Airline Passenger Satisfaction

Explainable Artificial Intelligence (XAI) techniques

Group PA2_DS3_GC: Athos Freitas up202108792 | Félix Martins up202108837 | Luís Du up202105385

Table of contents

01

Dataset and Learning Task

Definition of dataset and supervised learning algorithms; pre-modelling analysis

02

In-Modeling Explanations

Glass-box model on the case study and critically analyze its interpretability

03

Post-Modeling Explanations

Black-box model on our case study and then apply post-hoc XAI techniques

04

Quality of the explanations

Functionally-grounded metric to evaluate explanations generated by post-hoc XAI techniques

01

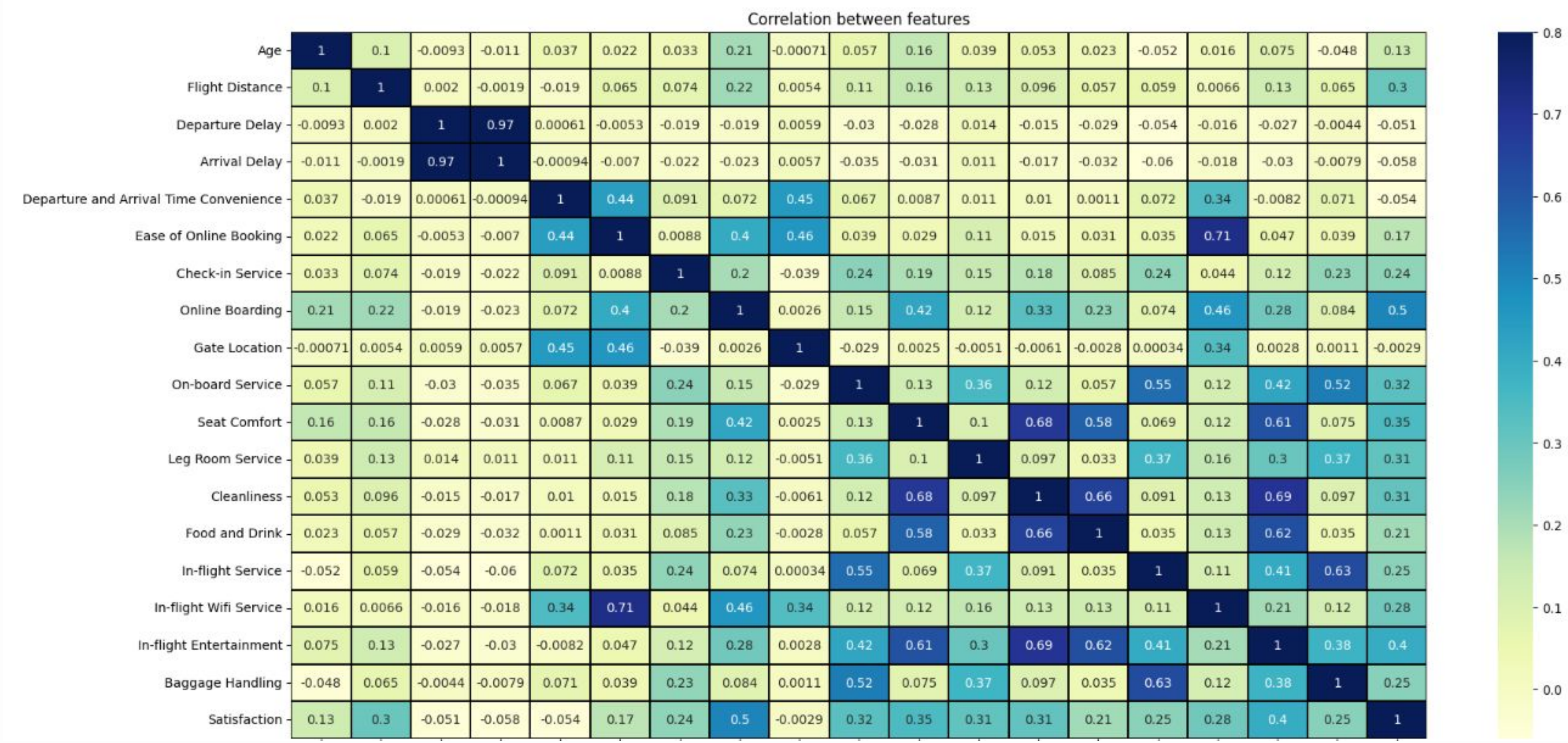
Pre-modelling

About this dataset

This dataset contains an airline passenger satisfaction survey.

Dimensionality	24 Columns and 129880 Rows.
Target class	**Neutral or Dissatisfied:** 56.6% **Satisfied:** 43.4%
Features	**Passenger data** - gender, age **Travel information** - business/personal, economy/business **Flight information** - flight distance, arrival/departure delay **Satisfaction levels (1-5)** related to specific travel aspects (service quality)
Missing values	0.013% Missing values; Dropped 393 rows with missing values.
Duplicate rows	There are no duplicate rows in the dataset.

Correlation matrix



Age	1	0.1	-0.0093	-0.011	0.037	0.022	0.033	0.21	-0.00071	0.057	0.16	0.039	0.053	0.023	-0.052	0.016	0.075	-0.048	0.13
Flight Distance	0.1	1	0.002	-0.0019	-0.019	0.065	0.074	0.22	0.0054	0.11	0.16	0.13	0.096	0.057	0.059	0.0066	0.13	0.065	0.3
Departure Delay	-0.0093	0.002	1	0.97	0.00061	-0.0053	-0.019	-0.019	0.0059	-0.03	-0.028	0.014	-0.015	-0.029	-0.054	-0.016	-0.027	-0.0044	-0.051
Arrival Delay	-0.011	-0.0019	0.97	1	-0.00094	-0.007	-0.022	-0.023	0.0057	-0.035	-0.031	0.011	-0.017	-0.032	-0.06	-0.018	-0.03	-0.0079	-0.058
Departure and Arrival Time Convenience	0.037	-0.019	0.00061	-0.00094	1	0.44	0.091	0.072	0.45	0.067	0.0087	0.011	0.01	0.0011	0.072	0.34	-0.0082	0.071	-0.054
Ease of Online Booking	0.022	0.065	-0.0053	-0.007	0.44	1	0.0088	0.4	0.46	0.039	0.029	0.11	0.015	0.031	0.035	0.71	0.047	0.039	0.17
Check-in Service	0.033	0.074	-0.019	-0.022	0.091	0.0088	1	0.2	-0.039	0.24	0.19	0.15	0.18	0.085	0.24	0.044	0.12	0.23	0.24
Online Boarding	0.21	0.22	-0.019	-0.023	0.072	0.4	0.2	1	0.0026	0.15	0.42	0.12	0.33	0.23	0.074	0.46	0.28	0.084	0.5
Gate Location	-0.00071	0.0054	0.0059	0.0057	0.45	0.46	-0.039	0.0026	1	-0.029	0.0025	-0.0051	-0.0061	-0.0028	0.00034	0.34	0.0028	0.0011	-0.0029
On-board Service	0.057	0.11	-0.03	-0.035	0.067	0.039	0.24	0.15	-0.029	1	0.13	0.36	0.12	0.057	0.55	0.12	0.42	0.52	0.32
Seat Comfort	0.16	0.16	-0.028	-0.031	0.0087	0.029	0.19	0.42	0.0025	0.13	1	0.1	0.68	0.58	0.069	0.12	0.61	0.075	0.35
Leg Room Service	0.039	0.13	0.014	0.011	0.011	0.11	0.15	0.12	-0.0051	0.36	0.1	1	0.097	0.033	0.37	0.16	0.3	0.37	0.31
Cleanliness	0.053	0.096	-0.015	-0.017	0.01	0.015	0.18	0.33	-0.0061	0.12	0.68	0.097	1	0.66	0.091	0.13	0.69	0.097	0.31
Food and Drink	0.023	0.057	-0.029	-0.032	0.0011	0.031	0.085	0.23	-0.0028	0.057	0.58	0.033	0.66	1	0.035	0.13	0.62	0.035	0.21
In-flight Service	-0.052	0.059	-0.054	-0.06	0.072	0.035	0.24	0.074	0.00034	0.55	0.069	0.37	0.091	0.035	1	0.11	0.41	0.63	0.25
In-flight Wifi Service	0.016	0.0066	-0.016	-0.018	0.34	0.71	0.044	0.46	0.34	0.12	0.12	0.16	0.13	0.13	0.11	1	0.21	0.12	0.28
In-flight Entertainment	0.075	0.13	-0.027	-0.03	-0.0082	0.047	0.12	0.28	0.0028	0.42	0.61	0.3	0.69	0.62	0.41	0.21	1	0.38	0.4
Baggage Handling	-0.048	0.065	-0.0044	-0.0079	0.071	0.039	0.23	0.084	0.0011	0.52	0.075	0.37	0.097	0.035	0.63	0.12	0.38	1	0.25
Satisfaction	0.13	0.3	-0.051	-0.058	-0.054	0.17	0.24	0.5	-0.0029	0.32	0.35	0.31	0.31	0.21	0.25	0.28	0.4	0.25	1

Preprocessing

One-Hot encode categorical features

Transform categorical values (*Gender*, *Customer Type, Type of Travel, Class*) into binary columns

Standardize numerical features

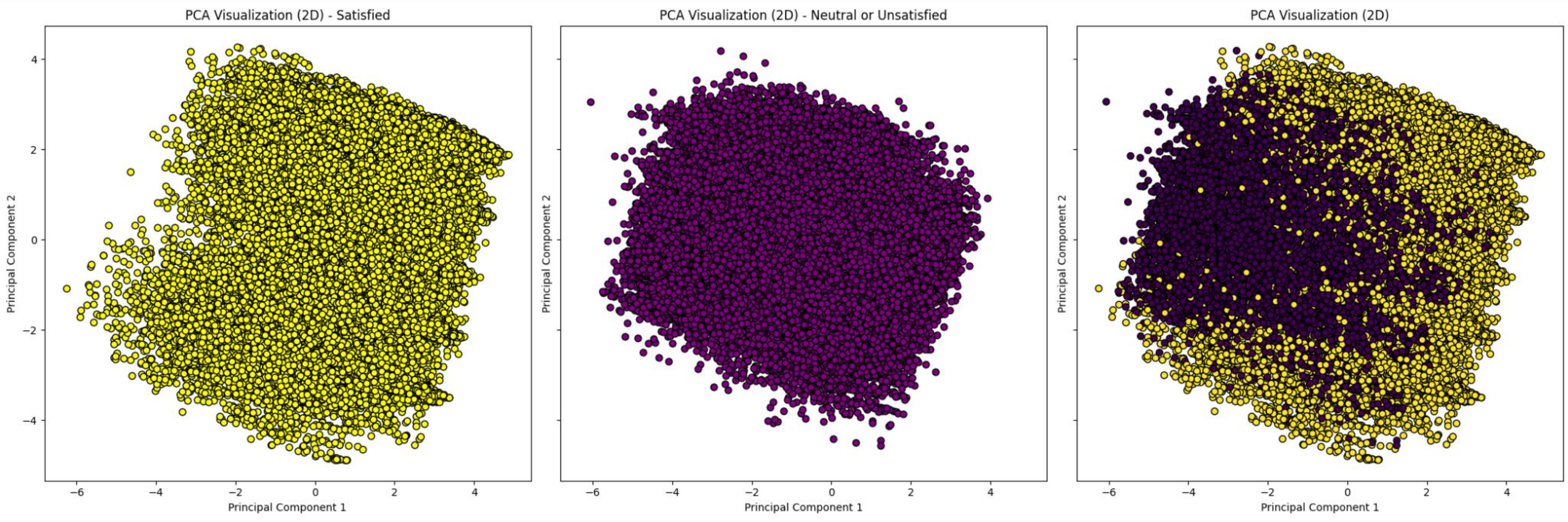
Transform numerical values to have a mean of 0 and a standard deviation of 1 - (same scale)

Remove highly correlated features

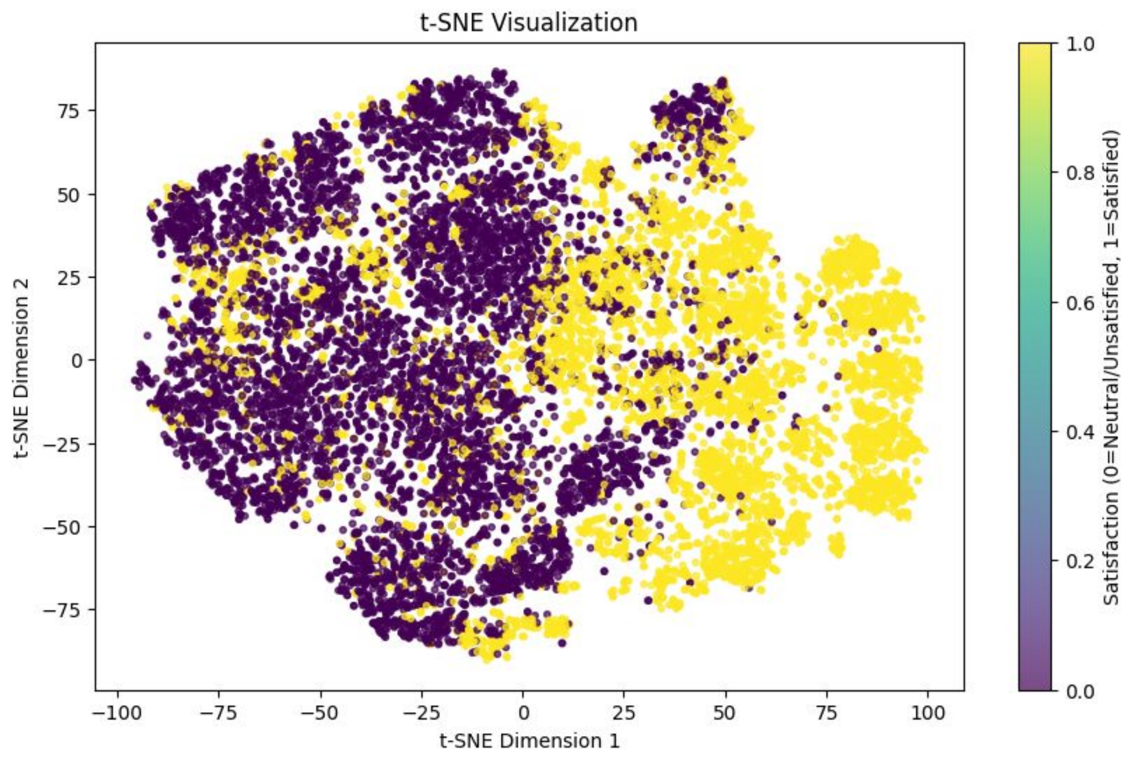
Remove features with high correlation (> 0.8), keeping only the one with the highest variance.

```
preprocessor = ColumnTransformer(
    transformers = [
        ('cat', OneHotEncoder(drop='first'), categorical_columns),
        ('num', StandardScaler(), numerical_columns),
        ('remove_corr', RemoveHighlyCorrelated(), numerical_columns)
    ]
)
```

Principal Component Analysis (PCA)



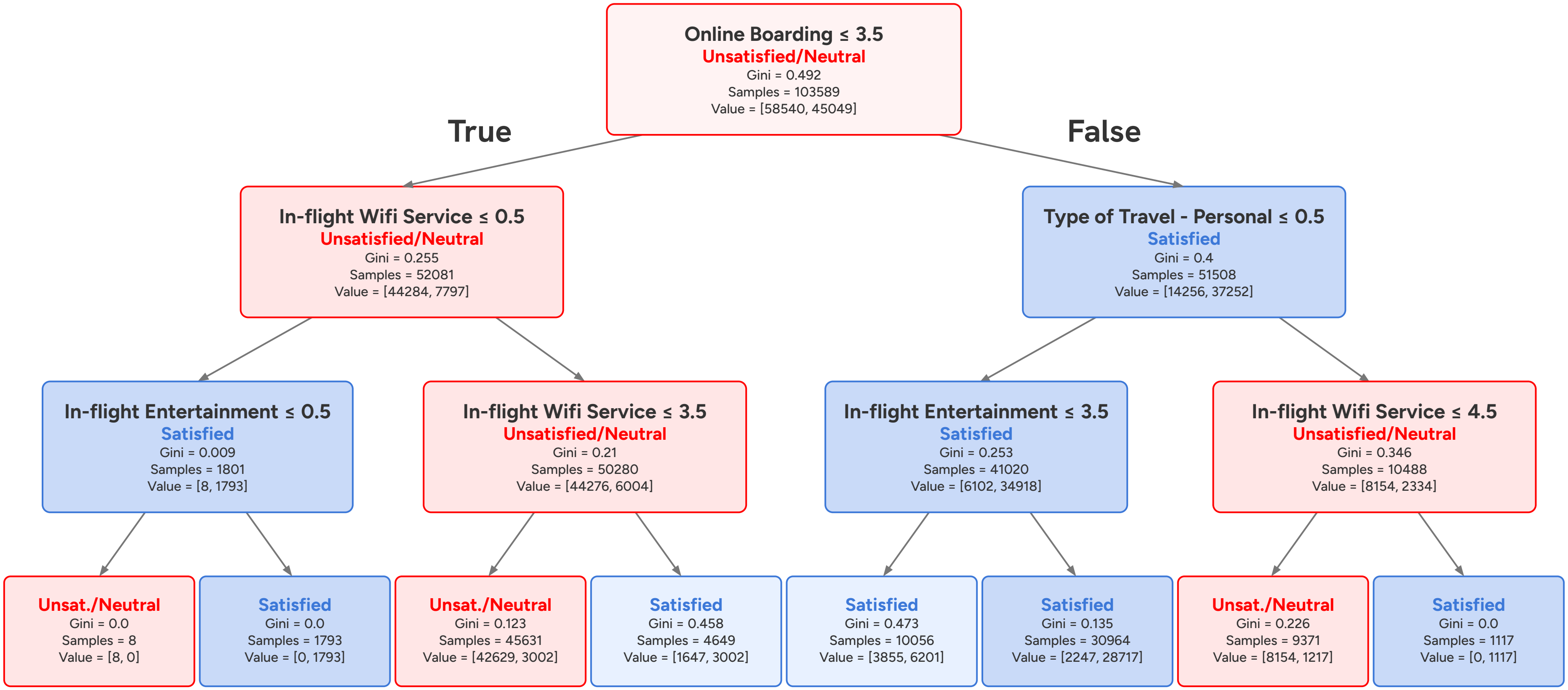
t-distributed Stochastic Neighbor Embedding



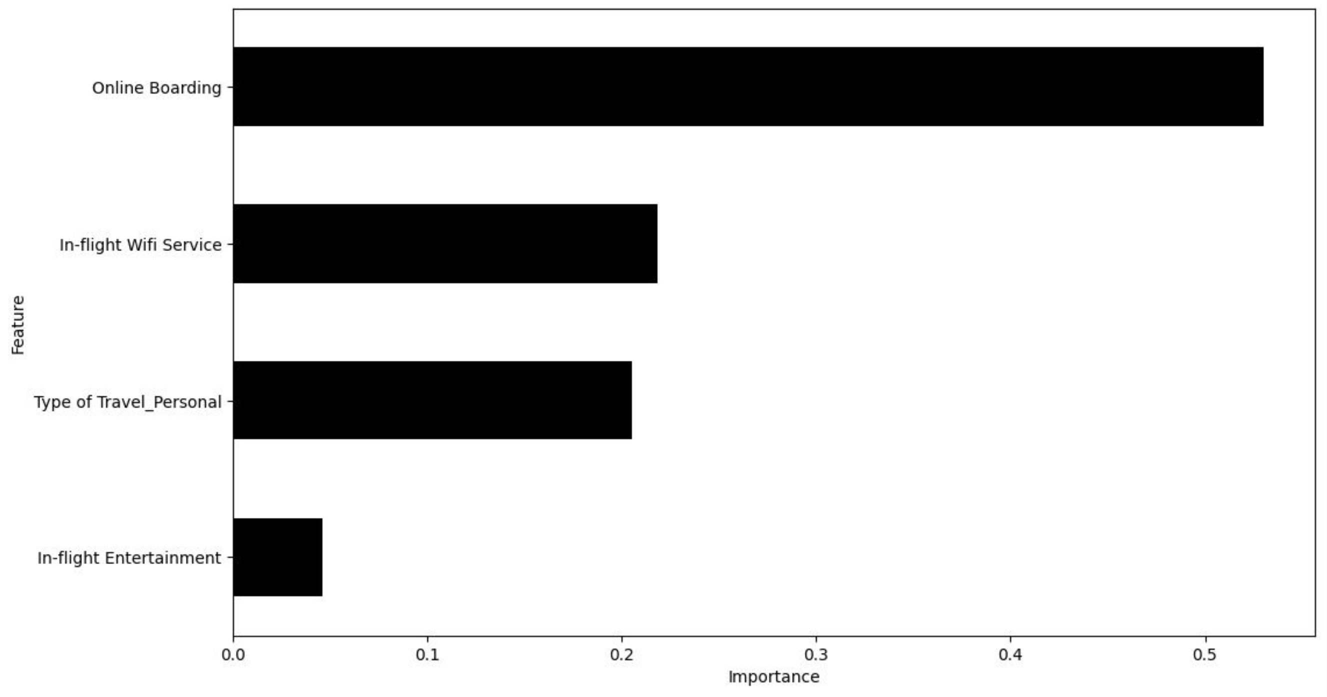
02

In-modelling

Glass-box model - *Decision Tree*



Feature importance



0.530056

0.218382

0.205443

0.045624

03

Post-modelling

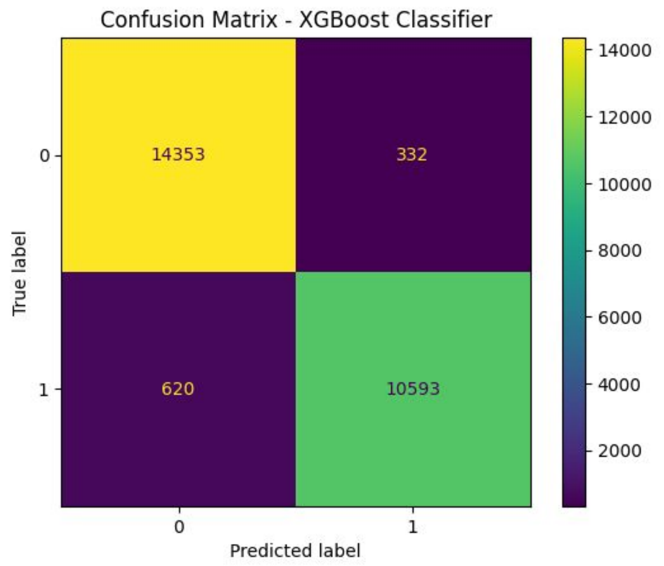
Black-box model - *XGBoost*

Accuracy: 96.3%

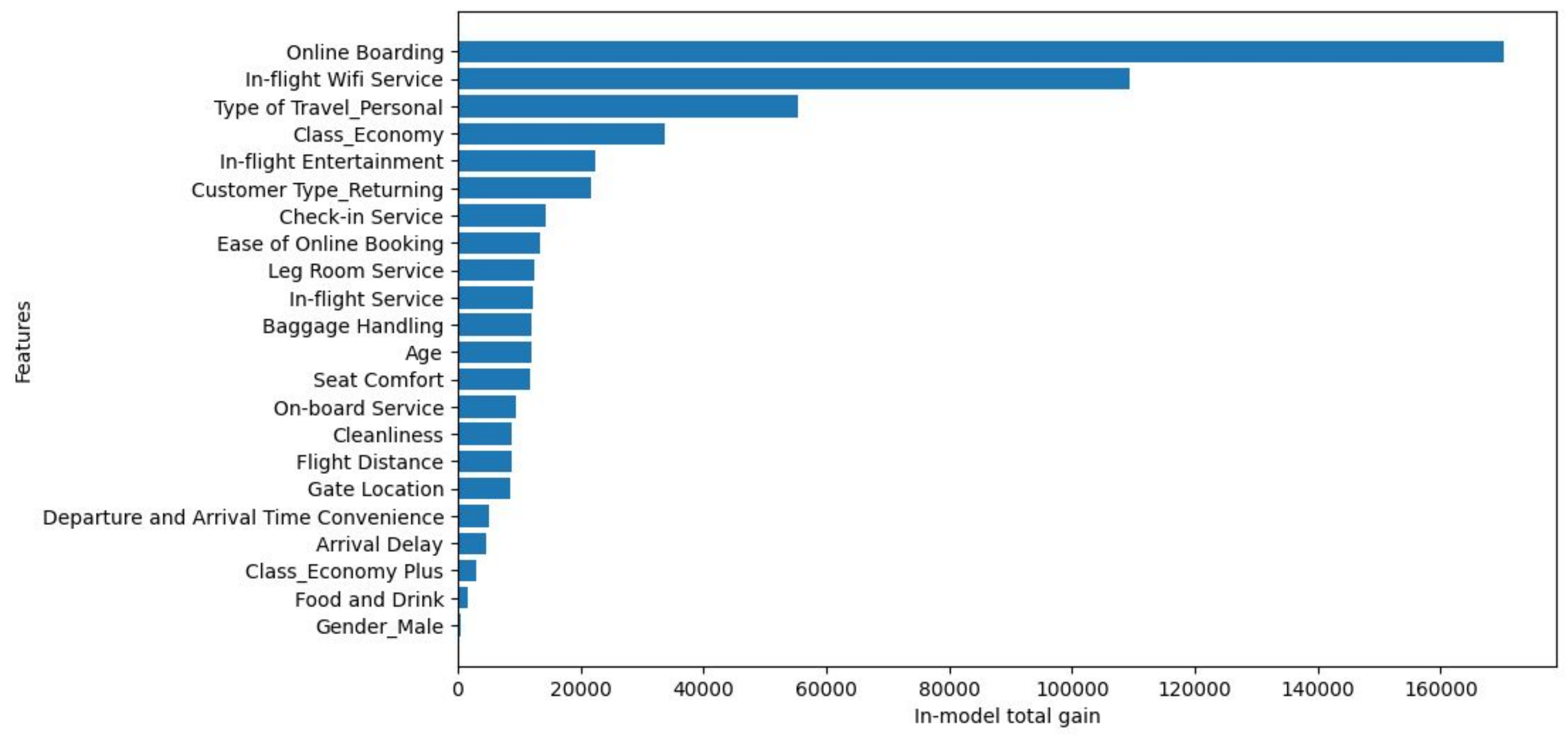
- Tree-based model
- Evaluation metric: Logarithmic loss function

Fine-tuned accuracy: 96.5%

- Evaluation metric: Logarithmic loss function
- Estimators : 200
- Max depth: 7



XGBoost In-model Feature Importance



Surrogate Model - Decision Tree

Small Decision Tree to approximate Black Box model

Decision Tree with maximum depth 3

Evaluation: Agreement Rate

Accuracy with black box predictions as ground truth and surrogate model predictions as predicted values

```
surrogate_model = DecisionTreeClassifier(max_depth=3)
surrogate_pipeline = Pipeline(steps = [
    ('preprocessor', preprocessor),
    ('classifier', surrogate_model),
])
y_train_surrogate = black_box.predict(X_train)
surrogate_pipeline.fit(X_train, y_train_surrogate)
agreement_rate = accuracy_score(y_true=y_pred, y_pred=y_pred_surrogate)
```

Surrogate Model - Decision Tree

The surrogate decision tree obtained was exactly the same as the Glass-Box model trained on the data for In-modelling explanations.

Results:

- **Agreement rate**, evaluates fidelity: 89.6%
- **Accuracy**, in the actual data: 88.5%

Black-Box and post-hoc surrogate Glass-Box

- Best possible accuracy
- Explanations: less precise, can lead to confusion

In-modelling Glass-Box

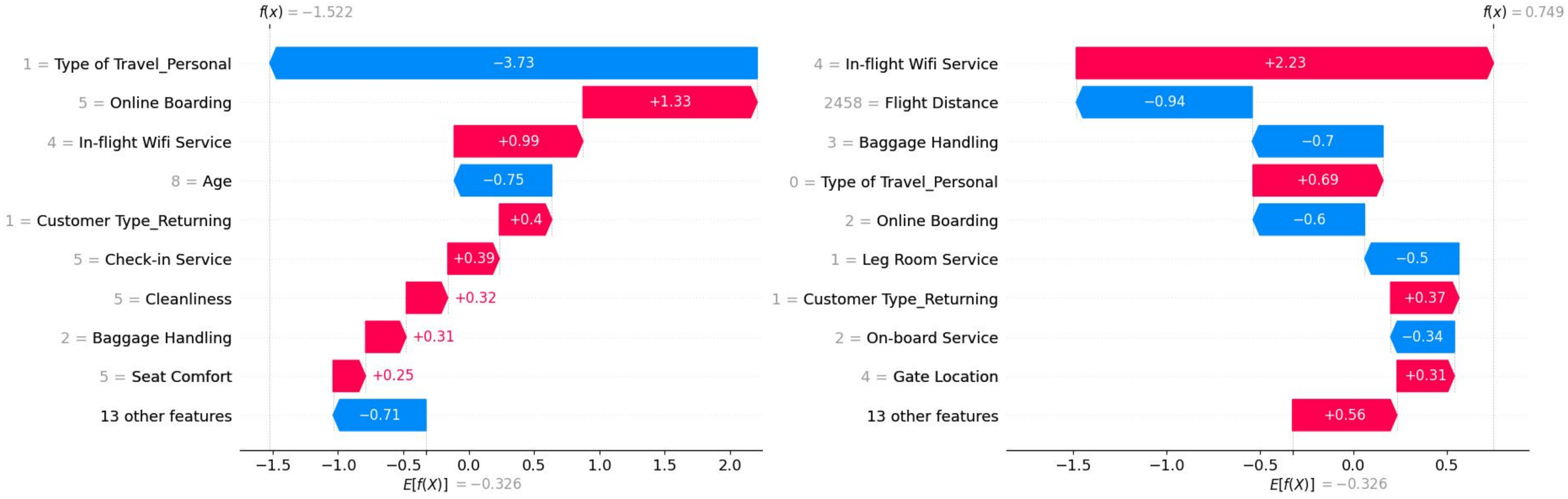
- Worse accuracy
- Precise explanations
- Better when explanations are valued above everything else

SHAP - Global explanation

- Type of travel - Personal
  - Personal travelling impacts negatively (towards Dissatisfied or Neutral)
  - Business travelling impacts positively (towards satisfied)

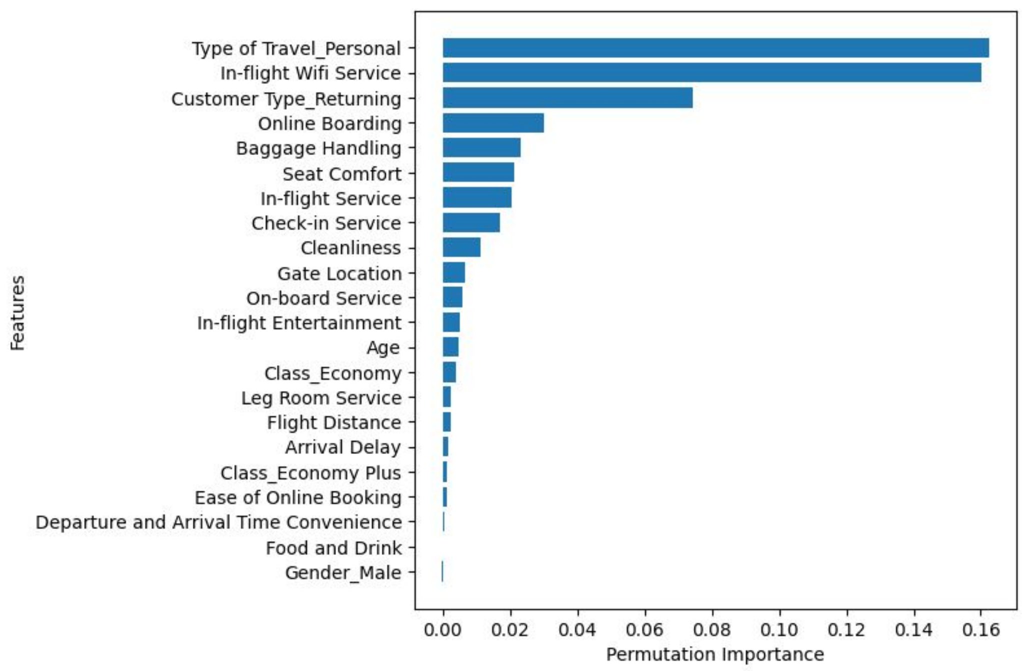
- Customer type - Returning
  - Returning customer impacts positively
  - First impact negatively

SHAP - Local explanation

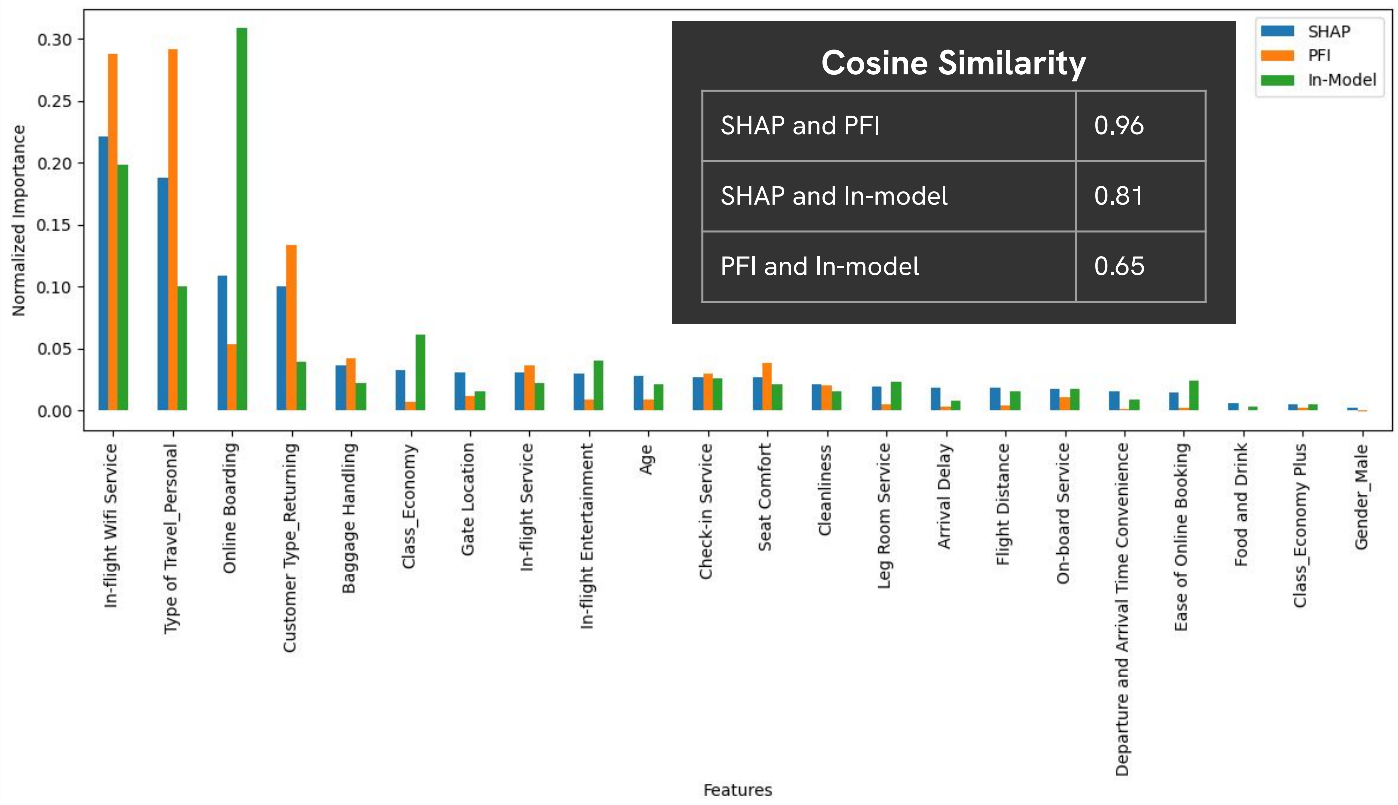


Permutation Feature Importance

- Does not include if the impact is positive or negative as SHAP does
- Highest Importance:
    - In-flight Wifi Service
    - Type of Travel - Personal
    - Online Boarding
    - Customer type - Returning



SHAP vs PFI - Global explanations



Cosine Similarity

SHAP and PFI	0.96
SHAP and In-model	0.81
PFI and In-model	0.65

Anchors

- Anchors for an instance of each label: obtained with a beam width of 5 and a precision threshold of 95%
- Only instances correctly classified by the black-box model were chosen to explain

Unsatisfied/Neutral

On-board Service ≤ 4
Online Boarding ≤ 3
Baggage Handling ≤ 3
Check-in Service ≤ 4

Precision: 0.963

Coverage: 0.190

Satisfied

Baggage Handling ≤ 5
On-board Service > 4
Online Boarding > 2
Departure and Arrival Time Convenience ≤ 3
In-flight Wifi Service ≤ 2
Leg Room Service > 2

Precision: 0.988

Coverage: 0.044

- The rule for the Satisfied instance is more complex and has lower coverage
- This instance is much closer to the decision boundary of the model

Anchors

The default beam size of 1 leads to bad results. For the same instance, a larger value for this parameter leads to much better coverage results, while keeping the precision.

Beam width = 1

In-flight Wifi Service ≤ 3
In-flight Entertainment ≤ 2
Online Boarding ≤ 3
Check-in Service ≤ 3
Ease of Online Booking ≤ 2

Precision: 0.955

Coverage: 0.062

Beam width = 5

On-board Service ≤ 4
Online Boarding ≤ 3
Baggage Handling ≤ 3
Check-in Service ≤ 4

Precision: 0.963

Coverage: 0.190

- Better anchors obtained at the cost of computing time
- Further increases to the beam width would continue to improve the results