

Machine Learning Term Paper

Title : Comparative Study of Parametric and Non-Parametric Classifiers for Binary Classification on Numerical Datasets

Abstract : This study compares four machine learning classifiers—Logistic Regression (parametric, gradient-based), Linear SVM (parametric, maximum-margin), Weighted k-NN (non-parametric, distance-based), and Decision Tree (CART) (non-parametric, rule-based)—on five binary classification datasets: Wine Quality (Red), Titanic, Pima Indians Diabetes, Heart Disease, and Adult Income. Performance is evaluated using accuracy, precision, and F1-score. Non-parametric models excel on complex, non-linear datasets, while parametric models perform well on linearly separable data. This analysis highlights their strengths for numerical binary classification tasks..

Group Members :

- **Roll No.** S20230010169 : **Saketh Pabbu (Leader)**
- **Roll No.** S20230010157 : **M Sai Pavan**
- **Roll No.** S20230010217 : **Bhargav Reddy**
- **Roll No.** S20230010229 : **Naga Sai Praveen**

Classifiers Selected :

Classifier	Description	Type	Linearity
Logistic Regression	Gradient-based linear classifier	Parametric	Linear
Linear SVM	Maximum-margin linear classifier	Parametric	Linear
Weighted k-NN	Distance-based, instance-driven classifier	Non-Parametric	Non-linear
Decision Tree (CART)	Recursive, rule-based hierarchical splits	Non-Parametric	Non-linear

Datasets Selected :

- **Wine Quality:** Good/Bad, [Kaggle - Wine Quality](#)
- **Diabetes:** Diabetic/Non-Diabetic, [Kaggle – Diabetes Dataset](#)
- **Banknote Authentication:** Authentic/Fake, [Kaggle - Banknote Authentication](#)
- **Breast Cancer:** Malignant/Benign, Kaggle, [Breast Cancer Wisconsin](#)
- **Bank Customer Churn:** Churned/Not Churned, [Kaggle- Bank Customer Churn Prediction](#)

Evaluation/Comparison Metrics :

- **Accuracy:** Proportion of correctly classified instances
- **Precision:** Fraction of predicted positives that are correct
- **Recall (Sensitivity):** Fraction of actual positives correctly identified
- **F1-Score:** Harmonic mean of precision and recall
- **Average Accuracy / Cross-Validation:** Accuracy averaged across folds