UNIVERSITÀ DI PISA

Artificial Intelligence and Data Engineering

Data Mining and Machine Learning

AirBnb Price Estimation using Machine Learning

Project Documentation

TEAM MEMBERS:
Edoardo Fazzari
Mirco Ramo

Academic Year: 2020/2021

Contents

1 — Introduction

Housing prices are an important reflection of the economy, and pricing a rental property on Airbnb, therefore, can be a challenging task for the owner as it determines the number of customers for the place. On the other hand, customers have to evaluate an offered price with minimal knowledge of an optimal value for the property.

1.1 Goals

The aim of our project is to compute the price of a **BnB** through the main feature that characterize a property (*e.g.*, position, facilities, property type), so as to propose to the owner and the possible host a fair amount for that particular real estate. In order to do that we divided the process in two phases:

1. Preprocessing:

2. Classification:

At the end of documentation, we propose an application in which a user can input some main features about a property for retrieving as output the appropriate price computed.

1.2 Initial Dataset

The fee of a real estate is strictly related to the city where it is located, thus would force us to recompute for every different city each step we make over and over, wasting time. Thus we decide to consider only one location (i.e, city), big enough to have a huge number of records. We chose *New York City*. The dataset, related to all the **BnB** situated in *NYC*, is taken from `http://insideairbnb.com/get-the-data.html` [1]. All the imported data has been modified, updated and preprocessed in prefer to satisfy the study needs as the chapter 2 will describe. The volume of the coarse data, considering 54k approximately entries (i.e., BnB), is more or less 100MB.

---

[1]In the case of the dataset form the website inside the *listing.csv* file, related to NYC, is updated, we stored the dataset used on Google Drive. It can be downloaded at:

2 — Preprocessing

2.1 What can be preprocessed now?

Nel senso che essendo unsupervised lo possiamo fare prima di splitting in train/test

2.2 Data Cleaning and Reduction

2.2.1 Removing Noisy and Irrelevant attributes

Metterci screenshot di excel o numbers

3 — Classification

3.1 Strategies

Come dividiamo train e test, quali classificatori scegliamo, quali algoritmi di attribute selection

3.1.1 Train and Test Splitting

3.1.2 Chosen Classifiers

3.2 Building Classification Models

Metterci anche il codice

3.2.1 Procedure

3.2.2 Implementation

3.3 Performance Evaluation and Effects od Attribute Selection

Metterci codice, screenshots, procedure di valutazione e un SACCO DI ROBA

3.4 Conclusions

4 — AirBnb Price Estimator App

4.1 Introduction

4.2 Functional Requirements

4.3 Application Guide