Leonell Puso
ID# 302164083
CSC 180-01 Intelligent Systems (Fall 2023)
October 27, 2023

PROJECT 3 REPORT

(1) Problem Statement:

The purpose of this project is to create a house price estimation system by using both textual and visual inputs. Each house will have four images and information in text which includes bedrooms, number of bathrooms, area of the house, zipcode and the price. Based on the images and textual data, the model will predict the price of a house.

(2) Methodology

Data Collection:

The data in https://github.com/emanhamed/Houses-dataset will be used. This dataset contains both visual and textual information.

Each house will have:

- Four images
    - Bedroom
    - Bathroom
    - Kitchen
    - Front of house
- Text file with each row consisting of
    - Number of bedrooms
    - Number of bathrooms
    - Area of the house
    - Zipcode
    - Price

Data Cleaning:

- Remove outliers by keeping only the houses with a price between $100K and $900K

Feature Preparation:

- Encode categorical features and normalize numeric features.
- Combine all four images into one image for each house
- Split data for training (80%) and testing (20%)

Model Building:

- Use training data to train your models and evaluate the model quality using test data
- Use TensorFlow Functional API
    - Fully Connected Neural Network (FCNN)
        - First input model
        - Use for the textual information
    - Convolutional Neural Network (CNN)
        - Second input model
        - Use for the images

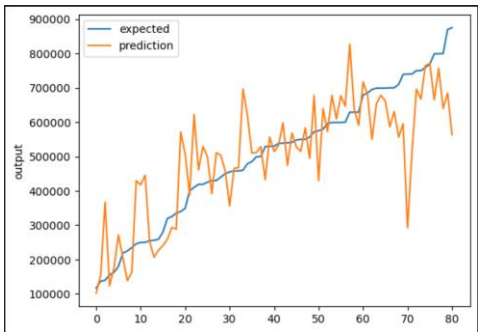
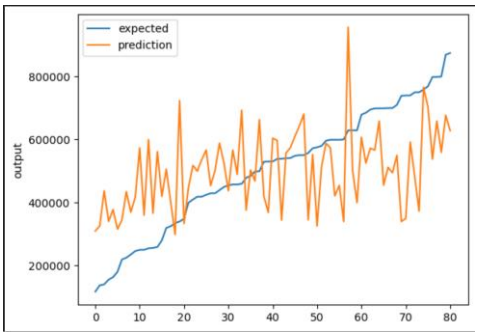
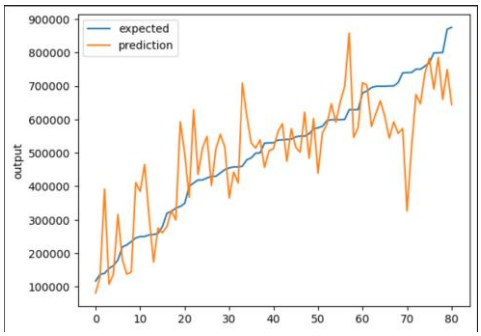
Training and Validation:

- Use EarlyStopping and ModelCheckpoint to detect and prevent overfitting. Use this technique in 10 iterations but only keep the best model.

Evaluation and Analysis:

- Evaluate the model's performance by showing the RMSE score and the lift chart
    - Use test data
- Extra Feature:
    - Remove zipcode
    - Zipcode as separate input

(3) Experimental Results and Analysis

Original model	Zipcode Remove	Zipcode as separate input
Score (RMSE): 115049.5859375	Score (RMSE): 181072.8125	Score (RMSE): 114855.4453125
		

Removing the zipcode yielded the worst result, while separating the zipcode as a third input model gave a better output.

(4) Task Division and Project Reflection

Only one member.

Project Reflection:

This project was a practical application of intelligent systems, focusing on house price estimation using textual and visual data. It highlighted the significance of data preprocessing, model development, and feature engineering. The use of TensorFlow Functional API, EarlyStopping, and ModelCheckpoint improved model performance. Treating the zipcode as a separate input enhanced accuracy. Overall, this project provided hands-on experience in solving real-world problems using machine learning techniques.

This project has been a valuable learning experience. It reinforced the importance of data quality in machine learning projects, underscoring the impact of thoughtful preprocessing. The use of neural network architectures for combining text and image data showcased the versatility of deep learning. The project's results emphasized that feature engineering, when done strategically, can significantly enhance model accuracy. This practical experience has deepened my understanding of intelligent systems and their practical applications in solving complex, real-world challenges.