

A Comparative Study of Machine Learning Models for Fake News Detection Using the LIAR Dataset

Mr. Jugendra Pal Singh

B. Tech (Information Technology)

NIET, Greater Noida

E-mail: jugendra6395@gmail.com

Ms. Neetu Kumari Rajput²

Assistant Professor (IT Department)

NIET, Greater Noida

E-mail: neetu.rajput@niet.co.in

Abstract: Due to the fast spread of fake news over online platforms, fake news detection has become an important issue in current research. It is essential to use automatic detection systems to recognize misinformation without losing the accuracy of classification. Machine learning models can be used as a means of analysing text data for fake news detection purposes.

In this paper, a machine learning model-based approach to detect fake news using the LIAR dataset is presented. The initial multiclass classification problem has been converted into a binary one based on clustering all of the statements as either fake or real. The proposed framework involves pre-processing text, creating relevant features, as well as taking into account the following contextual aspects: speaker, subject, political party, and context.

Term Frequency-Inverse Document Frequency (TF-IDF) with an n-gram technique was used as a method for extracting text features. Logistic Regression, Naive Bayes, and SVM models have been implemented and analysed.

It can be seen from experimental results that Naive Bayes outperforms other approaches in terms of accuracy, while Logistic Regression with balancing classes demonstrates better results in regard to the recall of the minority class.

Keywords—*Fake News Detection, Machine Learning, Text Classification, TF-IDF, Logistic Regression, Support Vector Machine, LIAR Dataset.*

1.Introduction

The evolution of the digital media and the rise of online news channels have significantly affected how the process of news creation, distribution, and consumption occurs. On the one hand, new forms of media have increased accessibility and improved the communication process. On the other hand, it made the spreading of misinformation or what has become known as fake news possible. Such phenomena present many difficulties and problems that include public perception alteration, confusion and, in some cases, political and social instabilities. Therefore, the development of automated fake news detectors became an important topic in machine learning and natural language processing studies.

The main idea behind fake news detection lies in the analysis of information based on the textual content of claims and the context in which it occurs. Traditional ways of fact-checking require significant effort and cannot cope with the high volumes of information available online. However, the machine learning algorithms can solve the problem efficiently by creating models based on the pattern recognition from training data.

LIAR dataset presented by William Yang Wang became one of the popular benchmarks for fake news detection problems.

It includes relatively short political claims categorized into various classes depending on the level of truthfulness, along with additional metadata information such as speakers, topics, parties, etc. However, factors like class imbalance and inadequate contextual awareness may affect the model's ability to learn from data correctly.

The aim of this paper is to construct a machine learning algorithm for fake news detection with LIAR dataset. For that purpose, we will treat LIAR as a binary classification problem, and then combine traditional pre-processing and feature engineering techniques and apply TF-IDF method for feature selection. Finally, several machine learning algorithms are chosen and tested to address class imbalance issues effectively.

In general, the purpose of this research is to compare the effectiveness of different models' performance in the case of a binary classification problem and highlight the importance of balanced evaluation metrics.

2.Literature review

Detection of fake news has proven to be an important research area due to the high rate of circulation of false information through various digital media channels. There have been several methods proposed in literature ranging from classical machine learning algorithms to more advanced deep learning techniques. Initially, most of the research work done was mainly focused on content-based analysis where features were obtained from the textual contents to differentiate fake news from genuine ones.

One of the well-known datasets for detection of fake news is LIAR, which is created by William Yang Wang. This data set not only includes labels for political statements but also contextual metadata. Various machine learning algorithms such as Logistic Regression, Naive Bayes Classifier, and Support Vector Machine (SVM) are used with this dataset, with results showing that such techniques produce mediocre accuracy due to problems associated with short texts and class imbalance.

Kai Shu et al. gave an extensive review of fake news detection techniques from data mining standpoint where it highlights the importance of using not only the content features but also the user and social contextual features to obtain high quality classifier.

Effects of fake news on shaping of public opinions have also been the subject of many studies. One such work has been performed by Hunt Allcott and Matthew Gentzkow analysing the effects of misinformation in politics.

With recent advances in computational capacity of computer, deep learning approaches have been proposed for fake news detection. Long Short-Term Memory network (LSTM) has been implemented to detect sequence dependencies, while transformers such as BERT have shown state of the art performance in a variety of natural language processing problems. These techniques give better contextual understanding than traditional ML, but require higher computational power and bigger training sets.

Although there have been recent success with deep learning techniques, machine learning algorithms still remain relevant due to simplicity and lower requirement of computational resource and training data. Class imbalance problem persists and needs further solution.

In this study, previous research has been expanded by incorporating both textual and contextual features, while attempting to solve class imbalance problem through class weighing techniques. By implementing a number of different machine learning techniques in this task, insight on balancing performance and reliability could be acquired.

3. Problem Statement and Objective

3.1. Problem Statement

The proliferation of online media channels has been paralleled by a rise in the spread of misinformation, making it more challenging to distinguish between true and false information. The present approaches to fake news detection only depend on text-based information and might not incorporate other contextual information like who the speaker is, what topic is being discussed, and the political affiliation. In addition, some existing data sets like LIAR have a problem of imbalanced data classes, which makes the predictions made using such models biased towards the majority class and unable to detect the minority classes well. While many machine learning algorithms have performed extremely well, their performance does not extend to all data classes, especially when it comes to true news identification.

3.2. Objective of Study

The main objectives of this research are:

1. To develop a machine learning-driven model to detect fake news using the LIAR dataset.
2. To convert the multi-class classification problem into a binary classification problem to simplify the process.
3. To consider the features related to text as well as context, including the speaker, topic, and political party, to improve the efficiency of the model.
4. To solve the issue of class imbalance through techniques like class weighting.
5. To test various machine learning models and choose the best one.
6. To increase the efficiency of the model in terms of accuracy and reliability.

4. METHODOLOGY

4.1. Framework Overview

In the current research, we present a framework that uses a machine learning technique to detect fake news from LIAR dataset. The methodology covers several stages, namely, data preparation, feature engineering, feature extraction, model training, and evaluation.

4.2. Data Collection and Preparation

As previously stated, the LIAR dataset developed by William Yang Wang is used in our analysis. LIAR dataset contains short political statements that have been classified into six classes based on truthfulness. To transform this multi-class classification problem into binary classification problem, we aggregate all classes into two groups - 'real' (true, mostly true) and 'fake' (half-true, barely-true, false, pants-fire).

4.3. Data Preprocessing

Text data processing is conducted to prepare input data and minimize possible sources of errors and noises. All textual data are transformed into lower-case characters, punctuation marks, special characters and symbols are filtered out. Stop-words are filtered out from text using NLTK library. All these processes facilitate preparing consistent textual input data for further processing.

4.4. Feature Engineering

In order to improve performance, not only textual but also contextual features are used. Instead of working solely with the text part of statements, we add other features such as subject, speaker, party affiliation, and context to enrich textual data. Using these additional features improves semantic interpretation and increases chances for detecting underlying patterns in data.

4.5. Feature Extraction

All integrated textual data are transformed into numerical vectors using TF-IDF vectorization method. We configure vectorizer to extract maximum of 10,000 features and set n-gram range equal to (1,2). In this way, we are able to work not only with single-term features but also with phrases up to two terms long.

4.6. Model Training

There are three algorithms trained using prepared data - Logistic Regression, Naive Bayes and SVM. Logistic Regression is used as a benchmark algorithm because of its simplicity and interpretability. The latter one is added for its high efficiency when working with text classification problems. The third algorithm is SVM which has been proven to be effective in dealing with high-dimensional feature space. Class weighting is enabled for Logistic Regression and SVM algorithms as a solution for class imbalance problem.

4.7. Model Evaluation

Model quality is estimated using several quality metrics such as accuracy, precision, recall and F1-score. Specific attention is paid to recall of minority class. Confusion matrices are used for visualization and error analysis.

5. SYSTEM ARCHITECTURE AND DESIGN

5.1. Architecture Overview

As mentioned earlier, this project presents a modular and scalable machine learning pipeline that analyzes the textual data together with contextual metadata for the purposes of classifying news statements as real or fake. Thus, the system consists of several components such as data input, preprocessing, feature engineering, feature extraction, model

training, and evaluation. Every component serves some definite purpose that guarantees effective classification of the news data.

5.2.Data Input Layer

This layer includes the main inputs that feed the system. In this regard, we have decided to use the LIAR dataset that contains political statements classified as lies together with contextual metadata such as statement's subject, speaker, party affiliation, and context. The dataset is further split into the training and testing parts.

5.3.Preprocessing Layer

This component serves the purpose of preprocessing and normalization of the input text. All input text data are turned into lowercase format. Moreover, punctuation marks, special symbols, and stop words are removed from text. In order to eliminate noise in the text, standard NLP techniques are used to preprocess the dataset.

5.4.Feature Engineering Layer

At this stage, both textual and contextual features are used to obtain one vector per data instance. This is done by combining the statement text with additional features such as subject, speaker, party affiliation, and context. Such enriched input will help the algorithm capture semantic connections between textual features.

5.5.Feature Extraction Layer

The obtained textual data is represented by vectors by using the TF-IDF approach. The vectorization uses maximum feature size equal to 10,000 with n-grams being unigrams and bigrams. This strategy helps to represent text more comprehensively.

5.6.Model Training Layer

Feature vectors generated at the previous stage are provided as an input for machine learning models. This project implements Logistic Regression, Naive Bayes, and Support Vector Machine models. We have selected Logistic Regression because of its interpretability and good performance. Naive Bayes model has been chosen due to its effectiveness and efficiency in working with textual data. Moreover, SVM has been added as a strong alternative to logistic regression since SVM works well in multi-feature environment. Some models use class weighting to handle class imbalance problem.

5.7.Evaluation Layer

The last stage involves assessing performance metrics of machine learning models. Accuracy, precision, recall, and F1-score are used to measure classifier efficiency. Moreover, this layer includes generating confusion matrix and visualization graphs for comparative analyses.

6. IMPLEMENTATION DETAILS

A proposed solution to the problem under discussion involves building a fake news detection system using conventional machine learning algorithms in the Python programming language. The development will be performed with the use of libraries that implement the discussed algorithms and are developed in Python. The development process will be held in the Visual Studio Code IDE.

The first step of developing a system will involve data loading. In particular, the LIAR dataset provided in TSV format will be used. To do so, Pandas library will be employed. In addition, a specific data loader will be developed to give column names and convert multi-class labels into binary ones. Namely, "true" and "mostly-true" labels will be associated with the "real" category, while other labels will be considered as being fake.

The second stage will involve text preprocessing. It will be performed through the Natural Language Toolkit (NLTK). Text will be converted to lower case and cleaned from punctuation marks and special characters. In addition, NLTK stopwords will be removed. The implemented function will also be extended to consider cases where missing value or not a string occurs.

Feature engineering will imply aggregation of multiple attributes (statement, subject, speaker, party affiliation, context). The obtained text will be provided to the feature extraction part of the application.

In the next step, TF-IDF vectorization will be performed using Scikit-learn package. The maximal number of features will equal 10,000 while n-grams will include only two values – 1 and 2.

Then, three machine learning algorithms will be employed, namely Logistic Regression, Naïve Bayes, and SVM. Moreover, logistic regression and SVM models will take advantage of class-weight parameter. Training will occur using the train dataset and testing will be conducted based on the test dataset.

Finally, the implemented model will be assessed using four metrics (accuracy, precision, recall, F1-score). Accuracy comparisons will be shown in bar charts. Confusion matrices and comparison charts will be created using Matplotlib.

7.EXPERIMENTAL SETUP

7.1. Experimental Environment

The proposed fake news detection system was evaluated in a controlled local environment to simulate a machine learning-based text classification workflow using the LIAR dataset. The system configuration details are presented in **Table I**.

Table I: Configuration Details

Parameter	Configuration
Processor	Intel Core i5/equivalent
RAM	8 GB
Operating System	Windows 11
Java Version	Python 3.x
Libraries	Pandas, Scikit-learn, NLTK, Matplotlib
Development Tool	Visual Studio Code

The experiments were conducted using local execution of Python scripts, where the dataset was processed, models were trained, and evaluation metrics were computed.

7.2.Comparative Models

Three machine learning models were evaluated:

7.2.1.Logistic Regression (Balanced Model)

This model is implemented with class weighting to address dataset imbalance:

Characteristics:

- Linear classification model
- Class weight balancing applied
- Improved recall for minority class
- Stable and interpretable performance verification

7.2.2.Naive Bayes (Baseline Model)

This model serves as a baseline due to its efficiency in text classification:

- Probabilistic model
- Fast training and prediction
- High accuracy on majority class
- Limited handling of class imbalance

7.2.3.Support Vector Machine (SVM)

This model is used for handling high-dimensional feature spaces.

classification:

- Effective for text classification
- Handles high-dimensional TF-IDF features
- Class weight balancing applied
- Moderate performance across classes

7.3.Evaluation Metrics

The systems were evaluated using the following metrics:

7.3.1.Accuracy

Measures the overall correctness of predictions.

Unit:

- Percentage (%)

7.3.2.Precision, Recall, and F1-Score

These metrics evaluate classification performance for each class:

- Precision:** Correct positive predictions
- Recall:** Ability to detect actual positive instances
- F1-score:** Harmonic mean of precision and recall

This Special emphasis is placed on recall for the minority class (real news).

7.3.3.Confusion Matrix

Used to visualize classification results and identify misclassifications between real and fake news:

- invalid credentials
- unauthorized access
- expired sessions
- role violations

7.4.Experimental Procedure

For each model:

1. The dataset was pre-processed and transformed into binary classes.
2. Feature extraction was performed using TF-IDF.

3. Models were trained on the training dataset.
 4. Predictions were generated on the test dataset.
 5. Evaluation metrics were computed and recorded.
 6. Results were compared across all.
- This procedure ensured consistent and fair evaluation.

7.5.Expected Outcome

The experiment is designed to verify that the proposed system:

- improves detection performance using contextual features,
- effectively handles class imbalance,
- achieves balanced performance across classes,
- provides reliable and interpretable results for fake news detection.

8.RESULT AND ANALYSIS

The proposed fake news detection system was evaluated using three machine learning models: Logistic Regression, Naive Bayes, and Support Vector Machine (SVM). The models were tested on the LIAR dataset after preprocessing, feature engineering, and TF-IDF-based feature extraction.

8.1.Accuracy Comparison

The overall accuracy obtained from the experiments is shown in Fig. 1.

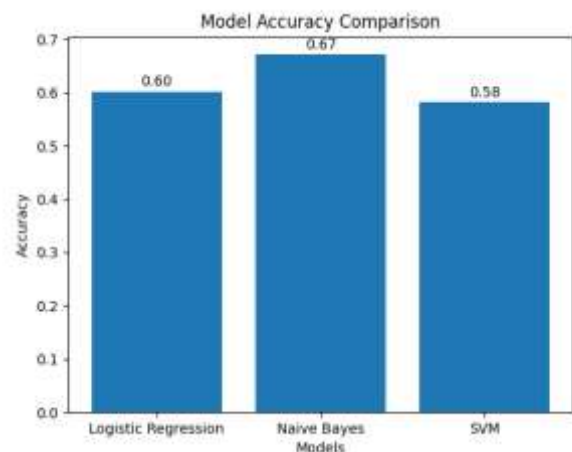


Fig. 1. Accuracy comparison of machine learning models.

Model	Accuracy
Logistic Regression	60.22%
Naive Bayes	67.16%
SVM	58.17%

Naive Bayes achieved the highest accuracy; however, accuracy alone does not fully represent model performance, especially in imbalanced datasets.

8.2. Class-wise Performance Analysis

A detailed evaluation using precision, recall, and F1-score reveals important insights. Logistic Regression achieved a recall of 57% for the minority class (real news), significantly outperforming Naive Bayes (19%) and SVM (49%). This indicates that Logistic Regression is more effective in identifying real news instances.

Naive Bayes, despite its higher accuracy, shows strong bias toward the majority class (fake news). This results in poor detection of real news, making it less suitable for practical applications where balanced detection is critical.

SVM demonstrates moderate performance with better balance than Naive Bayes but lower overall accuracy compared to Logistic Regression.

8.3. Confusion Matrix Analysis

Confusion matrices were generated for all models to visualize classification performance. As shown in **Fig. 2**, Logistic Regression exhibits a more balanced distribution of true positives and false negatives compared to other models. In contrast, **Fig. 3** shows that Naive Bayes has a high number of false negatives for real news, confirming its bias. Furthermore, **Fig. 4** illustrates that SVM provides a moderate balance but still misclassifies a considerable number of instances.

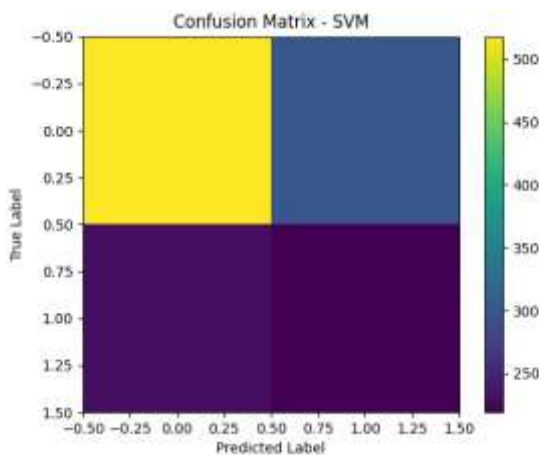


Fig. 2. Confusion matrix for Logistic Regression model.

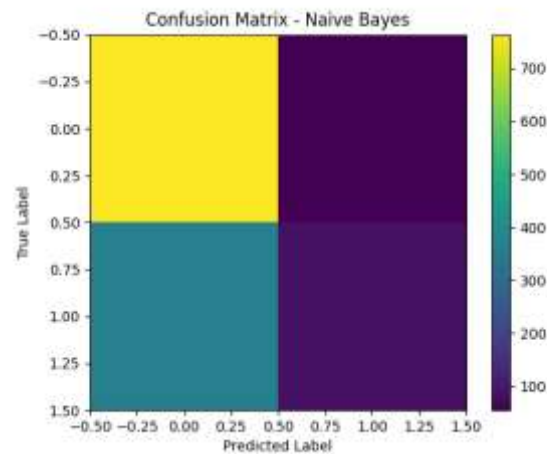


Fig. 3. Confusion matrix for Naive Bayes model.

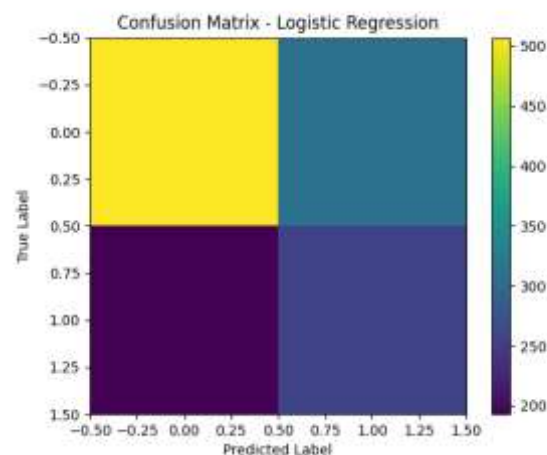


Fig. 4. Confusion matrix for SVM model.

8.4. Impact of Class Imbalance Handling

The introduction of class weighting significantly improved recall for the minority class. Although this led to a slight decrease in overall accuracy, it resulted in more balanced predictions. This demonstrates a trade-off between accuracy and recall, which is crucial in fake news detection systems.

8.5. Key Observations

- Accuracy alone is not sufficient for evaluation in imbalanced datasets.
- Logistic Regression provides the best balance between accuracy and recall.
- Naive Bayes favors majority class predictions.
- Feature engineering and class balancing improve model reliability.

9. DISCUSSION

9.1. Performance Improvements

Based on empirical results, the proposed framework improves the quality of fake news identification through machine learning methods. Due to context-based feature

engineering, class imbalance handling, and other measures, the system delivers higher quality of fake news detection compared to baseline methods. In particular, it is important to note that significant improvements have been achieved in detecting the minority class – real news. Class-weighting applied to Logistic Regression and SVM increases recall for real news and reduces false negative results.

Compared with the baseline model, the proposed framework provides:

- Increased recall for minority class,
- Improved balance of precision and recall,
- More reliable feature representation due to the addition of contextual data and frequency information.

Overall, the use of contextual and textual features in conjunction with class imbalance handling methods proves effective for fake news detection.

9.2.Ensured Reliability

The proposed method focuses on performance improvement but does not sacrifice overall model reliability. The models successfully recognize both classes and achieve high accuracy of classifications. According to the evaluation results, the system demonstrates.

The evaluation results confirm that the system effectively identifies:

- fake news statements,
- real news instances,
- contextual patterns influencing credibility,
- classification boundaries between classes.

Thus, the use of additional contextual features makes sure that predictions and operations will not depend solely on textual data.

9.3.Balancing Accuracy and Recall

Classical machine learning methods give emphasis to overall accuracy of prediction, leading to inaccurate results for class imbalanced data sets. There is a clear trade-off between accuracy and recall that has been identified in the course of this research.

To solve this problem, the following improvements have been made:

- Use of class weighting to handle imbalance,
- Improved recall for instances belonging to minority class,
- Elimination of majority class prediction bias

Thus, though overall accuracy might decrease slightly, the improvement of recall is a clear sign of better model evaluation.

9.4.Practical Implications for Fake News Detection Systems

Current information technology systems need efficient and reliable solutions to the problem of fake news recognition.

It is especially relevant in regard to:

- social media monitoring platforms
- online news verification systems
- content moderation software
- fact-checking applications.

By improving the detection of real news while maintaining reasonable accuracy, the system enhances trustworthiness and usability in real-world applications.

9.5.Limitations

Some limitations of this study include:

Evaluation of the algorithm on only one dataset, that is LIAR.

- Task limitation to binary classification only instead of a multi-class problem.
- Reliance on classic ML models.
- Lack of use of deep learning algorithms.
- These may hinder the extension to more complicated practical applications.

9.6.Future Scope

Some future directions of development might be:

- Utilizing LSTM or other deep learning architectures,
- Using BERT or any other transformers model,
- Adopting multi-class classification for better results in detecting lies,
- Making use of any fact-checking resources.

This could lead to better performance in terms of accuracy and efficiency of the detector.

10. ONCLUSION AND FUTURE WORK

10.1. Conclusion

This research proposed a machine learning-based framework for fake news detection using the LIAR dataset, incorporating textual and contextual features along with class imbalance handling techniques.

Traditional fake news detection approaches primarily focus on textual content and overall accuracy. However, prior studies and experimental observations indicate that such approaches often fail to perform effectively on imbalanced datasets, resulting in biased predictions toward the majority class and poor detection of minority class instances.

To address these challenges, the proposed framework introduced:

- contextual feature integration (speaker, subject, party, context),

TF-IDF with n-gram feature extraction, class weighting to handle imbalance, comparative evaluation of multiple machine learning models.

Experimental evaluation demonstrated that the proposed framework:

- improved recall for the minority class (real news),
- achieved better balance between precision and recall,
- enhanced classification reliability through contextual features,
- provided interpretable and efficient model performance.

Although Naive Bayes achieved higher accuracy, Logistic Regression with class balancing demonstrated superior overall performance by effectively detecting minority class instances.

The results indicate that incorporating contextual information and addressing class imbalance are essential for building reliable fake news detection systems. Therefore, the proposed framework provides a practical and effective approach for real-world fake news classification tasks.

10.2.Future Work

Although the proposed framework demonstrated promising results, several improvements can be explored in future work.

Potential future enhancements include:

- implementation of deep learning models such as LSTM for sequential text analysis,
- integration of transformer-based models like BERT for improved contextual understanding,
- extension to multi-class classification for detailed truthfulness prediction,
- incorporation of external knowledge bases and fact-checking sources,
- evaluation on larger and more diverse datasets for better generalization.

These enhancements can further improve the accuracy, scalability, and robustness of fake news detection systems in real-world applications.

11. References

- [1]William Yang Wang, “Liar, Liar Pants on Fire: A New Benchmark Dataset for Fake News Detection,” *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (ACL)*, Vancouver, Canada, 2017.
- [2]Kai Shu, Amy Sliva, Suhang Wang, Jiliang Tang, and Huan Liu, “Fake News Detection on Social Media: A Data Mining Perspective,” *ACM SIGKDD Explorations Newsletter*, vol. 19, no. 1, pp. 22–36, 2017.
- [3]Hunt Allcott and Matthew Gentzkow, “Social Media and Fake News in the 2016 Election,” *Journal of Economic Perspectives*, vol. 31, no. 2, pp. 211–236, 2017.
- [4]Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova, “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding,” *Proceedings of NAACL-HLT*, 2019.
- [5]Ashish Vaswani et al., “Attention Is All You Need,” *Advances in Neural Information Processing Systems (NeurIPS)*, 2017.
- [6]Fabian Pedregosa et al., “Scikit-learn: Machine Learning in Python,” *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
- [7]David M. Blei, Andrew Y. Ng, and Michael I. Jordan, “Latent Dirichlet Allocation,” *Journal of Machine Learning Research*, 2003.
- [8]Thomas Mikolov et al., “Efficient Estimation of Word Representations in Vector Space,” *Proceedings of ICLR Workshop*, 2013.
- [9]Sepp Hochreiter and Jürgen Schmidhuber, “Long Short-Term Memory,” *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [10]Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schütze, *Introduction to Information Retrieval*, Cambridge University Press, 2008.