belonging to the other class. The equidistant lines above line $w \cdot x + b > 0$ and below line $w \cdot x + b < 0$ of the hyperplane refer to the parallel lines that define the margin which is determined by support vectors or the data points that are closest to the hyperplane. The margin is the distance between the hyperplane and the closest data points of each class where a larger margin leads to better generalization performance of the classifier, the margin of SVM is shown in Eq. (4)

$$\rho = \frac{2}{|w|}, \tag{4}$$

where ρ is the margin of SVM and $|w|$ is the Euclidean norm of w.

SVM using a sigmoid kernel is a variant of the SVM algorithm that relies on a sigmoid function based on a logistic function and maps the input data into high-dimensional space, which is the non-linear kernel, enabling the model to capture complex relationships between the input features. The formula of the sigmoid function can be expressed as Eq. (5)

$$K(x, y) = \tanh(\gamma \cdot \mathrm{x} \cdot \mathrm{y} + \mathrm{r}) \tag{5}$$

where K is the kernel function, x and y are the input vectors, γ is the kernel parameter that controls the width of the Gaussian distribution, and r is the kernel parameter that represents the bias term.

In reality, not all characteristics of data can be linearly separated by one straight line efficiently which requires specified approaches to handle. The kernels come from a set of mathematical functions used in SVM which provide different approaches to manipulate non-linear relationship data by transforming those data into high-dimensional spaces where the data can be linearly separable. The SVM kernels comprise as follows: the linear kernel for linearly separable data, the polynomial kernel for capturing non-linear relationships in high-dimensional mapping, the radial basis function (RBF) kernel for infinite-dimensional mapping using Gaussian radial basis functions, and the sigmoid kernel, leveraging the hyperbolic tangent (tanh) function to capture non-linear relationships.

The sigmoid kernel computes the tanh of the dot product between the input vectors x and y, scaled by γ, and adds the bias term r to map the input data into a high-dimensional space. This enables SVM to find a decision boundary that separates the classes. In this research, the SVM model was trained on the training dataset using a sigmoid kernel with a regularization of 10, no class weight was applied, and the estimation of class probabilities for each prediction was enabled. These settings were selected to optimize the model's capacity to classify the data accurately, reflecting different aspects of the SVM algorithm.

• **K-Nearest Neighbors (KNN):** KNN [22] is a non-parametric supervised learning algorithm used for classification and regression tasks. The algorithm predicts the class or value of a new data point based on the majority class or average value of the k nearest neighbors in the feature space, determined using a distance metric, typically Euclidean distance. For classification, KNN assigns the new data point to the most frequent class among the k nearest neighbors, while for regression, the algorithm predicts the value as the average of the k nearest neighbors. The choice of k is crucial, as this value determines the balance between model complexity and generalization ability. A smaller k leads to a more complex decision boundary, potentially overfitting the training data, while a larger k results in a smoother decision boundary, possibly underfitting the data. In this research, the KNN model was trained with 7 neighbors with Euclidean distance of 2, and used weighted by inverse distance, chosen to tune the model's performance for classifying the dataset.

• **Extreme Gradient Boosting (XGBoost):** XGBoost has been developed to increase the effectiveness, adaptability, and scalability of an ML system for tree boosting by optimizing under gradient boosting algorithm [23]. The fundamental strategy of XGBoost involves aggregating a series of weak classifiers with low accuracy to construct a robust classifier capable of achieving better classification performance. Specifically, when the weak learner at each iteration is based on the gradient direction of the loss function, this methodology is referred to as Gradient Boosting Machines. The objective function (L) of XGBoost is tailored to minimize the combination of the loss function and regularization term, expressed as follows:

$$L = \sum_{\mathrm{i}=1}^{\mathrm{N}} l\left(\hat{y}_{\mathrm{i}}, y_{\mathrm{i}}\right) + \lambda \sum_{\mathrm{k}=1}^{\mathrm{M}} \left|w_{\mathrm{j}}\right|, \tag{6}$$

where $i = 1, 2, 3, \dots, N$. l is the loss function of predicted output ($\hat{y}_{\mathrm{i}}$) and output (y_{i}). In Eq. (6), the $L1$ regularization term (lasso regression) represents the magnitude of the coefficient values by summing the absolute values of the weights (w_{j}) in the model, where $j = 1, 2, 3, \dots, M$. λ is the penalty term. For the objective function applied with $L2$ regularization term (ridge regression) can be expressed as:

$$L = \sum_{\mathrm{i}=1}^{\mathrm{N}} l\left(\hat{y}_{\mathrm{i}}, y_{\mathrm{i}}\right) + \lambda \sum_{\mathrm{k}=1}^{\mathrm{M}} w_{\mathrm{j}}^{2}. \tag{7}$$

The weight penalty is calculated by squaring the magnitude of the coefficients in Eq. (7), resulting in the summation of squared weights. In this research, the XGBoost model was trained on the training dataset using 29 trees, each with a maximum depth of 30 splits. *L2* regularization was employed to mitigate over-fitting problem. The learning rate was set to 0.0001. For binary classification—distinguishing between “ham” and “spam” —the objective function was logistic. These parameter selections aimed to optimize the performance and accuracy of the XGBoost classifier in discerning real and fake news from the Thai news headlines dataset.

2.5 Ensemble Techniques

Ensemble techniques involve combining multiple predictions from various base ML models to create a single, accurate, and robust outcome. This approach leverages the strengths and diversity of individual models to improve prediction accuracy beyond what is achievable with a single model. This research leverages the power of ensemble techniques to enhance the accuracy of predictions for Thai news headlines. To achieve this goal, different ensemble methods, including fuzzy integral techniques (Choquet integral and Sugeno integral) and blending techniques, are experimented.

2.5.1 Fuzzy Integral

The fuzzy integral is a mathematical concept derived from fuzzy set theory, designed to facilitate decision-making in the context of uncertainty and ambiguity.

Unlike traditional sets or crisp sets, which define strict inclusion or exclusion, fuzzy sets allow for degrees of membership, providing a more flexible and nuanced way to quantify how much an element belongs to a given set. Fuzzy integrals serve as a mechanism to aggregate or combine values based on a fuzzy measure, offering a sophisticated approach to handling uncertainty, interactions, and the relative importance of different factors. A fuzzy measure, also known as a fuzzy set function, assigns a value to subsets of a given set, reflecting varying levels of significance or relevance. This research utilizes two fuzzy integral techniques, the Choquet integral and the Sugeno integral, to enhance the decision-making capabilities of the base models. The robustness and accuracy of the predictions are aimed to be improved by leveraging these fuzzy integral approaches.

• **Choquet Integral:** Choquet integral is used in the area of fuzzy measures and decision-making processes to aggregate information from multiple sources where each source has its importance without considering how these sources interact with each other. The Choquet integral is defined for a finite set X, a collection of information sources $x_1, x_2, \ldots, x_n$, and a fuzzy measure g on X. Given a confident function $f : X \rightarrow [0,1]$ that assigns a confidence value to each element in X, the Choquet integral of f over X with respect to g is calculated as follows [24]:

$$\int_{c} f \circ g = \sum_{i=1}^{n} f\left(x_i\right)\left[g\left(A_i\right) - g\left(A_{i-1}\right)\right], \tag{8}$$

where $g\left(X_0\right) = 0$, $A_i = \{x_1, x_2, \ldots, x_i\}$, and $f(x_1) \geq f(x_2) \geq f(x_3) \geq \ldots \geq f(x_n)$.

• **Sugeno Integral:** Sugeno integral is used in the area of fuzzy measure and decision-making processes to aggregate information from multiple sources where each source has its importance by considering how these sources interact between them. The Sugeno integral is defined for a finite set X, a collection of information sources $x_1, x_2, \ldots, x_n$, and a fuzzy measure g on X. Given a function $f : X \rightarrow [0, 1]$ that assigns a confidence value to each element in X, the Sugeno integral of f over X with respect to g is calculated as follows [25]:

$$\int f \circ g = \max_{1 \leq i \leq n} \left(min \left(f\left(x_i\right), g\left(A_i\right) \right) \right), \tag{9}$$

where $A_i = \{x_1, x_2, \ldots, x_i\}$, and $f(x_1) \geq f(x_2) \geq f(x_3) \geq \ldots \geq f(x_n)$.

2.5.2 Blending techniques

Blending, also known as meta-ensemble, is a ML ensemble technique that combines or aggregates predictions from multiple base models to improve overall performance. Unlike stacking, where a meta-model is trained on out-of-fold predictions from base models, blending involves training a meta-model on predictions generated by base models using out-of-sample data. This difference in training data can influence the effectiveness and generalization of the ensemble approach. A logistic regression model is used as the meta-model for blending in our research. To achieve a more accurate and robust final prediction result, the predictions of various base models are aggregated. This blending approach provides a complementary way to enhance performance and reduce overfitting compared to using a single mode.

The proposed approach for utilizing two ensemble techniques, including Fuzzy integral and Blending technique involves the following key steps. First, commencing with the training of individual classifiers, including MNB, SVM, KNN, and XGBoost, each model undergoes separate training. Second, predicted probabilities for each class are obtained from each model. Subsequently, these predicted probabilities from the four ML models serve as input for both the Fuzzy integral and Blending techniques. This process yields an aggregated set of prediction probabilities. Following this, the softmax function is applied to these aggregated probabilities, thereby converting them into a final set of probabilities for each class. Finally, the argmax function is employed to determine the class with the highest probability. This approach enables the assessment of the performance of different ensemble techniques based on accuracy.

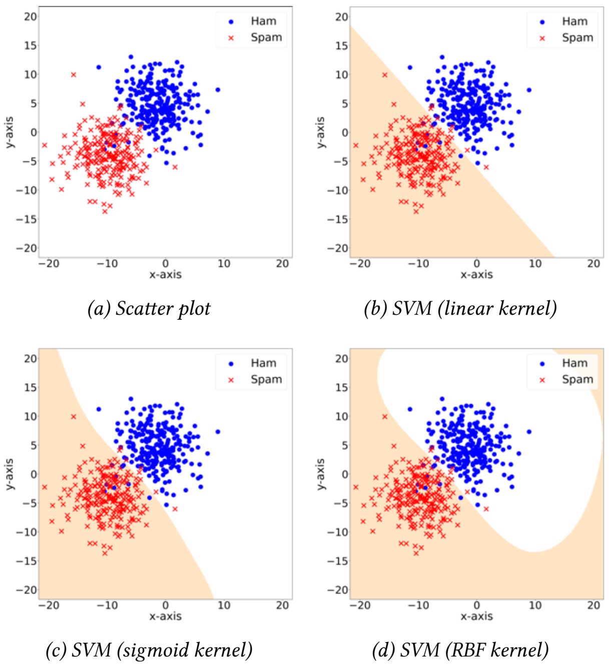
3. EXPERIMENTAL RESULT

Following the feature extraction step, the news headlines were transformed into a vectorized format suitable for classification using ML models. The results were derived after experimenting with the MNB, SVM, KNN, and XGBoost models utilizing the TF-IDF approach. The evaluation metrics used to assess the proposed models included the confusion matrix, accuracy, precision, recall, and F1-Score. The results from these metrics are presented in Table 2.

3.1 Evaluation of Multi-classifier Model

The comparison results from each classifier indicate that the SVM model with a sigmoid kernel achieved the highest accuracy, at a rate of 91%. The MNB model attained the second-highest accuracy, reaching 83%. Following, the XGBoost model ranked third with an accuracy rate of 82%, while the KNN model, with k equals to 7, secured the fourth-highest accuracy, at 73%. These findings underscore the relative performance of each ML model in classifying real and fake news from Thai news headlines.

The evaluation of the multi-classifier model also encompassed the F1-Score, precision, recall, and accuracy. To select the best representative from each model, this research investigated the performance of different kernels for SVM, including linear, sigmoid, and radial basis function kernels. Additionally, for KNN, the research explored the optimal k-value to select the most suitable representative. The results reveal that the



(a) Scatter plot *(b) SVM (linear kernel)*

(c) SVM (sigmoid kernel) *(d) SVM (RBF kernel)*

Fig. 2*: Decision boundaries of different SVM kernels. (a) represents the scatter plot of the dataset, while (b), (c), and (d) illustrate the SVM decision boundaries using linear, sigmoid, and RBF kernels, respectively.*

SVM with a sigmoid kernel and KNN with k equals to 7 emerge as the best representatives for forwarding to ensemble learning, as illustrated in Table 2. Furthermore, all metrics, including F1-Score, precision, recall, and accuracy (as presented in Table 2), highlight that the SVM model (sigmoid kernel) achieved the highest F1-Score, with a value of 0.91. The MNB model followed closely with the second-highest F1-Score of 0.82, while the XGBoost model achieved an F1-Score of 0.81. Lastly, the KNN model, using k set to 7, yielded an F1-Score of 0.73.

3.2 Decision Boundary of SVM Models

The data points and decision boundary of three SVM models are illustrated as follows, Figure 2(a) presents the data points of two classes of the data including ham as the blue circle mark and spam as the red 'x' mark, Figure 2(b) presents the decision boundary of SVM model using linear kernel, Figure 2(c) presents the decision boundary of SVM model using sigmoid kernel, and Figure 2(d) presents the decision boundary of SVM model using RBF kernel.

The linear kernel of SVM is the simplest kernel compared to the other kernels, it is useful for classifying the classes of data that are linearly separable. The sigmoid kernel of SVM is a non-linear kernel function that creates a non-linear decision boundary or a curve that attempts to separate classes by fitting a sigmoid-shaped curve to capture a logistic pattern in the data. The RBF kernel is another non-linear kernel function that transforms the input space into the higher-dimensional space where the data become linearly separable, it is useful for dealing with non-linearly separable data.

The data point (Fig. 2(a)) showed that there is a mix of both ham and spam classes near the point (-10, 0), and two clustered data points of both classes. The decision boundary of the SVM model using linear kernel, illustrated in Fig. 2(b) showed that the simple straight line can classify particular data points but is not efficient enough. The decision boundary of the SVM model using RBF kernel, presented in Fig. 2(d) showed that the model classifies the classes of data by creating the curve to cover ham data points and all beige color areas are the spam class, even though the model can classify the data efficiently, but the model still missed to capture some of data points near the boundary. The decision boundary of the SVM model using a sigmoid kernel, indicated in Fig. 2(c) showed that the model classifies the classes of data by fitting the sigmoid-shaped curve to capture the data, at this point the sigmoid kernel performed better to capture the boundary data points compared to RBF and linear kernel.

3.3 Confusion Matrix

The confusion matrix results, expressed as percentages for each classifier model, the confusion matrices are illustrated as follows, Figure 3(a) presents the confusion matrix of MNB model, Figure 3(b) presents the confusion matrix of SVM model with sigmoid kernel, Figure 3(c) presents the confusion matrix of KNN model, and Figure 3(d) presents the confusion matrix of XGBoost model.

From Tab. 2 and Fig. 3, the SVM model emerged as the most accurate model for identifying fake news headlines in the Thai language due to the TF-IDF value, as shown in Fig. 2. The TF-IDF data are more suitable for the sigmoid kernel than other kernels. Its high accuracy and F1-Score demonstrate its effectiveness in distinguishing between reliable and unreliable news headlines within the given dataset. This makes it a strong candidate for practical applications aimed at combating the spread of misinformation in the Thai context.

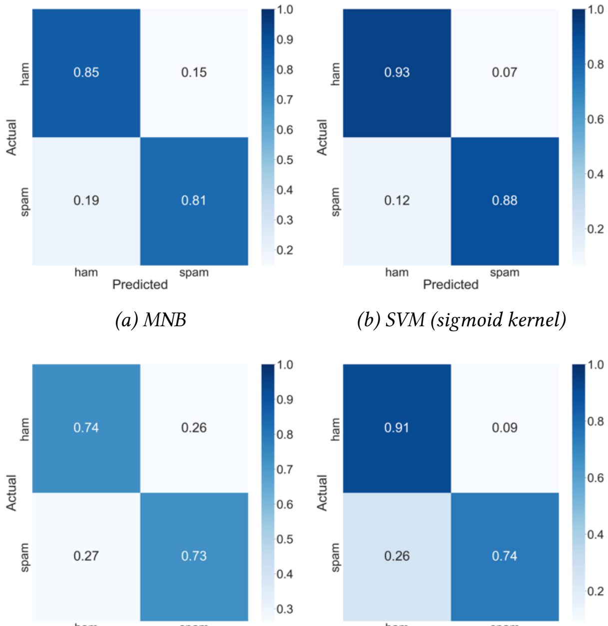
3.4 Ensemble Techniques Results

After collecting the classification results from the multi-classifier models, ensemble techniques are applied to enhance the research's performance. The comparison results of ensemble techniques indicate that the blending technique, using logistic regression as the meta-model, achieved the highest accuracy, with a rate of 97%. The Choquet integral attained the second-highest accuracy at 87%, while the Sugeno integral ranked third, with an accuracy rate of 86%. These findings underscore the relative performance of each ensemble technique in the decision-making of the ML model's probabilities. The comparison of the accuracy of individual classification models and the ensemble techniques is presented in Fig. 4. Figure 4 demonstrates that the TF-IDF values from the dataset align well with the sigmoid function due to the accuracy of SVM with the sigmoid kernel

***Table 2**: Comparison of performance matrices including accuracy, F1-Score, precision, and recall.*

Models	Accuracy (%)	F1-Score	Precision	Recall
MNB	82.87	0.82	0.81	0.85
SVM, linear	88.91	0.88	0.88	0.90
SVM, sigmoid	90.85*	0.91*	0.89*	0.93*
SVM, RBF	87.15	0.86	0.84	0.91
XGBoost	82.49	0.81	0.78	0.91
KNN, k=3	71.20	0.71	0.71	0.72
KNN, k=5	71.40	0.71	0.71	0.73
KNN, k=7	73.15	0.73	0.73	0.74

The optimal performance of a machine learning model is indicated by the highest values of accuracy, F1-Score, precision, and recall, which are denoted with an asterisk (*).



(a) MNB *(b) SVM (sigmoid kernel)* *(c) KNN (K=7)* *(d) XGBoost*

***Fig. 3**: Confusion matrices of four classifiers. (a), (b), (c), and (d) depict the confusion matrices for MNB, SVM with sigmoid kernel, KNN with K equals to 7, and XGBoost, respectively.*

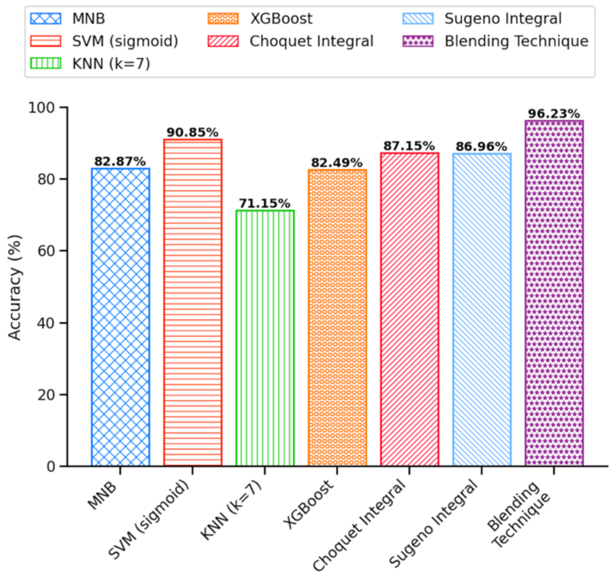


***Fig. 4**: Accuracy Comparison of Multi-Classifier (MNB, SVM with sigmoid kernel, KNN ($k = 7$), XGBoost) and Ensemble Learning (Choquet Fuzzy Integral, Sugeno Fuzzy Integral and Blending Technique).*

and the blending technique (where logistic regression is embedded with the sigmoid function).

4. CONCLUSION AND POSSIBLE FUTURE WORKS

This research introduced an ensemble machine learning (ML) technique that can classify Thai news headlines as either reliable (ham) or unreliable (spam). The dataset was collected from the Anti-Fake News Center Thailand. The proposed ensemble ML technique integrated predictions from four distinct ML models: multinomial naïve Bayes (MNB), support vector machine (SVM) with a sigmoid kernel, k-nearest neighbor (KNN) with k set to 7, and extreme gradient boosting (XGBoost). The choice of kernel in SVM and the value of k in KNN were determined based on their respective accuracy performances. The investigation of each classifier showed that SVM with a sigmoid kernel outperformed the others. Furthermore, blending techniques and fuzzy integrals (i.e., Choquet integral and Sugeno integral) were used to improve classification for each model. Results revealed that blending techniques with logistic regression achieved the highest accuracy rate of 97%. This suggests a strong relationship between the dataset, implemented term frequency-inverse document frequency (TF-IDF), and logistic function. The findings of this research could aid in the development of a fake news filtering system in Thai, mitigating the risks and harms associated with misinformation. Effective classification of real and fake news headlines plays a crucial role in upholding the integrity of information in the digital age.

Possible future works to enhance this research include expanding the dataset to incorporate diverse sources and multiple languages. Such an expansion could improve the model's robustness and generalizability across different linguistic contexts. Additionally, further investigation into the integration of advanced features in semantic analysis, including natural language processing (NLP) techniques such as word embeddings and contextual embeddings, represents a promising avenue for future research. Finally, addressing considerations for real-world applications is essential to ensure practical implementation and effectiveness in operational settings.

ACKNOWLEDGEMENT

This research was financially supported by the Faculty of Engineering at Chiang Mai University. We extend our gratitude to the Anti-Fake News Center Thailand, a division of the Ministry of Digital Economy and Society Thailand, for generously providing the Thai news headlines utilized in this research.

REFERENCES

[1] S. Lasse, "Official Truths in a War on Fake News: Governmental Fact-Checking in Malaysia, Singapore, and Thailand," *Journal of Current Southeast Asian Affairs*, vol. 40, no. 2, pp. 340-371, 2021.

[2] R. B. Smith and M. Perry, " "Fake News" Legislation in Thailand: The Good, the Bad and the Ugly," *Athens Journal of Law*, vol. 6, no. 3, pp. 243-264, 2020.

[3] A. A. Khan, M. Abou El-Ela and M. A. Al-Turaigi, "Current-mode Precision Rectification," *International Journal of Electronics*, vol. 79, no.2, pp.853-859, 1995.

[4] E. Broda and J. Strömbäck, "misinformation, disinformation, and fake news: lessons from an interdisciplinary, systematic literature review," *Annals of the International Communication Association*, vol. 48, no. 2, pp. 1-28, 2024.

[5] J. W. Cheng, H. Mitomo, A. Kamplean and Y. Seo, "Lesser evil? Public opinion on regulating fake news in Japan, South Korea, and Thailand A three-country comparison," *Telecommunications Policy*, vol. 45, no. 9, pp. 102-185, 2021.

[6] H. L. Dang, "Social Media, Fake News, and the COVID-19 Pandemic: Sketching the Case of Southeast Asia", *Austrian Journal of South-East Asian Studies (ASEAS)*, vol. 14, no. 1, pp. 37-52, 2021.

[7] A. Poncelas, W. Pidchamook, C.H. Liu, J. Hadley, and A. Way, "Multiple Segmentations of Thai Sentences for Neural Machine Translation," Proceeding of *1st Joint SLTU and CCURL Workshop (SLTU-CCURL 2020)*, May. 2020, Marseille, pp. 240-244.

[8] H. Linmei, W. Siqi, Z. Ziwang, and W. Bin, "Deep learning for fake news detection: A comprehensive survey," *AI open*, vol. 3, pp. 133-155, 2022.

[9] R. K. Kaliyar, A. Goswami, and P. Narang, "Fake-BERT: Fake news detection in social media with a BERT-based deep learning approach," *Multimedia tools and applications*, vol. 80, no. 8, pp. 11765-11788, 2021.

[10] M. F. Mridha, A. J. Keya, M. A. Hamid, M. M. Monowar, and M. S. Rahman, "A Comprehensive Review on Fake News Detection with Deep Learning," *IEEE Access*, vol. 9, pp. 156151-156170, 2021.

[11] T. Jiang, J. P. Li, A. U. Haq, A. Saboor, and A. Ali, "A Novel Stacking Approach for Accurate Detection of Fake News," *IEEE Access*, vol. 9, pp. 22626-22639, 2021.

[12] P. Boonyopakorn, N. Wisitpongphan, and U. Changsan, "Classifying Cybercrime and Threat on Thai Online News: A Comparison of Supervised Learning Algorithms," Proceeding of *2023 International Technical Conference on Circuits/Systems, Computers, and Communications (ITC-CSCC)*, Aug. 2023, Jeju, Korea, Republic of, pp. 1-6.

[13] K. Tiemtud and P. Leesutthipornchai, "Thai Fake News Detection Using Machine Learning Model," Proceeding of *2023 International Electrical Engineering Congress (iEECON)*, May. 2023, Krabi, Thailand, pp. 346-350.

[14] A. Noppakaow and O. Uchida, "Examinations on the Performance of Classification Models for Thai News Articles," Proceeding of *2019 11th International Conference on Information Technology and Electrical Engineering (ICITEE)*, Dec. 2019, Pattaya, Thailand, pp. 1-4.

[15] "Anti-Fake News Center Thailand", https://www.antifakenewscenter.com, Accessed: Mar. 18, 2024.

[16] W. Phatthiyaphaibun, K. Chaovavanich, C. Polpanumas, A. Suriyawongkul, L. Lowphansirikul, P. Chormai, P. Limkonchotiwat, T. Suntorntip, and C. Udomcharoenchaikit, "PyThaiNLP: Thai Natural Language Processing in Python," Proceeding of *3rd Workshop for Natural Language Processing Open Source Software (NLP-OSS 2023)*, Dec. 2023, Singapore, pp. 25-36.

[17] J. Rennie, L. Shih, J. Teevan, and D. Karger, "Tackling the Poor Assumptions of Naive Bayes Text Classifiers," Proceeding of *20th International Conference on Machine Learning (ICML-2003)*, Jul. 2003, Washington DC, USA, pp. 616-623.

[18] P. Nilpao, N. Nanta, N. Suetrong, and N. Promsuk, "Development of the recommended coffee shops application based twitter sentiment analysis," Proceeding of *37th International Technical Conference on Circuits/Systems, Computers and Communications (ITC-CSCC)*, Oct. 2022, Phuket, Thailand, pp. 305-308.

[19] V. Vapnik, Statistical Learning Theory, Wiley-Interscience, New York, 1998.

[20] C. Cortes and V. Vapnik, "Support-Vector Networks," *Machine Learning*, vol. 20, pp. 273-297, 1995.

[21] N. Promsuk, "Improving of the Interference Clas-

sification Techniques under the Smart Farming Environment using iSVM," Procceeding of *19th International Joint Conference on Computer Science and Software Engineering (JCSSE)*, Jul. 2022, Bangkok, Thailand, pp. 1-5.

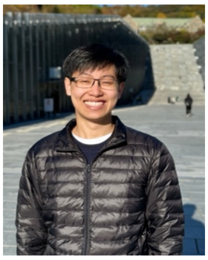
[22] T. Cover and P. Hart, "Nearest Neighbor Pattern Classification," *IEEE Transactions on Information Theory*, vol. 13, no. 1, pp. 21-27, 1967.

[23] C. Tianqi and G. PictureCarlos, "Xgboost: A scalable Tree Boosting System," Proceeding of *22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Aug. 2016, San Francisco, USA, pp. 785-794.

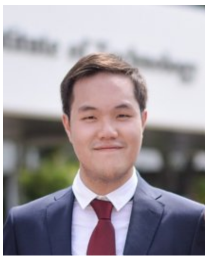
[24] S. Auephanwiriyakul, J. M. Keller, and P. D.Gader, "Generalized Choquet fuzzy integral fusion," *Information fusion*, vol. 3, no. 1, pp. 69-85, 2002.

[25] T. Murofushi and M. Sugeno, "Fuzzy t-conorm integral with respect to fuzzy measures: Generalization of Sugeno integral and Choquet integral," *Fuzzy sets and systems*, vol. 42, no. 1, pp. 57-71, 1991.

Komcharn Nitrat received a Bachelor of Engineering in Information System and Network Engineering in 2023 from Chiang Mai University, Chiang Mai, Thailand. Currently, he is studying for a Master of Engineering in Computer Engineering at Chiang Mai University. His research interests are natural language processing, cybersecurity, and machine and deep learning.



Nopparuj Suetrong received a Bachelor of Engineering in Information Systems and Network Engineering in 2021 and a Master of Engineering in Computer Engineering in 2024, both from Chiang Mai University, Chiang Mai, Thailand. He is currently a lecturer in the Department of Computer Engineering at Chiang Mai University. His research interests include cybersecurity, digital signal processing, wireless communication, and applications of machine and deep learning.



Natthanan Promsuk obtained a Bachelor of Engineering in Computer Engineering from Chiang Mai University (2014), a Master of Engineering (2017), and a Doctor of Engineering in Telecommunications (2020) from the Asian Institute of Technology (AIT), Pathum Thani, Thailand. Currently, he serves as an assistant professor in the Department of Computer Engineering at Chiang Mai University. His research focuses on digital signal processing, signal detection, interference suppression, Internet of Things, and machine/deep learning.

Kanok Konglar is a cybersecurity professional and educator with a focus on AI-driven security solutions for IoT systems. He holds Bachelor's and Master's degrees in Computer Engineering from Chiang Mai University. Kanok's industry experience as a Core Network Engineer at Huawei Technologies Thailand provided him with insights into real-world cybersecurity challenges. Currently, he lectures in Network Security and Information Security at Chiang Mai University. His research interests combine Cybersecurity, AI, and IoT, particularly in developing efficient machine learning algorithms for threat detection in IoT networks and enhancing AI system resilience against security threats.