MADRID AIRBNB PRICES

Camila Barbagallo,
Ryan Daher, Paula García
and Rocío González

SCHOOL OF
HUMAN SCIENCES
& TECHNOLOGY

TABLE OF CONTENTS

BUSINESS UNDERSTANDING 01

airbnb

ABOUT AIRBNB

- Began in 2008
- Two designers who had space to share hosted three travelers looking for a place to stay.
- List spaces and book unique accommodations anywhere in the world.
- Make sharing easy, enjoyable, and safe.

AirBed&Breakfast

Book rooms with locals, rather than hotels.

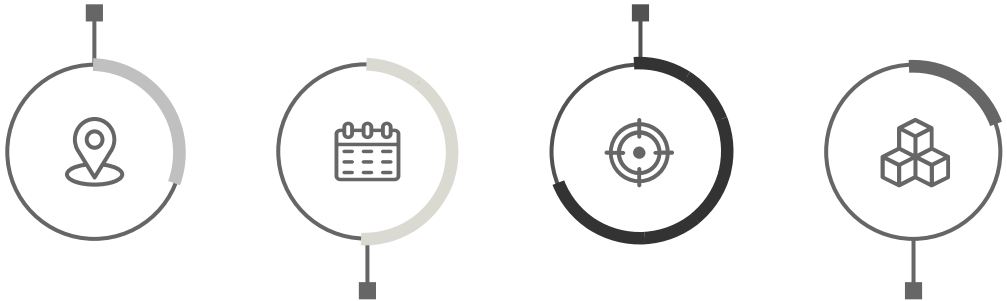
AIRBNB PRICE IMPACTORS

LOCATION

Areas closer to tourist hot-spots, or in prime locations take higher rent.

HOUSE DETAILS

Number of Rooms and Bathrooms and Size have direct impacts on the price.



DATE

Bookings closer to events or holidays tend to be overbooked and cause price spikes.

AMENITIES

Certain additional amenities increase the price.

ADDITIONAL PRICE INFLUENCERS

SuperHosts

SuperHosts may be less likely to post smaller, less expensive homes.

Overall Availability

Places that are booked often, may push the landlord to increase prices to reach optimum.

Guest Accommodations

Amount of guests accommodatable, and the addition of any extra guests.

Positive Reviews

Better reviews may encourage a landlord to boost prices.

02 DATA UNDERSTANDING

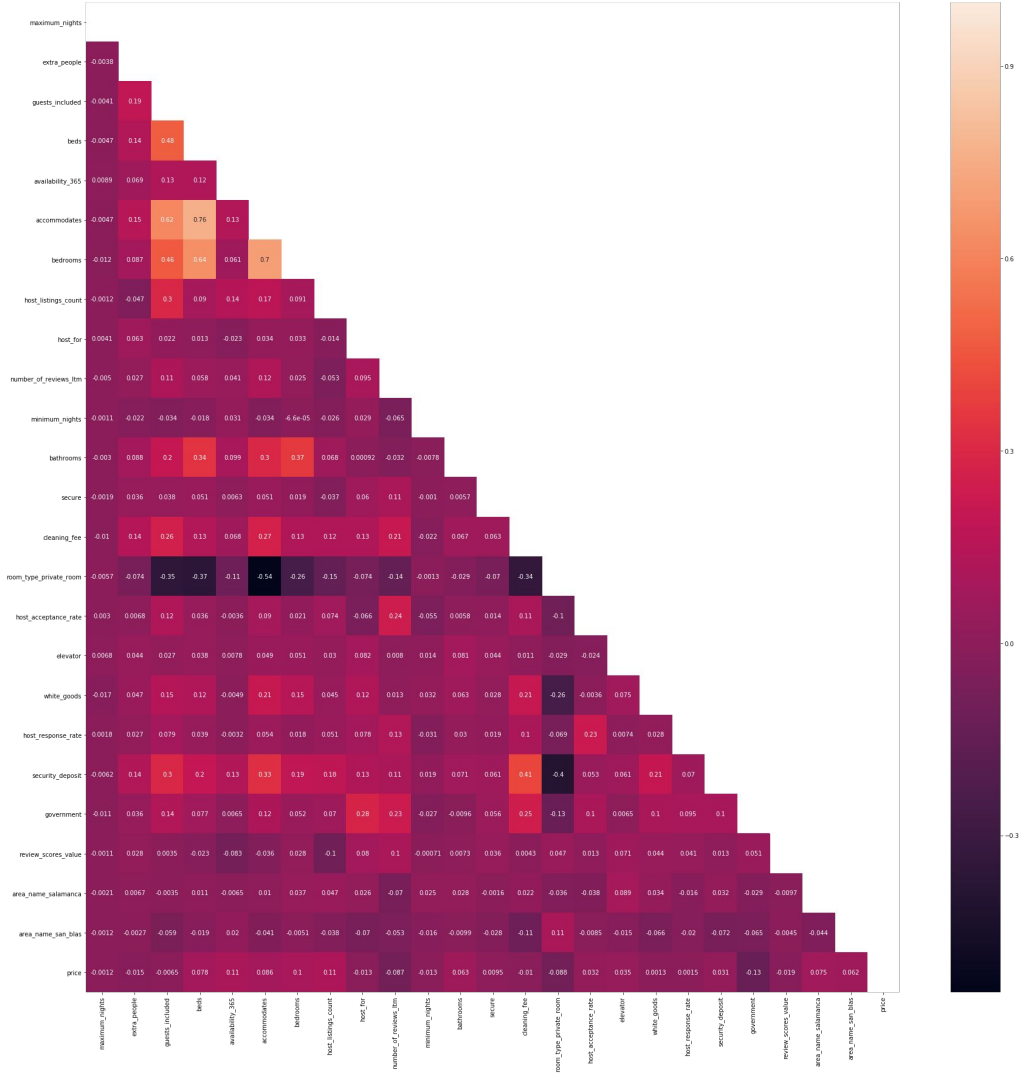
SCHOOL OF HUMAN SCIENCES & TECHNOLOGY

airbnb

DATASET OVERVIEW

Dataset Obtained from the airbnb website

107	21,845
Variables	Observations



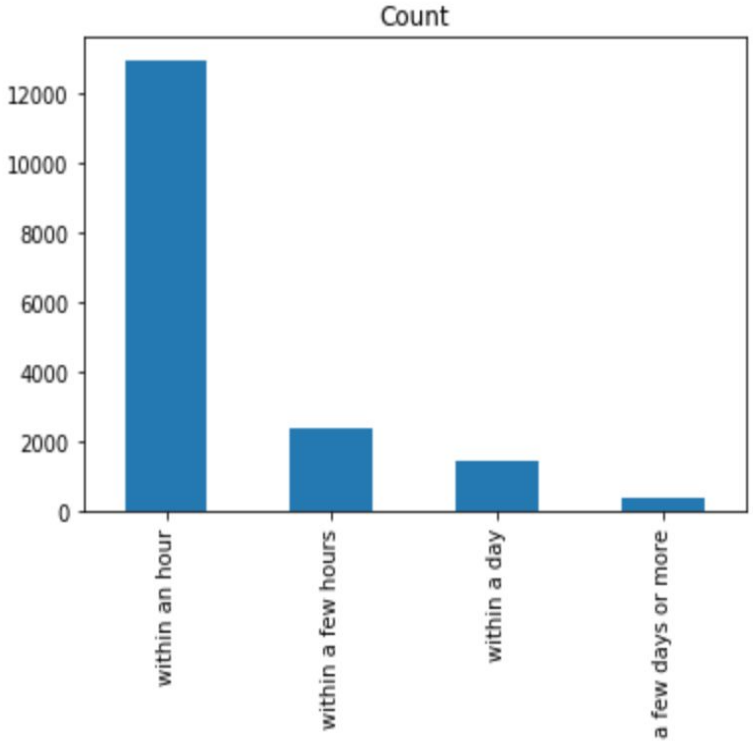
CORRELATIONS & MULTI-COLLINEARITY

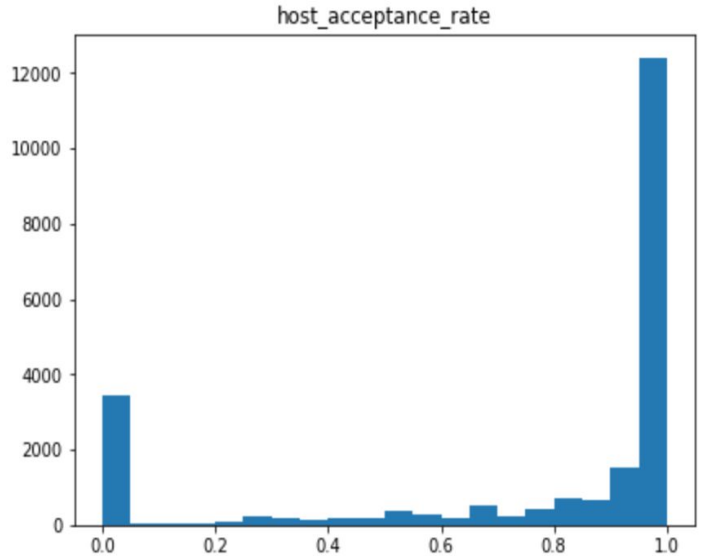
Potential Multicollinearity:

- Guests Included - Accommodates - Beds
- Private Room Type - Accommodates

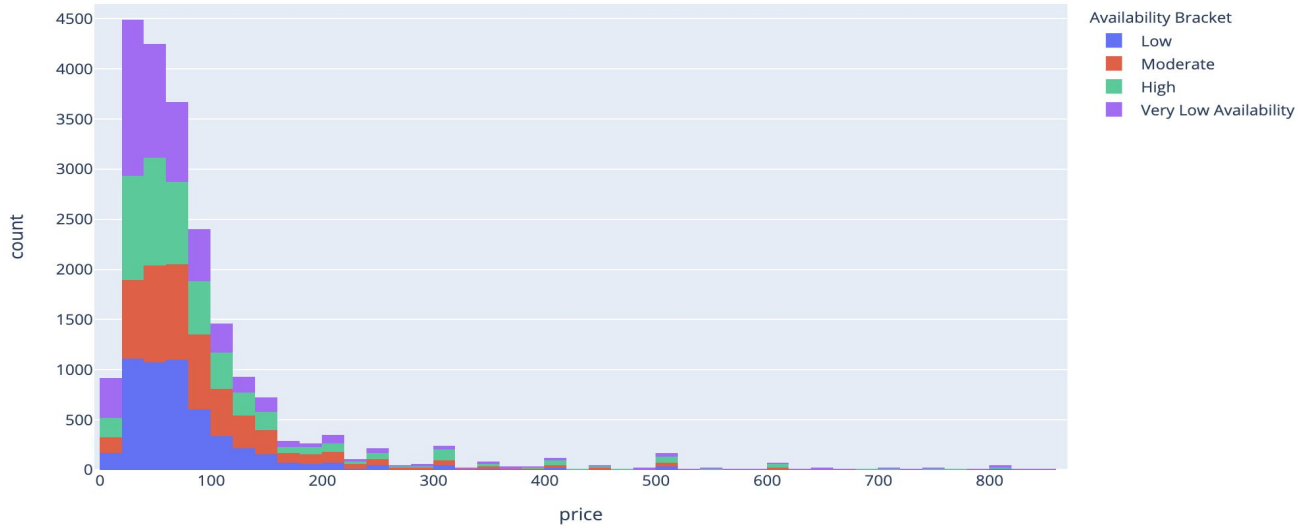
Low correlation between any variable and price

HIGH SKEWNESS & IMBALANCES

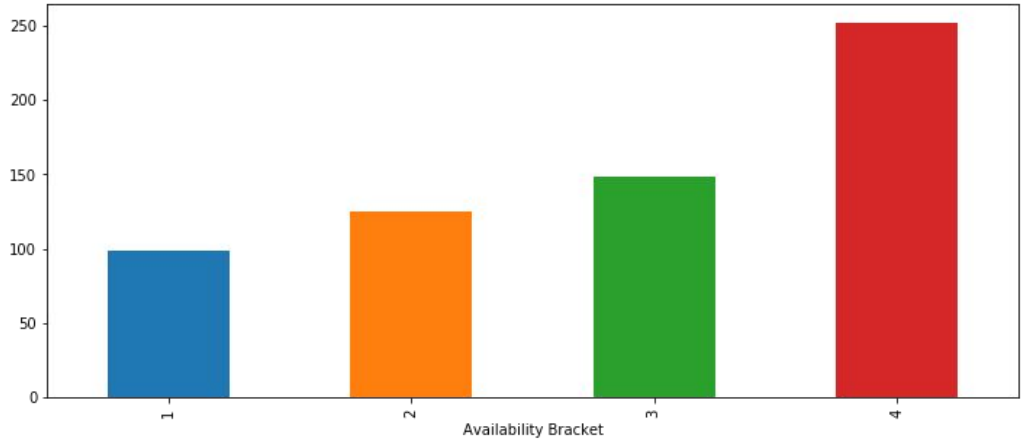




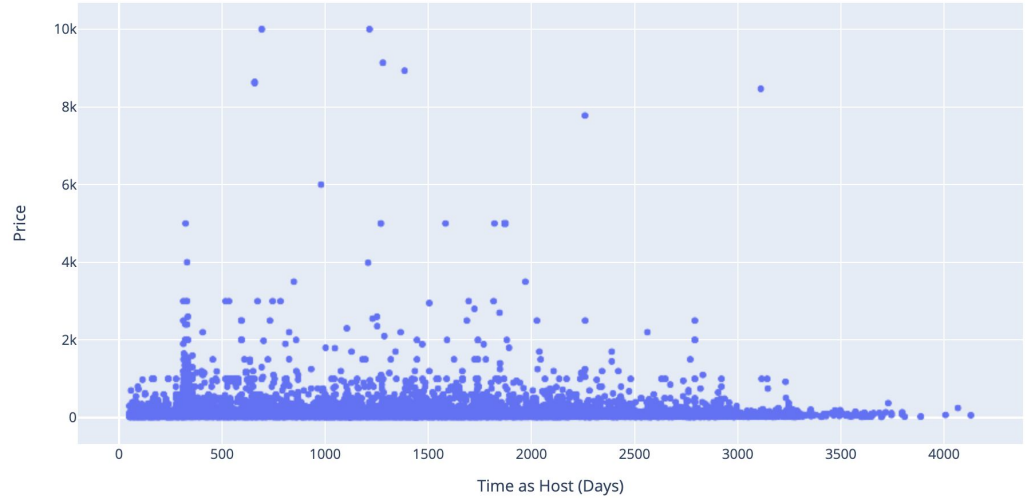
VARIABLE IMPACT ON PRICE



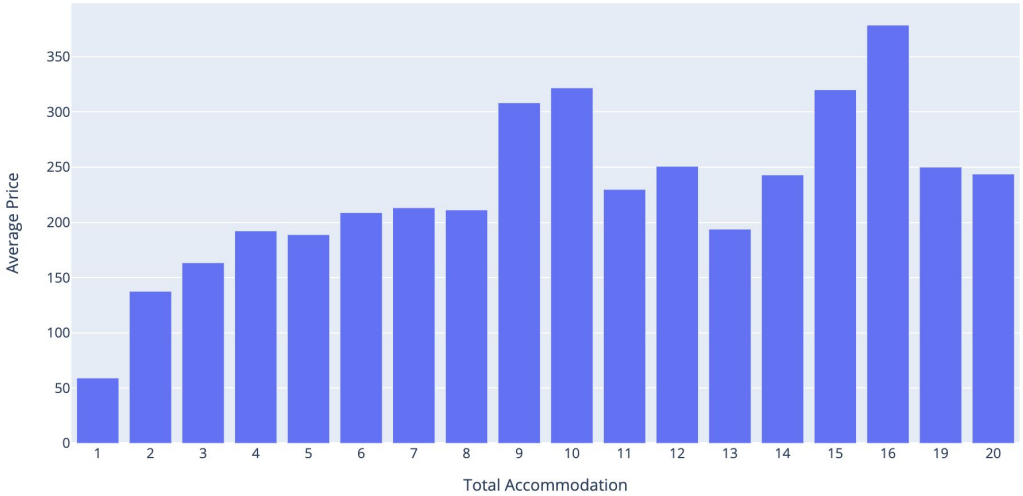
Availability Bracket



VARIABLE IMPACT ON PRICE



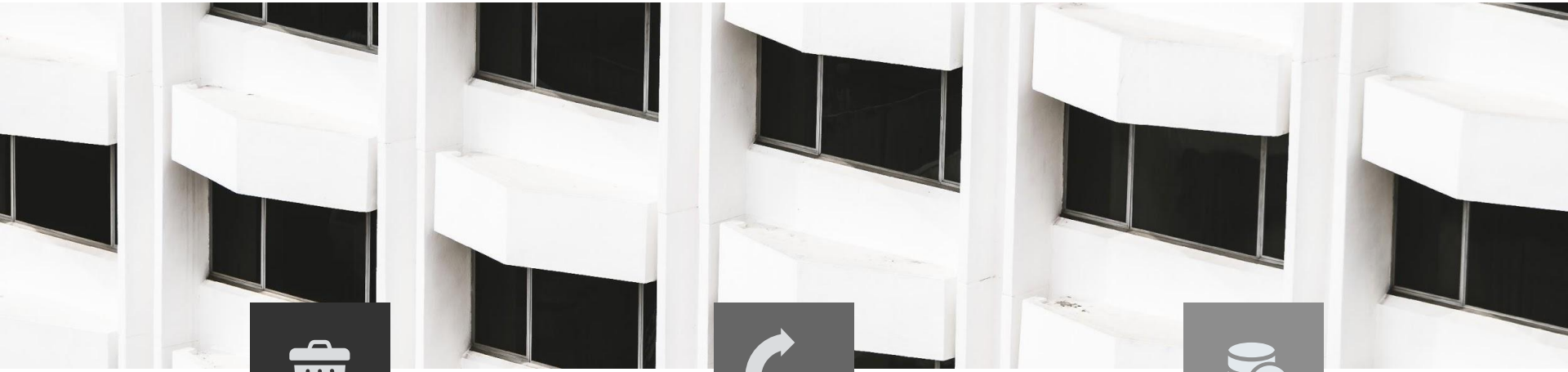
Time as Host



Total Accommodation

03 DATA PREPARATION

DATA CLEANING & TRANSFORMATIONS



Dropping Variables

- Re-Computed Variables
- Irrelevant Extra Information
- Inconsistent Data

Data Formatting

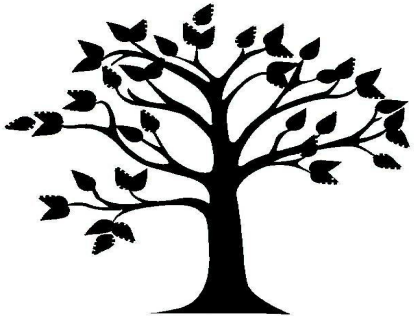
- Re-formatting inconsistent data
- Removal of improper strings

Cleaning And Imputing

- Manual Imputations
- Averaging
- Worst Case Scenario

MODELING 04

airbnb



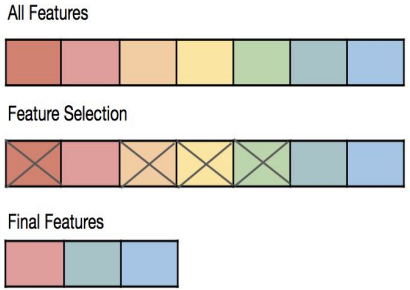
KBest: Chi-Square

Measure independence, look for ones that were dependent on target

FEATURE SELECTION

ExtraTrees Classifier

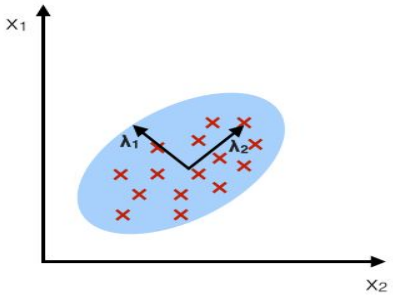
Feature Importance



LightGBM RFE

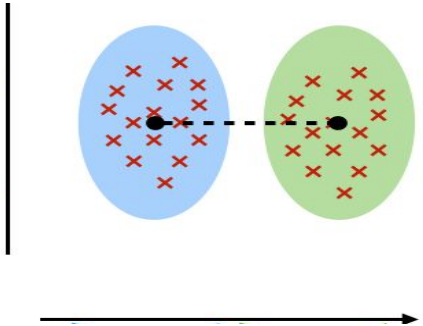
Feature Importance

24 Features



PCA + LDA

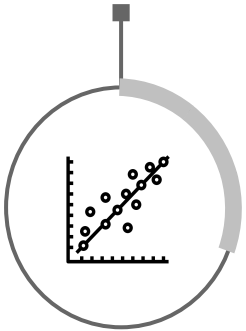
Number of components
-> Number of features



POOR PERFORMING MODELS

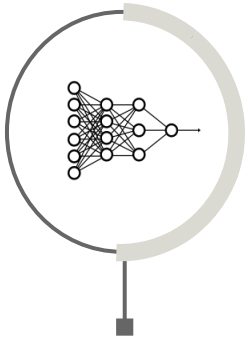
Multiple Linear Regression

Benchmark Model

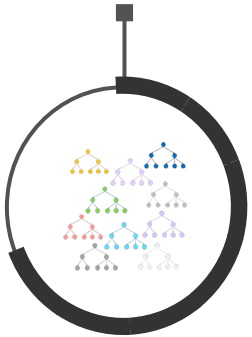


MLP Regressor

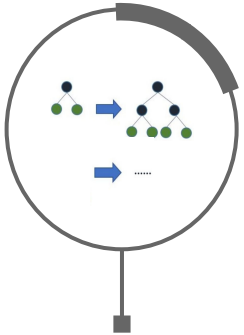
Neural Network



Random Forest



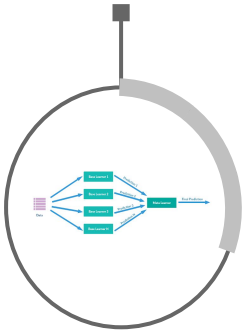
XGBoost



BETTER PERFORMING MODELS

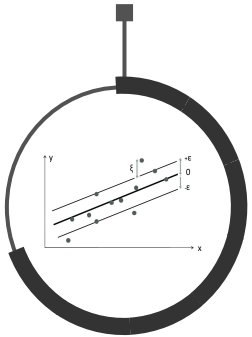
Voting Ensemble: SVR + XGBoost

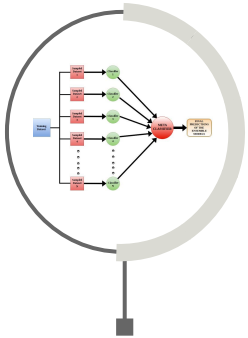
Train:51.79
Test: 57.22



SVR

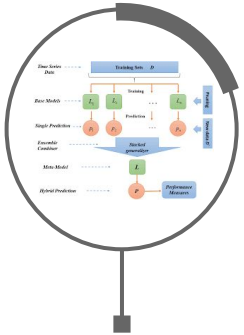
Train: 42.67
Test: 42.50





Average Ensemble: SVR, XGBoost, MLP

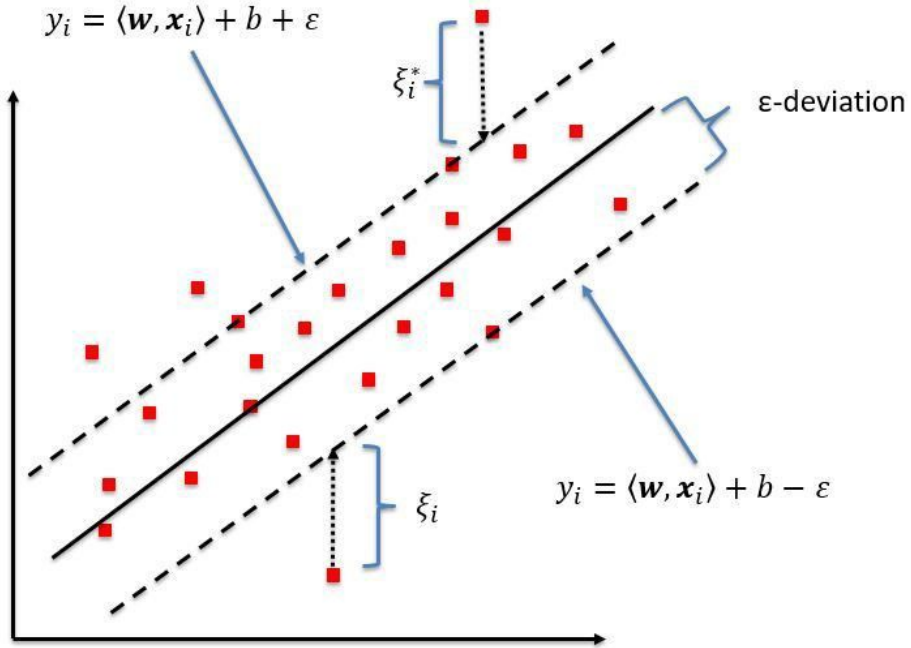
Train: 42.87
Test: 44.30



Bagging Regressor with SVR

Train: 42.54
Test: 42.40

SUPPORT VECTOR REGRESSOR



- Minimize l2-norm of the coefficient vector — not the squared error.
- Error term handled in the constraints, where the absolute error is less than or equal to a specified margin.

PARAMETER TUNING

Kernel

RBF

is a real-valued function whose value depends only on the distance between the input and some fixed point, either the origin or some other fixed point.

Epsilon

0.01

How much error we are willing to allow per training data instance.

C

1

Regularization parameter.

The strength of the regularization is inversely proportional to this value.

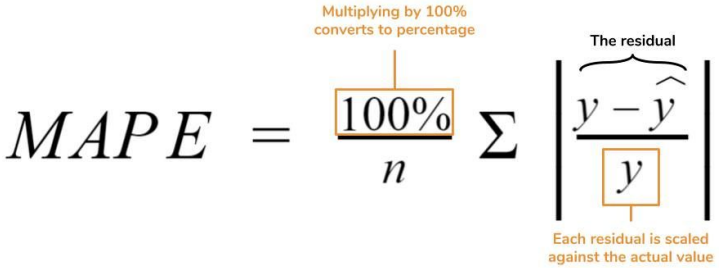
05 EVALUATION

VALIDATION TECHNIQUE

SCORING

MAPE

Average of absolute percentage errors



BAGGING REGRESSOR with SVR	42.40%
SVR	42.50%
AVERAGE ENSEMBLE: SVR + XGBOOST + MLP	44.30%
VOTING ENSEMBLE: SVR + XGBOOST	57.22%

06 DEPLOYMENT

RECOMMENDATIONS

New Features

Average price of hotels in the area, income per capita, distance to the city center/ airport/ train.

Ranking

Mantain Validity of the Model

Log(Price)

Standardized
Easy to interpret

THANK YOU

Does anyone have any questions?

Camila Barbagallo, Ryan Daher, Paula García and Rocío González