

Final Report - George Sorondo

Executive Summary

Explored machine learning across three datasets (Iris, Breast Cancer, Wine Quality) using scikit-learn. Performed preprocessing, model training, hyperparameter tuning, and evaluation.

Datasets

- Iris: Classify flower species.
- Breast Cancer: Diagnose tumors.
- Wine Quality: Predict wine scores.

Models

- Logistic Regression
- k-Nearest Neighbors
- Multi-Layer Perceptron (MLP)
- Linear Regression (for Wine)

Key Results

- Iris: 96%+ accuracy.
- Breast Cancer: 97-98% accuracy.
- Wine: RMSE ~0.68 (matches Cortez 2009 RMSE ~0.64).

Reflection

MLP outperformed other models. Classification tasks easier than regression. Project demonstrated

Final Report - George Sorondo

full ML pipeline application.