

DSC 412 Fall 2024 Exploring Machine Learning- Project Proposal

Student Name: August Pallesen

Student ID: 200618081

Email: apalles@ncsu.edu

Proposal:

As an foreign exchange student, I would like to travel as much as possible while I am here. Doing so is expensive and time consuming, f.ex. finding the right airbnb to stay at, while still traveling on a budget is difficult yet important. So why not create a machine learning model to predict the price of a given airbnb. This is the aim of this project, namely to build a ML model that within a relatively high degree of accuracy can predict the price of a given air bnb listing in New york (where I will be going soon, and spent a lot of time trying to find the right place). This will hopefully help other like minded people who want to travel to new york. In the development of this project, several challenges may arise, such as flawed or incomplete data, null values, large quantities of data. The novelty of this project lies in the problem, as I wont be using any other machine learning techniques or approaches than what is already available.

Plan:

The project will be developed using the dataset: New York Airbnb Data Mining from kaggle. The dataset consists of approx. 25k entries, each detailed through 22 features. The goal is to analyze the data through careful data preprocessing, and then to apply a regression model, in order to find any trends or patterns in the data that could help us gain insight into the volatile airbnb listing market in New york. The accuracy of the model will be measured through performance metrics such as confusion matrices, RMSE, and/or other accuracy metrics. Since the dataset is from kaggle, other data scientists may have tried to develop a similar model, which could prove useful as a baseline for improvement down the line.