Datasets Used

1. Airline Passenger Satisfaction Dataset

The airline passenger satisfaction dataset is taken from Kaggle. This is a classification problem, where the objective is to predict the passenger satisfaction level. The independent variables include "Gender", "Customer Type", "Age" and "Type of Travel", with "Satisfaction" being the binary variable to be predicted (neutral or dissatisfied, satisfied).

This dataset has an older version that contains all the data instead of splitting it into subsets for training and testing, and it has been used for clustering.(old version of airline passenger satisfaction))

2. Bank Churn Dataset

The bank customer churn dataset is taken from Kaggle. This is a classification problem with the objective to predict the customer churn of a bank. The independent variables include "credit_score", "country", "gender" and "age", with "churn" being the binary variable to be predicted (0-customer did not leave the bank during a certain period, 1-customer left the bank during the period).

Model development

1. Learning algorithm 1: Decision Tree

The Decision Tree algorithm was chosen for its interpretability, simplicity, and effectiveness in handling both numerical and categorical data. Decision trees work by recursively splitting the dataset into subsets based on feature values, optimising a chosen criterion such as Gini impurity or information gain (entropy). The algorithm's strengths lie in its ability to capture complex, non-linear relationships and its intuitive representation as a tree structure. However, decision trees are prone to overfitting, particularly when the tree grows too deep, and are sensitive to small changes in the data. These limitations were addressed in the experiments through hyperparameter tuning and cross-validation, as recommended in the MLDM lectures.

Hyperparameter optimization was performed using GridSearchCV with 5-fold cross-validation. The best parameters for each dataset are summarised in Table 3.1. These settings reflect the distinct characteristics of the datasets, with the Airline Passenger dataset benefiting from a larger tree due to its balanced classes, while the Customer Churn dataset required stronger regularisation to address imbalanced and noisier data.

Table 3.1: Hyperparameters Used in Decision Tree

Dataset	criterion	max_depth	min_samples_split	min_samples_leaf
Airline	Entropy	20	2	4
Churn	Gini	10	10	2

Model Evaluation:

Experiment: Optimizing Decision Tree Models for Churn and Satisfaction Prediction

1. NULL HYPOTHESIS 1

There is no significant difference in the predictive performance (AUC or classification metrics) between the base decision tree model and the fine-tuned decision tree model.

2. MATERIAL & METHODS

A train-test split of 75/25 was used for the Customer Churn dataset, while the Airline Passenger dataset used predefined train.csv and test.csv splits. Class imbalance was handled using SMOTE. A base model was trained, followed by hyperparameter tuning using GridSearchCV with 5-fold cross-validation. The performance was

evaluated using classification metrics (F1 etc.) and results were visualised through ROC curves and boxplots of predicted probabilities.

3. Results & Discussion

The results in Table 4.1.3 show a noticeable improvement in the fine-tuned models across both datasets. Hyperparameter tuning via 5-fold cross-validation enhanced model performance, particularly improving the F1-score and ROC AUC (also validated through ROC curve in figure 4.1.3.2), which reflect better balance in precision and recall and superior discrimination between classes. The boxplot of predicted probabilities in figure 4.1.3.1 further revealed narrower distributions in the fine-tuned model, indicating more consistent predictions compared to the base model. Based on these results, the null hypothesis is rejected, as fine-tuning significantly improved model performance for both datasets.

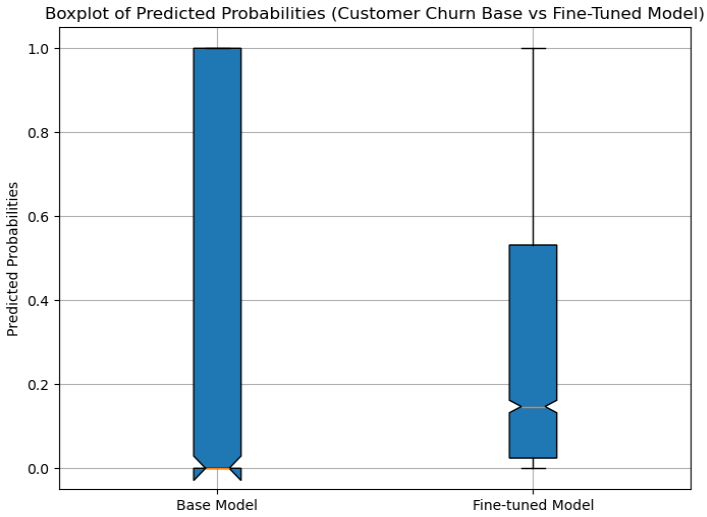

Figure 4.1.3.1: Boxplot of Predicted Probabilities (Base vs Fine-Tuned Model)

The fine-tuned model's boxplot shows improved prediction consistency with a narrower interquartile range (IQR), indicating tighter clustering of probabilities. The median, closer to the lower end, reflects confidence in identifying the majority class ("No Churn"). Shorter whiskers show reduced variability in extreme probabilities, while fewer outliers highlight improved stability after hyperparameter tuning. This aligns with the model's enhanced performance metrics, such as F1-score and ROC AUC.

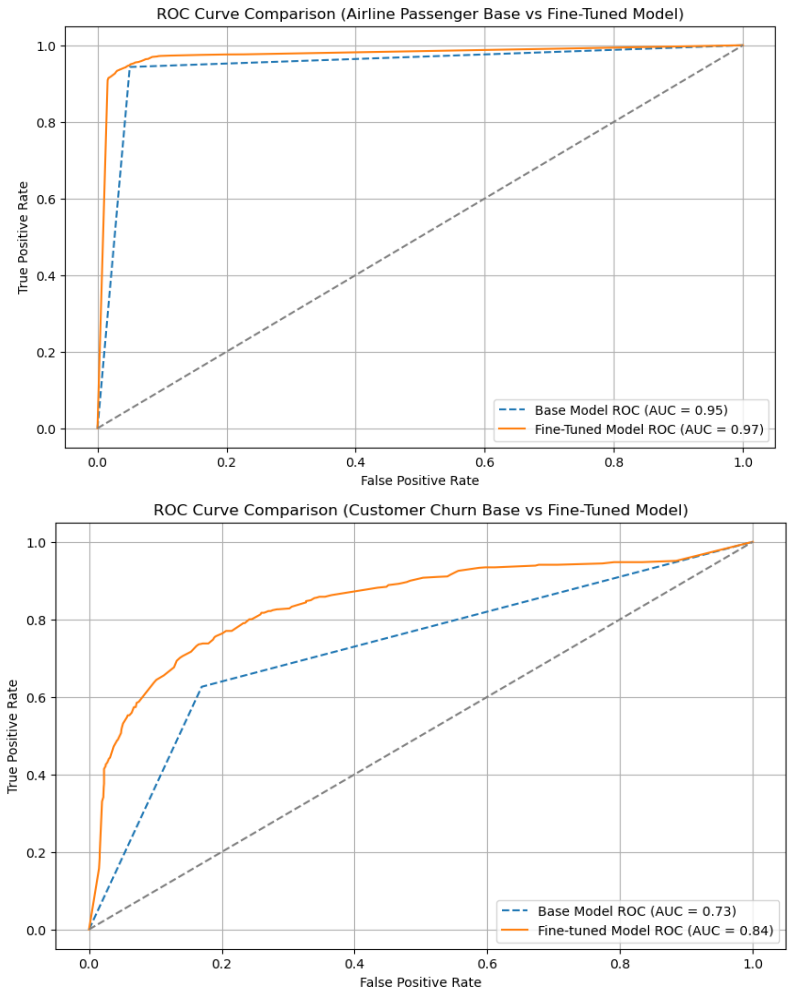

Figure 4.1.3.2: ROC Curves Comparison

Table 4.1.3: Decision Tree Results

Dataset	Model	Accuracy	F1 (Minority)	F1 (Weighted)	ROC AUC
Customer Churn	Base	77.0%	0.63	0.77	0.73
Customer Churn	Fine-Tuned	82.0%	0.69	0.82	0.84
Airline Passenger	Base	94.6%	0.94	0.95	0.95
Airline Passenger	Fine-Tuned	95.2%	0.95	0.95	0.97

The decision tree performed better on the Airline Passenger dataset, likely due to more balanced classes and richer feature diversity. The churn dataset's challenges included noisier features and class imbalance, which limited predictive accuracy. Despite these challenges, both experiments confirmed that hyperparameter tuning enhanced model performance, with higher AUC and F1-scores compared to baseline models. The results suggest that while decision trees are effective and interpretable, addressing class imbalance with techniques like oversampling or ensemble methods may further improve performance.