El Cezeri
Journal of Science and Engineering
ISSN: 2148-3736

Research Article

# Comparison of the Performance of Machine Learning Methods with Intuitionistic Fuzzy Decision Making for Effective Fake News Detection

**Feride Tuğrul[1a]**
[1] Munzur University, Faculty of Engineering, Department of Computer Engineering, Tunceli, Türkiye

feridetugrul@munzur.edu.tr

DOI : 10.31202/ecjse. 1830741
Received: 26.11.2025 Accepted: 16.01.2026
**How to cite this article:**
Tuğrul F., "Comparison of the Performance of Machine Learning Methods with Intuitionistic Fuzzy Decision Making for Effective Fake News Detection ", El-Cezeri Journal of Science and Engineering, Vol: 13, Iss: 2, (2026), pp.(286-300).
ORCID: [a]000-0001-7690-8080

**Abstract** In this paper, the success of machine learning models applied to two datasets consisting of fake news was examined. The performance of these methods was measured using Accuracy, Precision, Recall, and F1-Score. Evaluation metrics for all models were calculated using TF-IDF and N-gram TF-IDF for Dataset 1 and Dataset 2, respectively. In continuation of the study, a decision-making mechanism was created to measure the success of machine learning methods. The success of these models was compared by creating a decision-making mechanism using intuitionistic fuzzy sets. The PROMETHEE method was used here. In the first stage of the study, the dataset samples evaluated were expressed using classical sets, while in the second stage, the success of the models according to the metrics was expressed using intuitionistic fuzzy values. This was to minimize the uncertainty in the model success results. The results obtained in the first stage of the study were evaluated in the second stage using a decision-making mechanism. In this mechanism, machine learning models represent the alternatives, while classification metrics represent the criteria. When evaluating machine learning models, experts provide subjective opinions based on each classification metric to determine the successful model.

**Keywords:** Decision making, fake news detection, intuitionistic fuzzy set, machine learning.

## 1. INTRODUCTION

In recent years, with the rapid development of digital media and social networks, the spread of misinformation and disinformation has become a serious global problem. Sharing unverified content, particularly on social media platforms, leads to misleading individuals, manipulating public opinion, and undermining social trust. In this context, fake news detection has become an increasingly important research area in computer science and information technology. Fake news not only impacts individual information gathering processes but also negatively impacts political decision-making [1], financial markets [2], and even global health crises [3].

The literature demonstrates the extensive use of machine learning algorithms in fake news detection. Supervised learning methods such as Naive Bayes [2], Decision Trees, Random Forest, and Logistic Regression have achieved significant success in the classification of news texts. However, Support Vector Machines (SVM) stand out with their higher accuracy rates, especially in text classification problems with strong linear separability [4]. Similarly, feature extraction techniques such as Bag of Words, N-gram, and TF-IDF are widely preferred to improve the performance of machine learning-based algorithms. Furthermore, deep learning-based approaches have recently provided significant advantages in capturing semantic relationships in texts, and models such as long short-term memory (LSTM) and convolutional neural networks (CNN) have been reported to have shown promising results in fake news detection [5].

However, the dynamic and ever-changing nature of fake news is one of the most significant challenges facing studies in this field. The production of news in diverse contexts, linguistic diversity, and the inclusion of user comments directly impact the performance of machine learning-based systems [6]. Therefore, it is critical that systems for detecting fake news not only achieve high accuracy but also possess the flexibility to adapt to different contexts.

Consequently, machine learning is one of the most preferred methods in the field of fake news detection due to its high accuracy and scalability. The fake news detection architecture proposed in this study consists of preprocessing text data and then feeding the resulting TF-IDF and N-Gram TF-IDF vector representations into Random Forest (RF), Extreme Gradient Boosting (XGBoost), Logistic Regression (LR), Naive Bayes (NB), and LightGBM machine learning methods to detect fake news. Subsequently, a multi-criteria decision-making mechanism such as intuitionistic fuzzy PROMETHEE is used to reduce uncertainty in evaluating the metric results used to measure the success of these models. Successful results are obtained in the fake news detection task using statistical feature extraction methods. On the other hand, in irony or sarcasm detection tasks, methods that obtain contextual information in a multifaceted way may be more successful. The TF-IDF and N-gram TF-IDF methods used here are based on the principle of weighting words and word groups in a text according to their frequency. With this weighting, words and distributional patterns that distinguish real news from fake news can be reflected with weights. For example, TF-IDF identifies keywords from multiple texts, finding those that occur frequently in the analyzed text, but these keywords do not appear with the same frequency in other texts. Thus, they are given more weight in differentiating between texts. N-grams, on the other hand, take recurring word phrases into account, thereby capturing contextual information and improving the success of text classification.

In their study, Hu et al. developed a fuzzy logic and artificial intelligence-based approach to reduce uncertainty in the supplier evaluation and selection process and enable more reliable decision-making. To this end, the study combined both fuzzy data from expert judgments and the learning capabilities of artificial neural networks using the ANFIS (Adaptive Neuro-Fuzzy Inference System) method. The dataset was created from expert evaluations based on criteria such as price, quality, delivery time, flexibility, and reliability. The results show that the ANFIS-based model provides higher accuracy and flexibility than traditional methods and enables more reliable supplier rankings by minimizing the uncertainties in expert opinions [7].

In their study, Ahmad et al. aim was to develop a new approach to provide a more effective and reliable solution to the identified problem area. In this context, the method used in this study was chosen to overcome the shortcomings of classical methods in literature, enabling more accurate data processing and management of uncertainties. The dataset consisted of observations and measurements collected under different criteria in line with the purpose of the study and was used in the training and testing stages of the model during the analysis process. The results revealed that the proposed method was more successful than existing methods, providing superiority in terms of accuracy and flexibility, thus providing a more reliable solution for decision-makers [8].

In their study, Baarir and Djeffal aimed to develop an effective system for detecting fake news that spreads rapidly on social media and the internet. In this context, a machine learning-based approach was used to classify fake and real news. The dataset used in the study was created by combining two different data sources obtained from the Kaggle platform. In addition to text, these datasets also included additional attributes such as date, source, author, and sentiment analysis. The results showed that the proposed model achieved high accuracy, and in particular, attributes such as author, source, and date significantly increased classification success [9].

Manzoor et al. conducted a review of studies using machine learning approaches to detect fake news. They described the machine learning methods used [10].

In their study, Sharma et al. used concepts from Artificial Intelligence, Natural Language Processing, and Machine Learning to perform binary classification of various news articles available online. They aimed to provide users with the ability to classify news as fake or real, and also to verify the credibility of the website publishing the news [11].

In their study, Alghamdi et al. examined and compared representation models with classical algorithms for fake news detection. It was suggested that performance could be improved if context-based information was considered in addition to the news text and other factors such as style and sentiment were incorporated. The paper concluded that the experimental results demonstrated that no single technique could achieve the best performance scores across all the datasets used [12].

Alghamdi et al., in their study, thoroughly discussed how machine learning and deep learning techniques are used in fake news detection and their applications. They investigated the characteristics of fake news and their datasets [13].

Yakkundi et al. highlight the spread of fake news on social media and the potential challenges it may pose in the future. To mitigate its harmful impact, different approaches to fake news detection using machine learning are proposed [14].

Many researchers have studies in areas such as machine learning, social networking, and decision making [15,16,17,18,19].

## 2. MATERIALS AND METHODS

This section describes in detail the materials and methods used during the fake news detection task. First, two datasets shared on the Kaggle platform for fake news detection were used.

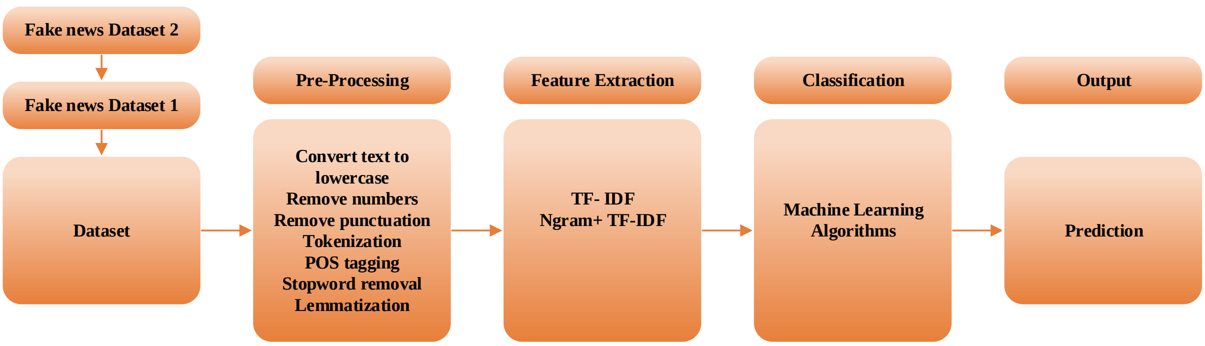


**Figure 1. General architecture of the fake news detection system**

Data preprocessing methods were applied to these datasets, followed by feature selection methods and feature engineering. The resulting document vectors were fed into five popular machine learning methods, and classification results were obtained. The general framework used in this study for fake news detection using machine learning methods is presented in Figure 1.

## 2.1. Dataset

### 2.1.1. Dataset 1

Dataset 1 consists of news text labeled as fake and true. Obtained from the Kaggle platform [20], this dataset consists of 44,898 rows in total. The dataset has four columns: title, text, subject, and date. A small sample of the dataset used is shown in Table 1.

**Table 1. A small sample from dataset 1**

| Title | Text | Subject | Date | Label |
|---|---|---|---|---|
| As U.S. budget fight looms, Republicans flip t... | WASHINGTON (Reuters) - The head of a conservat... | politicsNews | December 31, 2017 | 1 |
| U.S. military to accept transgender recruits o... | WASHINGTON (Reuters) - Transgender people will... | politicsNews | December 29, 2017 | 1 |
| Senior U.S. Republican senator: 'Let Mr. Muell... | WASHINGTON (Reuters) - The special counsel inv... | politicsNews | December 31, 2017 | 1 |
| FBI Russia probe helped by Australian diplomat... | WASHINGTON (Reuters) - Trump campaign adviser ... | politicsNews | December 30, 2017 | 1 |
| Trump wants Postal Service to charge 'much mor... | SEATTLE/WASHINGTON (Reuters) - President Donal... | politicsNews | December 29, 2017 | 1 |

When we examine the labeled data in the dataset, we see that it is largely balanced. The dataset contains 21417 fake and 23481 true labeled texts. When we examine the news texts by subject in Figure 1, the most frequent texts in the dataset are 11272 in the politicnews topic. Politicnews is followed by 10145 texts from worldnews, 9050 from news, 6841 from politics, 4459 from left-news, and 1570 from government news.

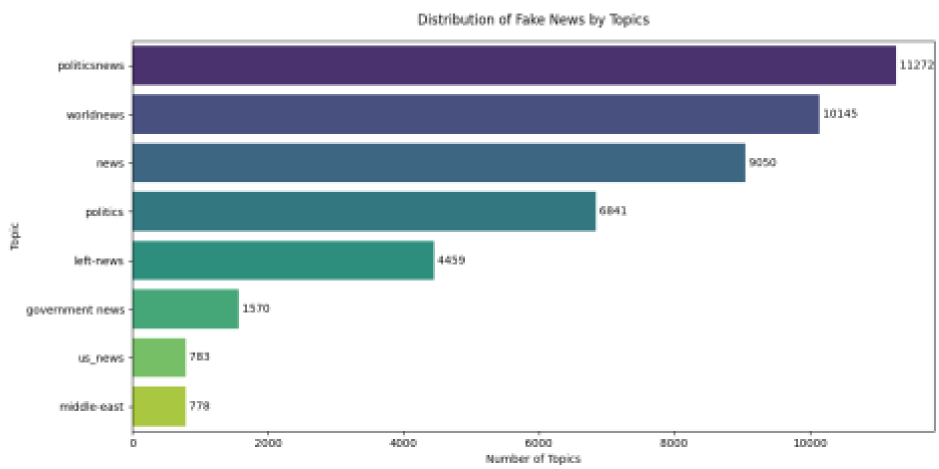


**Figure 2. Distribution of news in the dataset by topic**

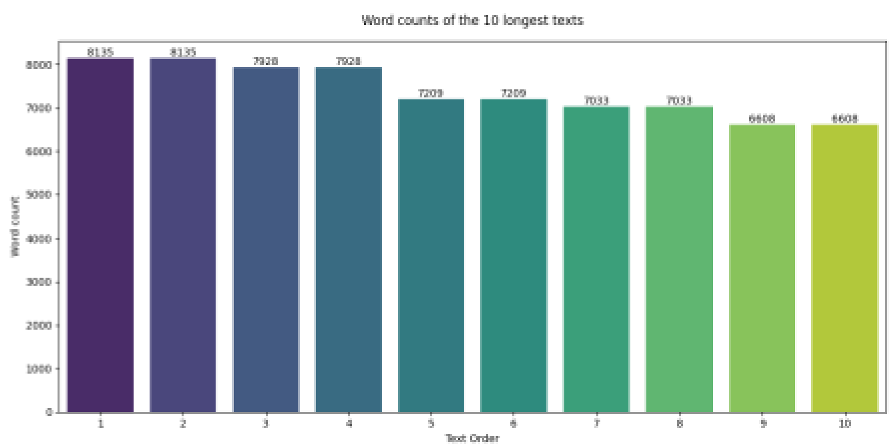


**Figure 3. Word lengths of the 10 longest news items in the dataset**

Figure 2 shows the word lengths of the ten longest texts. Accordingly, the longest news text consists of 8135 words. The following long news texts consist of 8135 and 7928 words.

### 2.1.2. Dataset 2

Dataset 2 is named WELFake [21]. This dataset consists of 72134 news articles, comprising 35028 real news articles and 37,106 fake news articles. For this, the authors combined four popular news datasets (i.e., Kaggle, McIntire, Reuters, BuzzFeed Political) to prevent overfitting of the classifiers and provide more text data for better machine learning training. Table 2 provides a small sample dataset from Dataset 2.

**Table 2. A small sample from dataset 2**

| Title | Text | Label |
|---|---|---|
| Jimmy Carter recovers from dehydration scare in Canada | WINNIPEG, Manitoba (Reuters) - Former U.S. President Jimmy Carter, appearing fully recovered from dehydration suffered while helping to build a home for charity in Canada, was released from an overnight hospital stay on Friday and addressed the project's closing ceremony……….. | 0 |
| White House: Obama May Leave the Country if Trump is Elected | The White house is refusing to deny the rumors that President Obama claimed he will relocate to Canada if Donald Trump is elected as his successor in November. The rumor began when Canadian satire website The Burred Street Journal ran a fictional story titled "Obama Declares His Family Will Move To Canada If Trump Is Elected." The story was complete with a fake statement from Obama which read, "It's something Michelle, the kids and I have discussed as a potential solution to the Donald…… | 1 |
| Vice President-elect Pence says 'new hope dawns' for Cuba | WASHINGTON (Reuters) - U.S. Vice President-elect Mike Pence said on Saturday that "new hope dawns" for Cuba after the death of Fidel Castro. "The tyrant Castro is dead. New hope dawns. We will stand with the oppressed Cuban people for a free and democratic Cuba. Viva Cuba Libre!" Pence said on Twitter. | 0 |
| Boy With Autism Makes His First Friend Ever And His Mom Can't Stop Crying | Approximately 1 in 68 children has an autism spectrum disorder, making the disorder more common than it used to be and a force to be reckoned with. There are even service dogs that are specially... | 1 |
| Trump to Republican senators: Don't leave town until pass healthcare legislation | WASHINGTON (Reuters) - U.S. President Donald Trump urged Republican senators on Wednesday to work out their differences and pass healthcare legislation before leaving Washington for their August recess. A day after the latest version of legislation to replace Obamacare fell victim to squabbling among Republicans who control …….. | 0 |

When we look at the labeled data in the dataset, we see that it is largely balanced. The average word count of the news articles in the dataset is around 302.

## 2.2. Data Preprocessing Steps

Data preprocessing is a fundamental step in making raw text suitable for analysis and modeling. This process involves processes such as tokenization, stopword removal, stemming, and lemmatization to reduce noise in the data and improve model accuracy. Textual data lacks structure, contains spelling and grammatical errors, and non-standard formatting, complicating the analysis process. In this study, the data were converted to lowercase letters, cleaned of numbers and punctuation, tokenized, and POS tagged. Then, lemmatization was performed by removing stopwords. Finally, the cleaned text was reconstructed to yield a normalized dataset for modeling and analysis.

## 2.3. Vectorization of Documents with TF-IDF and N-Gram TF-IDF Method

Computer processing of documents requires a numerical representation of text. This representation is a fundamental step in machine learning processes. Words in documents are represented vectorially using statistical methods such as TF-IDF (Term Frequency–Inverse Document Frequency). TF-IDF is the weight value that determines the importance of a term within a document. Term frequency represents the relative occurrence of a term in the text, while inverse document frequency represents the logarithm of the ratio of the number of documents containing that term to all documents. The TF-IDF value obtained by multiplying these two metrics numerically expresses the relative importance of each word within the context of the text. The TF-IDF method is expressed as follows:

$$a_{ij} = tf_{ij}idf_j = tf_{ij} * log_2(\frac{N}{d_{fj}}) \quad (1)$$

$a_{ij}$ is the weight of term $j$ in document $i$. N represents the total number of documents in the text set. $tf_{ij}$ is the frequency of term *j* in document *i*, and $d_{fj}$ is the document frequency of term *j*' in the corpus.

The N-gram method analyzes texts based on groups of consecutive words rather than individual words, revealing the contextual relationships between words. This approach allows for richer features to be obtained by considering the meaning of a word in conjunction with the words that follow or precede it. For example, unigrams represent single words ("big"), bigrams represent two-word sequences ("very big"), and trigrams represent three-word sequences ("not very big"). In the N-gram TF-IDF model, N-grams are identified across all documents, and the frequencies of these groups are weighted using the TF-IDF method to numerically express the contextual structure of each document.

### 2.4. Evaluation Metrics

In this study, the performance of classification models was evaluated using Accuracy, Precision, Recall, and F1-Score metrics. True Positive (TP) values represent instances where the model correctly predicted positives, False Positive (FP) represents negative instances that were falsely predicted as positives, True Negative (TN) represents instances where the model correctly predicted negatives, and False Negative (FN) represents instances where the model was positive but negative. Accuracy represents the ratio of all correct predictions to total predictions; Precision represents the fraction of positive predictions that were correct; and Sensitivity represents the fraction of true positives that were correctly predicted. The F1-Score is the harmonic mean of precision and sensitivity and more accurately reflects the overall performance of the model, especially in imbalanced datasets.

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+FN+TN} \tag{2}$$

$$\text{Precision} = \frac{TP}{TP+FP} \tag{3}$$

$$\text{Recall} = \frac{TP}{TP+FN} \tag{4}$$

$$\text{F1} - \text{Score} = \frac{2*\text{Precision}*\text{Recall}}{\text{Precision}+\text{Recall}} \tag{5}$$

## 3. EXPERIMENTAL RESULTS

In this study, five popular machine learning methods were used on two datasets consisting of fake news: Random Forest (RF), Extreme Gradient Boosting (XGBoost), Logistic Regression (LR), Naive Bayes, and LightGBM. Furthermore, hyperparameter optimization was performed for each model using the Gridsearch method. The success of these methods was measured using Accuracy, Precision, Sensitivity, and F1-score.

### 3.1. Dataset 1 Results

#### 3.1.1. Classification Results of RF, NB, LR, XGBoost and LightGBM Models with TF-IDF

The complexity matrices of the RF, NB, LR, XGBoost, and LightGBM methods obtained with TF-IDF are shown below. Gridsearch hyperparameter optimization and 5-fold cross-validation were applied to all methods to improve their performance. Table 3 below shows the complexity matrices of the RF and LR algorithms.

**Table 3. Complexity matrix results of RF and IR algorithms, respectively**

| | | RF | |
|---|---|---|---|
| | | Prediction | |
| | | Negative | Positive |
| True | Negative | 4680 | 16 |
| | Positive | 6 | 4278 |

| | | LR | |
|---|---|---|---|
| | | Prediction | |
| | | Negative | Positive |
| True | Negative | Negative | 32 |
| | Positive | 26 | 4258 |

According to the results obtained, the RF algorithm correctly classified 8958 out of 8980 data points and achieved an accuracy value of 99.76%. In addition, the hyperparameters of the best performing model were set as ‘max_depth’: None, ‘min_samples_split’: 5, and ‘n_estimators’: 100. The LR algorithm correctly classified 8922 out of 8980 data points and achieved an accuracy value of 99.35%. In addition, the hyperparameters of the best performing model were set as ‘C’: 10, ‘solver’: ’liblinear’. Table 4 below shows the complexity matrices of the NB and XGBoost algorithms.

**Table 4. Complexity matrix results of NB and XGBoost algorithms respectively**

| | | NB | |
|---|---|---|---|
| | | Prediction | |
| | | Negative | Positive |
| True | Negative | 4411 | 285 |
| | Positive | 322 | 3962 |

| | | XGBoost | |
|---|---|---|---|
| | | Prediction | |
| | | Negative | Positive |
| True | Negative | 4682 | 14 |
| | Positive | 6 | 4278 |

According to the results obtained, the NB algorithm correctly classified 8373 out of 8980 data points and achieved an accuracy value of 93.24%. Additionally, the hyperparameters of the best performing model were set as 'alpha': 0.1. Table 6 below shows the confusion matrices of XGBoost. On the other hand, the XGBoost algorithm correctly classified 8960 out of 8980 data points and achieved an accuracy value of 99.78%. Additionally, the hyperparameters of the best performing model were set as {'learning_rate': 0.2, 'max_depth': 3, 'n_estimators': 200}. Table 5 below shows the confusion matrices of LightGBM.

**Table 5. Complexity matrix results of the LightGBM algorithm**

| | | Prediction | |
|---|---|---|---|
| | | Negative | Positive |
| True | Negative | 4683 | 13 |
| | Positive | 6 | 4278 |

According to the results, the LightGBM algorithm correctly classified 8961 out of 8980 data points, achieving an accuracy of 99.79%. Furthermore, the best-performing model had hyperparameters of {'learning_rate': 0.2, 'max_depth': 3, 'n_estimators': 200}. Table 8 below compares all methods in terms of their success metrics.

**Tablo 6. Evaluation criteria of TF-IDF and RF, NB, LR, XGBoost and LightGBM models**

| Model | Accuracy | Precision | Recall | F1-Score |
|---|---|---|---|---|
| LightGBM | 99,79% | 99,78% | 99,79% | 99,79% |
| XGBoost | 99,78 % | 99,77% | 99,78% | 99,78% |
| RF | 99,76 % | 99,77% | 99,76 % | 99,75 % |
| LR | 99,35 % | 99,35% | 99,36% | 99,35% |
| NB | 93,24% | 93,24% | 93,21% | 93,22% |

According to Table 6 above, the most successful model is LightGBM. This model received the highest score in every classification metric. After LightGBM, the most successful algorithm is XGBoost, followed by RF, LR, and NB.

### 3.1.2. Classification Results of RF, NB, LR, XGBoost and LightGBM Models with N-Gram TF-IDF

The complexity matrices of the RF, NB, LR, XGBoost, and LightGBM methods obtained using N-Gram TF-IDF are shown below. Gridsearch hyperparameter optimization and 5-fold cross-validation were applied to all methods to improve their performance. Table 7 below shows the complexity matrices of the RF and LR algorithms.

**Table 7. Complexity matrix results of RF and IR algorithms respectively**

| | | RF | |
|---|---|---|---|
| | | Prediction | |
| | | Negative | Positive |
| True | Negative | 4685 | 11 |
| | Positive | 4 | 4280 |

| | | LR | |
|---|---|---|---|
| | | Prediction | |
| | | Negative | Positive |
| True | Negative | 4671 | 25 |
| | Positive | 24 | 4260 |

According to the obtained results, the RF algorithm correctly classified 8965 out of 8980 data points and achieved an accuracy value of 99.83%. In addition, the hyperparameters of the best performing model were set as ‘max_depth’: None, ‘min_samples_split’: 5, and ‘n_estimators’: 100. Table 9 below shows the confusion matrices of LR. The LR algorithm correctly classified 8931 out of 8980 data points and achieved an accuracy value of 99.45%. In addition, the hyperparameters of the best performing model were set as ‘C’: 10, ‘solver’: ’liblinear’. Table 8 below shows the confusion matrices of the NB and XGBoost algorithms.

**Table 8. Complexity matrix results of NB and XGBoost algorithms respectively**

| | | NB | |
|---|---|---|---|
| | | Prediction | |
| | | Negative | Positive |
| True | Negative | 4451 | 245 |
| | Positive | 214 | 4070 |

| | | XGBoost | |
|---|---|---|---|
| | | Prediction | |
| | | Negative | Positive |
| True | Negative | 4685 | 11 |
| | Positive | 5 | 4279 |

According to the obtained results, the NB algorithm correctly classified 4521 out of 8980 data points and achieved an accuracy value of 94.89%. Additionally, the hyperparameters of the best performing model were set as 'alpha': 0.1. Table 9 below shows the confusion matrices of XGBoost. According to the obtained results, the XGBoost algorithm correctly classified 8964 out of 8980 data points and achieved an accuracy value of 99.82%. Additionally, the hyperparameters of the best performing model were set as {'learning_rate': 0.2, 'max_depth': 3, 'n_estimators': 200}. Table 9 below shows the confusion matrices of LightGBM. According to the results, the LightGBM algorithm correctly classified 8961 out of 8980 data points, achieving an accuracy of 99.79%. Furthermore, the best-performing model had hyperparameters of {'learning_rate': 0.2, 'max_depth': 20, 'n_estimators': 200}. Table 9 below compares all methods in terms of their success metrics.

**Table 9. Evaluation criteria of RF, NB, LR, XGBoost and LightGBM models with N-gram TF-IDF**

| Model | Accuracy | Precision | Recall | F1-Score |
|---|---|---|---|---|
| RF | 99,83% | 99,83 % | 99,84 % | 99,83% |
| XGBoost | 99,82% | 99,82% | 99,82% | 99,82% |
| LightGBM | 99,79% | 99,78% | 99,79% | 99,79% |
| LR | 99,45 % | 99,45% | 99,45% | 99,45% |
| NB | 94,89% | 94,87% | 94,89% | 94,88% |

According to Table 9 above, the most successful model is RF. This model received the highest score in every classification metric. After RF, the most successful algorithm is XGBoost. This is followed by LightGBM, LR, and NB algorithms, respectively. Figure X shows the results obtained by the N-Gram TF-IDF method and the machine learning algorithms graphically. According to the results, the N-Gram TF-IDF method and the RF algorithm achieve the best results. As shown in the graphic, the methods yield balanced results across all performance metrics.

### 3.2. Dataset 2 Results

#### 3.2.1. Classification Results of RF, NB, LR, XGBoost and LightGBM Models with TF-IDF

The complexity matrices of the RF, NB, LR, XGBoost, and LightGBM methods obtained with TF-IDF are shown below. Gridsearch hyperparameter optimization and 5-fold cross-validation were applied to all methods to improve their performance. Table 10 below shows the complexity matrices of the RF and LR algorithms.

**Table 10. Complexity matrix results of RF and IR algorithms respectively**

| | | RF | |
|---|---|---|---|
| | | Prediction | |
| | | Negative | Positive |
| True | Negative | 6468 | 538 |
| | Positive | 264 | 7046 |

| | | LR | |
|---|---|---|---|
| | | Prediction | |
| | | Negative | Positive |
| True | Negative | 6572 | 434 |
| | Positive | 363 | 6947 |

According to the obtained results, the RF algorithm correctly classified 13514 data out of 14316 and achieved an accuracy value of 94.40%. In addition, the hyperparameters of the best performing model were set as ‘max_depth’: None, ‘min_samples_split’: 2, and ‘n_estimators’: 200. The LR algorithm correctly classified 13519 data out of 14316 and achieved an accuracy value of 94.43%. In addition, the hyperparameters of the best performing model were set as ‘C’: 10, ‘solver’: ’lbfgs’. Table 11 below shows the complexity matrices of the NB and XGBoost algorithms

**Table 11. Complexity matrix results of NB and XGBoost algorithms respectively**

| | | NB | |
|---|---|---|---|
| | | Prediction | |
| | | Negative | Positive |
| True | Negative | 5666 | 1340 |
| | Positive | 945 | 6365 |

| | | XGBoost | |
|---|---|---|---|
| | | Prediction | |
| | | Negative | Positive |
| True | Negative | 6580 | 426 |
| | Positive | 178 | 7132 |

According to the results obtained, the NB algorithm correctly classified 12031 data out of 14316 and achieved an accuracy value of 84.04%. In addition, the hyperparameters of the best performing model were set as 'alpha': 0.1. The XGBoost algorithm correctly classified 7006 data out of 14316 and achieved an accuracy value of 95.78%. In addition, the hyperparameters of the best performing model were set as {'learning_rate': 0.2, 'max_depth': 6, 'n_estimators': 200}. Table 12 below shows the confusion matrices of LightGBM.

**Table 12. Complexity matrix results of the LightGBM algorithm**

| | | Prediction | |
|---|---|---|---|
| | | Negative | Positive |
| True | Negative | 6668 | 338 |
| | Positive | 152 | 7158 |

According to the results, the LightGBM algorithm correctly classified 13826 out of 14316 data points, achieving an accuracy of 96.58%. Furthermore, the best-performing model had hyperparameters of {'learning_rate': 0.2, 'max_depth': -1, 'n_estimators': 200}. Table 13 below compares all methods in terms of their success metrics.

**Table 13. Evaluation criteria of TF-IDF and RF, NB, LR, XGBoost and LightGBM models**

| Model | Accuracy | Precision | Recall | F1-Score |
|---|---|---|---|---|
| LightGBM | 96,58% | 96,63% | 96,55% | 96,57% |
| XGBoost | 95,78 % | 95,86% | 95,74% | 95,77% |
| RF | 94,40 % | 94,49% | 94,35 % | 94,39 % |
| LR | 94,43 % | 94,44% | 94,42% | 94,43% |
| NB | 84,04 % | 84,16% | 83,97% | 84,00% |

According to Table 6 above, the most successful model is LightGBM. This model received the highest score in every classification metric. After LightGBM, the most successful algorithm is XGBoost, followed by RF, LR, and NB.

### 3.2.2. Classification Results of RF, NB, LR, XGBoost and LightGBM Models with N-Gram TF-IDF

The complexity matrices of the RF, NB, LR, XGBoost, and LightGBM methods obtained using N-Gram TF-IDF are shown below. Gridsearch hyperparameter optimization and 5-fold cross-validation were applied to all methods to improve their performance. Table 14 below shows the complexity matrices of the RF and LR algorithms.

**Table 14. Complexity matrix results of RF and IR algorithms respectively**

| | | RF | |
|---|---|---|---|
| | | Prediction | |
| | | Negative | Positive |
| True | Negative | 6475 | 531 |
| | Positive | 244 | 7066 |

| | | LR | |
|---|---|---|---|
| | | Prediction | |
| | | Negative | Positive |
| True | Negative | 6590 | 416 |
| | Positive | 316 | 6994 |

According to the obtained results, the RF algorithm correctly classified 13541 data out of 14316 and achieved an accuracy value of 94.59%. In addition, the hyperparameters of the best performing model were set as ‘max_depth’: None, ‘min_samples_split’: 5, and ‘n_estimators’: 200. The LR algorithm correctly classified 13584 data out of 14316 and achieved an accuracy value of 94.89%. In addition, the hyperparameters of the best performing model were set as ‘C’: 10, ‘solver’: ’liblinear’. Table 15 below shows the complexity matrices of the NB and XGBoost algorithms.

**Table 15. Complexity matrix results of NB and XGBoost algorithms respectively**

| | | NB | |
|---|---|---|---|
| | | Prediction | |
| | | Negative | Positive |
| True | Negative | 5725 | 1281 |
| | Positive | 859 | 6451 |

| | | XGBoost | |
|---|---|---|---|
| | | Prediction | |
| | | Negative | Positive |
| True | Negative | 6586 | 420 |
| | Positive | 170 | 7140 |

According to the obtained results, the NB algorithm correctly classified 12176 data out of 14316 and achieved an accuracy value of 85.05%. Additionally, the hyperparameters of the best performing model were set as 'alpha': 0.1. The XGBoost algorithm correctly classified 13726 data out of 14316 and achieved an accuracy value of 95.88%. Additionally, the hyperparameters of the best performing model were set as {'learning_rate': 0.2, 'max_depth': 6, 'n_estimators': 200}. Table 16 below shows the confusion matrices of LightGBM.

**Table 16. Complexity matrix results of the LightGBM algorithm**

| | | Prediction | |
|---|---|---|---|
| | | Negative | Positive |
| True | Negative | 6673 | 333 |
| | Positive | 158 | 7152 |

According to the results, the LightGBM algorithm correctly classified 13825 out of 14316 data points, achieving an accuracy value of 96.57%. Furthermore, the best-performing model had hyperparameters of {'learning_rate': 0.2, 'max_depth': 20, 'n_estimators': 200}. Table 17 below compares all methods in terms of their success metrics.

**Table 17. Evaluation criteria of RF, NB, LR, XGBoost and LightGBM models with N-Gram TF-IDF**

| Model | Accuracy | Precision | Recall | F1-Score |
|---|---|---|---|---|
| LightGBM | 96,57% | 96,62% | 96,54% | 95,87% |
| XGBoost | 95,88% | 95,96% | 95,84% | 95,87% |
| LR | 94,89 % | 94,91% | 94,87% | 94,88% |
| RF | 94,59% | 94,69% | 94,54 % | 94,58% |
| NB | 85,05% | 85,19% | 84,98% | 85,01% |

According to Table 17 above, the most successful model is LightGBM. This model received the highest score in every classification metric. After LightGBM, the most successful algorithm is XGBoost, followed by LR, RF, and NB.

## 4. COMPARISON OF MACHINE LEARNING METHODS WITH INTUITIONISTIC FUZZY DECISION MAKING

The dataset above was evaluated using five different machine learning methods. Using machine learning methods such as LightGBM, XGBoost, LR, RF, and NB, the samples in the dataset were evaluated based on the classification metrics Accuracy, Precision, Recall, and F1-Score. Models were individually analyzed and interpreted based on each metric. Establishing a mechanism that simultaneously evaluates all classification metric results will greatly facilitate the performance of machine learning methods. This is because a model that performs poorly according to the Accuracy metric may perform better than other models according to the Precision metric. Changing the model's performance across all metrics is a common occurrence, especially in imbalanced datasets. Based on the reasons mentioned above, this study aims to establish a system that simultaneously evaluates models based on all classification metrics. Furthermore, intuitionistic fuzzy sets were utilized to minimize uncertainty in the dataset. Unlike classical sets, intuitionistic fuzzy sets, which examine membership, non-membership, and sensitivity levels within a [0,1] unit interval, facilitate the work of experts, especially in sensitive situations. This study aims to establish a decision-making mechanism by considering intuitionistic fuzzy sets and decision-making methods together.

After fuzzy sets were established in 1965 to express uncertainty, intuitionistic fuzzy sets were defined [22,23]. Multi-criteria decision-making processes evaluate alternatives among many based on specific criteria. In this study, the PROMETHEE method, a decision-making method, was used in conjunction with intuitionistic fuzzy sets [24,25,26].

**Definition 2.1**. [22,25] Let $X \neq \emptyset$ . An IF set $A$ in $X$;

$$A = \{\langle x, \mu_A(x), \nu_A(x)\rangle |\ x \in X\} \tag{6}$$

$$\mu_A(x), \nu_A(x), \pi_A(x): X \to [0,1].$$

$$\mu_A(x) + \nu_A(x) + \pi_A(x) = 1.$$

IF value (IFV) described [25]. IFV is presented: $\tilde{a} = (\mu_{\tilde{a}}, \nu_{\tilde{a}}, \pi_{\tilde{a}})$, where $\mu_{\tilde{a}}, \nu_{\tilde{a}}, \pi_{\tilde{a}} \in [0,1]$.

For each IFS $\tilde{A}$;

$$\pi_{\tilde{A}} = 1 - \mu_{\tilde{A}} - \nu_{\tilde{A}} \tag{7}$$

For IFVs $\tilde{f}$ and $\tilde{t}$ the following operations have been carried out [27,28]:

$$\tilde{f} \oplus \tilde{t} = \left(\mu_{\tilde{f}} + \mu_{\tilde{t}} - \mu_{\tilde{f}}\mu_{\tilde{t}}, \nu_{\tilde{f}}\nu_{\tilde{t}}\right) \tag{8}$$

$$\tilde{f} \otimes \tilde{t} = \left(\mu_{\tilde{f}}\mu_{\tilde{t}}, \nu_{\tilde{f}} + \nu_{\tilde{t}} - \nu_{\tilde{f}}\nu_{\tilde{t}}\right) \tag{9}$$

$$\bigoplus_{j=1}^{m} \mathrm{f} = \left(1 - \prod_{j=1}^{m}\left(1 - \mu_j\right), \prod_{j=1}^{m} \nu_j\right) \tag{10}$$

$$\bigotimes_{j=1}^{m} \mathrm{f} = \left(\prod_{j=1}^{m} \mu_j, \prod_{j=1}^{m}\left(1 - \nu_j\right)\right) \tag{11}$$

A lot of researchers have suggested ways to compare IFVs [27,28,29]. The method that follows will be utilized in this paper [30]:

$$\rho(\alpha) = 0.5(1 + \pi_\alpha)(1 - \mu_\alpha) \tag{12}$$

### 4.1. The IF PROMETHEE

The preference for the intuitionistic fuzzy PROMETHEE method in this study provides significant methodological and analytical advantages, particularly in multi-criteria decision-making (MCDM) problems where uncertainty, ambiguity, and subjective judgments are prevalent. The main advantages of this method are as follows:

- Intuitionistic fuzzy sets represent the missing and uncertain information in decision-makers' evaluations more comprehensively, taking into account not only the degree of membership but also the degrees of non-membership and hesitation (uncertainty). This offers a model more suited to the nature of the decision-making process compared to classical and fuzzy PROMETHEE approaches.
- When the outranking logic of the PROMETHEE method, based on pairwise comparisons, is integrated with the intuitionistic fuzzy structure, the preference intensities, reservations, and hesitations of decision-makers can be included in the analysis simultaneously. Thus, preference relationships become more flexible and discriminatory.
- In problems where qualitative criteria are dominant, the ability to express expert opinions through intuitionistic fuzzy numbers without imposing numerical precision is a significant advantage of the method. This minimizes information loss in decision problems involving human judgments.
- Including the degree of hesitation in the modeling allows for a more precise distinction between close performance differences among alternatives. This increases the method's discriminative power, especially when ranking alternatives with similar characteristics.
- The intuitionistic fuzzy PROMETHEE method can be easily integrated with different preference functions, weighting approaches, and uncertainty levels. This flexibility allows the method to be used effectively in different disciplines and application areas.
- Explicitly modeling the hesitation information ensures that the resulting ranking and superiority results are more reliable and interpretable for decision-makers. Thus, the decision support process is strengthened both analytically and cognitively. In conclusion, the intuitionistic fuzzy PROMETHEE method has a significant advantage as a flexible decision support tool that more accurately reflects decision-maker behavior in uncertain, subjective, and multi-criteria decision problems.

Advantageous results can be achieved through the use of IF PROMETHEE. The weights could be figured as IFVs: $\widetilde{w}_j$ where $\mu_{\widetilde{w}_j} \in [0,1], \nu_{\widetilde{w}_j} \in [0,1], \mu_{\widetilde{w}_j} + \nu_{\widetilde{w}_j} \leq 1, j = 1,2,\ldots,m$. Decision makers can use different methods to determine fuzzy weights that are intuitionistic. This research utilized the V shape criterion type [31]:

$$P(d) = \begin{cases} 0, & d \leq w \\ \frac{d-w}{z-w}, & w < d \leq z \\ 1, & d > z \end{cases} \tag{13}$$

Thresholds $w$ and $z$ are presented as indifference and strict preference, respectively. Calculate the alternatives $x_i (i = 1,2,\ldots,n)$ relating to $c_j (j = 1,2,\ldots,m)$ and decide the deviations:

$$d_j(x,y) = c_j(x) - c_j(y) \tag{14}$$

**Definition 2.2** An IF preference relation $R$ on the set $X = x_1, x_2, \ldots, x_n$ is presented by $R = (r_{ik})_{n\times n}$, where $r_{ik} = \langle (x_i, x_k), \mu(x_i, x_k), \nu(x_i, x_k) \rangle$ for all $i, k = 1,2,\ldots,n$. $\mu_{ik} + \nu_{ik} \leq 1,\ \mu_{ik} = \nu_{ki},\ \mu_{ki} = \nu_{ik}$
$\mu_{ii} = \nu_{ii} = 0.5,\ \pi_{ik} = 1 - \mu_{ik} - \nu_{ik}$,
$\forall\ i, k = 1,2,\ldots,n$. The preference matrix is determined, then matrix of the IF preference relation is follows:

$$R^{(j)} = \left(r_{ik}^{(j)}\right)_{n\times n} = \begin{bmatrix} - & \left(\mu_{12}^{(j)}, \nu_{12}^{(j)}\right) & \cdots & \left(\mu_{1n}^{(j)}, \nu_{1n}^{(j)}\right) \\ \left(\mu_{21}^{(j)}, \nu_{21}^{(j)}\right) & - & \cdots & \left(\mu_{2n}^{(j)}, \nu_{2n}^{(j)}\right) \\ \vdots & \vdots & - & \vdots \\ \left(\mu_{n1}^{(j)}, \nu_{n1}^{(j)}\right) & \left(\mu_{n2}^{(j)}, \nu_{n2}^{(j)}\right) & \cdots & - \end{bmatrix} \tag{15}$$

There are a lot of aggregation operators for IF. The IFWA operator was utilized in research.

$$r(x_i, x_k) = r_{ik} = \bigoplus_{j=1}^{m} \left(\widetilde{w}_j \otimes r_{ik}^{(j)}\right) \tag{16}$$

Further, $r_{ik}$ is an IFV. $\widetilde{w}_j = \left(\mu_{\widetilde{w}j}, \nu_{\widetilde{w}j}\right)$, then according to;

$$\widetilde{w}_j \otimes r_{ik}^{(j)} = \left(\mu_{ik}^{(j)} \mu_{\widetilde{w}j}, \nu_{ik}^{(j)} + \nu_{\widetilde{w}j} - \nu_{ik}^{(j)} \nu_{\widetilde{w}j}\right) \tag{17}$$

$$r(x_i, x_k) = \bigoplus_{j=1}^{m} \left(\widetilde{w}_j \otimes r_{ik}^{(j)}\right)$$

$$= \begin{pmatrix} 1 - \prod_{j=1}^{m} \left(1 - \mu_{ik}^{(j)} \mu_{\widetilde{w}j}\right), \\ \prod_{j=1}^{m} \left(\nu_{ik}^{(j)} + \nu_{\widetilde{w}j} - \nu_{ik}^{(j)} \nu_{\widetilde{w}j}\right) \end{pmatrix} \tag{18}$$

Overall IF preference relationship is determined:

$$R = (r_{ik})_{n\times n} = \begin{bmatrix} - & (\mu_{12}, \nu_{12}) & \dots & (\mu_{1n}, \nu_{1n}) \\ (\mu_{21}, \nu_{21}) & - & \dots & (\mu_{2n}, \nu_{2n}) \\ \vdots & \vdots & - & \vdots \\ (\mu_{n1}, \nu_{n1}) & (\mu_{n2}, \nu_{n2}) & \dots & - \end{bmatrix} \tag{19}$$

Every alternative is checked to option $(n-1)$. Consequently IF outranking flows could be obtained:
(1) The IF positive outranking flow:

$$\tilde{\varphi}^{+}(x_i) = \frac{1}{n-1} \bigoplus_{k=1,k\neq i}^{n} r(x_i, x_k) = \frac{1}{n-1} \bigoplus_{k=1,k\neq i}^{n} r_{ik} \tag{20}$$

(2) The IF negative outranking flow:

$$\tilde{\varphi}^{-}(x_i) = \frac{1}{n-1} \bigoplus_{k=1,k\neq i}^{n} r(x_k, x_i) = \frac{1}{n-1} \bigoplus_{k=1,k\neq i}^{n} r_{ki} \tag{21}$$

The relationship between $\tilde{\varphi}^{+}(x_i)$ and $\tilde{\varphi}^{-}(x_i)$ is hereinbelow:

$$\rho\big(\varphi(x_i)\big) = \rho\big(\tilde{\varphi}^{+}(x_i)\big) - \rho\big(\tilde{\varphi}^{-}(x_i)\big) \tag{22}$$

The steps for the IF PROMETHEE have been designed:
**Step 1:** Alternatives $X = \{x_1, x_2, \dots, x_n\}$ are decided. And criteria $C = \{c_1, c_2, \dots, c_m\}$ that are important to evaluate alternatives are decided.
**Step 2:** The importance degree of the criteria is decided $\tilde{w}_j (j = 1,2, \dots, m)$.
**Step 3:** $w$ and $z$ thresholds are specified. Deviations $d_j(x, y)$ using Equation (10), the preferences $\mu_{ik}^{(j)}$ are calculated. Then, the preference matrix $U^{(j)} (j = 1,2, \dots, m)$ are created.
**Step 4:** The IF preference relation $R^{(j)} = \left(r_{ik}^{(j)}\right)_{n\times n}$ is created.
**Step 5:** The overall IF preference relation $R = (r_{ik})_{n\times n}$ utilizing Equation (18) is calculated.
**Step 6:** $\tilde{\varphi}^{+}(x_i)$ and $\tilde{\varphi}^{-}(x_i)$ are calculated.
**Step 7:** The connection of $\tilde{\varphi}^{+}(x_i)$ and $\tilde{\varphi}^{-}(x_i)$ is decided.

In this section, the classification results of machine learning models are evaluated using the intuitionistic fuzzy PROMETHEE method. The criteria set is determined as {Accuracy, Precision, Recall, F1-Score} and the alternatives set is {LightGBM, XGBoost, RF, LR, NB}. While creating the decision matrix, the values of each model according to the evaluation criteria are expressed as intuitionistic fuzzy values.

**Table 18. Performance of models for TF-IDF on Dataset 1 evaluated with intuitionistic fuzzy PROMETHEE**

| Alternatives/Criteria | K1/Accuracy | K2/Precision | K3/Recall | K4/F1-Score |
|---|---|---|---|---|
| LightGBM | 0,9979 | 0,9978 | 0,9979 | 0,9979 |
| XGBoost | 0,9978 | 0,9977 | 0,9978 | 0,9978 |
| RF | 0,9976 | 0,9977 | 0,9976 | 0,9975 |
| LR | 0,9935 | 0,9935 | 0,9936 | 0,9935 |
| NB | 0,9324 | 0,9324 | 0,9321 | 0,9322 |

Expert opinions were consulted when determining the criteria weights, and the criteria weights were determined as Medium, Important, Important, and Very Important, respectively. The primary reason for consulting expert opinions in determining criterion weights in this study is that the classification success metrics (Accuracy, Precision, Recall, and F1-Score) used in the fake news detection problem can have different levels of importance depending on the dataset and problem structure. Especially in text-based classification problems and datasets with different balance levels, the power of each metric to represent model success varies. Therefore, instead of assuming fixed and universal relative importance, it was preferred to determine them through expert evaluations based on domain knowledge. Experts, drawing on their experience in fake news detection and machine learning, evaluated the discriminative power and reliability of each classification metric within the context of the relevant dataset. These evaluations were expressed using intuitionistic fuzzy values instead of precise numerical weights. Thus, not only the experts' opinions regarding the importance of a criterion but also the level of uncertainty and hesitation in these opinions were included in the decision-making process. The numerical equivalents of these criteria weights are {(0.50,0.45),(0.75,0.20),(0.75,0.20),(0.90,0.10)}.The decision matrix is created in the table below. First, the classification results of the RF, NB, LR, XGBoost, and LightGBM models using TF-IDF on Dataset 1 samples were evaluated.

**Table 19. Value of performance of models for TF-IDF on Dataset 1 evaluated with intuitionistic fuzzy PROMETHEE**

| Alternatives | Positive Outranking Flow | Negative Outranking Flow | Net Outranking Flow |
|---|---|---|---|
| LightGBM | 0,945514463 | 1 | -0,054485537 |
| XGBoost | 0,945975138 | 0,999863147 | -0,053888009 |
| RF | 0,946874837 | 0,999155358 | -0,05228052 |
| LR | 0,953200841 | 0,988346183 | -0,035145343 |
| NB | 1 | 0,8425684 | 0,1574316 |

Secondly, the classification results of the Ngram-TF-IDF and RF, NB, LR, XGBoost and LightGBM Models of Dataset 1 samples were evaluated.

**Table 20. Performance of models for N-gram TF-IDF on Dataset 1 evaluated with intuitionistic fuzzy PROMETHEE**

| Alternatives/Criteria | K1/Accuracy | K2/Precision | K3/Recall | K4/F1-Score |
|---|---|---|---|---|
| LightGBM | 0,9979 | 0,9978 | 0,9979 | 0,9979 |
| XGBoost | 0,9982 | 0,9982 | 0,9982 | 0,9982 |
| RF | 0,9983 | 0,9983 | 0,9984 | 0,9983 |
| LR | 0,9945 | 0,9945 | 0,9945 | 0,9945 |
| NB | 0,9489 | 0,9487 | 0,9489 | 0,9488 |

**Table 21. Value of performance of models for N-gram TF-IDF on Dataset 1 evaluated with intuitionistic fuzzy PROMETHEE**

| Alternatives | Positive Outranking Flow | Negative Outranking Flow | Net Outranking Flow |
|---|---|---|---|
| LightGBM | 0,946874837 | 0,999155358 | -0,05228052 |
| XGBoost | 0,945975138 | 0,999863147 | -0,053888009 |
| RF | 0,945514463 | 1 | -0,054485537 |
| LR | 0,953200841 | 0,988346183 | -0,035145343 |
| NB | 1 | 0,8425684 | 0,1574316 |

When comparing the models based on the results obtained with the intuitionistic fuzzy PROMETHEE method: With TF-IDF, the most successful model is LightGBM, followed by XGBoost, RF, LR, and NB, respectively. With N-gram TF-IDF, the most successful model is RF, followed by XGBoost, LightGBM, LR, and NB, respectively.
Secondly, the classification results of TF-IDF and RF, NB, LR, XGBoost and LightGBM Models of Dataset 2 samples were evaluated.

**Table 22. Performance of models for TF-IDF on Dataset 2 evaluated with intuitionistic fuzzy PROMETHEE**

| Alternatives/Criteria | K1/Accuracy | K2/Precision | K3/Recall | K4/F1-Score |
|---|---|---|---|---|
| LightGBM | 0,9658 | 0,9663 | 0,9655 | 0,9657 |
| XGBoost | 0,9578 | 0,9586 | 0,9574 | 0,9577 |
| RF | 0,944 | 0,9449 | 0,9435 | 0,9439 |
| LR | 0,9443 | 0,9444 | 0,9442 | 0,9443 |
| NB | 0,8404 | 0,8416 | 0,8397 | 0,84 |

**Table 23. Value of performance of models for TF-IDF on Dataset 2 evaluated with intuitionistic fuzzy PROMETHEE**

| Alternatives | Positive Outranking Flow | Negative Outranking Flow | Net Outranking Flow |
|---|---|---|---|
| LightGBM | 0,850659144 | 1 | -0,149340856 |
| XGBoost | 0,871266658 | 0,991446023 | -0,120179365 |
| RF | 0,901363623 | 0,962979493 | -0,061615871 |
| LR | 0,901002962 | 0,963643834 | -0,062640871 |
| NB | 0,999999999 | 0,738040003 | 0,261959997 |

Classification results of RF, NB, LR, XGBoost and LightGBM Models with N-gram TF-IDF of Dataset 2 samples were evaluated.

**Table 24. Performance of models for N-gram TF-IDF on Dataset 2 evaluated with intuitionistic fuzzy PROMETHEE**

| Alternatives/Criteria | K1/Accuracy | K2/Precision | K3/Recall | K4/F1-Score |
|---|---|---|---|---|
| LightGBM | 0,9657 | 0,9662 | 0,9654 | 0,9587 |
| XGBoost | 0,9588 | 0,9596 | 0,9584 | 0,9587 |
| RF | 0,9459 | 0,9469 | 0,9454 | 0,9458 |
| LR | 0,9489 | 0,9491 | 0,9487 | 0,9488 |
| NB | 0,8505 | 0,8519 | 0,8498 | 0,8501 |

**Table 25. Value of performance of models for N-gram TF-IDF on Dataset 2 evaluated with intuitionistic fuzzy PROMETHEE**

| Alternatives | Positive Outranking Flow | Negative Outranking Flow | Net Outranking Flow |
|---|---|---|---|
| LightGBM | 0,868125603 | 1 | -0,131874397 |
| XGBoost | 0,88111433 | 0,994909929 | -0,113795599 |
| RF | 0,90831742 | 0,965542532 | -0,057225112 |
| LR | 0,903693013 | 0,97409388 | -0,070400867 |
| NB | 1 | 0,747998406 | 0,252001594 |

When comparing the models based on the results obtained with the intuitionistic fuzzy PROMETHEE method: With TF-IDF, the most successful model is LightGBM, followed by XGBoost, LR, RF, and NB. With N-gram TF-IDF, the most successful model is LightGBM, followed by XGBoost, LR, RF, and NB.

## 5. CONCLUSION AND COMPARISON

In this study, five popular machine learning methods—LightGBM, XGBoost, RF, LR, and NB—were used on two datasets consisting of fake news. The performance of these methods was measured utilizing Accuracy, Precision, Recall, and F1-Score. Complex matrices of RF, NB, LR, XGBoost, and LightGBM obtained with TF-IDF and N-gram TF-IDF are shown. Gridsearch hyperparameter optimization and 5-fold cross-validation were applied to all methods to improve their performance. Evaluation metrics for all models were calculated using TF-IDF and N-gram TF-IDF for Dataset 1 and Dataset 2, respectively. In continuation of the study, a decision-making mechanism was created to measure the success of machine learning methods. In this mechanism, machine learning methods represent alternatives, and classification metrics represent criteria. All values are expressed as intuitionistic fuzzy values to reduce sensitivity. The importance of the criteria varies depending on the datasets used and was determined by expert opinions in this study. Using the PROMETHEE method, a decision-making method, positive and negative flow values were calculated among the models, and success rankings were obtained based on net flow values.

For dataset 1, the most successful model is LightGBM in the calculations performed using TF-IDF, followed by the XGBoost, RF, LR, and NB algorithms, respectively. For dataset 1, the most successful model is LightGBM in the calculations performed using the intuitionistic fuzzy PROMETHEE method with TF-IDF, followed by the XGBoost, RF, LR, and NB models, respectively. A comparison of the results obtained reveals consistency. In the calculations performed using N-gram TF-IDF for dataset 1, the most successful model is RF, followed by the XGBoost, LightGBM, LR, and NB algorithms, respectively. In the calculations performed using N-gram TF-IDF and the intuitionistic fuzzy PROMETHEE method for dataset 1, the most successful model is RF, followed by the XGBoost, LightGBM, LR, and NB models, respectively. A comparison of the results obtained reveals consistency.

For dataset 2, the most successful model was LightGBM in the calculations performed using TF-IDF, followed by the XGBoost, RF, LR, and NB algorithms, respectively. For dataset 2, the most successful model was LightGBM in the calculations performed using the TF-IDF and intuitionistic fuzzy PROMETHEE method, followed by the XGBoost, LR, RF, and NB models, respectively. Comparing the results resulted in the observation that the rankings of the RF and LR models changed. When interpreting machine learning models based on the success of the classification metrics, the successes of the models were compared using the intuitionistic fuzzy PROMETHEE method, taking into account their success in all classification metrics. This is the reason for the difference in model success rankings. In the calculations performed using N-gram TF-IDF for dataset 2, LightGBM is the most successful model, followed by the XGBoost, LR, RF, and NB algorithms, respectively. In the calculations performed using N-gram TF-IDF and the intuitionistic fuzzy PROMETHEE method for dataset 2, LightGBM is the most successful model, followed by the XGBoost, LR, RF, and NB models, respectively. A comparison of the results obtained reveals consistency.

As a result, the success of machine learning models was compared by creating a decision-making mechanism using intuitionistic fuzzy sets. In the first stage of the study, classical sets were used to evaluate the dataset samples, while in the second stage, the success of the models according to the metrics was expressed using intuitionistic fuzzy values. This was to minimize the uncertainty in the model success results. The results obtained in the first stage of the study were evaluated in the second stage using a decision-making mechanism. When evaluating machine learning models, experts provide subjective opinions based on each classification metric to determine the successful model. However, especially in imbalanced datasets, the success of each classification metric can vary. To eliminate subjectivity and enable more general and objective interpretation, a mechanism was established that evaluates each classification metric simultaneously. When comparing the results of the first and second stages revealed that generally consistent results were obtained. The reason for the consistency is that the dataset samples studied are balanced and the model results are very close to each other according to the metrics. However, if the dataset samples are unbalanced or the model metric results are not close to each other, the results obtained with the established decision-making mechanism will differ. The novel aspect of this study is the use of five different machine learning models for fake news detection and the evaluation and comparison of the success of the models using the intuitionistic fuzzy PROMETHEE method. Given that the importance levels of the classification metrics will vary depending on expert opinions when applied to different dataset samples, objective results will be obtained for each dataset sample and each model. Evaluating model success using the decision-making mechanism will provide experts with a new perspective on interpretation.

## Authors' Contributions

In this study, author was responsible for all aspects of the study, including conceptualization, methodology, data collection and analysis, and manuscript preparation.

## Competing Interests

The author declare that they have no conflict of interest.

## References

[1] Conroy, N., Rubin, V. and Chen, Y., Automatic deception detection: Methods for finding fake news, Proceedings of the Association for Information Science and Technology, 2015, 52(1), pp.1-4.

[2] Granik, M., and Mesyura, V., Fake news detection using naive Bayes classifier, In 2017 IEEE first Ukraine conference on electrical and computer engineering (UKRCON), 2017, pp. 900-903.

[3] WHO, 2022. https://www.who.int/about/accountability/results/who-results-report-2020-2021

[4] Ahmed, H., Traore, I., and Saad, S., Detection of online fake news using n-gram analysis and machine learning techniques. In International conference on intelligent, secure, and dependable systems in distributed and cloud environments, 2017, pp. 127-138). Cham: Springer International Publishing.

[5] Wang, W. Y., " Liar, liar pants on fire": A new benchmark dataset for fake news detection, arXiv preprint arXiv:1705.00648, 2017.

[6] Khan, W., Daud, A., Khan, K., Nasir, J. A., Basheri, M., Aljohani, N., and Alotaibi, F. S., Part of speech tagging in urdu: Comparison of machine and deep learning approaches. IEEE Access, 2019, 7, pp.38918-38936.

[7] Hu, B., Mao, Z., and Zhang, Y., An overview of fake news detection: From a new perspective, Fundamental Research, 2025, 5(1), pp.332-346.

[8] Ahmad, I., Yousaf, M., Yousaf, S. and Ahmad, M. O., Fake news detection using machine learning ensemble methods, Complexity, 2020, 1, 8885861.

[9] Baarir, N. F., and Djeffal, A., Fake news detection using machine learning, In 2020 2nd International workshop on human-centric smart environments for health and well-being (IHSH), 2021, pp. 125-130.

[10] Manzoor, S. I. and Singla, J., Fake news detection using machine learning approaches: A systematic review. In 2019 3rd international conference on trends in electronics and informatics (ICOEI), 2019, pp. 230-234.

[11] Sharma, U., Saran, S. and Patil, S. M., Fake news detection using machine learning algorithms. International Journal of creative research thoughts (IJCRT), 2020, 8(6), pp.509-518.

[12] Alghamdi, J., Lin, Y. and Luo, S., A comparative study of machine learning and deep learning techniques for fake news detection, Information, 2022, 13(12), 576.

[13] Alghamdi, J., Luo, S. and Lin, Y., A comprehensive survey on machine learning approaches for fake news detection, Multimedia Tools and Applications, 2024, 83(17), pp.51009-51067.

[14] Yakkundi, S., Patil, R., Sangani, S., Goudar, R. H., Goudar, S. I. and Qazi, A., Exploring machine learning for fake news detection: techniques, tools, challenges, and future research directions, Discover Applied Sciences, 2025, 7(8), 873.

[15] Can, Ü. and Alatas, B., A new direction in social network analysis: Online social network analysis problems and applications, Physica A: Statistical Mechanics and its Applications, 2019, 535, 122372.

[16] Can, Ü. and Alataş, B., Review of Sentiment Analysis and Opinion Mining Algorithms, International Journal of Pure and Applied Sciences, 2017, 3(1), 75-111.

[17] Akinsola, J. E. T., Awodele, O., Kuyoro, S. O., and Kasali, F. A., Performance evaluation of supervised machine learning algorithms using multi-criteria decision making techniques, In Proceedings of the International Conference on Information Technology in Education and Development (ITED), 2019, pp. 17-34.

[18] Tuğrul, F., Examination of Recycled Papers Using an Intuitionistic Fuzzy Set-Based MCDM. In Advanced Mathematics for the Modeling and Solution of Challenging Problems in Engineering, 2025, pp 165-178. Cham: Springer Nature Switzerland.

[19] Tuğrul, F., An innovative application on supermarket selection through using intuitionistic fuzzy TOPSIS method, Sakarya University Journal of Science, 2022, 26(5), pp.1029-1039.

[20] URL-1: https://www.kaggle.com/code/abaghyangor/fake-news#Introduction

[21] URL-2: https://www.kaggle.com/datasets/saurabhshahane/fake-news-classification/data

[22] Atanassov, K.T., Intuitionistic fuzzy sets, Fuzzy Sets and Systems, 1986, 20(1), 87-96.

[23] [Zadeh, L.A., Fuzzy Sets. Information and Control, 1965, 8, pp.338-353.

[24] Brans, J.P., Lingenierie de la decision lelaboration dinstruments daidea la decision la methode PROMETHEE. In: Laidea la decision Nature, Instruments et Perspectives d Avenir (Edt. Nadeau, R., Landry, M), Les Presses de l'universite Laval, ISBN:2-7637-7084-3, Quebec, Canada.1982, pp.183-213.

[25] Atanassov, K.T., Intuitionistic fuzzy sets, International Journal Bioautomation, 2016, 20(S1), pp.1-6.

[26] Brans, J.P., Mareschal, B., and Vincke, P., PROMETHEE: a new family of outranking methods in multicriteria analysis. Operational Research, 1984, 84, pp.477-490.

[27] Xu, Z.S., and Yager, R.R., Some geometric aggregation operators based on intuitionistic fuzzy set, International Journal of General Systems, 2006, 35, 417-433.

[28] Xu, Z.S., Intuitionistic fuzzy aggregation operators, IEEE Transactions on Fuzzy Systems, 2007, 15, 1179-1187.

[29] Xu, Z.S., Intuitionistic preference relations and their application in group decision making, Information Sciences, 2007, 177(11), 2363-2379.

[30] Szmidt, E., and Kacprzyk, J., Amount of information and its reliability in the ranking of Atanassov's intuitionistic fuzzy alternatives. In: Recent advances in decision making (Edt. Rakus-Andersson, E., Yager, R.R., Ichalkaranje, N., Jain, L) Springer, ISBN: 978-3-642-02187-9, Berlin, Heidelberg, 2009, pp.7-19.

[31] Vincke, P., and Brans, J.P., A preference ranking organization method. The PROMETHEE method for MCDM. Management Science, 1985, 31(6), pp.647-656.