# Decision Tree Classifier Project

Advanced Programming Course

Instructor: Dr. Carlo Abnousian

Semester: Spring 1403-1404

Submission Deadline: Tir 10, 1404

## 1. Introduction

This project aims to implement a complete Decision Tree algorithm in Python without using built-in decision tree functions. Decision Trees are widely used for classification problems due to their simplicity, interpretability, and effectiveness.

## 2. Dataset and Preprocessing

The dataset used for this project is an airline passenger satisfaction dataset with over 100,000 records and 25 features. During preprocessing, unnecessary columns such as IDs were removed. Continuous features like Age and Flight Distance were discretized using quartile binning. Missing values were handled, and the dataset was fully transformed into string values suitable for tree splitting.

## 3. Decision Tree Implementation

The algorithm was implemented from scratch using the following components:  
- Entropy Calculation  
- Gini Index Calculation  
- Information Gain for feature selection  
- Recursive tree construction  
- Leaf nodes represent the final predictions

## 4. Python Code Overview

The implementation is structured as follows:  
- `entropy()` and `gini\_index()` calculate purity measures.  
- `info\_gain()` chooses the best feature.  
- `Node` class represents the tree.  
- `build\_tree()` builds the tree recursively.  
- `predict()` makes predictions for new data.  
- `print\_tree()` prints the tree structure textually.

## 5. Results and Evaluation

The tree was trained using 80% of the data and tested on the remaining 20%. The printed tree structure allows clear interpretation of splits and predictions. The final accuracy on the test set is approximately reported by the code during execution.

## 6. Conclusion

This project demonstrates a manual implementation of a Decision Tree classifier, including all essential steps from preprocessing to training and evaluation. The project enhances understanding of the decision-making process in tree-based models.