Machine Learning Model Performance Study

Abstract:
This study evaluates the performance of various machine learning models on a customer churn prediction dataset. We trained and tested five different algorithms using cross-validation and measured their accuracy, precision, recall, and F1 scores. The Random Forest model achieved the highest overall performance with 94.2% accuracy.

Model Performance Comparison

Model	Accuracy	Precision	Recall	F1 Score	Training Time (min)
Logistic Regression	87.3%	85.1%	86.2%	85.6%	2.1
Decision Tree	89.5%	88.0%	89.1%	88.5%	3.5
Random Forest	94.2%	93.8%	94.5%	94.1%	12.3
XGBoost	93.8%	93.2%	94.0%	93.6%	15.7
Neural Network	91.6%	90.5%	91.8%	91.1%	28.4

Dataset Characteristics

Characteristic	Value
Total Samples	50,000
Training Set	40,000 (80%)
Test Set	10,000 (20%)
Number of Features	23
Class Balance	30% churn, 70% retained
Cross-validation Folds	5