AGH UNIVERITY OF SCIENCE AND TECHNOLOGY
FACULTY OF APPLIED MATHEMATICS

The Selection of a Model for Airlines Customer Satisfaction

Joanna Krężel, Anna Matysek, Piotr Mikler, Adam Szczerba

12/29/2021

Abstract

The project aims to analyze the data about satisfaction of Investico Airlines passengers. Having customer-granular observations about the cruise and a reported satisfaction level for particular aspects of the flight we try to fit models which predict whether they are satisfied with the service or not. The binary classification models used during the project are {} {} {}, out of which our recommendation is {} based on {}. This document describes the process we undertook and presents the results of data preprocessing, model selection and model validation.

Contents

1 Introduction

1.1 Introduction

2 The Data

The data is downloaded from www.kaggle.com and delivered by an airline organization. The dataset consists of the details of customers who have already flown with them. The feedback of the customers on various context and their flight data has been consolidated. The main purpose of this dataset is to predict whether a future customer would be satisfied with their service given the details of the other parameters values. Also the airlines need to know on which aspect of the services offered by them have to be emphasized more to generate more satisfied customers. The data consists of 129880 rows and 23 columns.

Below we list all column names with explanations of the variables' meaning. For categorical variables in the dataset describing satisfaction level, 0 means *Not Applicable*.

Feature	Description	Values
Satisfaction	Airline satisfaction level	satisfied/dissatisfied
Gender	Gender of the passenger	male/female
Customer type	The customer type	loyal / disloyal customer
Age	The age of a passenger	[7; 85] years
Type of travel	Purpose of the flight	personal / business travel
Class	Travel class in the plane	business / eco / eco plus
Flight distance	The flight distance of the journey	[50; 6951] miles
Seat comfort	Satisfaction level of seat comfort	{0-5}
Departure/arrival	Satisfaction level of departure/arrival time	{0-5}
Food and drink	Satisfaction level of food and drink	{0-5}
Gate location	Satisfaction level of gate location	{0-5}

Feature	Description	Values
Inflight WiFi service	Satisfaction level of the inflight wifi service	{0-5}
Inflight entertainment	Satisfaction level of inflight entertainment	{0-5}
Online support	Satisfaction level of online support	{0-5}
Ease of online booking	Satisfaction level of online booking	{0-5}
On-board services	Satisfaction level of on-board service	{0-5}
Leg room service	Satisfaction level of leg room service	{0-5}
Baggage handling	Satisfaction level of baggage handling	{0-5}
Checkin service	Satisfaction level of check-in service	{0-5}
Cleanliness	Satisfaction level of cleanliness	{0-5}
Online boarding	Satisfaction level of online boarding	{0-5}
Departure delay in minutes	Delay upon departure	[0; 1592] minutes
Arrival delay in minutes	Delay upon arrival	[0; 1584] minutes

2.1 Data Preprocessing

As both academia and business point out, the data-related operations typically constitute about 80 of the whole effort of a modeling pipeline. The performance of any model is heavily driven by the quality of it's inputs. It can be easily proven in a simple trial by combat that even a suboptimal model running on high quality data can oftentimes bring a sophisticated one with poor inputs to it's knees.

For that reason it is of utmost importance to pay extra care and attention to the data which is fed to the decision making models. As the first step in our modeling pipeline we are going to look at the dataset to gain additional

insight about it's statistics and information it conveys. We are going to perform quality checks on the data, such as outlier detection and treatment of not-available values (*NA*s). Lastly in this pre-modeling step we will look at the actual *information* carried by particular features to remove variables of low added value.

2.1.1 Exploratory Data Analysis

Overview like histograms & counts of the data. To be done...

2.1.2 Feature Selection

The more is not always the better. Every model has a certain computational complexity that increases with the number of additional explanatory variables. The feature selection in a pre-modeling environment serves identifying groups of variables which carry repeated or very similar informational value. *Filtering feature selection methods* allow one to discard redundant features in a model independent way. By reducing the number of variables they simplify the model and increase it's interpretability. It is also a step which tackles multicollinearity (high linear codependency of explanatory variables) which kills stability and predictive power of some models.

2.2 Categorical Variables

2.2.1 The information Value

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

2.3 The Continous Variables

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu

fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

2.3.1 Decide which Continuous Variable to Use

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

2.4 Data Binning

2.4.1 The Categorical Variables

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

2.4.2 The Continuous variables

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

3 The Logisic Regresssion

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu

fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

4 The performance of the Model

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

5 Validation of the Model

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

5.1 Monte Carlo Cross Validation

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

6 The Challenger Models

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu

fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

6.1 Neural Network

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

6.2 Another logistic regression: logistic 2

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

7 Conclusion

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

8 Bibliography