Predicting Airline Passenger Satisfaction Using Decision Trees

Abstract

This project focuses on predicting passenger satisfaction for an airline using a Decision Tree model. The goal was to identify key factors influencing satisfaction and build a model that classifies whether a passenger is satisfied based on survey and flight features. The final model achieved high accuracy and interpretability, making it a suitable choice for decision support in the airline industry.

Problem Statement

Passenger satisfaction is a key metric in the airline industry that directly impacts brand loyalty, retention, and revenue. Airlines need data-driven insights to optimise services and target improvements. This project aims to develop a machine learning model to predict satisfaction levels using Decision Trees. The dataset includes variables such as inflight service, seat comfort, check-in experience, and personal demographics. The objective is to build a model that is both accurate and explainable.

1. Dataset Overview

- Source: Kaggle – Airline Passenger Satisfaction Dataset
- Features: age, gender, travel class, seat comfort, inflight service, cleanliness, etc.
- Target variable: satisfaction (binary: satisfied vs. neutral/dissatisfied)
- Dataset was split into predefined train.csv and test.csv files.

2. Data Preparation & Preprocessing

- Removed null entries and restricted analysis to adult passengers (age > 18)
- Applied dummy encoding to categorical features
- Selected relevant variables for modeling
- Dataset was well-balanced; no SMOTE required

3. Model Building – Decision Tree Classifier

- Implemented DecisionTreeClassifier from scikit-learn
- Performed hyperparameter tuning using GridSearchCV with 5-fold cross-validation
  - Criterion: Entropy
  - Max Depth: 20
  - Min Samples Split: 2
  - Min Samples Leaf: 4

4. Model Evaluation

- Accuracy: 95.2%
- F1 Score: 0.95
- ROC AUC: 0.97
- Evaluated using confusion matrix, ROC curve, and boxplots of predicted probabilities
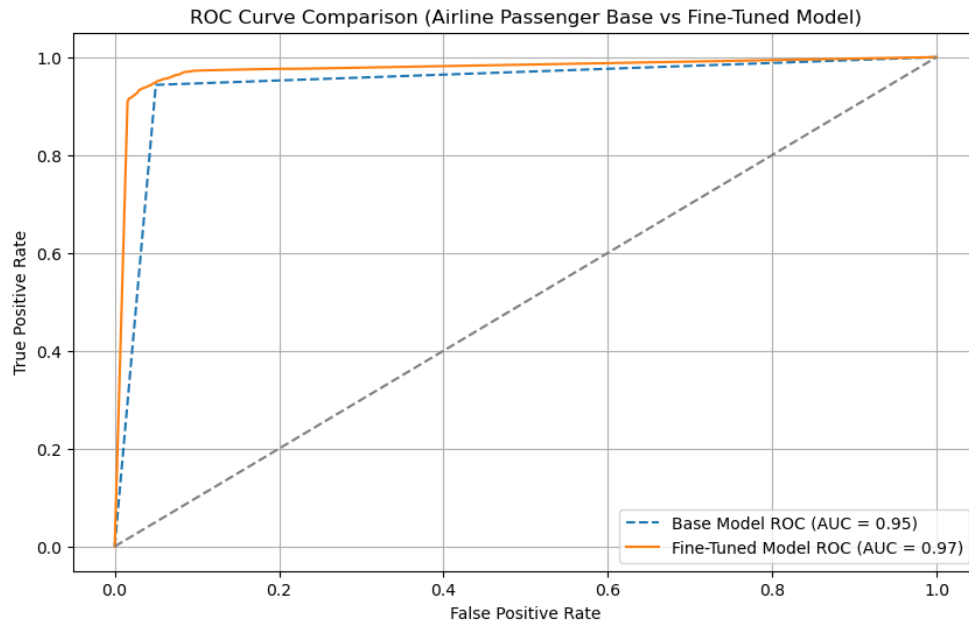
5. Insights & Discussion
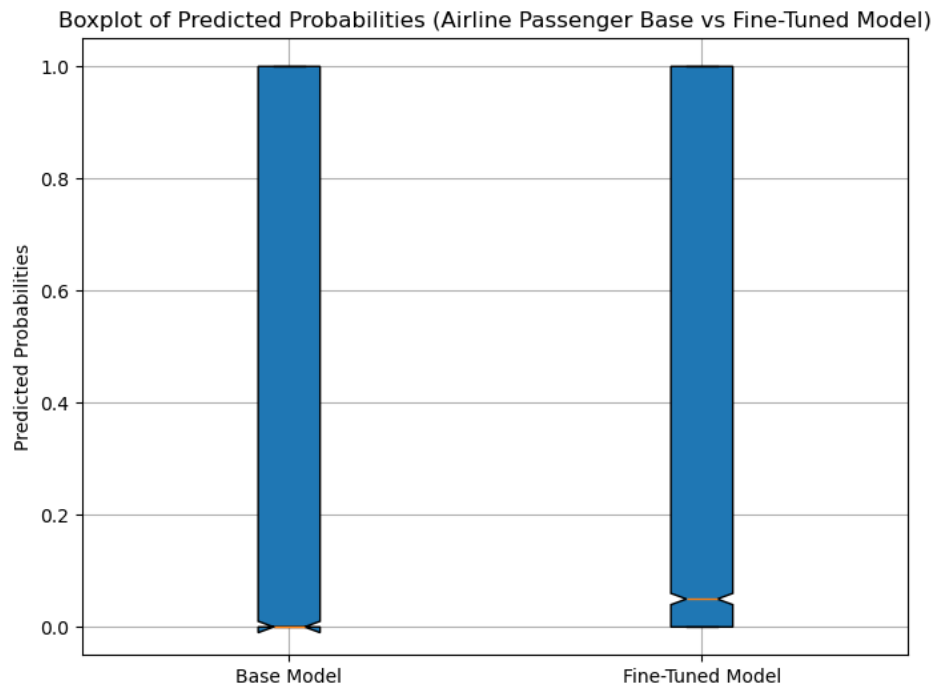

Figure 5.1: ROC Curves

Figure 5.2: Boxplots

The Decision Tree model revealed the most important factors for predicting satisfaction, such as inflight Wi-Fi, seat comfort, and check-in service. The tree-based model performed with high reliability and provided transparency in how predictions were made. This is crucial for business stakeholders to act on specific service components.

6. Conclusion

Decision Trees offer a balance of performance and interpretability. The high F1 score and ROC AUC indicate strong predictive power. The model is deployable in service dashboards or customer feedback analysis pipelines. Future enhancements could include ensemble models and integration with real-time survey systems.