Machine Learning Algorithms (MAAI)

Practical Work – **Phase 1**

Student Name: Mina Ezach Naeem Faltos **Student Number: 34388**

A Predictive Model for Airbnb Rental Pricing

1. Introduction and Problem Definition

A) Background

The sharing economy is growing fast and the hospitality industry can’t stand longer in the shadow of platforms like Airbnb. One of the key problems hosts (owners of the property) face is determining the correct nightly price for their property. Setting the price too high results in a very low occupancy whilst setting it too low means you do not receive the full potential income. The process is manual, subjective and prone to mistakes.

B) Problem Statement

The problem that the project will attempt to solve is to calculate the "fair market value" of a short-term rental listing based on its objective attributes. The objective is to build a system that can automate this valuation using past data.

2. Machine Learning Problem

A) Problem Type

This problem consists of a Supervised Learning problem, more precisely a Regression problem.

B) Rationale

The problem is regression because the target variable (Price) is a non-discrete numerical variable (Like €85.50). Based on that, the model will learn what relationship between the input features and the continuous output variable, rather than divide data into discrete classes.

3. Goal & Output

A) Goal

The goal is to develop, train, and validate a Machine Learning model able to predict the price of a nightly rental of an Airbnb listing with great accuracy.

B) Output

The deliverable will be a trained model that when given certain characteristics of a property will output a predicted price.

Input : features such as location (neighborhood/latitude/longitude), room type (Like entire home, private room), accommodation capacity, number of bathrooms, etc.

Output : a float representing the predicted price per night.

4. Data strategy

A) Data Source

InsideAirbnb or a curated version of the dataset which is hosted on Kaggle ("Airbnb Open Data") are the datasets I’m planning to use. This data is appropriate to train on as it was created to be a continuous historical record of properties with associated metadata and prices.

B) Candidate Predictors

From the very first pass over the data, a handful of signals look worth keeping in the model:

- Where it is: the broader neighborhood chunk, plus the exact lat/lon pair.

- What it is: room category, availability (days available per year), and the shortest bookable stay.

- How guests react: total review count and reviews per month.

5. References

Zervas, G., Proserpio, D., & Byers, J. (2017). The Rise of the Sharing Economy: Estimating the Impact of Airbnb on the Hotel Industry.