AIRBNB for AI PREDICTION

Presented by Mostafa Ayman , Seif Ahmed , Abdullah Hossam

Your text here

Your text here

Contents

[Introduction to AI project 1](#_Toc153113705)

[Introduction to Airbnb 1](#_Toc153113706)

[Dataset attributes 1](#_Toc153113707)

[Project phases: 2](#_Toc153113708)

[First: data analysis 3](#_Toc153113709)

[Distribution 3](#_Toc153113710)

[Feature statistics 4](#_Toc153113711)

[Correlation 4](#_Toc153113712)

[Scatter plot 5](#_Toc153113713)

[Free viz 5](#_Toc153113714)

[Rank 6](#_Toc153113715)

[Second: data modeling 7](#_Toc153113716)

[Impute 7](#_Toc153113717)

[Preprocess 8](#_Toc153113718)

[Outliers 8](#_Toc153113719)

[Data sampler 9](#_Toc153113720)

[Test and score 10](#_Toc153113721)

[Confusion matrix 10](#_Toc153113722)

[Prediction 11](#_Toc153113723)

[Third: data prediction 12](#_Toc153113724)

[File 12](#_Toc153113725)

[Image processing 13](#_Toc153113726)

[import images (1) 13](#_Toc153113727)

[Test and score 13](#_Toc153113728)

[Import image (2) 14](#_Toc153113729)

[Image gride 14](#_Toc153113730)

[References: 1](#_Toc153113731)

Figure 1 (distribution) 3

Figure 2 (Feature statistics) 4

Figure 3 (Correlation) 4

Figure 4 (Scatter plot) 5

Figure 5 (Free viz) 5

Figure 6 (Rank) 6

Figure 7 (Impute) 7

Figure 8 (preproccess) 8

Figure 9 (Outliers) 8

Figure 10 (Data sampler) 9

Figure 11 (Test and score) 10

Figure 12 (Confusion matrix) 10

Figure 13 (Prediction) 11

Figure 14 (File) 12

Figure 15 (Import image 1) 13

Figure 16 (Test and score) 13

Figure 17 (Import image 2) 14

Figure 18 (Image gride) 14

# Introduction to AI project

Artificial Intelligence is a method of making a computer or a software think intelligently like the human mind. AI is accomplished by studying the patterns of the human brain and by analyzing the cognitive process.

# Introduction to Airbnb

Airbnb (ABNB) is an online marketplace that connects people who want to rent out their homes with people looking for accommodations.

*How Airbnb Works*

It matches local people with a spare room or entire home to rent to others. Airbnb can provide a place to stay often at a lower price than hotels charge.

*History of Airbnb*

In 2008, Brian Chesky (the current CEO), Nathan Blecharczyk, and Joe Gebbia, established the company now known as Airbnb. The idea blossomed after two of the founders started renting air mattresses in their San Francisco home to conference visitors.

# Dataset attributes

This dataset contains 1999 instances, 24 features like:

* Host names.
* Room type which consists of private room and entire room/apt.
* Cancellation which consists of strict, moderate and flexible.
* Price.
* Service fee.

The target is cancelation policy (flexible, moderate, strict).



# Project phases:

**Project phase 1**

# First: data analysis

## Distribution

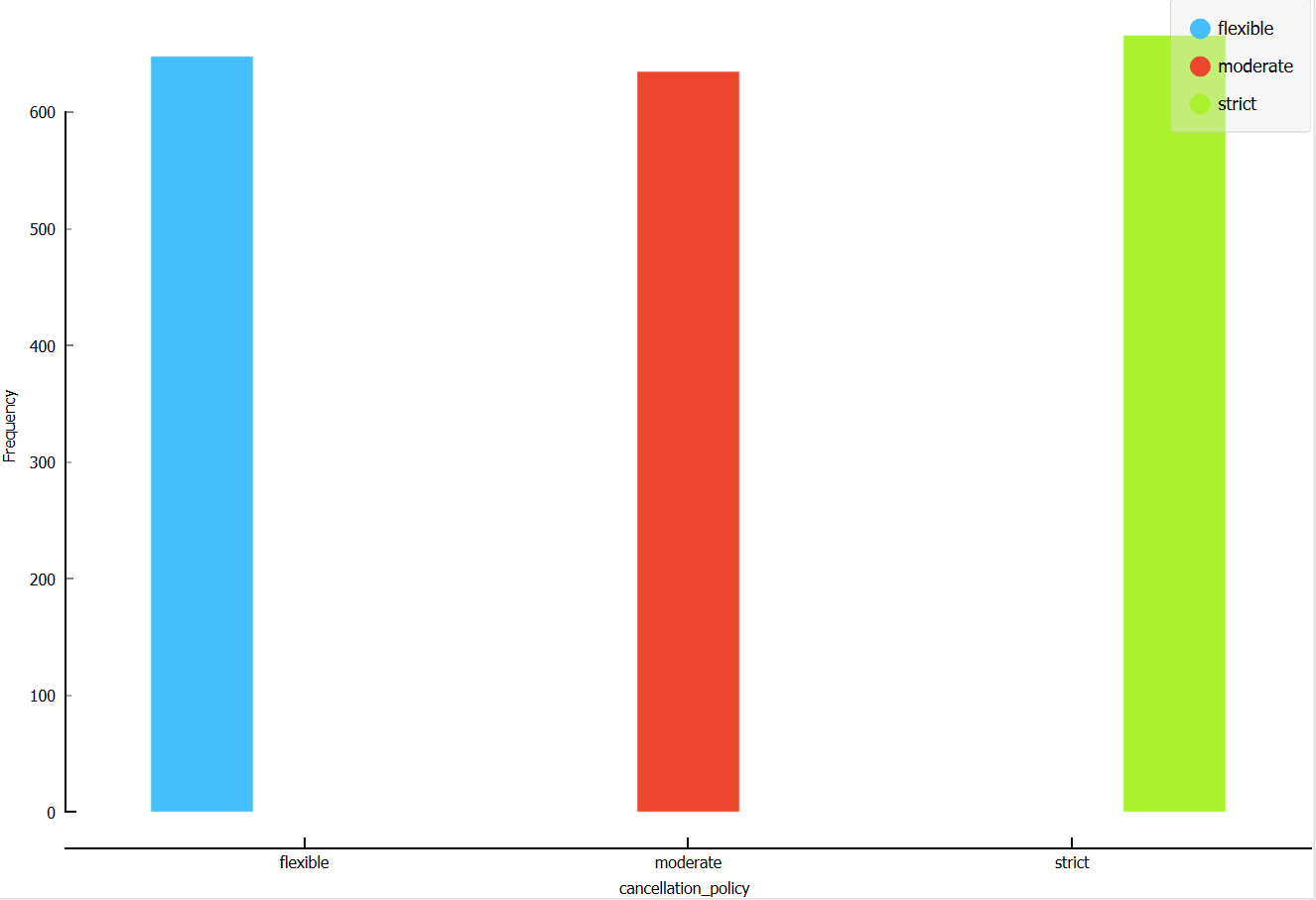


Figure (distribution)

*Distribution widget:* represents features in charts used to check if there is balance in target or not. In the dataset there is a balance.

## Feature statistics



Figure (Feature statistics)

*Feature statistics widget*: shows statistics for each feature like: max, min. In this dataset maximum of availability is (540). and the minimum of availability is (-10)

## Correlation

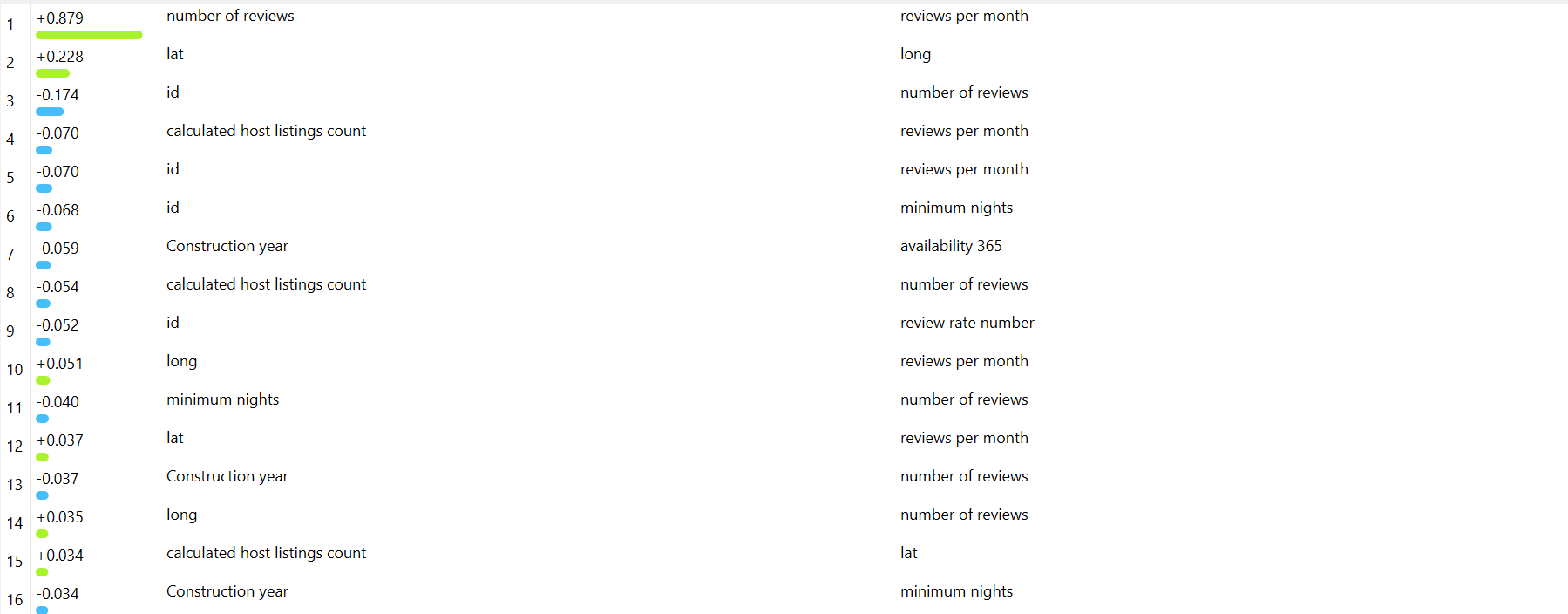


Figure (Correlation)

*Correlation widget*: used to determine the relation between features. In this dataset there is a positive relation between number of reviews and reviews per month. There is negative relation between: Id and number of reviews.

## Scatter plot

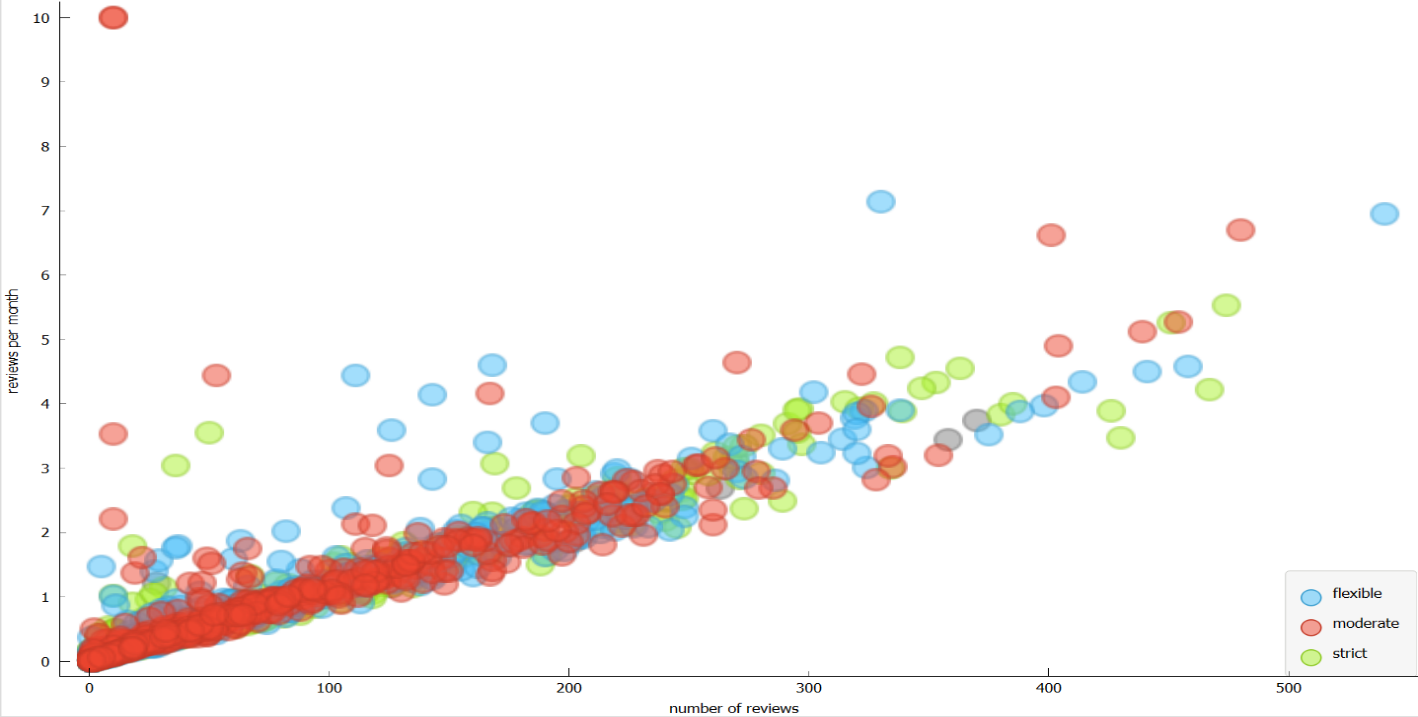


Figure (Scatter plot)

*Scatter plot widget*: displays the distributions of any numeric features on x-y axis.

## Free viz

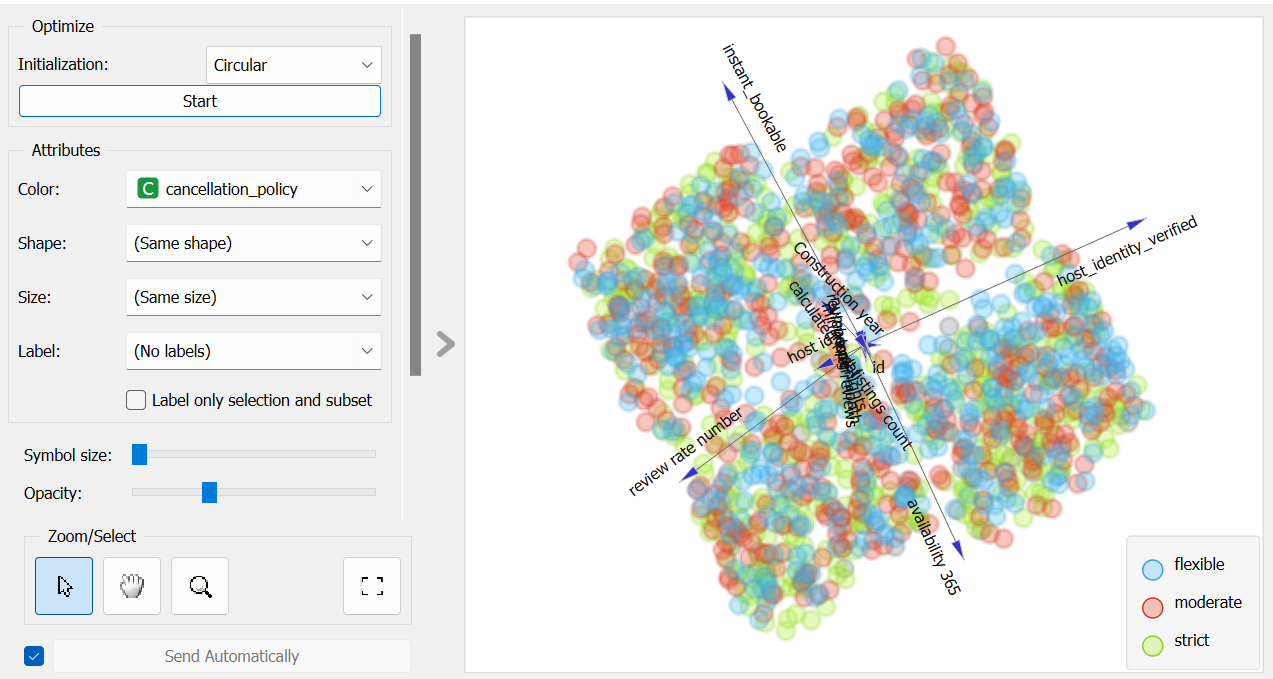


Figure (Free viz)

*Free viz widget*: display the most effective features on target graphically.

## Rank

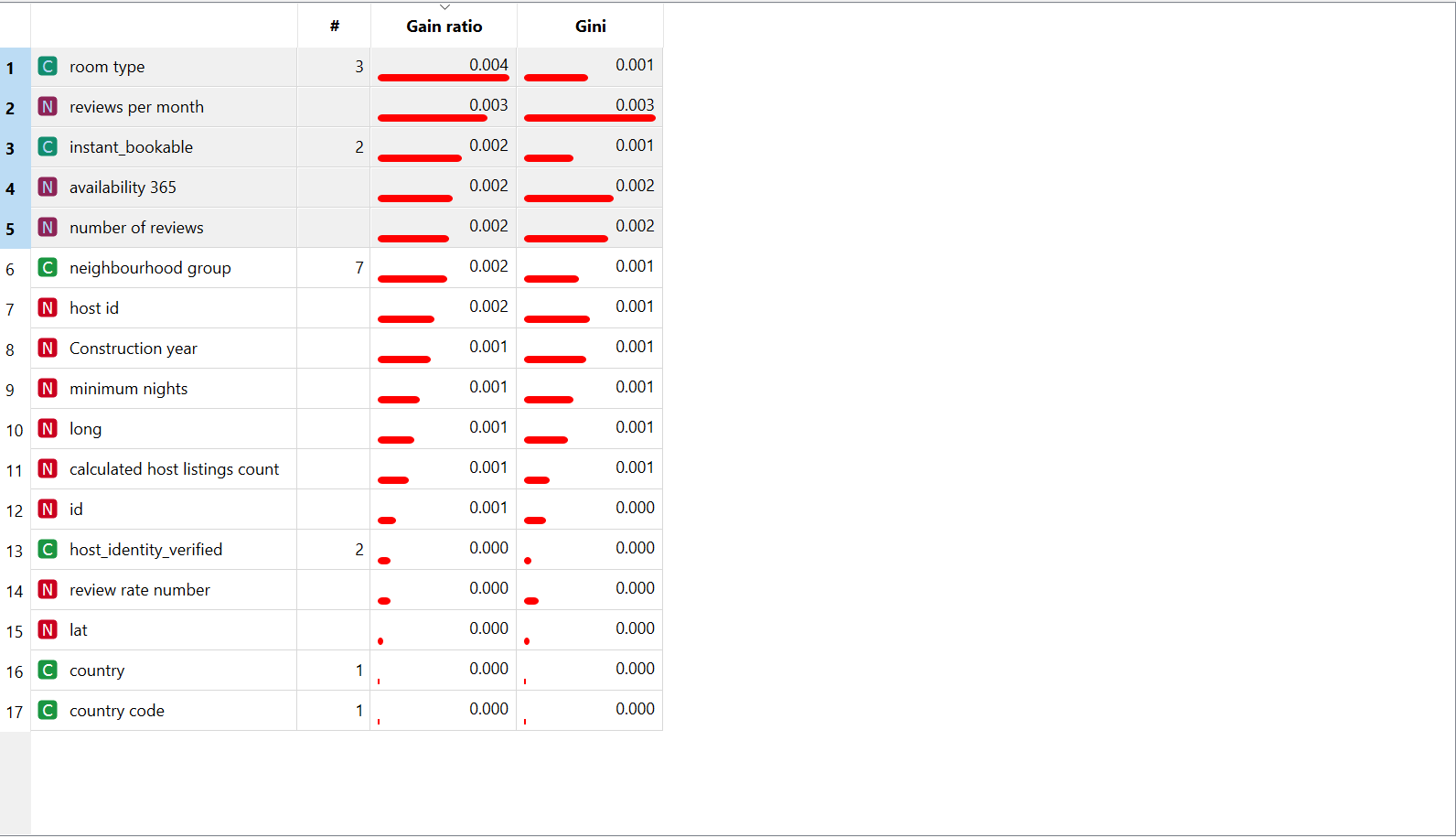


Figure (Rank)

*Rank widget*: display the most effective feature on target numerically.

In this dataset the most effective features are:

1. Room type
2. Reviews per month
3. Instant bookable
4. Availability 365
5. Number of reviews

# Second: data modeling

## Impute

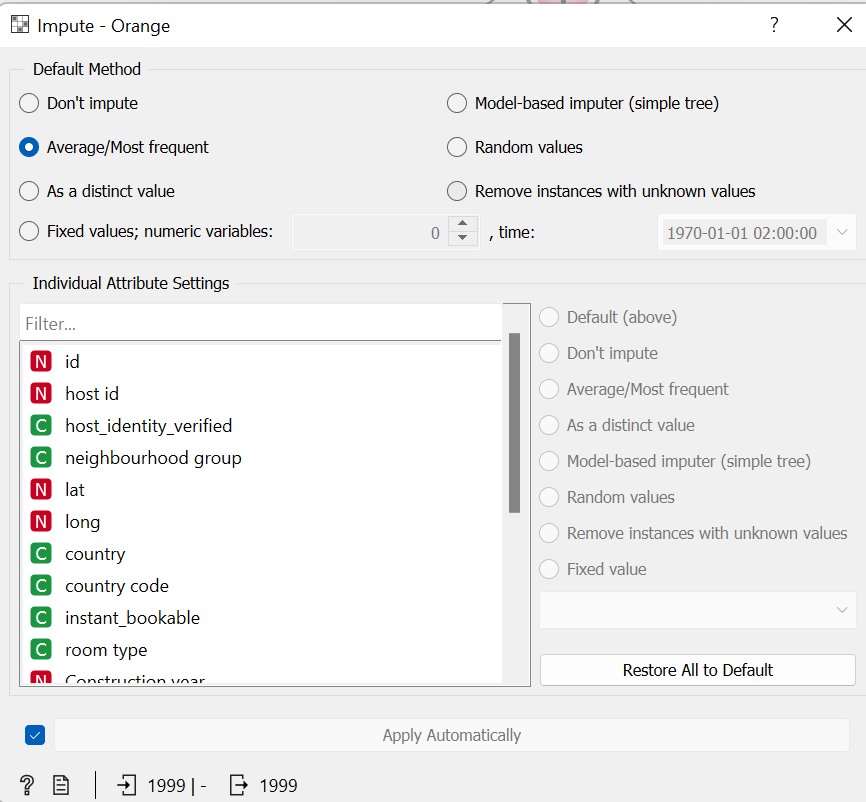


Figure (Impute)

*Impute widget*: substitute any missing value using several methods. This step will help not to lose any record. Method of imputing used in this dataset is the average/most frequent.

## Preprocess

Figure (preproccess)

*Preprocessing widget*: normalize any numeric field to range from 0 to 1. This step will increase the efficiency of analyzing the dataset and increase the accuracy of prediction.

## Outliers

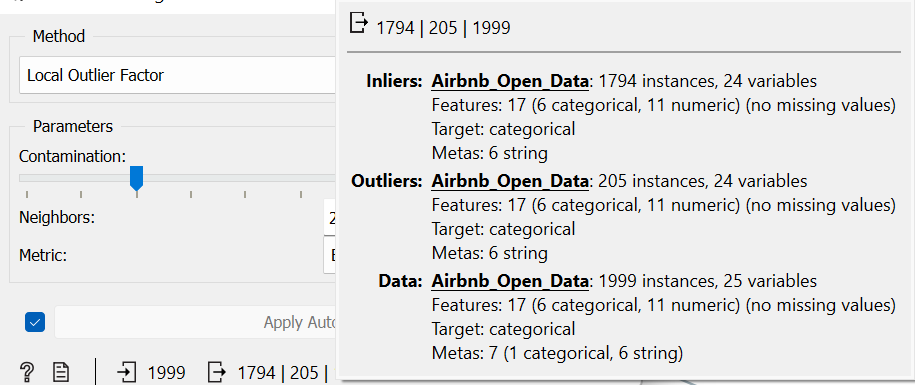


Figure (Outliers)

*Outliers' widget*: removes any abnormal records that will affect both training and prediction phases.

# 

## Data sampler

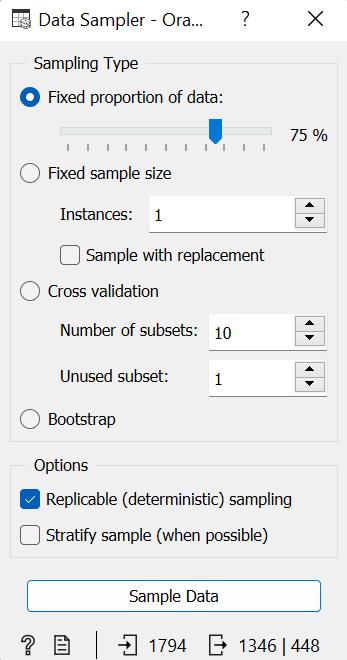
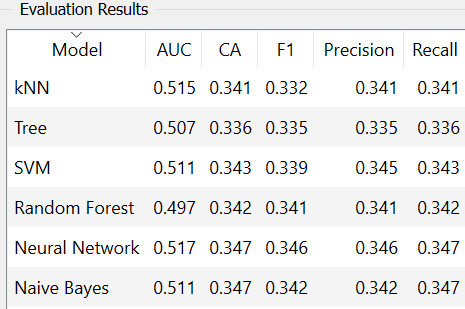


Figure (Data sampler)

*Data sampler widget*: divides the dataset into two proportions, sample and remaining. In this dataset (75%) will be used in test and score and the remaining (25%) will be used in prediction.

## Test and score



or

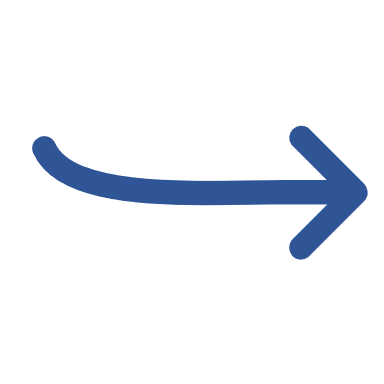


Figure (Test and score)

*Test and score widget*: used to check the training results. First, we check the **CA** for those models and take the highest. In this dataset neural network and naïve bayes models achieved the same ***CA***, so we checked the ***precision***. The neural network model has the highest **precision** that's why we are relying on this model in training.

## Confusion matrix

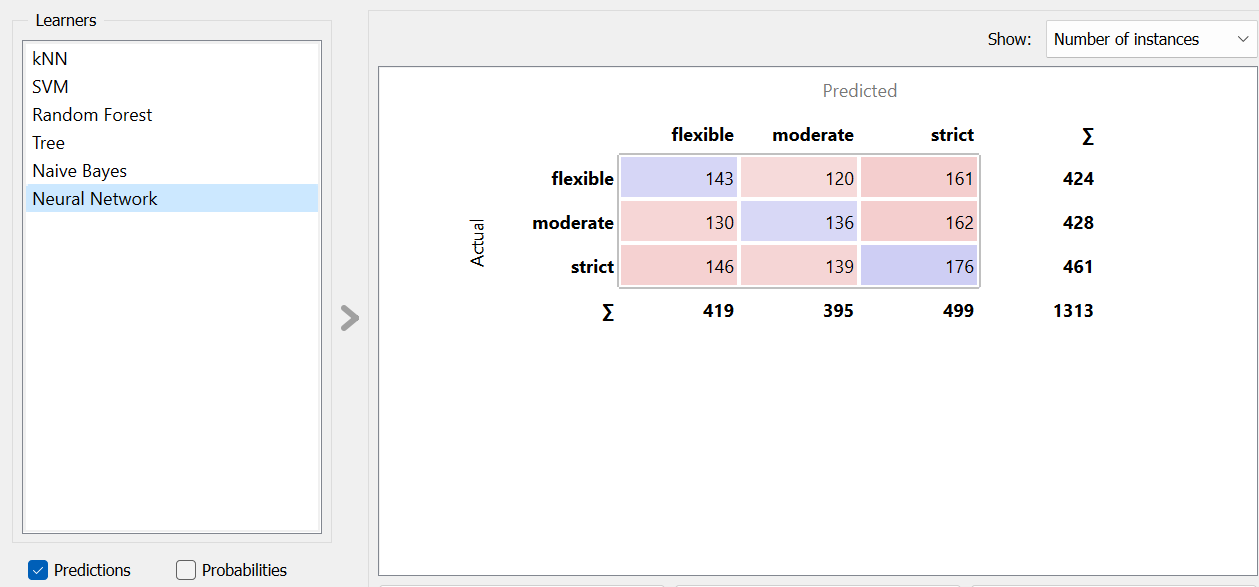


Figure (Confusion matrix)

*Confusion matrix widget*: states the confusion between the actual data and predicted data. It displays how many instances were classified and misclassified. In this dataset there are 143 instances were actually flexible while they were predicated by neural network model as 'flexible' this means that they are ***classified*.**  Also, there are 161 instances were actually flexible while they were predicated by neural network model as 'strict' this means that they are ***misclassified***.

## Prediction

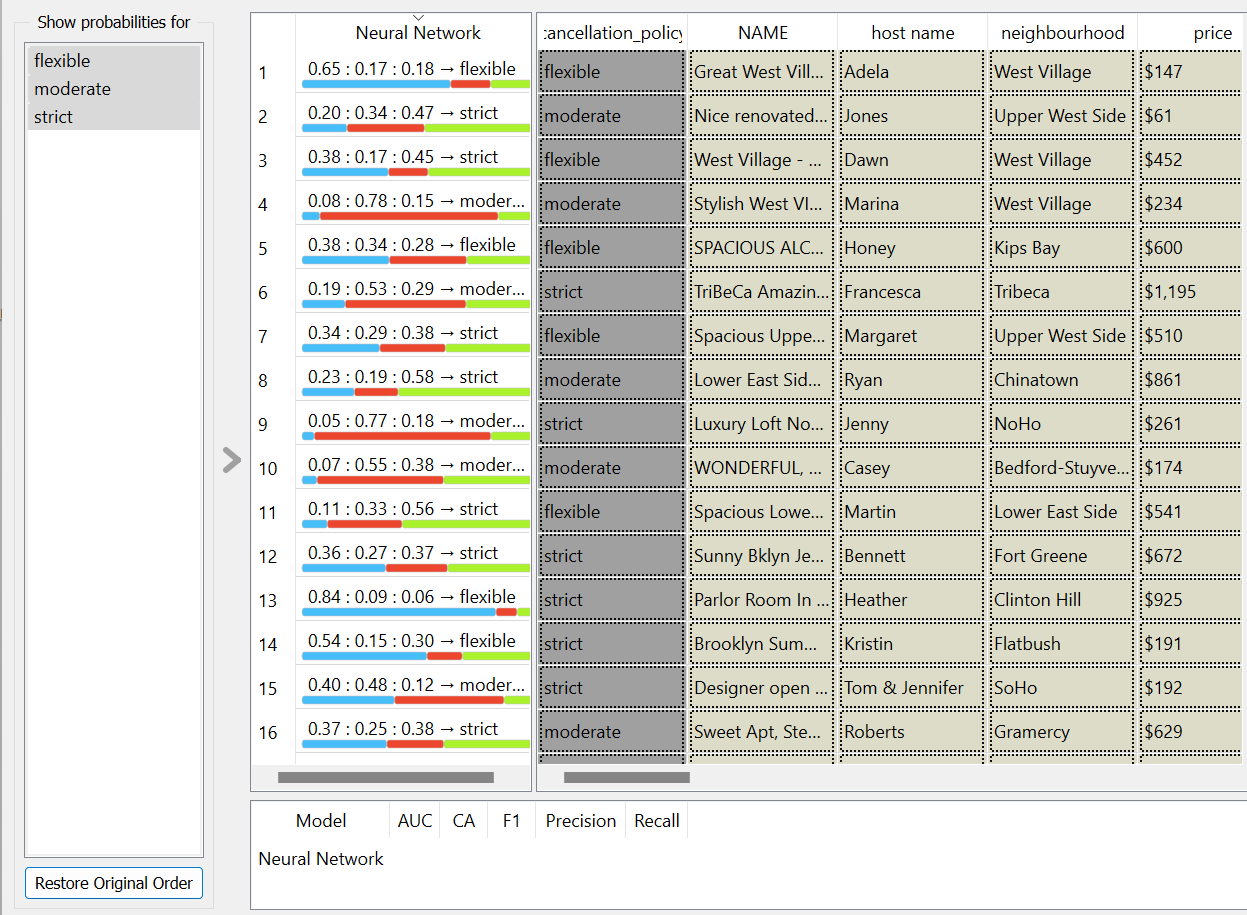


Figure (Prediction)

*Prediction widget:* used to predict the remaining instances whether the cancelation policy is flexible, moderate or strict. Prediction widget must be connected with data sampler widget and send only the remaining data

# Third: data prediction

## File

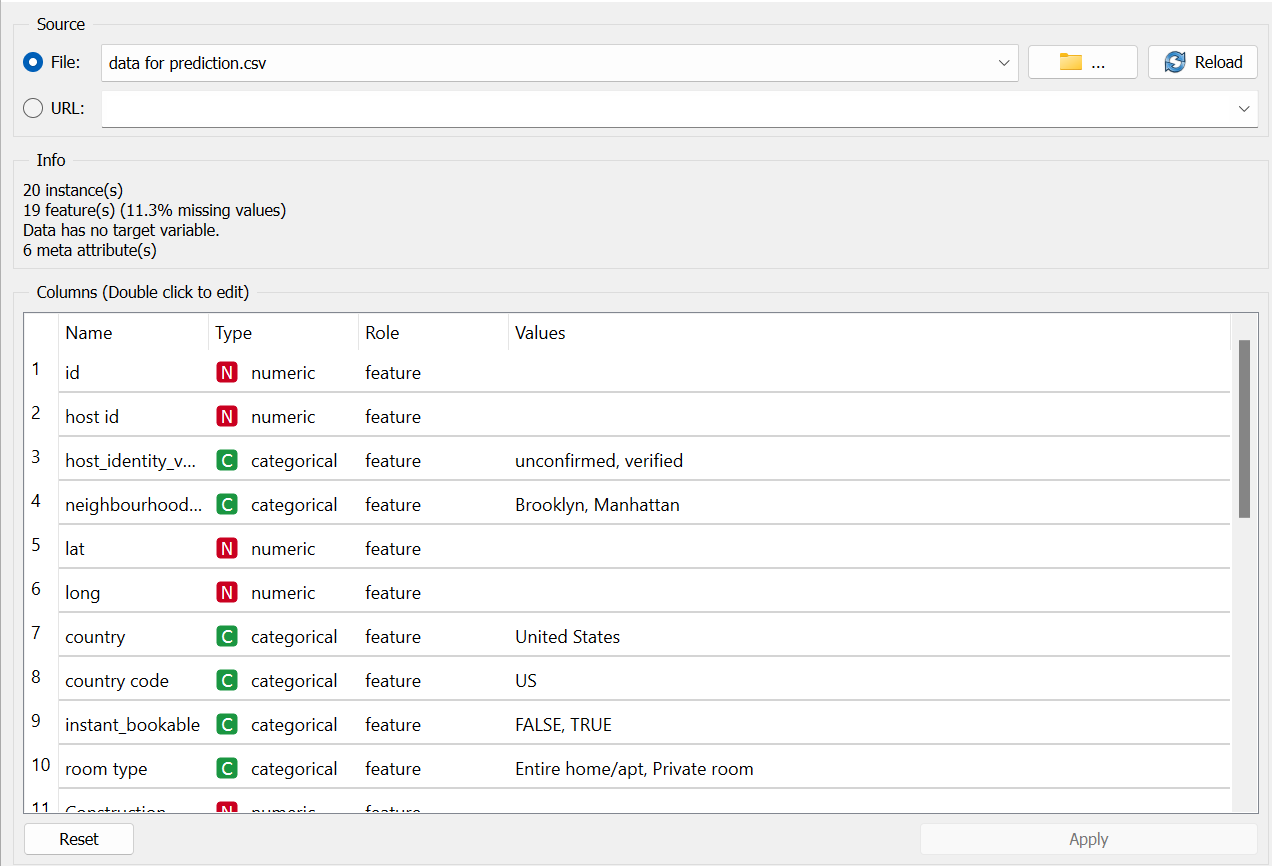


Figure (File)

*File widget:* in this widget we use different dataset without the target to predict the new dataset.

*Load model*: reload model that we save from data modeling.

*Prediction widget*: use to predict the new dataset that we removed the target from it.

**Project phase 2**

# Image processing

## import images (1)

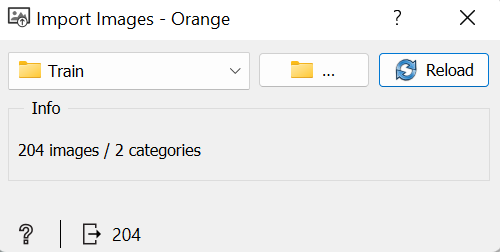


Figure (Import image 1)

*Import image widget*: in test and score phase we should categorize the images into two categories. In the image dataset there are two files: Ants and Bees

## Test and score

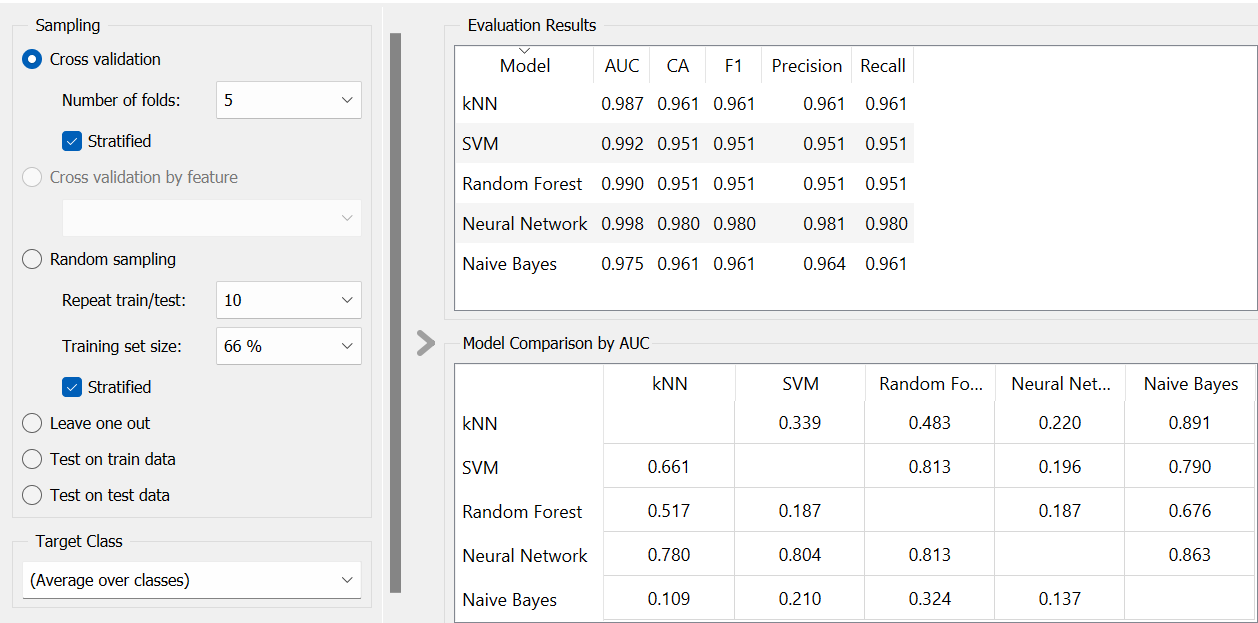


Figure (Test and score)

*Test and score widget*: used to check the training results. In this dataset neural network achieved the highest CA.

## Import image (2)

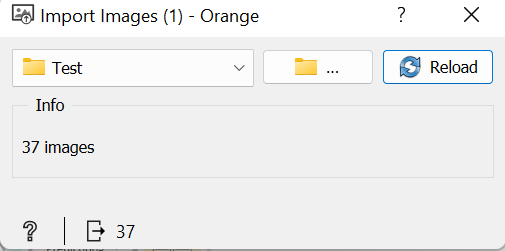


Figure (Import image 2)

*Import image widget*: in prediction phase we should get new pictures that don't exist in test and score and not to categorize them.

## Image gride

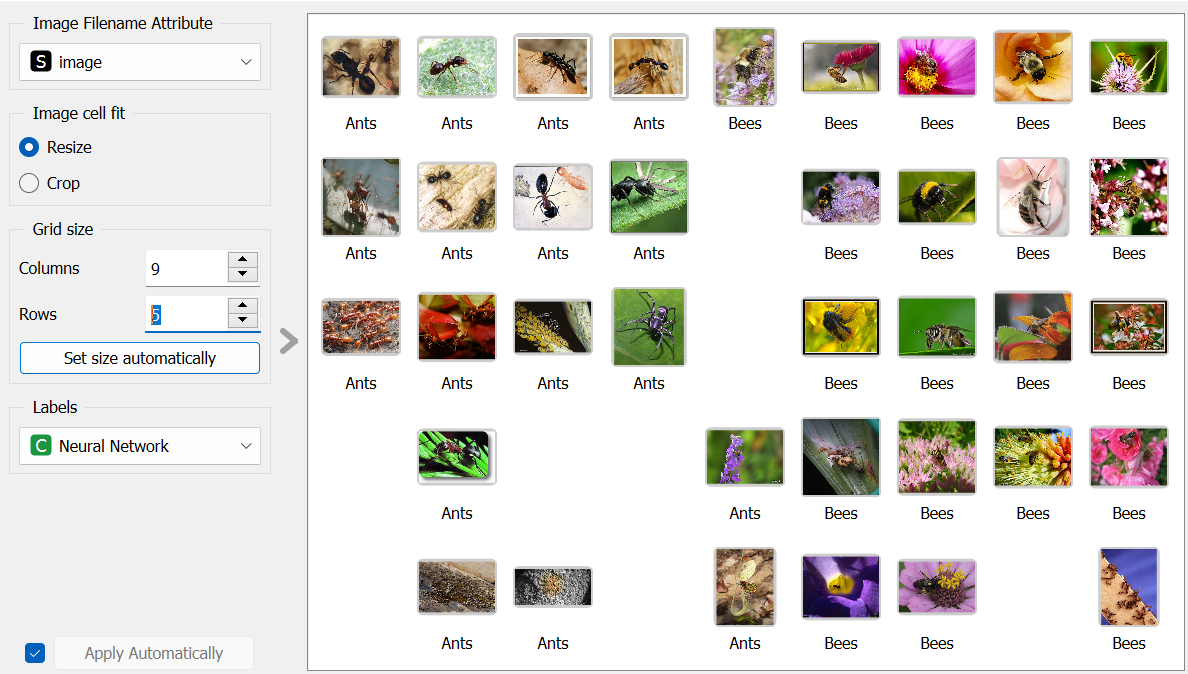


Figure (Image gride)

*Image gride widget:* used to show the prediction result as a photo and the name of the insect below it

# References:

Goldbloom, A. (2010, April ). *kaggle.com*. Retrieved from https://www.kaggle.com/datasets/arianazmoudeh/airbnbopendata

Kumar, K. (2010). *Simplilearn*. Retrieved from https://www.simplilearn.com/tutorials/artificial-intelligence-tutorial/what-is-artificial-intelligence#:~:text=Artificial%20Intelligence%20is%20a%20method,develops%20intelligent%20software%20and%20systems.

Silver, C. (1999). *Investopedia* . Retrieved from https://www.investopedia.com/articles/personal-finance/032814/pros-and-cons-using-airbnb.asp#toc-how-airbnb-works