Artificial Intelligence
University of the Philippines

FILIPINO FACT CHECK: AN ANALYSIS OF MACHINE LEARNING CLASSIFIERS FOR FILIPINO FAKE NEWS CLASSIFICATION

Jemar Laag | Alister Marc Domilies

AI 201 Mini Project

Introduction

Fake news, originating from internet, social media, and television sources, poses a pervasive challenge in the Philippines. A Social Weather Stations survey revealed that 51% of Filipinos struggle to identify misinformation. This issue is exacerbated by the vast volume of content, making it challenging for fact-checking organizations like Verafiles and NUJP to manually verify everything.

Many studies focus on developing fake news detection models however most are tailored exclusively for the English language. Moreover, a prior successful study by Cruz et al. achieved high accuracy in fake news detection in the Filipino language using deep learning techniques, but these methods present complexities that hinder their easy deployment for consumer-level applications.

Objective

This study aims to create a simplified model for predicting fake news in Filipino, targeting an accuracy level comparable to Cruz et al.'s findings, with the following objectives:

- Utilizing available algorithms in scikit-learn, including individual/base learners, ensemble learners, and neural network classifiers.
- Conducting hyperparameter tuning to optimize predictive performance.
- Selection of the top-performing model based on achieving the highest overall accuracy.
- Deployment of the chosen model for consumer-level applications."

Methodology

1 Dataset Acquisition and Preprocessing

The dataset utilized in this study originated from the dataset utilized in Cruz et al's study. Both training and test data set underwent the preprocessing phase which involves the removal of stop words, converting text to lowercase and excluding of punctuations marks.

2 Feature Selection and Extraction

Both training and test data set underwent feature extraction with the Term Frequency-Inverse Document Frequency (TF-IDF) technique. TF-IDF directly processes text data to generate numerical representations from the documents Feature selection involved exploring different values for tfidf_max_df and various settings for ngram_range

3 Model Training and Hyperparameter Tuning

The extracted features underwent model training using various classifiers sourced from the scikit-learn library. Additionally, different hyperparameter values were selected and tested for each specific classifier during the hyperparameter tuning process. A gridsearchCV methodology was employed to identify the optimal combination of features and hyperparameters for each classifier under assessment.

4 Model Evaluation, Selection and Deployment

Using the test dataset, the performance of these tuned models were compared to determine the highest accuracy achieved. The model with the highest accuracy was chosen for model deployment.

Methodology Flowchart

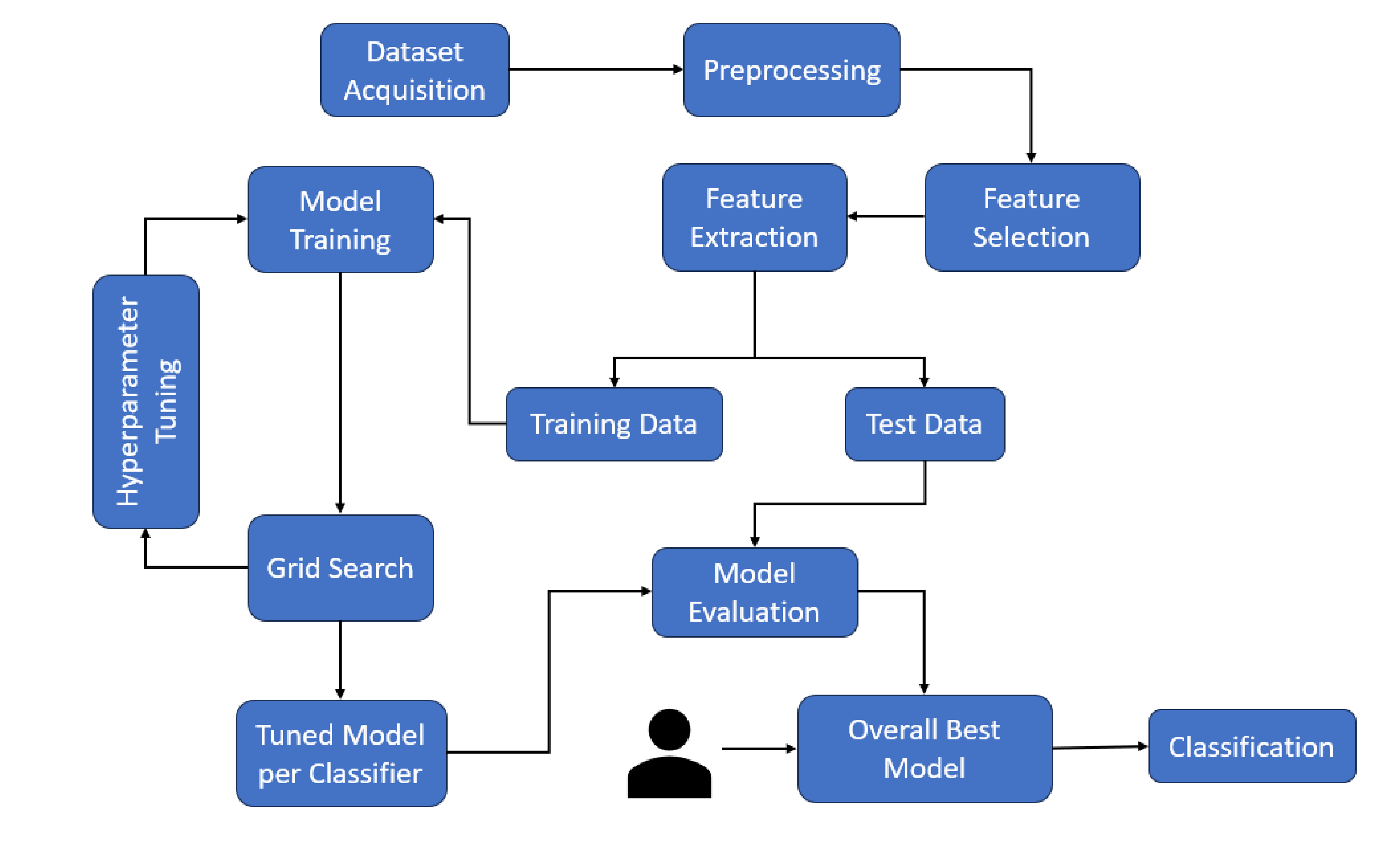
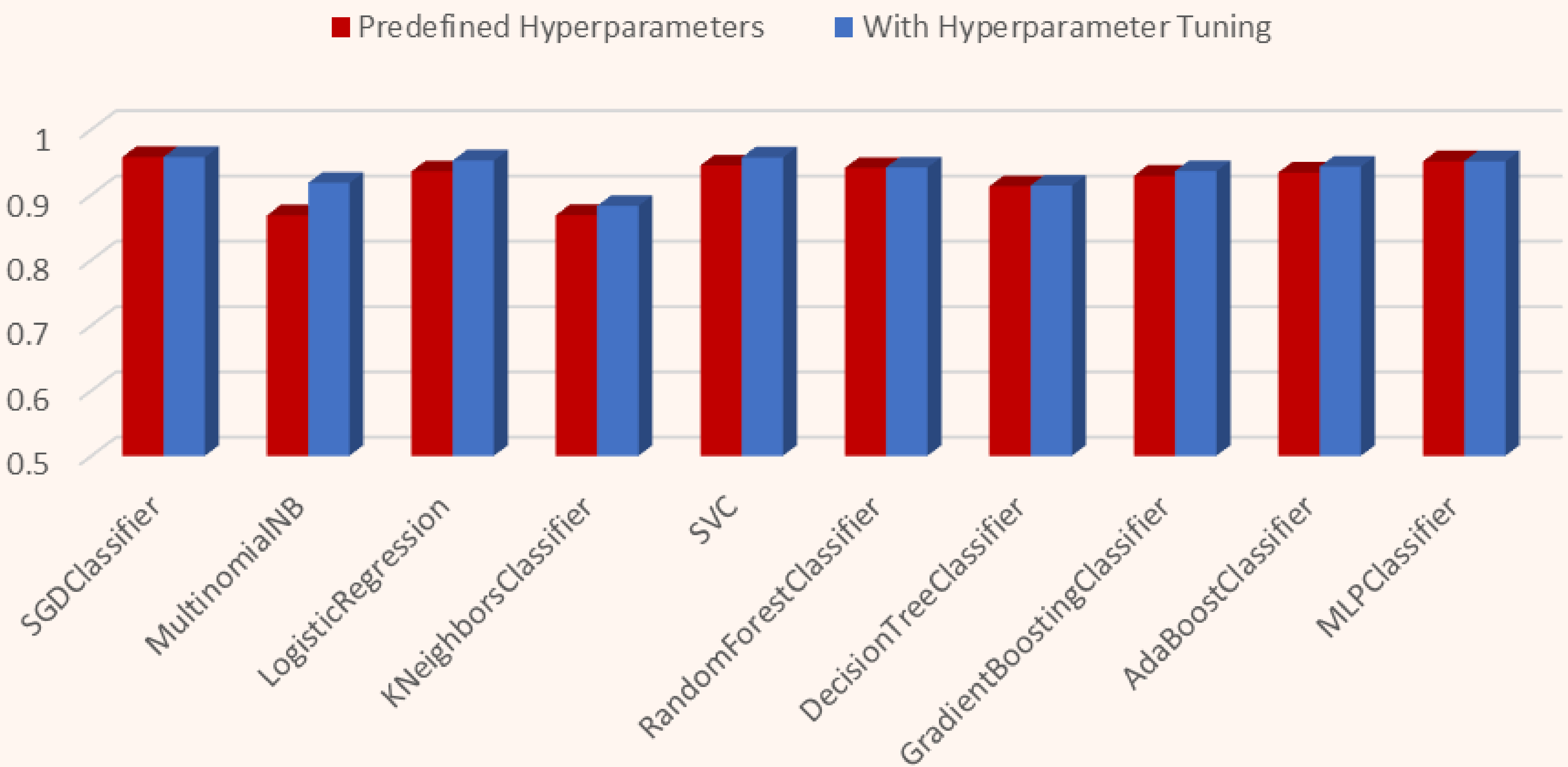


Figure 1: Flowchart for Determination of Best Model Per Classifier

Results and Discussion



Highest Accuracy with Predefined Hyperparameters

SGD Classifier - 95.8%

Highest Accuracy with Hyperparameter Tuning

SGD Classifier - 95.8%

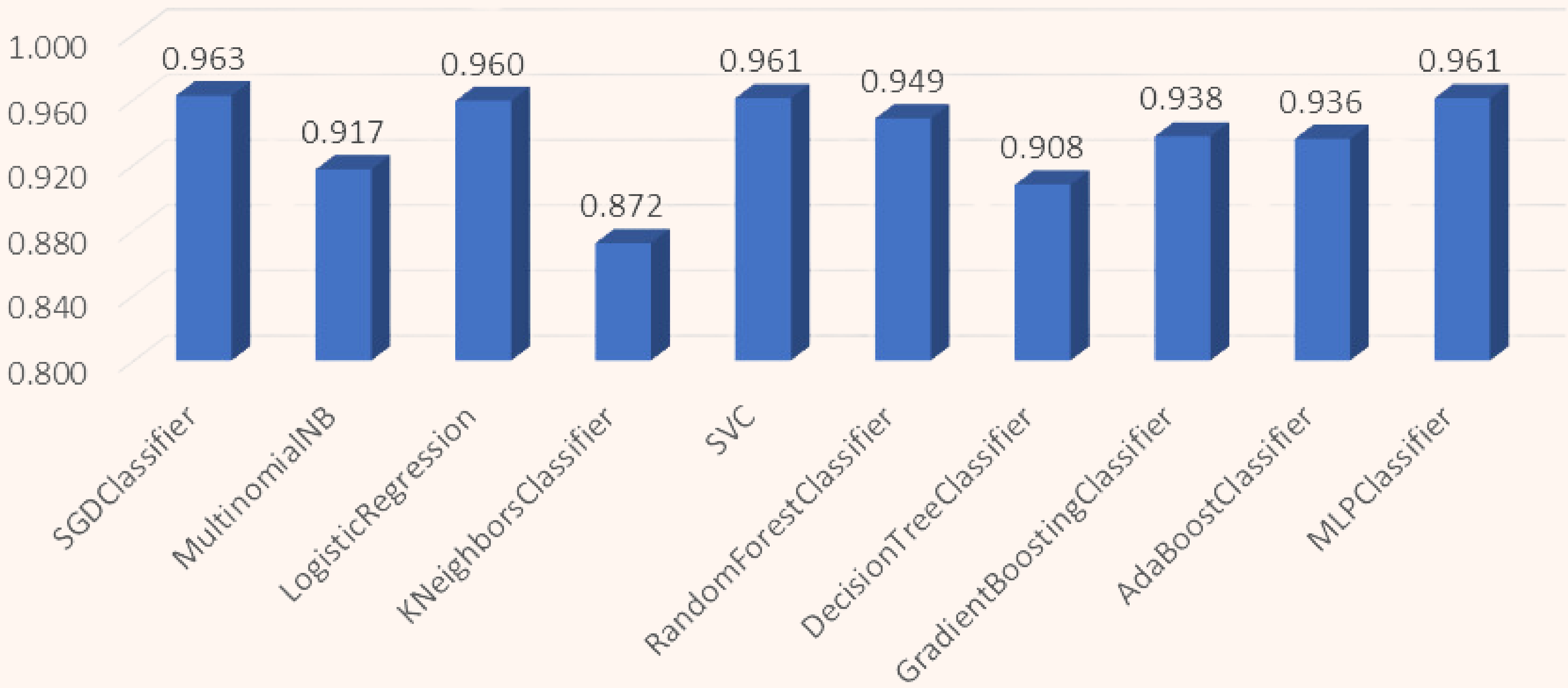
Highest Increase in Accuracy after Hyperparameter Tuning

MNB Classifier - 5.0 %

Figure 3: Training Accuracies: Predefined Hyperparameter vs. Hyperparameter Tuning

Initially, the models trained with predefined settings already demonstrated high accuracy. By implementing hyperparameter tuning, several classifiers exhibited increased training accuracies primarily affecting the individual learners among the classifiers which highlights their sensitivity to hyperparameter changes. In contrast, ensemble methods and neural networks, showcase more robust behavior or lesser sensitivity to certain hyperparameters, leading to relatively less impact on their performance through tuning.

Results and Discussion



Tuned Models with Highest Test Accuracuies

1. **SGD Classifier - 96.3%**
2. **MLP Classifier - 96.1%**
3. **SVC - 96.1%**
4. **Logistic Regression - 96.0%**

Figure 3: Test Accuracies of Tuned Models

Notably, four classifiers namely Stochastic Gradient Descent, Logistic Regression, Support Vector Machine (SVC), and the Multilayer Perceptron (MLP) Classifier achieved accuracies exceeding 95% while most of the remaining classifiers has achieved at least 90% accuracies suggesting great performance of these various models to the used dataset.

Result and Discussion

Individual learners typically demonstrate shorter training times, ranging from 1.2 seconds to 11.2 seconds. Meanwhile, ensemble learners require comparatively longer training times, spanning from 8.8 seconds to 46.8 seconds while the Multilayer Perceptron Classifier, a neural network classifier, exhibited the longest training time of 292.7 seconds.

This difference in training times shows the varying complexities among classifier types. Individual learners, being simpler models, generally demand less computational time for training. In contrast, ensemble methods and neural network classifiers inherently involve more complex operations, requiring increased computational resources and resulting in longer training durations.

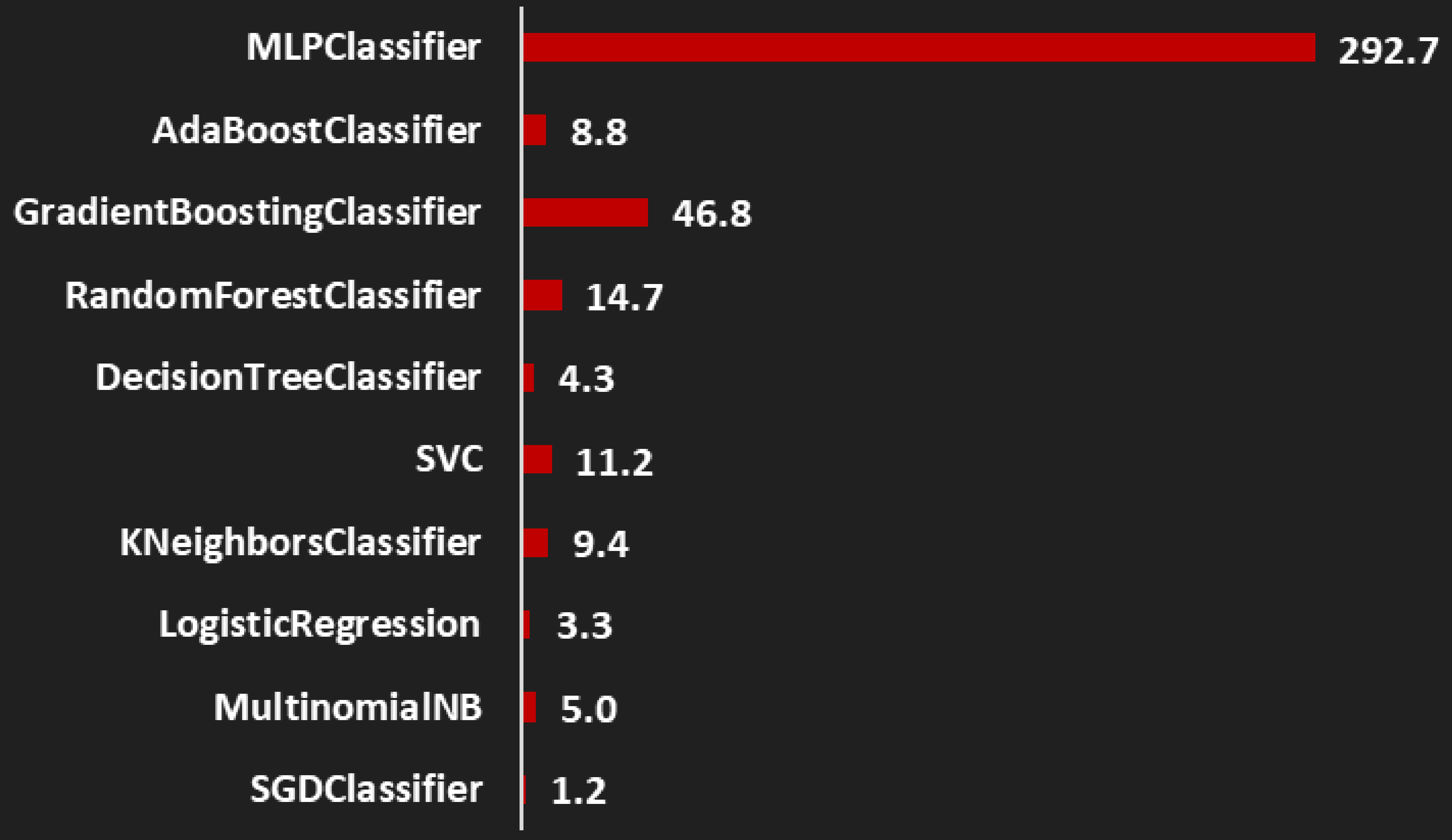


Figure 4: Training Times of the Various Classifiers

Results and Discussion

Highest Test Accuracy

96.26 %

The stochastic gradient descent classifier emerges as the top-performing model, boasting the highest test accuracy. Additionally, consistent high values across various performance metrics (ranging from 94% to 98%) signify the robustness and generalizability of the model.

Shortest Training Time

1.21 sec

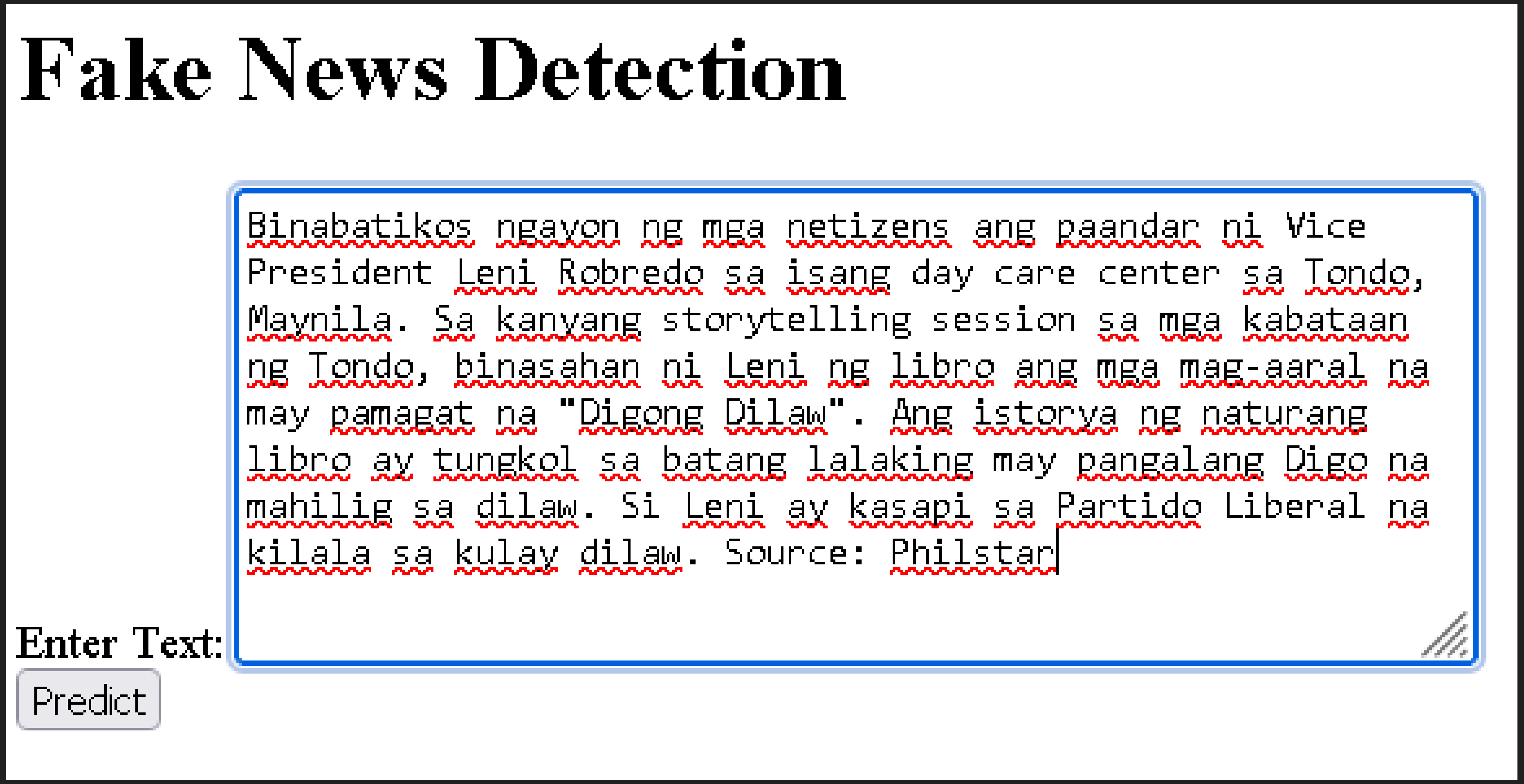
Despite its simplicity compared to other classifiers, it not only excelled in accuracy but also demonstrated the shortest training time for 1 combination of hyperparameters.

Deployment

Easy

Finally, the superior model was easily deployed via a Flask application which allows users to input text and predict whether it is classified as "fake" or not with a simple click. The application provides a streamlined user experience, enabling effortless interaction and immediate predictions.

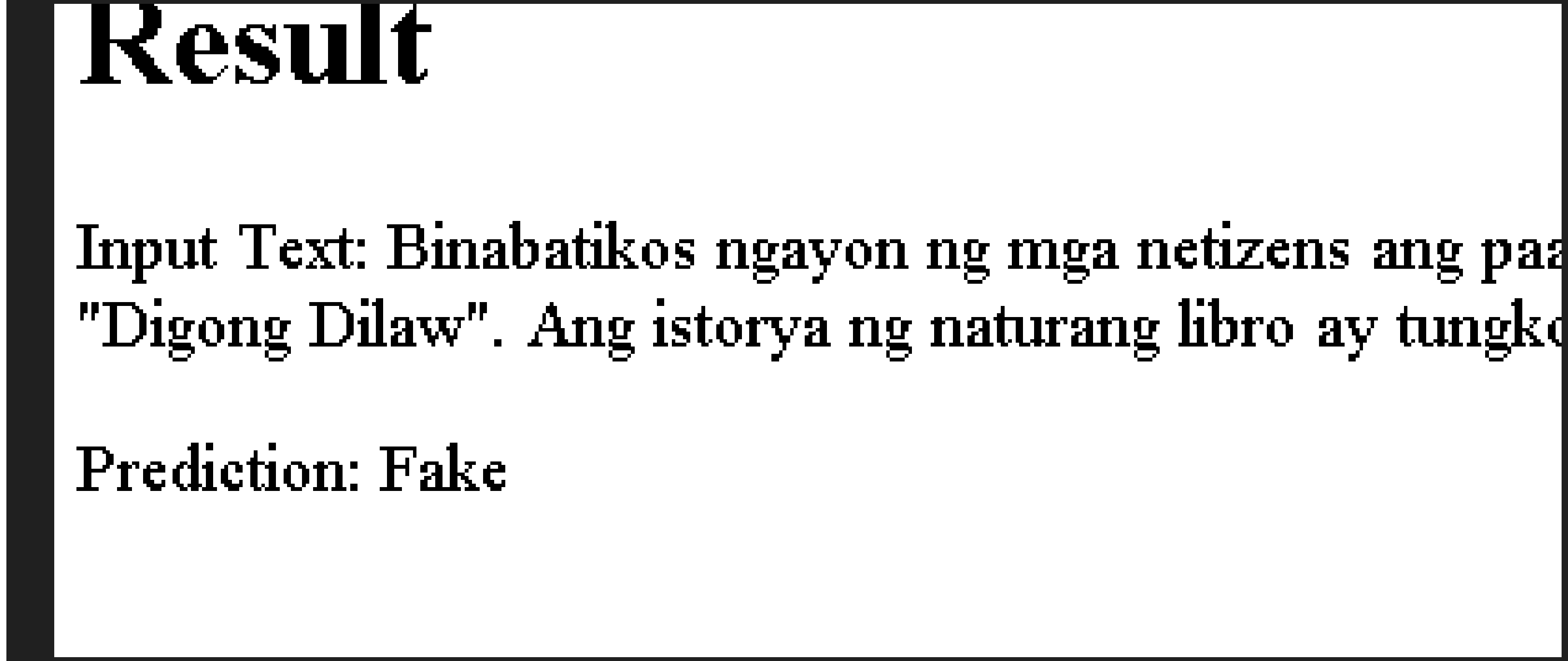
Model Deployment Sample



Fake News Detection

Enter Text: Binabatikos ngayon ng mga netizens ang paandar ni Vice President Leni Robredo sa isang day care center sa Tondo, Maynila. Sa kanyang storytelling session sa mga kabataan ng Tondo, binasahan ni Leni ng libro ang mga mag-aaral na may pamagat na "Digong Dilaw". Ang istorya ng naturang libro ay tungkol sa batang lalaking may pangalang Digo na mahilig sa dilaw. Si Leni ay kasapi sa Partido Liberal na kilala sa kulay dilaw. Source: Philstar

Figure 5: Input



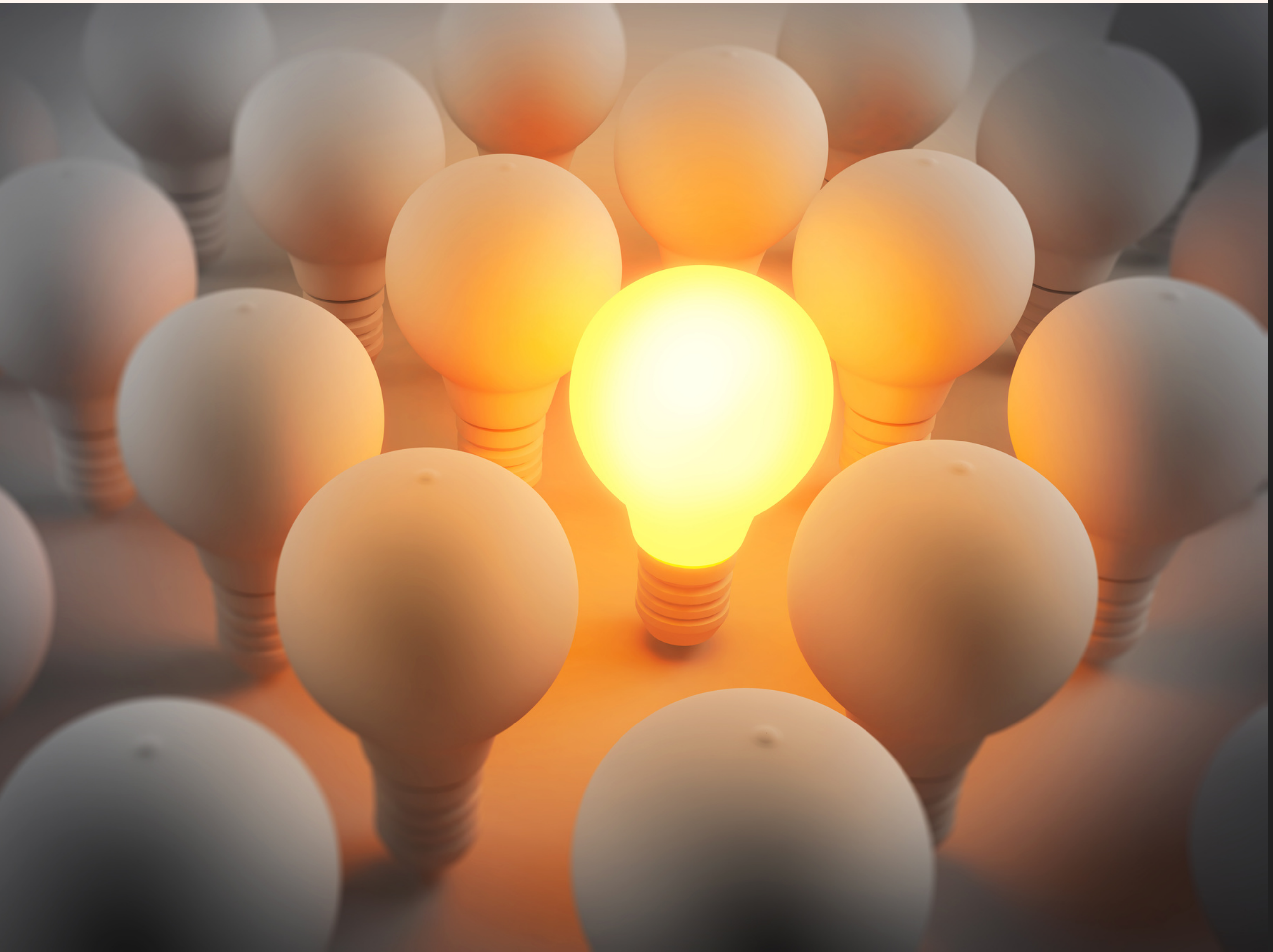
Result

Input Text: Binabatikos ngayon ng mga netizens ang paa
"Digong Dilaw". Ang istorya ng naturang libro ay tungko

Prediction: Fake

Figure 5: Result

Conclusions



Despite being an individual classifier, the stochastic gradient descent classifier emerged as the optimal model, displaying both the highest accuracy and remarkably shortest training times among all tested classifiers.

This study achieved a significant milestone by easily deploying the optimal model which marks a considerable advancement from the deployment challenges faced in Cruz et al.'s study.

Despite the simplified model architecture, this study achieved a remarkably high accuracy level that closely aligns with the findings of prior study conducted by Cruz et al. This initial implementation serves as a promising starting point for developing a consumer-level application.

Recommendations

1. Implement continuous updates on the news articles in the dataset to ensure the model remains adaptive to evolving patterns and trends in information

2. Explore additional datasets, particularly those containing localized posts on social media platforms such as Facebook and YouTube.

3. Inclusion of other classifiers not used in the study and broadening the scope of hyperparameter tuning.

THANK YOU SO MUCH!