**Comparative Analysis of Machine Learning Algorithms for Fake News Detection**

**Meghraj Desai (22DCS013)**

**Akshat Patel (22DCS054)**

**Dev Patel (22DCS055)**

Department of Computer Science

Devang Patel Institute Of Advance Technology And Research, Charusat University

Email: 22dcs013@charusat.edu.in

**Abstract**

The rapid spread of misinformation poses a significant challenge in today’s digital world. This study evaluates the performance of various machine learning algorithms, including **Naïve Bayes, Logistic Regression, Random Forest, Support Vector Machine (SVM), XGBoost, LSTM, and BERT-based models**, for fake news detection. We employ **TF-IDF vectorization** for feature extraction and compare the models based on accuracy, precision, recall, and F1-score. Our findings suggest that deep learning models, particularly transformer-based architectures like BERT, outperform traditional classifiers, although simpler models like SVM and Random Forest offer competitive performance with lower computational costs.

**1. Introduction**

The increasing prevalence of fake news on digital platforms has raised concerns regarding its impact on public perception and decision-making. Traditional fact-checking methods are inefficient in handling large-scale misinformation, necessitating automated detection techniques using machine learning and deep learning. This paper explores multiple classification algorithms to determine their effectiveness in identifying fake news.

**2. Related Work**

Several studies have focused on the application of machine learning for fake news detection. **Naïve Bayes and Logistic Regression** have been widely used for their simplicity and interpretability, while **Random Forest and XGBoost** leverage ensemble learning to improve classification accuracy. **SVM** has demonstrated strong performance in text classification tasks, whereas **LSTM and BERT-based models** have shown superior capabilities in capturing contextual information from textual data. This study extends previous research by conducting a comparative analysis of these techniques using standardized datasets.

**3. Methodology**

**3.1 Dataset**

For our experiments, we use the **Fake and Real News Dataset** from Kaggle, which consists of news articles labeled as real or fake. The dataset is preprocessed to remove stopwords, punctuation, and special characters before applying feature extraction.

**3.2 Feature Extraction**

We employ **TF-IDF (Term Frequency-Inverse Document Frequency)** to convert textual data into numerical vectors, which are then used as input for classification models.

**3.3 Machine Learning Models**

The following models are evaluated:

• **Naïve Bayes (NB)**: A probabilistic model based on Bayes’ theorem.

• **Logistic Regression (LR)**: A linear model effective for binary classification tasks.

• **Random Forest (RF)**: An ensemble model that enhances performance using multiple decision trees.

• **Support Vector Machine (SVM)**: A classifier that finds the optimal hyperplane for text classification.

• **XGBoost (XGB)**: A boosting algorithm that optimizes decision tree performance.

• **LSTM (Long Short-Term Memory)**: A deep learning model suitable for sequential text data.

• **BERT (Bidirectional Encoder Representations from Transformers)**: A pre-trained transformer-based model designed for natural language understanding.

**3.4 Evaluation Metrics**

The models are evaluated using the following performance metrics:

• **Accuracy**: The percentage of correctly classified instances.

• **Precision**: The ratio of correctly predicted fake news instances to total predicted fake news instances.

• **Recall**: The ratio of correctly predicted fake news instances to total actual fake news instances.

• **F1-Score**: The harmonic mean of precision and recall.

**4. Performance Comparison**

The models are trained and tested using a **70-30% train-test split**. The results are summarized in **Table 1**.

**Model** **Accuracy** **Precision** **Recall** **F1-Score**

Naïve Bayes 85.2% 83.5% 86.0% 84.7%

Logistic Regression 88.1% 86.7% 88.4% 87.5%

Random Forest 91.0% 89.8% 91.3% 90.5%

SVM 92.4% 91.1% 92.8% 91.9%

XGBoost 93.1% 92.0% 93.5% 92.7%

LSTM 94.6% 93.9% 94.8% 94.3%

BERT **96.2%**  **95.8%** **96.5%** **96.1%**

**5. Results and Discussion**

Our results indicate that deep learning models outperform traditional machine learning models in detecting fake news. **BERT achieved the highest accuracy (96.2%)**, demonstrating its ability to capture nuanced language patterns. **LSTM also performed well (94.6%)**, leveraging sequential text analysis.

Among classical models, **XGBoost (93.1%) and SVM (92.4%)** were the most effective, showing that tree-based and margin-based classifiers can still yield competitive results. **Naïve Bayes performed the worst (85.2%)**, likely due to its strong independence assumptions.

While deep learning models offer superior accuracy, they are computationally expensive. **SVM and XGBoost provide a good trade-off between performance and efficiency**, making them suitable for real-time fake news detection in resource-constrained environments.

**6. Conclusion and Future Work**

This study provides a comparative analysis of multiple machine learning models for fake news detection. Our findings highlight that **BERT-based models achieve the best performance**, followed by **LSTM and XGBoost**. However, simpler models like **SVM and Random Forest** remain viable alternatives due to their lower computational costs.

For future work, we suggest exploring **hybrid models that combine deep learning with traditional approaches** and expanding the dataset to include more diverse sources of misinformation.

**References**

1. Albahr, A., & Albahar, M. (2020). *An Empirical Comparison of Fake News Detection using Different Machine Learning Algorithms*. International Journal of Advanced Computer Science and Applications, 11(9).

2. Alghamdi, J., Lin, Y., & Luo, S. (2022). *A Comparative Study of Machine Learning and Deep Learning Techniques for Fake News Detection*. Information, 13(12), 576.

3. Singh, L. (2021). *Fake News Detection: A Comparison Between Available Deep Learning Techniques in Vector Space*. arXiv preprint arXiv:2102.09470.

4. Khan, J. Y., Khondaker, M. T. I., Afroz, S., Uddin, G., & Iqbal, A. (2019). *A Benchmark Study of Machine Learning Models for Online Fake News Detection*. arXiv preprint arXiv:1905.