# Machine Learning-Based Solution for Classification and Clustering

## Abstract

This report presents the implementation of a machine learning-based solution to perform classification and clustering tasks on business datasets. Using scikit-learn, various algorithms were evaluated based on their performance metrics, and hyperparameter tuning was applied to optimize selected models. A user-friendly interface was also developed to simplify interactions with the system.

## Introduction

Machine learning offers powerful tools to analyze and extract insights from data. In this project, we focus on two key tasks: classification and clustering, using publicly available datasets. The goal is to evaluate the performance of multiple algorithms and provide a robust solution with a simple user interface.

## Methodology

The steps undertaken in this project are as follows:

1. Selection of two datasets from public repositories.

2. Preprocessing and scaling of data for classification tasks.

3. Implementation of 10 classification algorithms.

4. Implementation of 3 clustering algorithms.

5. Hyperparameter tuning to optimize selected models.

6. Development of a simple user interface using Streamlit.

## Results

Classification results for 10 algorithms are shown below:

Algorithm Accuracy Precision Recall F1 Score Time (s)  
Logistic Regression 1.000000 1.000000 1.000000 1.000000 0.012088  
 Random Forest 1.000000 1.000000 1.000000 1.000000 0.188319  
 SVM 1.000000 1.000000 1.000000 1.000000 0.003657  
 KNN 1.000000 1.000000 1.000000 1.000000 0.002923  
 Naive Bayes 0.977778 0.979365 0.977778 0.977745 0.001359  
 Decision Tree 1.000000 1.000000 1.000000 1.000000 0.000539  
 Gradient Boosting 1.000000 1.000000 1.000000 1.000000 0.253618  
 AdaBoost 1.000000 1.000000 1.000000 1.000000 0.105038  
 MLP 1.000000 1.000000 1.000000 1.000000 0.161577

Clustering results for 3 algorithms are shown below:

Algorithm Clusters Found Time (s)  
 KMeans 4 0.806751  
Agglomerative 4 0.017740  
 DBSCAN 5 0.002272

Tuned classification results for Random Forest and SVM are shown below:

Algorithm Accuracy Precision Recall F1 Score Time (s)  
Random Forest (Tuned) 1.000000 1.000000 1.000000 1.000000 0.143278  
 SVM (Tuned) 0.977778 0.979365 0.977778 0.977745 0.001457

## Conclusion

The project demonstrates the effectiveness of machine learning algorithms in solving classification and clustering problems. By comparing multiple techniques, we identified the best-performing models for our datasets. Hyperparameter tuning further enhanced their performance. The developed UI ensures ease of use and accessibility for end users.