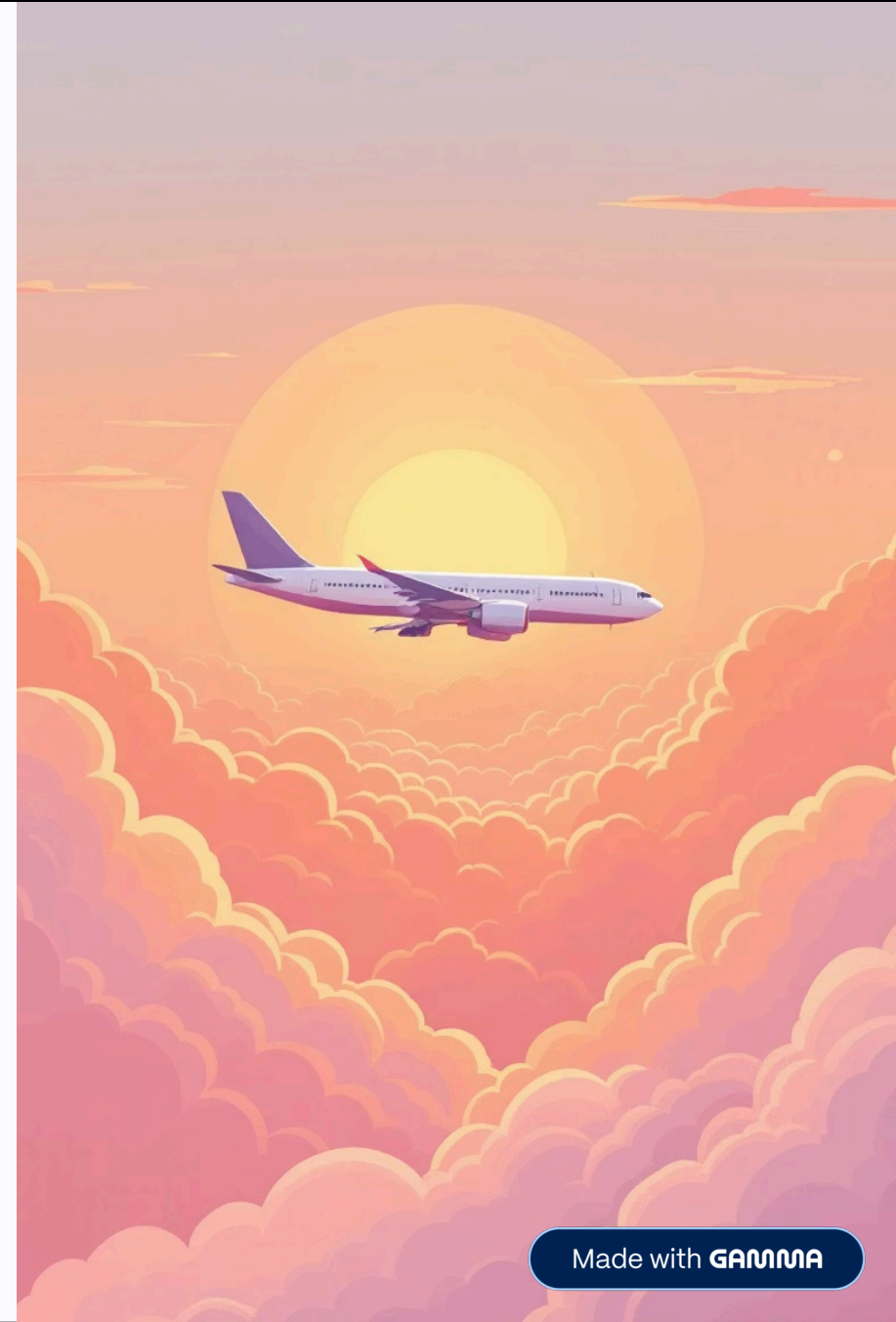


# Airline Customer Satisfaction Prediction

Using Decision Tree Classification to predict passenger satisfaction based on service quality, demographics, and travel details



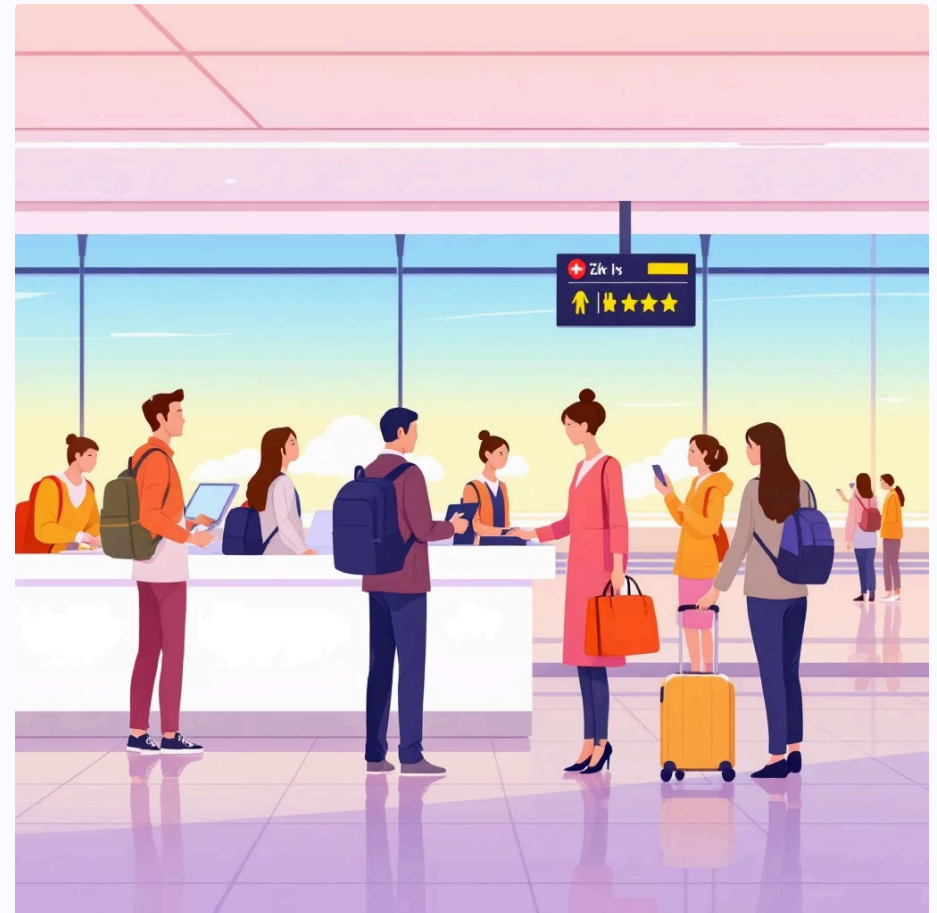
# Predicting Customer Satisfaction

## Business Challenge

Customer satisfaction drives airline competitiveness and retention. Understanding which factors influence satisfaction enables data-driven service improvements.

## Our Approach

- Supervised learning with binary classification
- Decision Tree algorithm for interpretability
- 10,000 passenger records analyzed



 DATASET

# Data Foundation

10K

Passenger  
Records

Comprehensive  
customer feedback  
data

23

Features

Demographics, travel  
details, service  
ratings

57%

Dissatisfied

Baseline satisfaction  
rate



# Data Processing Pipeline

01

## Data Cleaning

Removed unnecessary columns, standardized naming conventions, handled 26 missing values

02

## Exploratory Analysis

Examined distributions, identified patterns, analyzed satisfaction by demographics

03

## Quality Assurance

Validated data integrity, checked for outliers, ensured completeness

# Feature Engineering

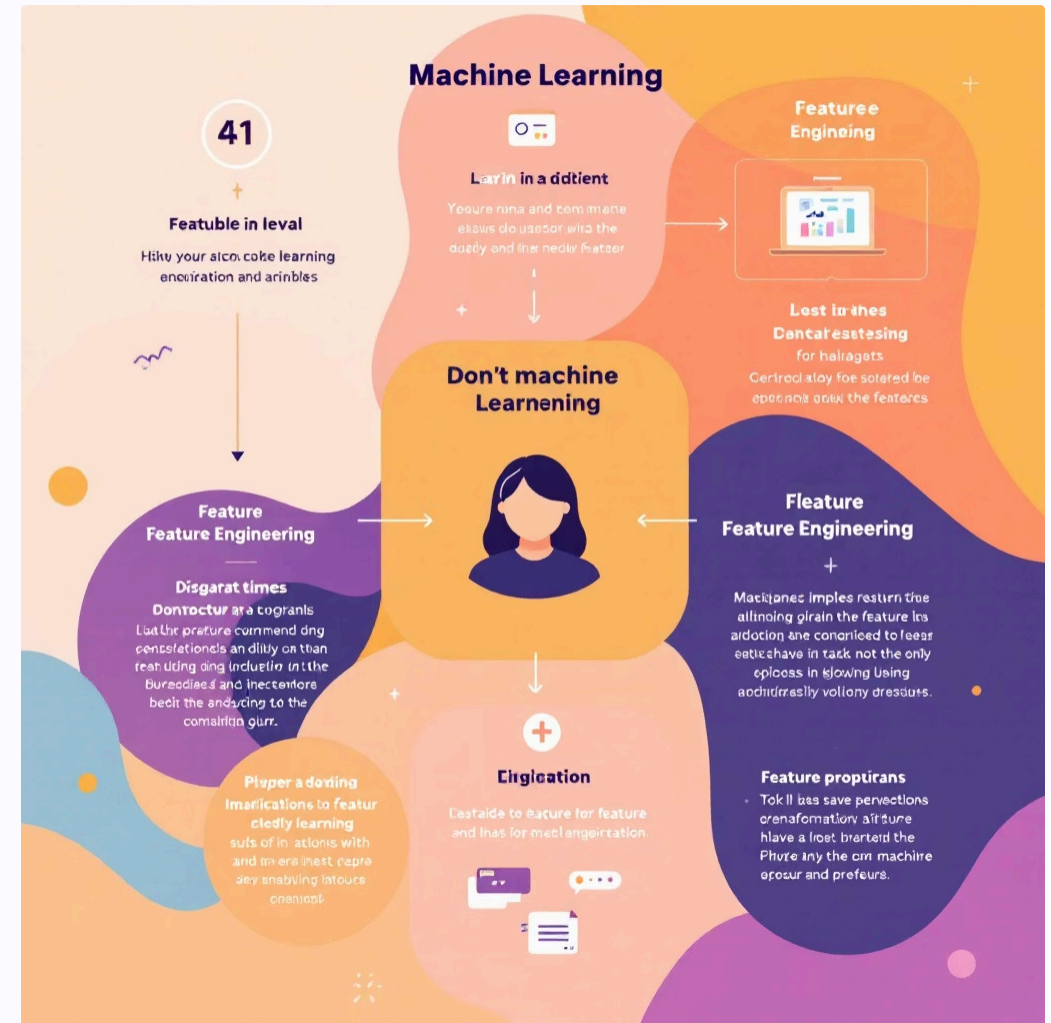
# Categorical Encoding

Transformed text features into numerical format using Label Encoding:

- Gender (Male/Female)
- Customer Type (Loyal/Disloyal)
- Travel Type (Business/Personal)
- Class (Business/Eco/Eco Plus)

## Target Variable

Satisfied  $\rightarrow 1$  | Neutral/Dissatisfied  $\rightarrow 0$



# Model Preparation Steps



## Feature Separation

Split predictors (X) from target variable (y)



## Feature Scaling

Standardized ranges using StandardScaler



## Train-Test Split

80% training, 20% testing  
(random\_state=42)



 TRAINING

# Decision Tree Model Training

## Algorithm Choice

Decision Tree Classifier  
selected for interpretability  
and performance

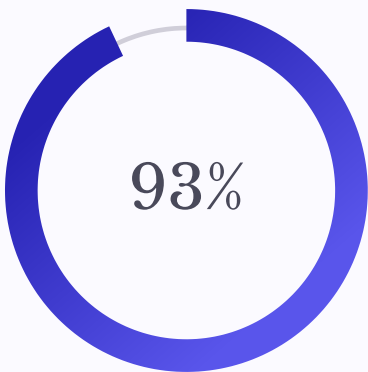
## Key Parameters

Gini Index criterion, max  
depth of 10 to prevent  
overfitting

## Training Process

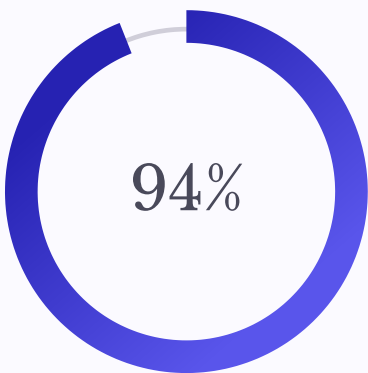
Model learned patterns from 7,979 training samples

# Model Performance Metrics



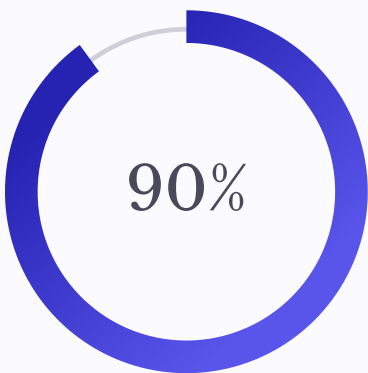
Accuracy

Overall prediction correctness



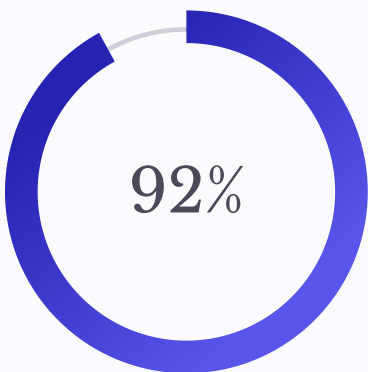
Precision

Positive prediction reliability



Recall

True positive detection rate



F1 Score

Harmonic mean of precision and recall



# Confusion Matrix Analysis

## Strong Classification

Model demonstrates excellent discrimination between satisfied and dissatisfied customers

## Key Findings

- 1,080 true negatives (95% specificity)
- 774 true positives (90% sensitivity)
- Only 141 total misclassifications
- Balanced performance across both classes





# Future Enhancements

## Hyperparameter Tuning

Optimize model parameters using GridSearchCV for peak performance

## Advanced Algorithms

Experiment with Random Forest and XGBoost for improved accuracy

## Production Deployment

Build Flask or Streamlit app for real-time satisfaction prediction