

Machine Learning Assignment 1 Summary Report

Project Overview:

This project aimed to practice Python for machine learning by replicating a study using a new simulated dataset generated with `make_blobs` from the `sklearn.datasets` library.

Dataset Description:

- **Generated Dataset:** 1000 samples divided into 4 clusters.
 - **Visualization:** A scatter plot displayed the 4 distinct clusters in the dataset, confirming that the `make_blobs` function successfully created well-separated classes.
-

Model Training:

- **Algorithm Used:** Logistic Regression (with `max_iter=1000` for convergence).
 - **Data Split:** 80% for training and 20% for testing using `train_test_split`.
-

Results:

- **Accuracy Score:** Approximately 96%.
 - **Classification Report:**
 - **Precision, Recall, and F1-Score** for each class were high, indicating the model's strong performance.
 - No significant misclassifications were observed, and the confusion matrix showed minimal errors.
-

Conclusion:

The project successfully demonstrated the use of Python for machine learning by generating a simulated dataset, training a model, and evaluating its performance. The Logistic Regression model achieved high accuracy, making it a suitable choice for this classification task. Future improvements could include experimenting with different algorithms, tuning hyperparameters, and exploring additional visualizations.

Prepared by: SIMRANJEET KAUR

Course: MSBA at Kent State University