Assignments List

1. Compare Logistic Regression vs SVM on the Diabetes Dataset

- **Task:** Train Logistic Regression and SVM classifiers on the PIMA Diabetes dataset and compare accuracy, precision, recall, and F1-score.
- **Dataset:** PIMA Indians Diabetes Dataset (UCI)

2. Compare Decision Tree vs Logistic Regression

- **Task:** Train Decision Tree and Logistic Regression models on a classification dataset and compare their performance.
- **Dataset:** Titanic Survival Dataset (Kaggle)

3. Compare Naive Bayes vs Logistic Regression for Classification

- **Task:** Build text classification models using Naive Bayes and Logistic Regression, compare results.
- **Dataset:** SMS Spam Collection Dataset

4. Compare k-means and Classification Algorithms

- **Task:** Apply k-means clustering on a labeled dataset, then compare results with supervised classifiers (Logistic Regression, Random Forest).
- **Dataset:** Iris Dataset (UCI)

5. Study Effect of Feature Scaling (Standardization vs Normalization)

- **Task:** Apply scaling methods and train models (Logistic Regression, KNN) to see effect on accuracy.
- **Dataset:** Wine Quality Dataset (UCI)

6. Study Effect of Removing Irrelevant Features on Accuracy

- **Task:** Train models (ANY 2 MODELS) with all features, then with selected relevant features, and compare performance.

- **Dataset:** Student Performance Dataset

7. Study Effect of Feature Selection (Chi-square, Mutual Information)

- **Task:** Apply Chi-square and Mutual Information for feature selection, then evaluate classification accuracy (any 2 models).

- **Dataset:** Breast Cancer Wisconsin Dataset (UCI)

8. Study Effect of Regularization (L1 vs L2) in Logistic Regression

- **Task:** Train Logistic Regression with L1 and L2 penalties, compare performance and sparsity of coefficients.

- **Dataset:** Heart Disease Dataset (UCI)

9. Study Effect of Train-Test Split Ratios (70–30 vs 80–20 vs 60–40)

- **Task:** Train classifiers (ANY 2 MODELS) with different train-test split ratios, compare performance stability.

- **Dataset:** Ionosphere Dataset (UCI)

10. Study Effect of Cross-Validation vs Hold-out Validation

- **Task:** Compare model (ANY 2 MODELS) evaluation results using hold-out validation and k-fold cross-validation.

- **Dataset:** Banknote Authentication Dataset (UCI)

11. Study Effect of Overfitting vs Underfitting in Decision Trees

- **Task:** Train Decision Trees with different max_depth and min_samples_split values, show overfitting/underfitting.

- **Dataset:** Mushroom Classification Dataset (UCI)

12. Study Effect of k Value in KNN on Accuracy

- **Task:** Vary the value of k in KNN and analyze effect on classification accuracy.
- **Dataset:** Seeds Dataset (UCI)

13. Study Effect of min_samples_split in Decision Trees

- **Task:** Vary min_samples_split in Decision Trees and compare accuracy vs generalization.
- **Dataset:** Glass Identification Dataset (UCI)

14. Study Effect of max_depth in Decision Trees

- **Task:** Train Decision Trees with increasing max_depth and study accuracy and overfitting.
- **Dataset:** Breast Cancer Wisconsin Dataset (UCI)

15. Study Effect of Number of Trees in Random Forest

- **Task:** Vary number of estimators in Random Forest and compare performance and runtime.
- **Dataset:** Sonar Dataset (UCI)

16. Study Effect of Learning Rate in Gradient Boosting

- **Task:** Train Gradient Boosting models with different learning rates and analyze accuracy vs convergence.
- **Dataset:** Adult Income Dataset (UCI)

17. Study Effect of Training Size on Classification Accuracy (Learning Curves)

- **Task:** Train models (ANY 2 MODELS) with progressively larger subsets of training data and plot learning curves.

- **Dataset:** Iris Dataset (UCI)

18. Study Performance with Imbalanced Datasets (Before vs After SMOTE)

- **Task:** Train classifiers (Any 2 models) on imbalanced data, apply SMOTE to rebalance, compare results.

- **Dataset:** Credit Card Fraud Detection Dataset (Kaggle)

19. Compare GridSearchCV vs RandomSearchCV for Hyperparameter Tuning

- **Task:** Train models (ANY 2 MODELS) using GridSearchCV and RandomSearchCV, compare efficiency and accuracy.

- **Dataset:** Wine Quality Dataset (UCI)

20. Compare Logistic Regression vs Decision Tree on Imbalanced Dataset

- **Task:** Apply both Logistic Regression and Decision Tree classifiers, analyze which handles imbalance better.

- **Dataset:** Breast Cancer Dataset (UCI)