Predicting Airline Passenger Satisfaction

MURAT AKDERE - 210717005
UYGAR SEVİM - 210717055

Why did we choose this topic?

The Passenger Perspective

- Airlines provide standard services, but passenger needs are unique.
- Example:
  - Business Traveler: Prioritizes Punctuality (Time).
  - Daily Traveler: Prioritizes Wi-Fi (Connectivity).

The Company Perspective

- Ignoring these differences leads to customer loss.
- Key Question: Which passenger is unhappy and will leave the company?

Project Objectives

BUİLD A ROBUST PREDİCTİON MODEL

Develop a machine learning model to predict passenger satisfaction with high accuracy.

Minimize the risk of unpredicted customer churn.

UNCOVER BEHAVİORAL PATTERNS

Go beyond simple metrics using Feature Engineering.

REAL-WORLD SOLUTION

To provide management with the answer to the question, "Where should you invest the money to increase customer satisfaction?"

Dataset Overview

```
RangeIndex: 103904 entries, 0 to 103903
Data columns (total 25 columns):
 #   Column                             Non-Null Count   Dtype
---  ------                             --------------   -----
 0   Unnamed: 0                         103904 non-null  int64
 1   id                                 103904 non-null  int64
 2   Gender                             103904 non-null  object
 3   Customer Type                      103904 non-null  object
 4   Age                                103904 non-null  int64
 5   Type of Travel                     103904 non-null  object
 6   Class                              103904 non-null  object
 7   Flight Distance                    103904 non-null  int64
 8   Inflight wifi service              103904 non-null  int64
 9   Departure/Arrival time convenient  103904 non-null  int64
 10  Ease of Online booking             103904 non-null  int64
 11  Gate location                      103904 non-null  int64
 12  Food and drink                     103904 non-null  int64
 13  Online boarding                    103904 non-null  int64
 14  Seat comfort                       103904 non-null  int64
 15  Inflight entertainment             103904 non-null  int64
 16  On-board service                   103904 non-null  int64
 17  Leg room service                   103904 non-null  int64
 18  Baggage handling                   103904 non-null  int64
 19  Checkin service                    103904 non-null  int64
...
 23  Arrival Delay in Minutes           103594 non-null  float64
 24  satisfaction                       103904 non-null  object
dtypes: float64(1), int64(19), object(5)
```

DATASET COMPOSİTİON

Demographics & Profile: Includes passenger details like Gender, Age, and Customer Type (Loyal vs. Disloyal).

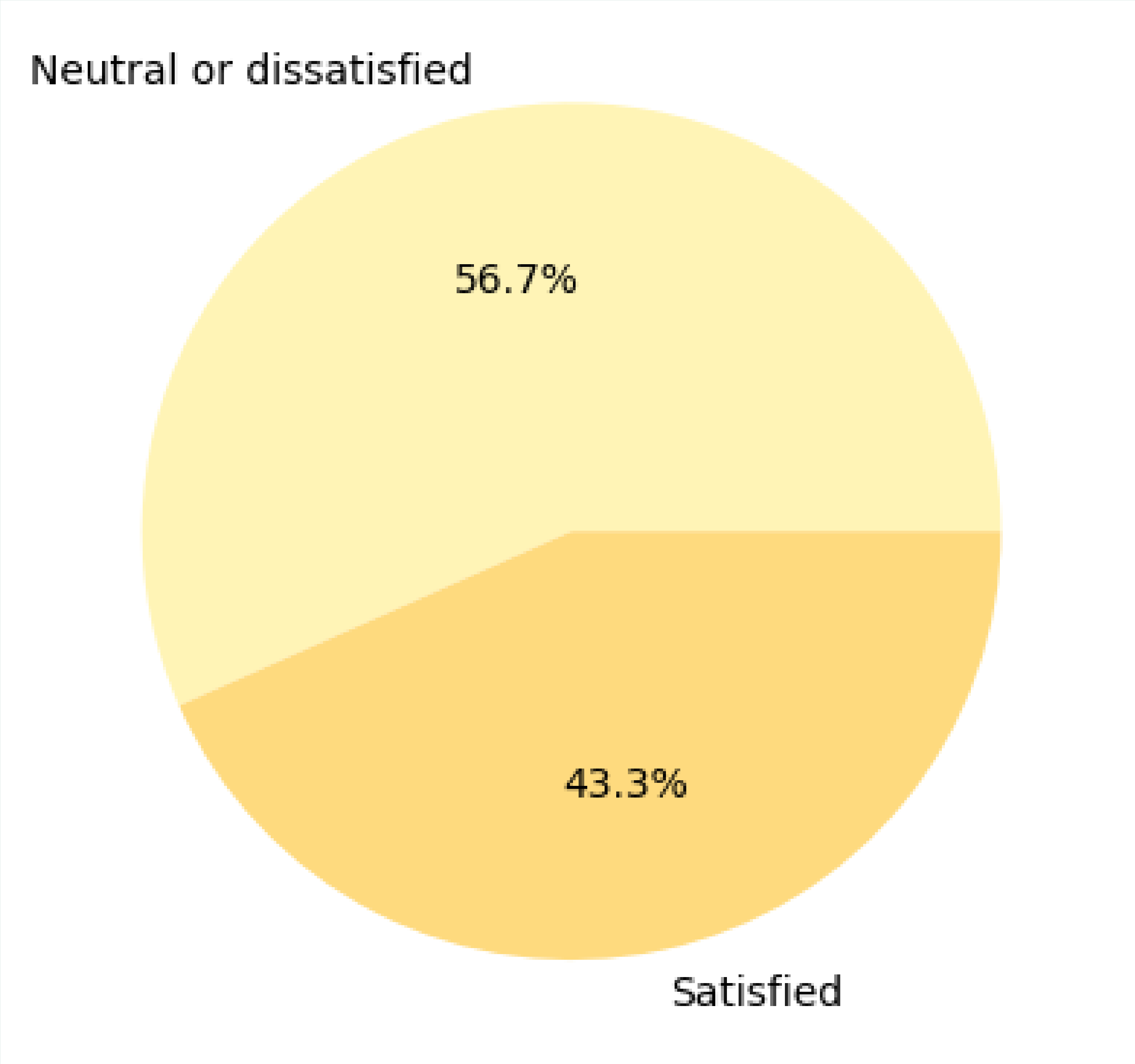
Flight Operations: Objective metrics such as Flight Distance, Class, and Delays (Departure/Arrival).

Service Ratings (1-5 Scale): Subjective scores for 14 services including Wifi, Seat Comfort, and Online Boarding.

Target: Binary Classification (Satisfied vs. Neutral/Dissatisfied).

Data Distribution

In this chart, we can see the distribution of our target variable, 'Satisfaction,' within the training dataset. 'Neutral or Dissatisfied' passengers make up 56.7% of the data, while 43.3% are 'Satisfied.' Since there is no significant class imbalance, our model will have a fair opportunity to learn both classes effectively.



Data Cleaning & Preprocessing

HANDLİNG MİSSİNG VALUES:

- **Action:** Imputed 310 missing values in Arrival Delay using the Median.
- **Reason:** Median is not affected by the outliers.

FEATURE SELECTİON (DROPPED COLUMNS)

- **Unnamed: 0 & id & Gender:** Removed as irrelevant identifiers.
- **Departure Delay:** Removed due to Multicollinearity (96% correlation with Arrival Delay).

```
Arrival Delay in Minutes  Departure Delay in Minutes    0.960247
```

```
Unnamed: 0                             0
id                                     0
Gender                                 0
Customer Type                          0
Age                                    0
Type of Travel                         0
Class                                  0
Flight Distance                        0
Inflight wifi service                  0
Departure/Arrival time convenient      0
Ease of Online booking                 0
Gate location                          0
Food and drink                         0
Online boarding                        0
Seat comfort                           0
Inflight entertainment                 0
On-board service                       0
Leg room service                       0
Baggage handling                       0
Checkin service                        0
Inflight service                       0
Cleanliness                            0
Departure Delay in Minutes             0
Arrival Delay in Minutes             310
satisfaction                           0
```

Feature Engineering Experiments

Hypothesis:

- Can we maintain high accuracy by grouping 14 specific service features into 3 macro-categories (Digital, Comfort, Staff) to simplify the model?

Result:

- Baseline Accuracy (All Features): %96.45 (Selected Model)
- Grouped Accuracy (3 Features): %92.00

Analysis:

- Information Loss: Aggregating features (taking averages) caused a loss of granular signals.
- Model Behavior: Performs better when it can split on specific "pain points" (e.g., Wifi specifically) rather than generalization.

TEST SET REPORT

	precision	recall	f1-score	support
0	0.9576	0.9801	0.9687	14573
1	0.9738	0.9446	0.9590	11403
accuracy			0.9645	25976
macro avg	0.9657	0.9623	0.9638	25976

TEST SET REPORT

	precision	recall	f1-score	support
0	0.9179	0.9417	0.9296	14573
1	0.9229	0.8923	0.9073	11403
accuracy			0.9200	25976
macro avg	0.9204	0.9170	0.9185	25976
weighted avg	0.9201	0.9200	0.9198	25976

Overfitting Analysis

CLASSIFICATION REPORT

TRAIN SETİ RAPORU

	precision	recall	f1-score	support
0	0.97	0.99	0.98	58879
1	0.98	0.96	0.97	45025
accuracy			0.98	103904
macro avg	0.98	0.97	0.98	103904
weighted avg	0.98	0.98	0.98	103904

TEST SETİ RAPORU

	precision	recall	f1-score	support
0	0.96	0.98	0.97	14573
1	0.97	0.95	0.96	11403
accuracy			0.96	25976
macro avg	0.97	0.96	0.96	25976
weighted avg	0.96	0.96	0.96	25976

- **Validation Strategy**: 20% of the training set was reserved for validation to strictly prevent Data Leakage.
- **Generalization**: The minimal gap (<2%) between Train and Test accuracy confirms no overfitting.

Model Selection & Hyperparameter Tuning

Candidate Algorithms

We benchmarked 5 different algorithms to find the best fit:

- Logistic Regression (Baseline)
- Random Forest
- LightGBM
- CatBoost
- XGBoost

Optimization Strategy (Optuna)

- We adopted Optuna (Bayesian Optimization) to overcome the inefficiencies of traditional Grid Search. By learning from previous trials rather than searching blindly, it allowed us to extract maximum performance from our models with minimal computational cost.

Baseline Model: Logistic Regression Performance

TRAINING SET REPORT

	precision	recall	f1-score	support
0	0.8776	0.9053	0.8912	58879
1	0.8708	0.8348	0.8524	45025
accuracy			0.8747	103904
macro avg	0.8742	0.8701	0.8718	103904
weighted avg	0.8746	0.8747	0.8744	103904

TEST SET REPORT

	precision	recall	f1-score	support
0	0.8735	0.9027	0.8879	14573
1	0.8701	0.8329	0.8511	11403
accuracy			0.8721	25976
macro avg	0.8718	0.8678	0.8695	25976
weighted avg	0.8720	0.8721	0.8717	25976

- **Baseline Choice:** Selected for its simplicity to serve as a reference point.
- **Performance Expectation:** We anticipated lower accuracy compared to tree-based models.
- **Limitation:** Its linear nature cannot fully capture complex, non-linear relationships in customer satisfaction data.

Random Forest

```
TRAINING SET REPORT
              precision    recall  f1-score   support

           0     0.9764    0.9942    0.9852     58879
           1     0.9922    0.9686    0.9803     45025

    accuracy                         0.9831    103904
   macro avg     0.9843    0.9814    0.9828    103904
weighted avg     0.9833    0.9831    0.9831    103904

TEST SET REPORT
              precision    recall  f1-score   support

           0     0.9568    0.9788    0.9677     14573
           1     0.9721    0.9435    0.9576     11403

    accuracy                         0.9633     25976
   macro avg     0.9644    0.9612    0.9626     25976
weighted avg     0.9635    0.9633    0.9632     25976
```

- **Robust All-Rounder:** We selected Random Forest because it is a versatile model that adapts well to almost any dataset with minimal tuning.
- **Step Up from Baseline:** As expected, it significantly outperformed Logistic Regression by capturing non-linear relationships that linear models miss.
- **Performance Expectation:** However, we anticipated it would fall slightly behind advanced Gradient Boosting models (like XGBoost), as it averages predictions rather than iteratively correcting errors.

LightGBM

```
TRAINING SET REPORT
              precision    recall  f1-score   support

           0     0.9633    0.9856    0.9743     58879
           1     0.9806    0.9508    0.9655     45025

    accuracy                         0.9705    103904
   macro avg     0.9719    0.9682    0.9699    103904
weighted avg     0.9708    0.9705    0.9705    103904

TEST SET REPORT
              precision    recall  f1-score   support

           0     0.9564    0.9798    0.9679     14573
           1     0.9733    0.9429    0.9579     11403

    accuracy                         0.9636     25976
   macro avg     0.9648    0.9613    0.9629     25976
weighted avg     0.9638    0.9636    0.9635     25976
```

- **Efficiency & Speed:** We selected LightGBM for its lightning-fast training speed and low memory usage, which is ideal for larger datasets.
- **Leaf-Wise Growth:** Unlike Random Forest, it uses a "leaf-wise" growth strategy that often achieves higher accuracy by minimizing errors more aggressively.
- **Boosting Advantage:** We anticipated it would surpass Random Forest by iteratively correcting errors (Boosting) rather than just averaging them, leading to sharper decision boundaries.

CatBoost

```
TRAINING SET REPORT
              precision    recall  f1-score   support

           0     0.9676    0.9867    0.9770     58879
           1     0.9821    0.9567    0.9693     45025

    accuracy                         0.9737    103904
   macro avg     0.9749    0.9717    0.9732    103904
weighted avg     0.9739    0.9737    0.9737    103904

TEST SET REPORT
              precision    recall  f1-score   support

           0     0.9583    0.9783    0.9682     14573
           1     0.9715    0.9456    0.9584     11403

    accuracy                         0.9640     25976
   macro avg     0.9649    0.9620    0.9633     25976
```

- **Native Categorical Support:** We selected CatBoost specifically for its superior ability to handle categorical features (like Class and Travel Type) without complex preprocessing.
- **Stability Focus:** It utilizes symmetric decision trees, which are known to be exceptionally stable and resistant to overfitting.
- **Efficiency Trade-off:** Although slightly more accurate than LightGBM, its slow training speed (200s) makes it less efficient than our eventual champion, XGBoost, which offers a better balance of speed and accuracy.

XGBoost

```
TRAINING SET REPORT
              precision    recall  f1-score   support

           0     0.9647    0.9868    0.9756     58879
           1     0.9822    0.9527    0.9672     45025

    accuracy                         0.9720    103904
   macro avg     0.9734    0.9698    0.9714    103904
weighted avg     0.9723    0.9720    0.9720    103904

TEST SET REPORT
              precision    recall  f1-score   support

           0     0.9576    0.9801    0.9687     14573
           1     0.9738    0.9446    0.9590     11403

    accuracy                         0.9645     25976
   macro avg     0.9657    0.9623    0.9638     25976
```

- **State-of-the-Art Choice:** We selected XGBoost as our strongest candidate because it is the industry standard for structured data, utilizing advanced regularization to prevent overfitting better than other boosting methods.
- **Peak Performance:** As anticipated, it delivered the best results among all algorithms, achieving the highest test accuracy of 96.45% and a remarkable 97.38% precision for the target class.
- **The Champion Rationale:** It earned the **"Champion"** title by striking the perfect balance: it is not only the most accurate model but also significantly faster (47.6s) than its closest rival, CatBoost (200.7s), making it the most efficient and safe choice for deployment.

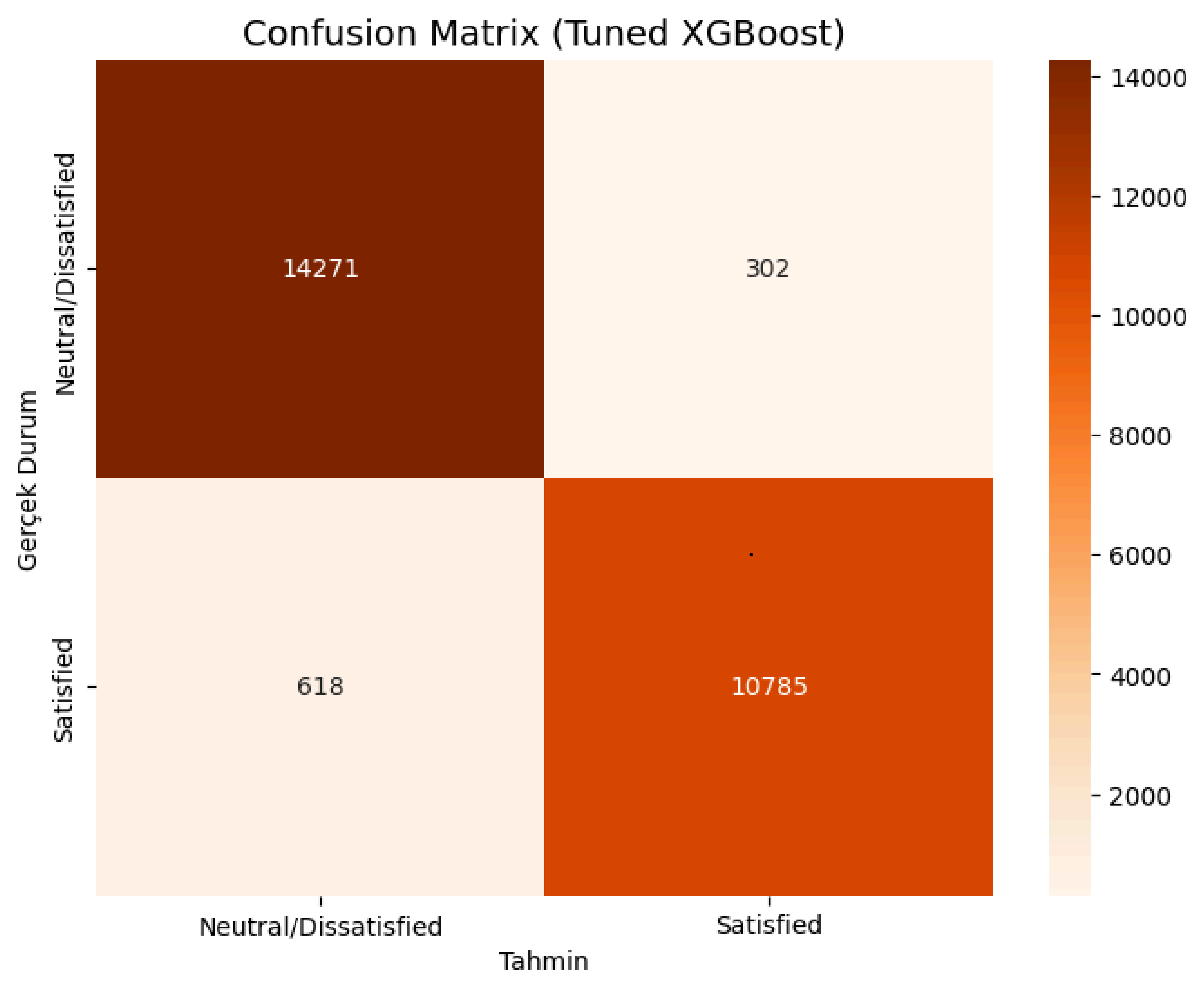
Model Performance & Training Time Comparison

Model Performance & Training Time Comparison

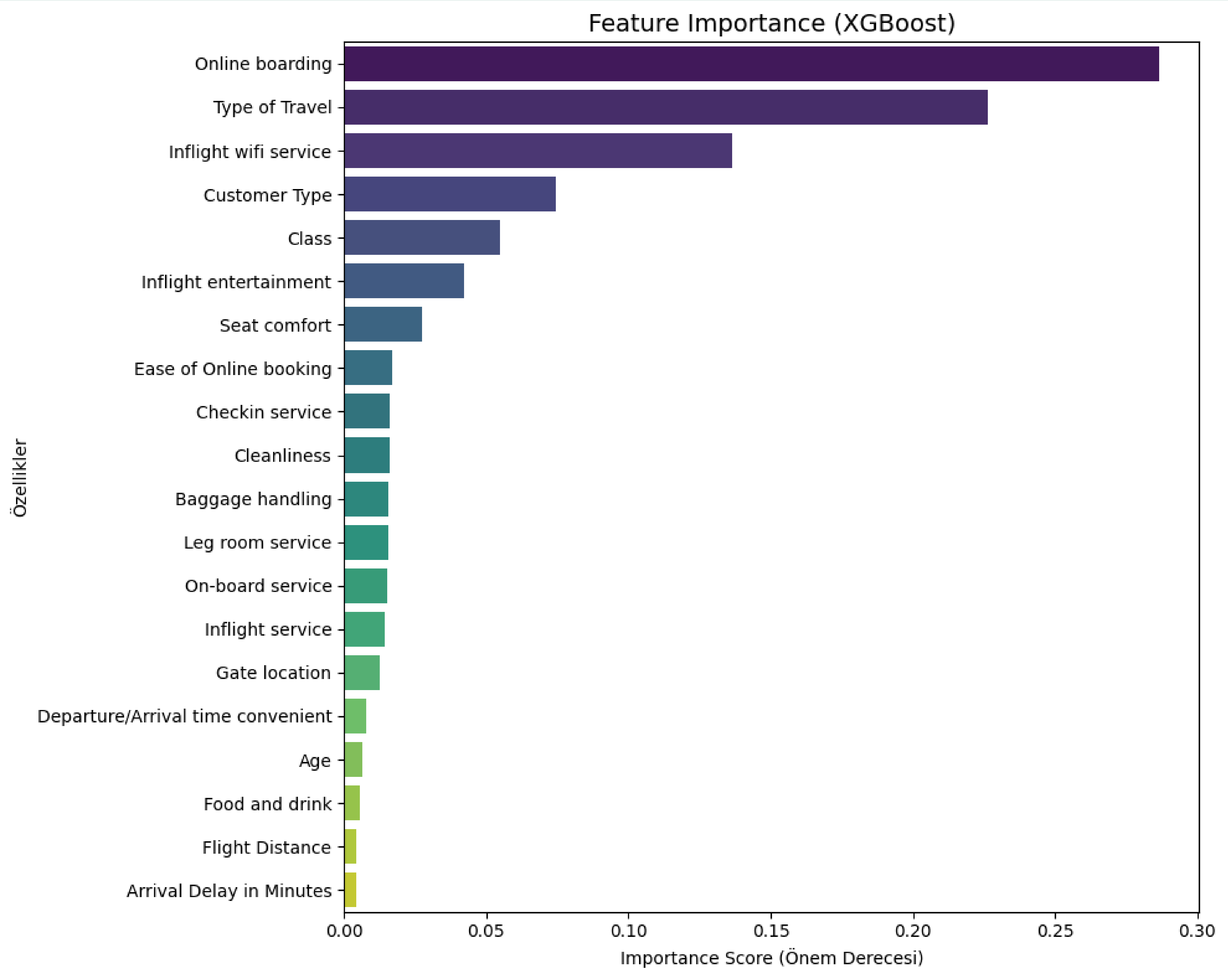
Model	Test Accuracy	Training Time (s)
XGBoost	**96.45%**	**47.6 s**
CatBoost	96.40%	200.7 s
LightGBM	96.36%	61.2 s
Random Forest	96.33%	65.0 s
Logistic Regression (Baseline)	87.21%	6.7 s

- **Champion Selection**: XGBoost is selected as the final model, achieving the highest test accuracy of 96.45%.
- **Efficiency Trade-off**: While CatBoost delivered comparable accuracy compared to XGBoost, it required significantly longer training time (200.7s vs. 47.6s). XGBoost offers the optimal balance of speed and performance.
- **Model Complexity Justified**: All ensemble methods (tree-based models) significantly outperformed the Logistic Regression baseline (87.21%), proving that the dataset requires complex, non-linear solutions.

XGBoost Confusion Matrix



XGBoost Feature Importance



SHAP ANALYSIS

RECOMMENDATIONS

1. Adopt a "Digital First" Strategy

- **Insight:** Wifi & Online Boarding impact satisfaction more than physical comfort.
- **Action:** Prioritize IT infrastructure investments over seat upgrades.

2. Focus on "Personal Travelers"

- **Insight:** Unlike business travelers, leisure travelers are the highest churn risk.
- **Action:** Launch targeted loyalty campaigns specifically for personal travel.

3. Optimize Budget Allocation

- **Insight:** Food and Drink has negligible impact on the model decisions.
- **Action:** Cut catering budget and reallocate funds to improve connectivity.

Thank you

for listening