Tyler Burguillos

Austin Nolte

Alan Lei

Nishan Majarjan

October 21, 2024

Professor Haiquan Chen

CSC 180 - Intelligent Systems

Due Date: October 23, 2024

Project 3 Report

Problem Statement: In this project, the goal is to predict house prices using a combination of visual and textual features. The dataset includes images of houses (bedroom, bathroom, kitchen, and frontal views) along with numerical data such as the number of bedrooms, bathrooms, area, and zip code. We will build models that integrate both Convolutional Neural Networks (CNN) for image data and Fully-Connected Neural Networks (FNN) for numerical data to estimate house prices. Evaluation will be conducted to optimize the models, using metrics such as Root Mean Square Error (RMSE), R-squared (R^2), loss curves and lift charts to provide additional insights into the models' accuracy.

Methodology:

In this project we used two different types of models for two different types of reasons. Firstly, we used a Fully-Connected Neural Network to detect the textual information from the number of bedrooms,bathrooms, and zip codes. This was used to make a prediction of its own for the housing price. The next type of model was a Convolutional Neural Network (CNN). This model was used to determine characteristics of the images of the houses. These characteristics were used to determine a dense layer. Lastly, we put the best models together to make hybrid models where the outputs of each were combined to make the best possible guess it could. We used each combination of activations and optimizers we have been learning about to come up with the best model. We used the given cheat sheet for data pre processing and also expanded upon it later. We expanded upon it later by conducting a paper on SURF and ORB. ORB was the one we decided to go with as SURF was causing issues in Google colab and ORB was free and fast. Our models used Early Stopping and Checkpointing to save the best results of each model before they became overfitted. We experimented with VGG16 transfer learning to see if it improved results. The results were good, but did not see any noticeable improvements.

Experimental Results and Analysis:

Best FNN

Architecture Column Format: # of Layers, Layer(B= Batch Normalization, D= Dropout, Number = Neurons in Dense Layer), Activation(R= ReLu, S= Sigmoid, T= Tanh), Optimizer(A= Adam, G= SDG)

Names	Architecture	RMSE	R-squared
Tyler Burguillos	5, B, 50, 25, 15, 1, T, A	109457.34375	0.72
Austin Nolte	2, 64, 1, S, A	121361.8359375	0.66
Alan Lei	9, 50, 40, D, B, 30, 20, 10, D, 1, R, A	112001.3203125	0.71
Nishan Maharjan	4, 128, 64, 32, R, A	117740.46875	0.68

Best CNN

Names	RMSE	R-squared
Tyler Burguillos	240200.421875	-0.35
Austin Nolte	208861.21875	-0.01767075
Alan Lei	229399.84375	-0.23
Nishan Maharjan	433494.625	-3.38

Our Best Hybrid Model:

Names	RMSE	R-squared
Tyler Burguillos	251005.8125	-0.47
Austin Nolte	953482.5625	-20.21
Alan Lei	276643.53125	-0.79

Task Division and Project Reflection:

For this project, we divided the workload into several tasks to ensure that every team member could focus on specific areas of expertise and contribute effectively. Each member explored different model architectures and contributed to the overall methodology, data preprocessing, and evaluation.

Austin Nolte - Created own CNN, FNN and Hybrid models with sigmoid activation and adam/sgd optimizer

Tyler Burguillos - Created own CNN, FNN and Hybrid models with and performed keras tuner hyper tuning on models. Created additional features and data pre processing as well

Alan Lei - Created own CNN and FNN models with varying amounts of layers and hyperparameters

Nishan Maharjan - Created own CNN and FNN models with varying amounts of layers and hyperparameters

Our hybrid models and CNN’s did not achieve the best results, we seemed to get the best results with just FNN models. Creating additional features to try and pre-process the images or data better did not achieve better results. Transfer learning with CNN’s did not achieve much better results either. We initially did not normalize the price column and that led to abnormalities. We normalized the price column and it more consistent outcomes.