COE 379L Project 4 - Initial Proposal

James Grant Robinett (jgr2722) - Hung-Yi Tseng (ht7796)

Introduction and Problem Statement

As we have seen during our learning in this class, there are many different ways and methods to process different data types, from one hot encoding to image classifications. However, we wanted to see if there was a way to combine both image and categorical/numeric data when training a machine learning model. For this project, we propose the development of a multimodal machine learning model that predicts AirBnB rental prices by analyzing both images and textual descriptions of Airbnb listings.

Data Source

The primary dataset for this project is a dataset from Hugging Face that consists of AirBnB listings with various data points (https://huggingface.co/datasets/MongoDB/airbnb_embeddings). The dataset contains 5555 rows and 43 columns, with each row representing a single AirBnB listing. Some of the most important features we will be looking at are columns like name, summary, images, property_type and price. We plan to use this data to train a multimodal model that can both use image data as well as numerical and categorical data.

Methods and Techniques

- Data Preprocessing: tokenization of text and image data.
- Baseline Models: text embeddings or image embeddings of baseline models.
- Multimodal Model Development: implement a multimodal neural neural network architecture that combines text and image, and find best performance of price prediction.
- Model Evaluation: Analyze the model performance on different other variables of the data (property type, location, price range, .etc).

Products to be delivered

Our project hopes to deliver a few tools that we can use to accurately predict AirBnB rental prices through our multimodal model. First, we hope to train a Convolutional Neural Network (CNN) to extract features from property images, which will allow us to analyze visual elements of a home that influences price. Next, we will fine tune a BERT model, tuning its performance for the textual descriptions associated with Airbnb listings. These two models will combine into a multimodal system with image and text data to deliver accurate price estimates. As a stretch goal, we hope to make a small UI to make it easy to send our model image and description data, and the model will return a price estimate. Lastly, we will deliver a detailed report documenting the methodology we used, experiment results, and evaluation of the models using metrics like Mean Squared Error (MSE) for price prediction accuracy.