Term Project Report: The classification of Airline Passenger Satisfaction

CSIS3290-003 Fundamental of Machine Learning

Instructor:

[redacted]

Submitted By:

[redacted]

Natibundit Suntraranusorn

[redacted]

Date:

December 6, 2023

Table of Contents

1.Introduction and discovery

I. Introducing the business domain:

The dataset under analysis associates with customer satisfaction scores from over 120,000 airline passengers, which includes additional information about each passenger, details about their flight, and the type of travel, along with their evaluations of various factors such as cleanliness, seat comfort, regular services and extra services, and the overall travel experience.

II. Framing the Problem:

Understanding the specific satisfactory factors of passengers can drive customer satisfaction in the airline industry is crucial to enhance the overall travel experience, strategically allocate resources, and gain a competitive advantage. Additionally, recognizing patterns within different types of travel (business vs. personal) enables the tailored marketing strategy and contributing to sustained success in a competitive market. These are the questions to address:

- What are the key factors that contribute to customer satisfaction in the airline industry?
- Which specific factors (cleanliness, comfort, service, etc.) are the most significant in determining overall satisfaction?
- How do different aspects of the flight, such as duration, class, and type of travel, influence passenger satisfaction?
- Can we identify patterns indicating specific preferences within different types of travel. For example, the most consideration of satisfaction for business could be services and the most consideration of satisfaction for personal could be comfort.

III. Developing Initial Hypotheses:

- We hypothesize that cleanliness and comfort will be major contributors to overall satisfaction that impact on passenger experience.
- We anticipate the passengers' evaluations of service quality will strongly correlate with the overall satisfaction scores.
- We expect that the higher-class and shorter duration flights will lead to higher passenger satisfaction scores.

2.Data Preparation

I. Data inventory:

This comprehensive data set contains 24 columns and 129,881 records. The major columns include type of airline customer (first-time or returning), purpose of the flight (business or personal), type of passenger seat, flight distance, fight delay information, the sentiment of passengers in term of time, online booking, check-in service, seat comfort, cleanliness, in-flight services, food and drink, facilitation, baggage handling, and overall satisfaction. The dataset is downloaded from https://www.kaggle.com/datasets/mysarahmadbhat/airline-passenger-satisfaction?select=airline_passenger_satisfaction.csv

II. Data processing:

The summary statistics describe the ratings for various activities related to airline services. Each activity has been assessed by passengers, which includes the average satisfactory, the dispersion in a dataset, minimum, and maximum ratings. Here's a brief overview of the data:

No.	Activities	Mean	Std	Min	Max
1	Online-Booking	2.74	1.42	0	5
2	Check-in Service	3.32	1.25	1	5
3	Online Boarding	3.24	1.37	0	5
4	Gate Location	2.99	1.28	1	5
5	On-board Service	3.42	1.28	1	5
6	Seat Comfort	3.46	1.32	1	5
7	Leg Room Service	3.37	1.32	0	5
8	Cleanliness	3.30	1.30	1	5
9	Food and Drink	3.20	1.33	0	5
10	In-flight Service	3.65	1.17	1	5
11	In-flight Wifi Service	2.70	1.34	0	5
12	In-flight Entertainment	3.37	1.34	1	5
13	Baggage Handling	3.67	1.17	1	5

Data Preparation

- Record Selection: due to dataset containing over 100,000 transactions that could affect the long-time loading process. Therefore, it is necessary to downsize the dataset strategically. This will involve a random selection process aimed at maintaining the same proportions as the overall dataset. Specifically, 2,000 transactions will be randomly selected based on the gender column, preserving the original gender distribution (51% female and 49% male). Additionally, a similar approach will be applied to the Class column, with 2,000 transactions randomly chosen while maintaining the original distribution between business and economy classes (50% each). Lastly, the sentiment column will undergo a similar process, with 2,000 transactions randomly selected to mirror the initial satisfaction proportions (57% dissatisfaction and 43% satisfaction). The selection process ensures that the downsized dataset accurately reflects the diversity and distribution patterns present in the overall dataset.
- Create dummy variables: to convert categorical values into numerical format, making them compatible with machine learning models that require numerical input.
- Handle missing value and duplicates: for ensuring data quality and preventing biased or inaccurate model training.
- Apply scaling with Robust Scaler: to standardize the range of numerical features and ensuring that all features contribute equally to the model's training process.
- Remove outliers: to helps improve the performance of the models that are sensitive to extreme values.

Data Exploration Analysis

- Creating the heatmap correlation matrix to observe the degree of relationship between independent features and the target column. These are high correlation that value greater than 0.3 or -0.3.

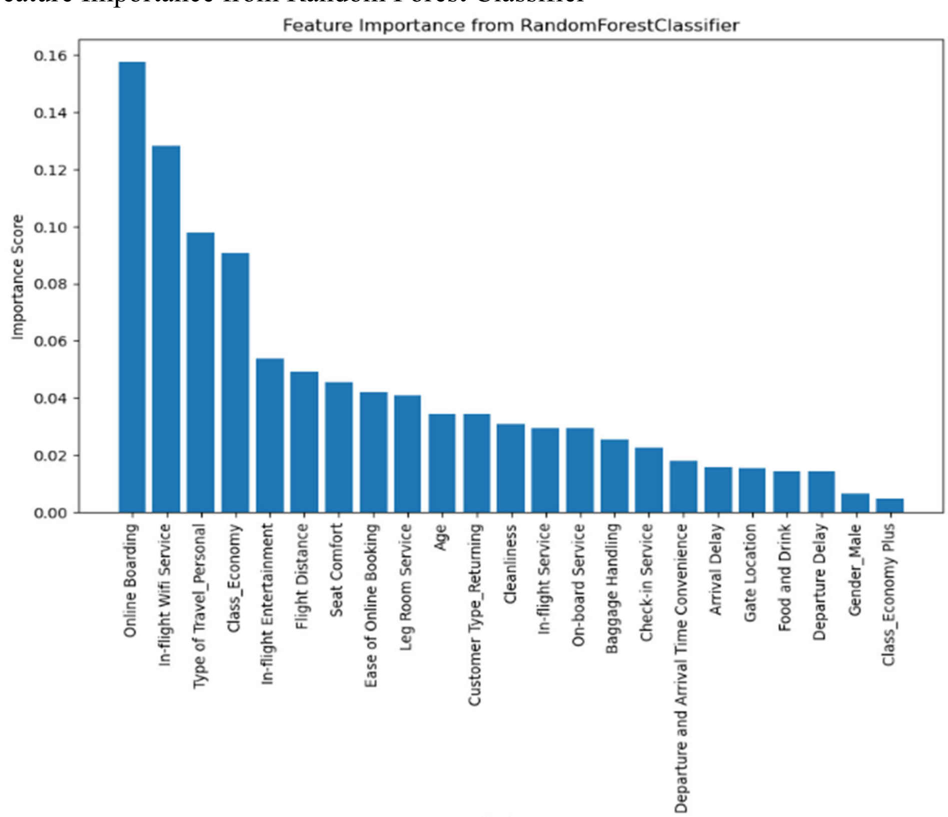
Independent Features	Correlation
Flight Distance	0.33
Online Boarding	0.49
Seat Comfort	0.33
Leg Room Service	0.32

Cleanliness	0.31
In-flight Entertainment	0.39
Type of Travel	-0.45
Class	-0.48

- Features selection: after testing 5 different methods of feature selection. Method 5 is selected because of giving the highest testing accuracy, related to the high correlation tables and minimizes the number of columns that effectively save the resources. These are the results from testing 5 different methods of feature selection.

Method No	Feature selection method	Number of Column Selection	Testing Accuracy for Logistic model
Method 1	Variance threshold	16	0.830
Method 2	SelectFromModel and Logistic Regression	7	0.859
Method 3	Generalized Linear Model (Binomial) from stats models	19	0.870
Method 4	Recursive Feature Elimination (RFE)	10	0.872
Method 5	SelectFromModel and Random Forest classifier	7	0.907
Maximum value		24	1

- Feature Importance from Random Forest Classifier



3.Model Planning and Implementation

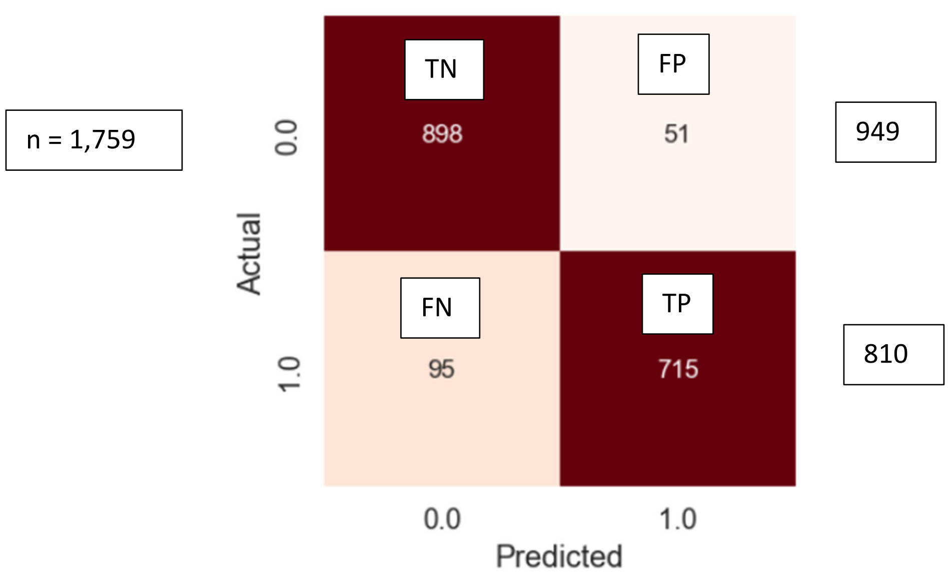
We explored several machine learning models to predict the target variable for selecting the best algorithm by using pipelines that facilitate the different models of experiment and hyperparameter settings to develop the classifiers. After evaluating various performance, the selected model is the Support Vector Machine with a radial basis function (RBF SVM). This choice is justified by its superior accuracy among the tested models, achieving an accuracy of 91.25% on the test set. These are the result from comparing different algorithms.

	Classifier	Accuracy
0	Logistic Regression	0.841956
1	Nearest Neighbors	0.899375
2	Linear SVM	0.848778
3	RBF SVM	0.912450
4	Decision Tree	0.905060
5	Naive Bayes	0.847641
6	Random Forest	0.903354

Hyperparameter tuning was performed by using grid search and cross-validation to identify the optimal combination of hyperparameters for the RBF SVM classifier. The best-performing hyperparameters were found to be {'C': 5, 'gamma': 0.95}, resulting in an improved accuracy of 92.03% on the training and validation sets and an improved accuracy of 91.7% on the test set.

4.Results Interpretation and Implications

Confusion Matrix allows us to assess the model's performance in terms of true positives, true negatives, false positives, and false negatives. This is a confusion matrix result.

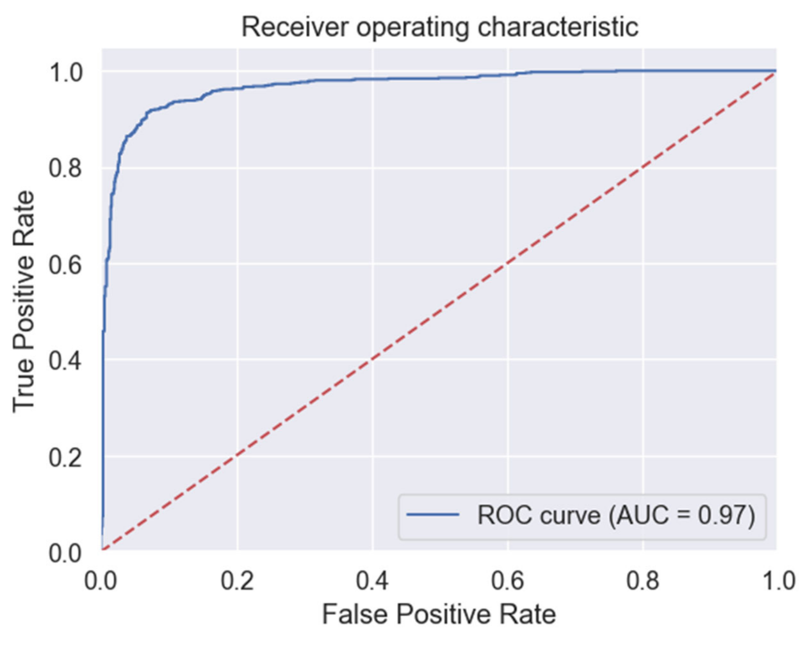


The accuracy of the tuned RBF SVM model on the test set is approximately 91.7%, indicating the proportion of correctly predictive.

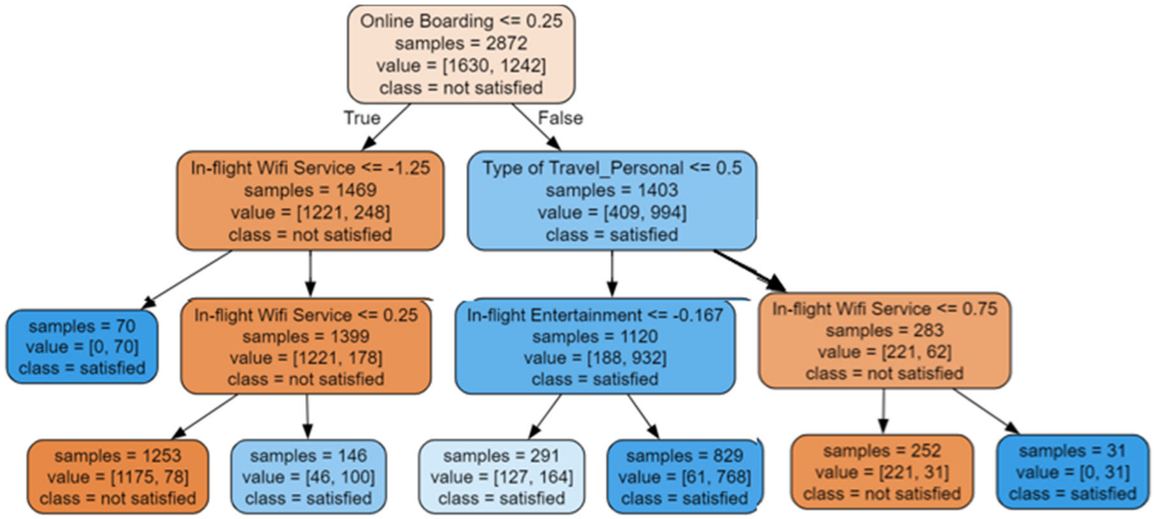
The classification report offers additional metrics, including precision, recall, and F1-score. Precision reflects the accuracy of positive predictions, recall represents the ability to capture all positive class, and the F1-score balances precision and recall. As a result of the classification report, high precision and high recall indicate a robust and reliable model. It implies that the model is effective in both minimizing false positives and capturing the majority of true positive cases. This is the result of the classification report.

	precision	recall	f1-score	support
0.0	0.90	0.95	0.92	949
1.0	0.93	0.88	0.91	810
accuracy			0.92	1759
macro avg	0.92	0.91	0.92	1759
weighted avg	0.92	0.92	0.92	1759

An AUC of 0.9679 is indicative of a highly accurate and well-performing model with strong discriminatory capabilities. It indicates that the model is reliable predictions and effectively distinguishing between positive and negative labels. This is the result of ROC diagram.



After we found the best classifier, we applied the decision tree model to understand the logic behind customer satisfaction based on the selected features. The tree has a depth of 3 level, these are the insight from decision tree structure:



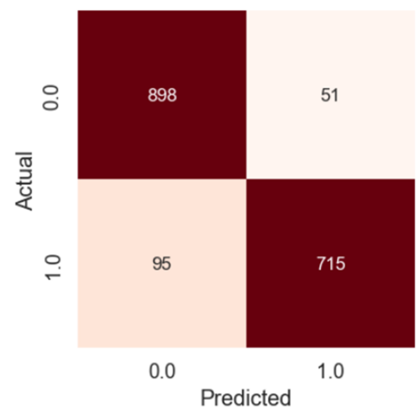
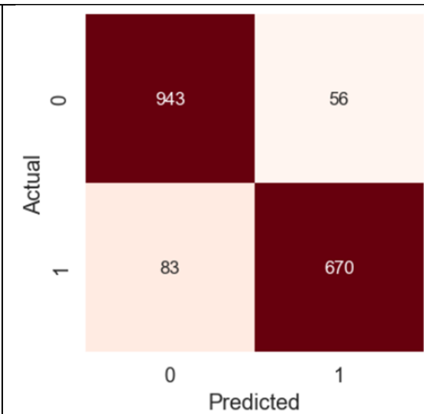
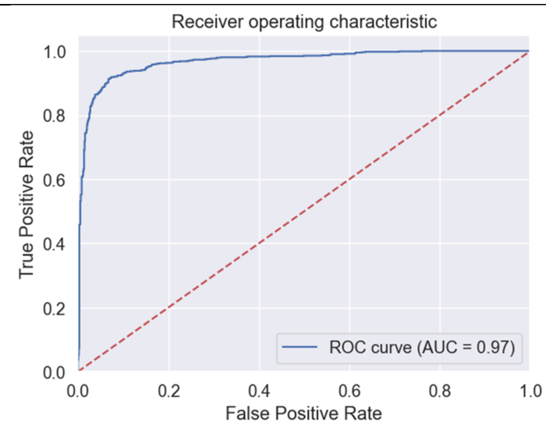
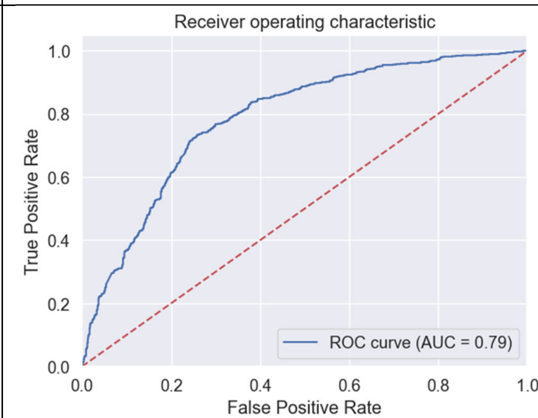
- "Online Boarding" indicates significance in predicting customer satisfaction. Customers with a low Online Boarding score (<= 0.25) are likely to be classified as "not satisfied."
- For customers with low Online Boarding satisfaction, the In-flight Wifi Service plays a crucial role. If this service is rated low (<= -1.25), the prediction is "not satisfied."
- If Online Boarding satisfaction is high (> 0.25), the Type of Travel becomes important. Personal travelers (<= 0.5) are predicted to be "satisfied."
- The further branches based on In-flight Entertainment and In-flight Wifi Service, capturing detailed patterns in customer satisfaction.

The insights gained from the tree structure can guide strategic decision-making, helping to prioritize areas to develop services such as Online Boarding, In-flight, In-flight Entertainment and Wifi Service to enhance customer satisfaction. The model's high accuracy on the test set will benefit its predictive power in the long-term.

5.Out-of-sample Predictions

A new dataset was created to simulate out-of-sample predictions, consisting of 6,000 records. The selection of records was based on previous criteria (gender, class, and satisfaction) but different number of random statement. After combining and shuffling the samples, the resulting dataset had 24 columns, with 5 categorical features and 19 numerical features. Applying the same steps of data preprocessing and pipeline models with a new dataset. These are the compare result between the previous dataset and out-of-sample dataset.

Indicators	Previous Dataset	Out-of-sample Dataset
Accuracy	91.7%	92.1%

<table>
<tr><td>Confusion Matrix</td><td>
</td><td>
</td></tr>
<tr><td>Classification Report</td><td><table>
<tr><td></td><td>precision</td><td>recall</td><td>f1-score</td><td>support</td></tr>
<tr><td>0.0</td><td>0.90</td><td>0.95</td><td>0.92</td><td>949</td></tr>
<tr><td>1.0</td><td>0.93</td><td>0.88</td><td>0.91</td><td>810</td></tr>
<tr><td>accuracy</td><td></td><td></td><td>0.92</td><td>1759</td></tr>
<tr><td>macro avg</td><td>0.92</td><td>0.91</td><td>0.92</td><td>1759</td></tr>
<tr><td>weighted avg</td><td>0.92</td><td>0.92</td><td>0.92</td><td>1759</td></tr>
</table></td><td><table>
<tr><td></td><td>precision</td><td>recall</td><td>f1-score</td><td>support</td></tr>
<tr><td>0</td><td>0.92</td><td>0.94</td><td>0.93</td><td>999</td></tr>
<tr><td>1</td><td>0.92</td><td>0.89</td><td>0.91</td><td>753</td></tr>
<tr><td>accuracy</td><td></td><td></td><td>0.92</td><td>1752</td></tr>
<tr><td>macro avg</td><td>0.92</td><td>0.92</td><td>0.92</td><td>1752</td></tr>
<tr><td>weighted avg</td><td>0.92</td><td>0.92</td><td>0.92</td><td>1752</td></tr>
</table></td></tr>
<tr><td>ROC</td><td>
</td><td>
</td></tr>
</table>

The overall evaluation, the out-of-sample dataset shows slightly improved accuracy. This suggests that the model generalizes well to unseen data and maintains high predictive performance. Slightly different results for confusion matrix and classification report indicate that model remained consistent across both datasets and reliable prediction. However, AUC noticeably decreased. It might imply a reduction in the model's discriminatory capability.

6.Concluding Remarks

In conclusion, our exploration into airline passenger satisfaction classification has discovered actionable insights for the airline industry. The model's impressive accuracy on the test set with its robust performance on out-of-sample data, signifies its real-world potential. The decision tree analysis identifies specific improvement areas such as "Online Boarding," "In-flight Entertainment," and "Wifi Service," empowering airlines to strategically enhance services and elevate overall travel experiences.

The RBF SVM model has tuned for optimal performance, not only demonstrates immediate effectiveness but also sustained predictive power. With a high AUC of 0.9679, the model's strong discriminatory capabilities make it a valuable asset for targeted interventions to distinguish between satisfied and dissatisfied passengers. As the airline industry evolves, continuous data-driven insights and further research into traveler preferences can ensure the model remains current and adaptable in the dynamic landscape of the airline industry.