



Effective detection of Thai fake news using machine learning method

Nopphadol Sumleeroung¹, Komsan Kanjanasit² and Nithizethe Mhuadthongon^{1*}

¹Information and Communication Technology, School of Science and Technology, Sukhothai Thammathirat Open University, Nonthaburi 11120, THAILAND

²College of Computing, Prince of Songkla University, Phuket Campus, Phuket 83120, THAILAND

*Corresponding author: Nithizethe.Mhu@stou.ac.th

ABSTRACT

The rapid spread of misinformation on digital platforms has become a critical challenge affecting public trust, social stability, and policy communication in Thailand. The aim of this study is to develop and evaluate a reproducible machine learning framework for Thai fake news detection from real world verified data from the Anti-Fake News Center Thailand during 2019 - 2024. The data set comprises four key data groups including governmental policies, health products, financial and stocks as well as disaster news. The suggested research methodology entails thorough data cleaning and text normalization, sentiment and emotions analysis through the usage of AI for Thai, dimension reduction by selecting relevant features and dealing with unbalanced classes through Synthetic Minority Oversampling Technique (SMOTE). For evaluation, three supervised learning algorithms (Naïve Bayes, Support Vector Machine [SVM] and neural networks) were tested using 80:20 and 90:10 train-test split ratios, K-Fold Cross-Validation and hyperparameter optimization. Based on the experimental findings, the neural network classifier achieved better accuracy rates, that is, 92.39%, with other performance metrics including precision, recall, F1-score, ROC-AUC and AUC-PR at 92.64%, 92.39%, 92.38%, 96.79%, 96.07% respectively. Therefore, the methodological framework was effective in identifying false news stories in the Thai language due to the use of linguistic features and oversampling strategies. The main contribution of this study is an interpretable baseline framework that integrates linguistic, affective, and metadata features for Thai fake news detection in a low-resource language setting.

Keywords: Fake news, Machine learning, Sentiment and emotion analysis

INTRODUCTION

The tremendous growth of internet-based platforms for information gathering has resulted in a huge shift in how information is produced and shared. These sources have been effective in improving the availability of information, yet they have at the same time facilitated the circulation of misinformation that is not backed up by evidence. In today's world, misinformation spreads rapidly due to algorithms designed to promote sensationalist or emotional information [1]. This means that besides being a communications issue, fake news has become a growing menace in society that causes misunderstanding, distrust of institutions, economic harm, and even risks to health and safety.

Fake news is not a recent phenomenon. However, due to the digitization and rapid mobile penetration, the impact of fake news has become significant in society. The findings of previous natural language processing (NLP) and machine learning (ML) studies show promise in detecting fake news automatically using linguistic, semantic,

and statistical textual features [2-4]. As early as 2015, the main approach to detect hate speech from online data sources was traditional text classification-based approaches. However, recent work has utilized advanced machine-learning and deep-learning architectures to enhance detection accuracy and robustness [5, 6]. Yet, there remains a limitation. Most of the existing in-house LLMs are developed and tested on English datasets, thus limiting their usability in non-English and low-resource language settings.

The phenomenon of fake news also poses problems for Thailand since people are now more inclined to access information through social media and messaging apps. Issues related to public health, economic policy, and disaster mitigation are among those being affected. Unfounded claims that involve government activities, dubious health products, scams, and others are frequently shared on popular platforms and sometimes lead to public hysteria, thus reducing public trust in communication from the government side. To address this problem, the Thailand Anti-Fake News Center was established.

The center is responsible for monitoring, verifying, and correcting fake news. As part of its operations, a systematic Thai-language verification dataset has been created.

There are some issues associated with fake news detection within the Thai language. Word boundaries are not easily distinguishable, and word order is quite flexible in this language; moreover, Thai is a highly contextual language. This adds up to the complexity of its automatic processing and feature extraction [7]. Segmentation mistakes can cause errors at the stage of feature extraction, leading to poor feature representation and performance. Moreover, this will add an extra layer of difficulty during feature engineering, which may potentially affect the model's performance and capacity to generalize knowledge. Another issue worth mentioning is the lack of annotated Thai corpora that consist of a substantial number of instances, and there is currently no benchmark or framework for evaluation.

Despite advancements in fake news detection in popular languages, the field of fake news detection in the Thai language still requires much attention and remains fragmented. A number of solutions depend on the task of text classification and are created specifically for small datasets, which does not fit well with the requirements of an online environment in Thailand. At the same time, there is no use of features related to the sentiments and emotions involved in posts by just a few researchers, even though the importance of the use of the content with affective properties was proven long ago. There have also been no systematic efforts to integrate national natural language processing tools like PyThaiNLP and AI for Thai into a unified model for fake news detection.

The suggested work introduces a machine learning approach to detecting fake news in Thai language, combining several kinds of features' representation with structural and contextual awareness. A fully developed and validated Thai dataset of fake news is used in the experiments; this database was created using data obtained from the Thailand Anti-Fake News Center. This database stands out because it is one of the first to consider public policy topics along with common ones. In addition to the database, a universal data processing pipeline is introduced and includes creating TF-IDF representations as well as integrating NLP models specific to Thai, for sentiment and emotion analysis, and contextual features like publication features. Moreover, a series of experiments on comparing the performance of Naïve Bayes, SVM, and Neural Network algorithms in various conditions, including feature selection and data balance.

The objective of this study is to develop and evaluate the machine learning framework for the Thai fake news detection based on linguistics, affective,

and metadata features extracted from the verified news data provided by the Anti-Fake News Center Thailand. In particular, the study compares the performance of Naïve Bayes, Support Vector Machine, and Neural Network classifiers under different train-test split configurations and examines the contribution of feature selection and class balancing to classification performance [8].

MATERIALS AND METHODS

Conceptual framework

The present research study design was constructed in such a way that various feature domains can be united into one modeling pipeline for fake news detection among Thai texts using an integrative approach. According to Figure 1, the framework used in the study can be described in three phases (i.e., Data Preparation, Processing, and Evaluation).

First of all, during the Data Preparation phase, textual data was obtained from a reliable source to assure the high quality of the data and its relevance to the real world. Noise reduction and data cleaning took place to enhance the overall quality. Moreover, affective features were also incorporated into the data set as feature representations. The structured data can be used for further processing and analysis.

The main objective of the second phase (i.e., Processing) is to transform the already structured data into feature representations suitable for classification purposes. During the initial stage of data processing, standardization of the data took place. Furthermore, after feature selection, the problem of class imbalance in the data set was solved by using oversampling methods. Thus, the framework was capable of facilitating the work of various classifiers.

The performance of the model was evaluated in the Evaluation stage with the validation strategy. The evaluation process involved splitting the data into training and testing sets, performing K-fold cross-validation, and conducting hyperparameter tuning. The system used evaluation criteria including accuracy, precision, recall, and F1-score to evaluate the model. The evaluation outputs of this stage were produced according to the model results in the framework.

Dataset

The dataset employed for this study was obtained from open data provided by the Anti-Fake News Centre Thailand (AFNC), which contained confirmed misinformation cases that were gathered by authoritative organizations based in Thailand [9]. The data retrieved contain news that has been published between the year 2019 and 2024. All these news items were individually labelled for their verification status, which is either real or fake. This allowed the use of supervised learning methodologies for carrying out the classification task.

The datasets covered various fields of public interest and were further divided into four main categories, namely, government policy and public

affairs, health products, finance and stock, and disaster-related information, as summarized in the Table 1.

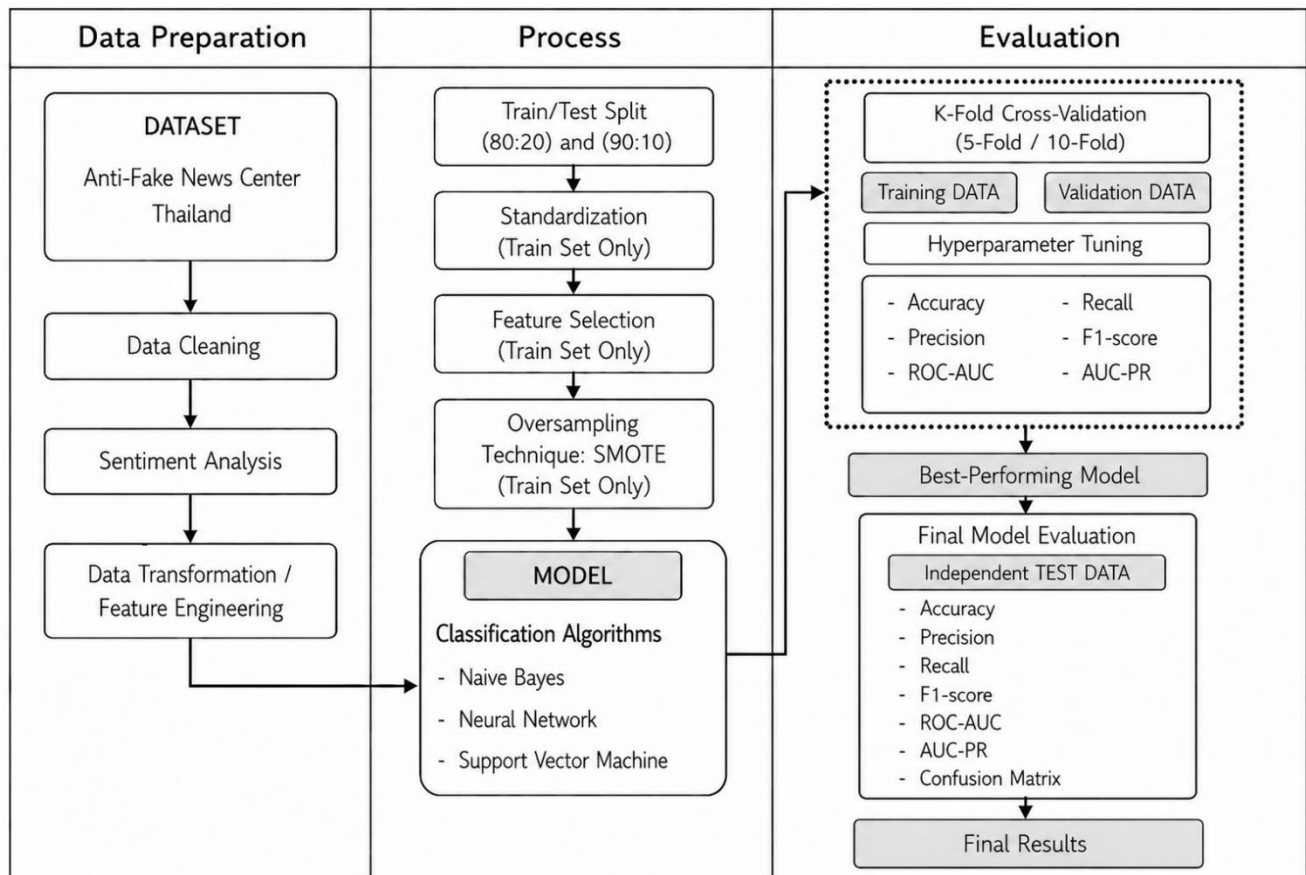


Figure 1 Proposed research framework.

Table 1 Number of news items by category.

Category	Number of Records
Government Policy and Public Affairs	2,737
Health Products	2,636
Finance and Stock	1,385
Disaster	256
Total	7,014

To ensure the quality of the data and reliability of the analyses, the datasets were initially screened to remove any incomplete, inconsistent, or irrelevant records. Moreover, any category with too few cases was excluded to avoid bias and instability in model training and assessment. Through filtering, a total of 7,014 news items were selected for model development.

Along with the textual data, the dataset also contains other attributes such as the date-time of publication, news URL, news headline, full news content, verifying agency, class label (real/fake), news category, and news hashtags, as given in Table 2 [9].

Data preprocessing and feature generation

The acquired dataset underwent data preprocessing and feature generation techniques to enhance its quality and suitability for subsequent analysis. First, the original text data underwent cleaning procedures through removal of unnecessary characters, solving any inconsistencies in the text format, and eliminating noise from the online news articles. Considering that the Thai language does not have distinct word boundaries, word segmentation was performed on the text by the PyThaiNLP library [10].

In addition, to incorporate semantic and affective aspects of the text, sentiment analysis was conducted using the AI for Thai web service through the use of its publicly available API [11, 12]. The output from the API included multiple variables such as sentiment polarity, sentiment score, and number of positive and negative words, which were employed as input features.

Sentiment analysis was followed by the conversion of the textual data to numerical form through the implementation of the Term Frequency-Inverse Document Frequency (TF-IDF) technique [13-15].

Furthermore, the sentiment values and text features were converted to structured numeric inputs ready for use in machine learning algorithms. Specifically, the sentiment values were mapped to three possible classes as follows: -1 for negative class, 0 for neutral class, and +1 for positive class. The number of positive

words and the number of negative words were captured as numeric values, while token-based features were summarized using numeric representation. A comprehensive list of the engineered features used in this work is described in Table 3.

Table 2 Information extracted from the Thailand Anti-Fake News Center website.

No	Extracted Data	Description
1	Publication Date and Time	The date and time when the news item was created or published
2	News URL	The web link where the news item is published on the Thailand Anti-Fake News Center website
3	Headline	The title of the news article
4	News Content	The full textual content of the news article
5	Verifying Agency	The organization responsible for verifying the news item
6	News Classification	The classification of the news item (real or fake)
7	News Category	The assigned category of the news item
8	Hashtag	Hashtags associated with the news item

Table 3 Description of engineered features used in model development.

No.	Attribute	Description	Data Type	Values / Definition
1	create_day	Day of publication	Polynomial	Sun = 1, Mon = 2, Tue = 3, Wed = 4, Thu = 5, Friday = 6, Sat = 7
2	create_time	Time period of publication	Polynomial	Midnight = 1, Morning = 2, Afternoon = 3, Evening = 4
3	isVerified	Verification status	Binomial	Verified by = 1, Not verified by agency = 2
4	Label	News classification	Binomial	Real news = 1, Fake news = 0
5	Category	News category	Polynomial	Government policy and public affairs = 1, Health products = 2, Finance and stock = 3, Disaster = 4
6	has_hashtag	Presence of hashtag	Binomial	Present = 1, Not present = 2
7	count_hashtags	Number of hashtags	Numeric	Integer ≥ 0
8	Sentiment	Sentiment polarity	Polynomial	Positive = 1, Neutral = 0, Negative = -1
9	sentiment_score	Sentiment score	Numeric	Range: 0-100
10	count_pos	Number of positive	Numeric	Integer ≥ 0
11	count_neg	Number of negative	Numeric	Integer ≥ 0
12	content_length	News content length	Numeric	Integer ≥ 0

Train-test data splitting

To evaluate the performance of the model, the dataset was divided into two subsets a training set and a test set. The training set was used to fit the model, while the test set was kept separate and used only for final evaluation. This approach helps ensure that the model is assessed on data it has never seen before, providing a more reliable estimate of its real-world performance.

The split was performed carefully to avoid data leakage between training and testing. In cases where the dataset was imbalanced, resampling techniques was applied only to the training set, while the test set remained unchanged. This allowed the

evaluation results to reflect the model's true ability to generalize to previously unseen data

Standardization and feature selection

After constructing the features for the dataset, feature scaling and feature selection processes were employed for consistency in measurements within the features and for stable training of models.

Among some features in the dataset were the count_hashtags, sentiment_score, count_pos, count_neg, and content_length features which were all numerical and whose units were on various scales and ranges. This can be seen in Table 4 below, showing the need for data standardization to avoid disparity in scales among the features.

Table 4 Statistical summary of numerical features before standardization.

Attribute	Max	Min	Mean	SD.
count_hashtags	10	0	3.16	1.03
sentiment_score	99.98	0	34.65	36.78
count_pos	5	0	0.24	0.50
count_neg	9	0	0.54	0.91
content_length	44	3	13.44	4.77

Table 5 Statistical summary of numerical features after standardization.

Attribute	Max	Min	Mean	SD.
count_hashtags	6.67	-3.09	0	1
sentiment_score	1.78	-9.42	0	1
count_pos	9.54	-4.76	0	1
count_neg	9.30	-5.89	0	1
content_length	6.40	-2.19	0	1

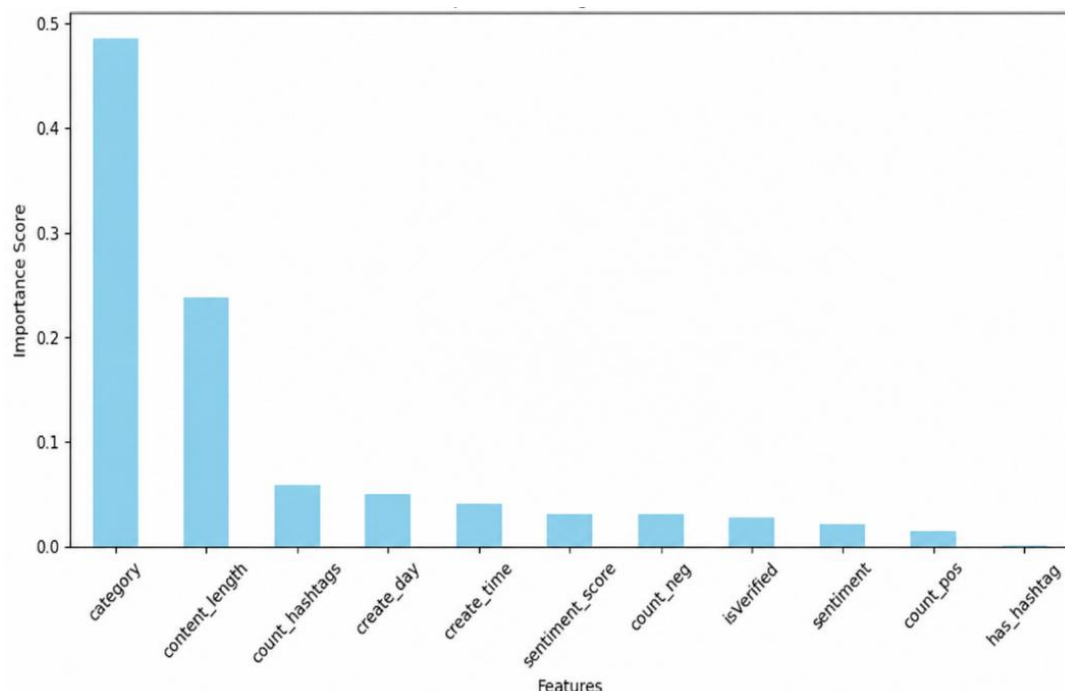


Figure 2 Feature importance scores obtained from the extra tree's classifier.

The StandardScaler technique was applied to transform the numerical features into a common scale with zero mean and unit variance [16]. This normalization process was applied to standardize feature distributions prior to model training. The transformed feature distributions after standardization are presented in Table 5.

Apart from the scaling, feature selection was done to select the most useful features and minimize redundancy among the features. Feature importance was calculated by the Extra Trees Classifier that calculates the feature's importance from impurity decrease while building trees [16].

The feature importance values obtained were then used to rank the importance of the features, as shown in Figure 2.

Moreover, correlation analysis was applied to analyze the relationship between the features and the target variable through a correlation matrix. Correlation analysis was applied to study the relationship between the features and the target variable, as indicated in Figure 3.

In addition to the above, Backward Feature Elimination technique was applied to determine the significance of each feature while training the model [17]. This procedure involves a systematic elimination of features in order to study their impact on the model. The outcomes of this analysis are outlined in Table 6. From the above processes, the features were selected for model development. Comparative analysis of feature selection techniques is presented in Table 7.

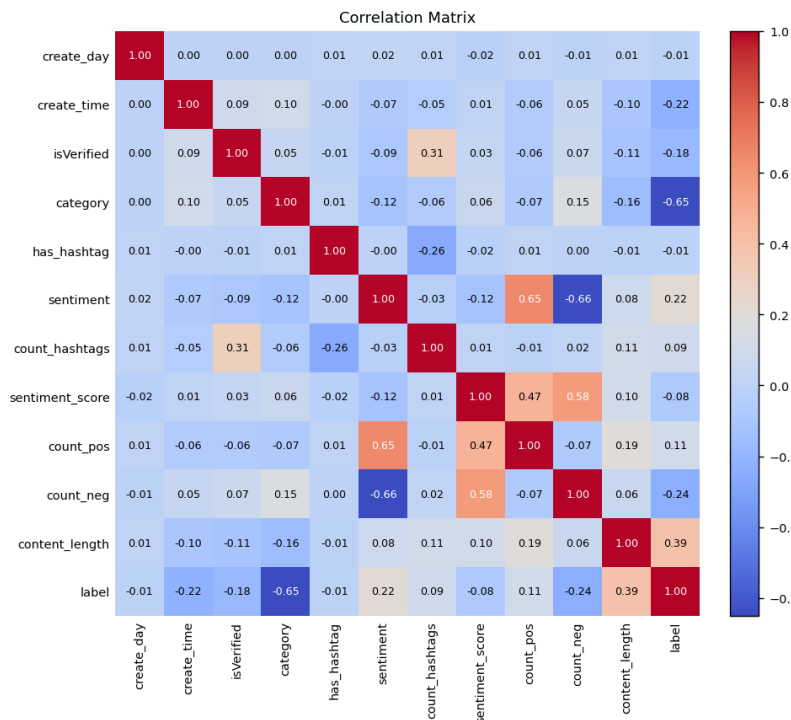


Figure 3 Correlation matrix showing relationships between features and the target variable.

Table 6 Impact of backward feature elimination on model accuracy.

Feature	NN	NB	SVM	Comparison with Baseline			
				NN	NB	SVM	Summary
Baseline Model	0.8978	0.8579	0.8308	-	-	-	-
Drop create_day	0.8978	0.8574	0.8308	-	Decreased	-	Decreased
Drop create_time	0.8971	0.8487	0.8308	Decreased	Decreased	-	Decreased
Drop isVerified	0.8962	0.8566	0.8308	Decreased	Decreased	-	Decreased
Drop category	0.8359	0.8111	0.8308	Decreased	Decreased	-	Decreased
Drop has_hashtag	0.8980	0.8509	0.8308	Increased	Decreased	-	-
Drop sentiment	0.9003	0.8637	0.8308	Increased	Increased	-	Increased
Drop count_hashtags	0.8942	0.8514	0.8308	Decreased	Decreased	-	Decreased
Drop sentiment_score	0.9019	0.8657	0.8948	Increased	Increased	Increased	Increased
Drop count_pos	0.8956	0.8603	0.8308	Decreased	Increased	-	-
Drop count_neg	0.8909	0.8774	0.8308	Decreased	Increased	-	-
Drop content_length	0.8613	0.8065	0.8308	Decreased	Decreased	-	Decreased

Table 7 Comparative results of feature selection methods.

Feature	Rank (Feature Importance)	Rank (Correlation Matrix)	Backward Feature Elimination	Final Selection
create_day	4	9	✓	✓
create_time	5	4	✓	✓
isVerified	8	5	✓	✓
Category	1	1	✓	✓
has_hashtag	11	9	-	✗
Sentiment	9	4	✗	✓
count_hashtags	3	7	✓	✓
sentiment_score	6	8	✗	✓

Feature	Rank (Feature Importance)	Rank (Correlation Matrix)	Backward Feature Elimination	Final Selection
count_pos	10	6	-	✓
count_neg	7	3	-	✓
content_length	2	2	✓	✓

Class imbalance

SMOTE was applied to the dataset to resolve the class imbalance issue in the dataset. In real-world fake news datasets, it is common for class distributions to be imbalanced, where one class can be much larger than the other.

The SMOTE was used to create synthetic observations of the minority class by interpolation of feature space observations [18]. The method was utilized to enhance the presence of the minority class without duplicating of existing representations.

Before developing the model for Thai fake news detection, the training data were resampled. Table 8 presents the distribution of the classes before and after using SMOTE.

Table 8 Comparison of training dataset distribution before and after SMOTE-based oversampling.

Split	Class	Before Over-sampling	After Over-sampling
80:20	Real	950	4,661
	Fake	4,661	4,661
	news		
90:10	Real	1,068	5,244
	Fake	5,244	5,244
	news		

This procedure was applied prior to model training to address class imbalance in the training dataset. SMOTE was applied only to the training set, while the test dataset was left unchanged.

Classification algorithms

The following machine learning methods were selected for implementation of the above-mentioned detection approach: Naïve Bayes (NB), Support Vector Machine (SVM), and Neural Networks (NN). Each algorithm was used to present alternative approaches to classifying data. All algorithms were trained based on the feature vector extracted from textual content and sentiment-emotion attributes.

Naïve Bayes (NB) is a probabilistic model based on Bayes' theorem with an additional conditional independence assumption of attributes [19]. In this study, Gaussian Naïve Bayes was used as the classification model.

Support Vector Machine (SVM) is a discriminative classifier that aims at choosing the best possible hyperplane by finding a maximal margin between

classes [20]. Thus, some transformations, including kernels, are needed for better separation of nonlinear patterns. In this study, the SVM model was configured with an RBF kernel, $C = 100$, and $\gamma = \text{auto}$, with probability estimation enabled.

Neural Network (NN) is a feedforward network model, which helps to find a nonlinear relation between feature vectors. Parameters of NN are optimized iteratively based on backpropagation [21, 22]. In this study, the NN model was implemented as a multilayer perceptron with two hidden layers containing 128 and 64 neurons, respectively. The tanh activation function and the Adam optimizer were used, and the model was trained for a maximum of 1,000 iterations.

Experimental setup and model evaluation

To evaluate the proposed Thai fake news detection framework, a systematic experimental setup was established. The dataset was partitioned using two train-test split strategies, namely 80:20 and 90:10. We also used K-fold cross-validation with $k = 5$ to check the model [23].

Table 9 Model configuration and hyperparameters.

Model	Key Parameters
NB	Gaussian Naïve Bayes
SVM	RBF kernel, $C = 100$, $\gamma = \text{auto}$, probability = True
NN	Hidden layers = 128 and 64, tanh activation, Adam optimizer, max_iter = 1,000

Table 10 Experimental configuration.

Component	Description
Train-Test Split	80:20, 90:10
Cross-Validation	5-Fold, 10-Fold
Evaluation Metrics	Accuracy, Precision, Recall, F1-score, ROC-AUC, AUC-PR
Oversampling Method	SMOTE
Tuning Method	Grid Search

We changed the hyperparameters for each model that did the classification. A grid search method was used to carefully look at different combinations of important parameters. Table 9 gives a brief overview of this. We used four standard evaluation metrics

accuracy, precision, recall, and F1-score to see how well the model worked. All of the experiments were done in a controlled computer environment.

All experiments were conducted within a controlled computational setting. The overall experimental configuration, including data splitting

strategies, validation methods, evaluation metrics, and tuning procedures, is summarized in Table 10.

The complete workflow of preprocessing, feature construction, model training, and evaluation is illustrated in Figure 4.

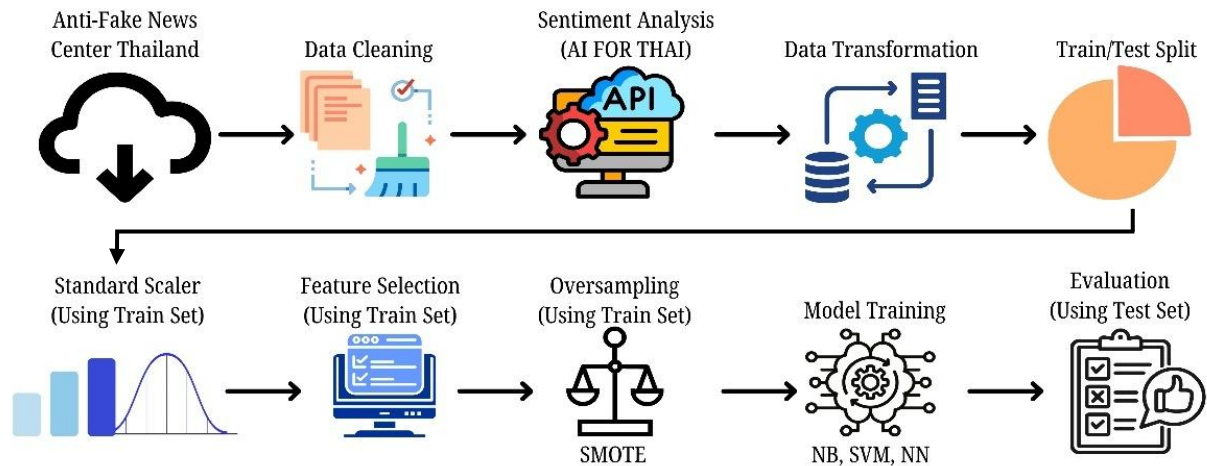


Figure 4 System architecture.

RESULTS AND DISCUSSION

Results

Neural Network (NN) model showed the highest values for all the evaluation parameters according to the experimental results. With the 90:10 train-test configuration, NN achieved an accuracy of 92.39%, precision of 92.64%, recall of 92.39%, F1-score of 92.38%, ROC-AUC of 96.79%, and AUC-PR of 96.07%, respectively (Table 12). According to Table 11, with the 80:20 split, the NN had 92.38% accuracy and 96.77% ROC-AUC, while SVM and NB had 92.09% and 87.31% accuracy, and 95.63% and 92.88% ROC-AUC, respectively. In terms of AUC-PR, the NN also achieved the highest value at 96.05%, followed by SVM at 93.46% and NB at 91.74%.

On the various train-test configurations, the NN model generated the highest values for all evaluation metrics, SVM generated the second highest

value, and NB generated the lowest value among the three models. The AUC-PR results further confirm that NN provided the best precision-recall balance in both experimental settings.

As indicated by the information in Table 11, the three models were evaluated using a 80:20 train-test configuration. NB achieved 87.31% accuracy and 92.88% ROC-AUC, while SVM and NN achieved 92.09% and 92.38% accuracy, and 95.63% and 96.77% ROC-AUC, respectively. All models exhibited uniformity in terms of precision, recall, F1-score, ROC-AUC, and AUC-PR.

The outcomes of the 90:10 train-test configuration, in other words, are summarized in Table 12. The NN model achieved 92.39% accuracy and 96.79% ROC-AUC, SVM followed with 92.01% accuracy and 95.71% ROC-AUC, and NB achieved 87.28% accuracy and 92.86% ROC-AUC, respectively. The precision, recall, F1-scores, ROC-AUC, and AUC-PR values are consistently distributed across the three models.

Table 11 Model performance on the independent test set using a 80:20 train-test split.

Model	Accuracy	Precision	Recall	F1-score	ROC-AUC	AUC-PR
NB	87.31	88.26	87.31	87.23	92.88	91.74
SVM	92.09	92.63	92.09	92.06	95.63	93.46
NN	92.38	92.64	92.38	92.37	96.77	96.05

Table 12 Model performance on the independent test set using a 90:10 train-test split.

Model	Accuracy	Precision	Recall	F1-score	ROC-AUC	AUC-PR
NB	87.28	88.24	87.28	87.20	92.86	91.74
SVM	92.01	92.55	92.01	91.98	95.71	93.67
NN	92.39	92.64	92.39	92.38	96.79	96.07

In this study, the Real News class was defined as the positive class for performance evaluation. Tables 13 and 14 present the confusion matrices of the NN model under the 80:20 and 90:10 train-test split configuration on the testing dataset. In both configurations, the numbers of True Positives and True Negatives were higher than those of False Positives and False Negatives.

Table 13 Confusion matrix of the best-performing neural network model using a 80:20 train-test split configuration on the testing dataset.

	Predicted Real	Predicted Fake
Real	209 (TP)	28 (FN)
Fake news	79 (FP)	1,087 (TN)

Table 14 Confusion matrix of the best-performing neural network model using a 90:10 train-test split configuration on the testing dataset.

	Predicted Real	Predicted Fake
Real	105 (TP)	14 (FN)
Fake news	39 (FP)	544 (TN)

However, although high classification performance was achieved by the NN model, some false positives and false negatives were still observed in the confusion matrices. False positives can be produced when fake news articles are written in a neutral, formal, or official style of language, and therefore linguistically closer to verified news. On the other hand, real news articles that contain emotional expressions, warnings and sensational wording, particularly in health- or disaster-related content, can increase false negatives. Moreover, the short content length and the Thai word segmentation errors might deteriorate the quality of the feature representation that can influence model decisions in some cases.

Discussion

Based on the results of this paper, it can be stated that non-linear analysis of the interaction between features is one of the factors in fake news detection in Thai language. The NN model was better than NB and SVM for both train-test splits; thus, the model was able to discover the correlation between TF-IDF, sentimental and metadata features. This is consistent with existing literature, which proves the efficiency of using neural networks for text classification in high-dimensional spaces [4, 21, 22].

The inclusion of affective and emotional features positively impacted classification performance by detecting affective features within the content of fake news. Therefore, it can be argued that affective features increase the separability between fake and real news as stated before in other works [2, 11].

The use of SMOTE as the balancing technique increased classification stability by improving the

representation of minority samples. This result is consistent with previous findings proving the benefits of oversampling [15].

Conclusions derived from feature selection include category, content length, time-based variables, and sentiment score as significant features affecting classification decisions. The removal of less relevant features does not impair accuracy, demonstrating that fewer but informative features will yield good predictive performance.

This conclusion follows several other investigations reporting the capacity of ensemble tree-based learning to detect significant predictors in highly dimensional feature spaces [24, 25].

The conclusions drawn from the confusion matrix output also reveal that the NN classifier performs equally well in predicting classes in either of the two categories. Nonetheless, the presence of a few classification errors implies the existence of a trade-off between precision and recall that ought to be modified according to the demands of a particular task.

The affective features involved in the classification process included sentiment_score, count_pos and count_neg, which resulted in capturing emotional cues in news content. Fake news often employs persuasive, emotional or attention-grabbing phrases, whereas verified news is typically written in neutral and factual language. Thus, these affective features helped to increase the separability between real and fake news, especially when combined with linguistic and metadata features.

Nevertheless, the analysis is constrained to a small number of machine-learning models and a single dataset. This study focused on an interpretable conventional machine-learning baseline. therefore, transformer-based models such as WangchanBERTa and mBERT were not included in the current experiments. Future research should compare the proposed framework with these models to evaluate technical competitiveness and potential performance improvement.

CONCLUSIONS

This research paper has presented a machine learning system aimed at identifying Thai-language fake-news using multiple approaches such as content, sentiment, and context analysis. The results of the study indicate that a neural network system has demonstrated the best classification performance in terms of the test results with the maximum accuracy score of 92.39% in the 90:10 setting.

The importance of the use of both content and sentiment features in Thai-language settings cannot be understated in terms of improving the accuracy rate. Specifically, the inclusion of affective attributes leads to higher class separation in terms

of fake vs. real news while SMOTE helps to boost the stability of the classifier.

This work has delivered one of the early structured frameworks used for detecting misinformation in low-resource languages. This research delivers an initial structured framework for creating a model capable of identifying Thai-language misinformation. As far as the practical implications of the study are concerned, the proposed approach allows for the creation of monitoring systems for fake news detection purposes.

REFERENCES

- Vosoughi S, Roy D, Aral S. The spread of true and false news online. *Science*. 2018;359(6380): 1146-51.
- Shu K, Sliva A, Wang S, Tang J, Liu H. Fake News Detection on Social Media. *ACM SIGKDD Explorations Newsletter*. 2017;19(1):22-36.
- Conroy NJ, Rubin VL, Chen Y. Automatic deception detection: Methods for finding fake news. *Proceedings of the Association for Information Science and Technology*. 2015;52(1):1-4.
- Jurafsky D, Martin JH. *Speech and Language Processing*. 3rd ed. Hoboken (NJ): Pearson; 2020.
- Zhou X, Zafarani R. A Survey of Fake News. *ACM Computing Surveys*. 2020;53(5):1-40.
- Kaliyar RK, Goswami A, Narang P. FakeBERT: fake news detection in social media with a BERT-based deep learning approach. *Multimed Tools Appl*. 2021;80(8):11765-88.
- Mookdarsanit P, Mookdarsanit L. The COVID-19 fake news detection in Thai social texts. *Bulletin of Electrical Engineering and Informatics*. 2021; 10(2):988-98. doi:10.11591/eei.v10i2.2745.
- Bird S, Klein E, Loper E. *Natural Language Processing with Python*. Sebastopol (CA): O'Reilly Media; 2009.
- Ministry of Digital Economy and Society. Thailand Anti-Fake News Center [Internet]. Bangkok: Ministry of Digital Economy and Society; 2025 [cited 2026 Mar 16]. Available from: <https://www.antifakenewscenter.com/>.
- Phatthiyaphaibun W, Chaovavanich K, Polpanumas C, Suriyawongkul A, Lowphansirikul L, Chormai P, et al. PyThaiNLP: Thai natural language processing in Python. In: *Proceedings of the 3rd Workshop for Natural Language Processing Open Source Software (NLP-OSS 2023)*; 2023. p. 25-36.
- Pang B, Lee L. Opinion Mining and Sentiment Analysis. *Foundations and Trends in Information Retrieval*. 2008;2(1-2):1-135.
- NECTEC. AI for Thai: natural language processing services for Thai language [Internet]. Pathum Thani: National Electronics and Computer Technology Center; 2025 [cited 2026 Mar 16]. Available from: <https://aiforthai.in.th/>.
- Zhou JL, Ye Z, Zhang S, Zhao G, Han N, Yang T. Investigating response behavior through TF-IDF and Word2vec text analysis: A case study of PISA 2012 problem-solving process data. *Heliyon*. 2024;10(16):e35945.
- Khamphakdee N, Seresangtakul P. An efficient deep learning for Thai sentiment analysis. *Data*. 2023;8(5):90. doi:10.3390/data8050090.
- Dhiman P, Kaur A, Gupta D, Juneja S, Nauman A, Muhammad G. GBERT: A hybrid deep learning model based on GPT-BERT for fake news detection. *Heliyon*. 2024 Aug 30;10(16).
- Géron A. *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*. 3rd ed. Sebastopol (CA): O'Reilly Media; 2022.
- Meesad P. Thai fake news detection based on information retrieval, natural language processing and machine learning. *SN Computer Science*. 2021 Nov;2(6):425.
- Dablain D, Krawczyk B, Chawla NV. DeepSMOTE: Fusing Deep Learning and SMOTE for Imbalanced Data. *IEEE Transactions on Neural Networks and Learning Systems*. 2022;34(9):6390-404.
- McCallum A, Nigam K. A comparison of event models for Naïve Bayes text classification. In: *Proceedings of AAAI-98 workshop on learning for text categorization*; 1998 Jul 27; Madison (WI). Menlo Park (CA): AAAI Press; 1998. p. 41-8.
- Cortes C, Vapnik V. Support-vector networks. *Machine Learning*. 1995;20(3):273-97.
- Goodfellow I, Bengio Y, Courville A. *Deep learning*. Cambridge (MA): MIT Press; 2016.
- Rumelhart DE, Hinton GE, Williams RB. Learning representations by back-propagating errors. *Nature*. 1988;323(6088):533-6.
- Kohavi R. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*; 1995. p. 1137-43.
- Breiman L. Random forests. *Machine Learn*. 2001; 45(1):5-32.
- Geurts P, Ernst D, Wehenkel L. Extremely randomized trees. *Machine Learning*. 2006;63(1):3-42.