Bank Marketing Campaign Prediction

using ID3, CART & Naive Bayes

Abstract

This project presents a classification system built from scratch to predict whether a customer will subscribe to a term deposit product. The implementation involves three algorithms: ID3, CART, and Naive Bayes. The dataset used is the Bank Marketing dataset from UCI, and models were evaluated using accuracy, precision, recall, and F1-score.

Introduction

Classification is a supervised learning technique used to predict the category of a data point. In this project, we address a real-world marketing use case—predicting the success of telemarketing campaigns by banks. This is crucial for reducing costs and increasing effectiveness.

Problem Statement

Given a set of features such as age, job, marital status, and previous marketing outcomes, we aim to predict if a customer will subscribe to a term deposit (target: 'yes' or 'no').

Dataset Description

We used the Bank Marketing dataset containing 4521 records and 17 features. The data was preprocessed by encoding categorical variables and splitting into training (4000), validation (400), and test (121) sets. The final predictions were made without access to the true labels.

Algorithms Used

- **ID3**: Builds a decision tree using information gain.
- **CART**: Splits data using Gini Index and supports binary splits.
- **Naive Bayes**: A probabilistic classifier based on Bayes' Theorem with strong independence assumptions.

Implementation

All algorithms were implemented from scratch without using libraries like scikit-learn. Data was encoded manually, trees were built recursively, and Naive Bayes was trained using frequency tables. Validation was performed on 400 samples to compute performance metrics.

Results

```
ID3 Accuracy: 85.0%
CART Accuracy: 88.5%
Naive Bayes Accuracy: 78.0%
```

```
ID3 Metrics:
Accuracy : 85.00%
Precision: 23.81%
Recall   : 10.20%
F1 Score : 14.29%
```

```
CART Metrics:
Accuracy : 88.50%
Precision: 61.54%
Recall   : 16.33%
F1 Score : 25.81%
```

```
Naive Bayes Metrics:
Accuracy : 78.00%
Precision: 27.59%
Recall   : 48.98%
F1 Score : 35.29%
```

Final Predictions

The best-performing model was used to predict the final 121 rows. Predictions were converted to original labels (yes/no) and printed sequentially.

Conclusion

The CART model achieved the best accuracy on validation data and was used to make predictions. This demonstrates the effectiveness of decision trees in bank marketing prediction problems.