Satellite Imagery-Based Property Valuation

Chikoti Sai Tejaswini[24115050]

INTRODUCTION

The objective of this project is to predict house prices using both a combination of structured tabular data and satellite images . traditional house price prediction models only use numerical tabular data which does give a lot of information about the house such as its size , number of bedrooms , lot size , age etc , but it does not take into account of the effect of features such as the type of surroundings , the neighborhood in which the house is located . Factors like the population density , surrounding infrastructure etc.

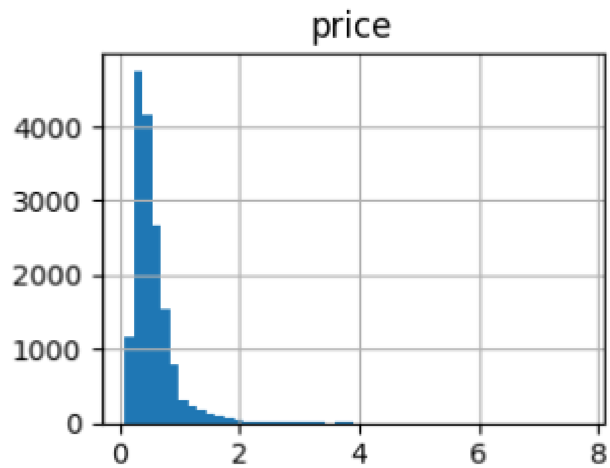
To address this limitation we try to build a multi modal regression model which integrates both satellite imagery and the tabular information to predict the prices .

This report is divided into 5 sections from here namely

1. Data analysis and exploration
2. Extraction of satellite imagery
3. Analysis of Baseline models
4. Architecture and working of multi modal regression model
5. Prediction of prices for test dataset and GradCam analysis
6. Issues faced
7. Conclusion
8. References

1. Data Analysis and Exploration

- The dataset consists of structured tabular information of features related to the house properties along with geographical ones like the coordinates of the location of the house and zip code .
- The target feature is ‘price’ which is right skewed , that is most house prices are in the mid region while only a few belong to the higher price category , we can see this through the histogram of the price column .



- The dataset has structural , geographical , temporal features a few of which were modified such as the date feature which was of object datatype and indicated at what date and time a house was sold , since our models cannot process object datatypes we extracted useful features such as the year sold and the month sold .
- Also we have a few temporal features such as year built and year renovated which alone are informative but the difference between them [year built - year renovated] gives us information about how old the house is , and the condition of the house , hence a feature house age was added which is the difference between [year built - year renovated] .
- Below is the list of features with their correlation with the target variable price , which shows that features like sqft_living , grade show high correlation , which proves that larger houses with better construction quality cost more .

Sqft_living	0.700255
grade	0.668470
sqft_above	0.599769
sqft_living15	0.586629
bathrooms	0.530931
view	0.894552
sqft_basement	0.320422
lat	0.309012
bedrooms	0.300996
Floors	0.249966
waterfront	0.222474

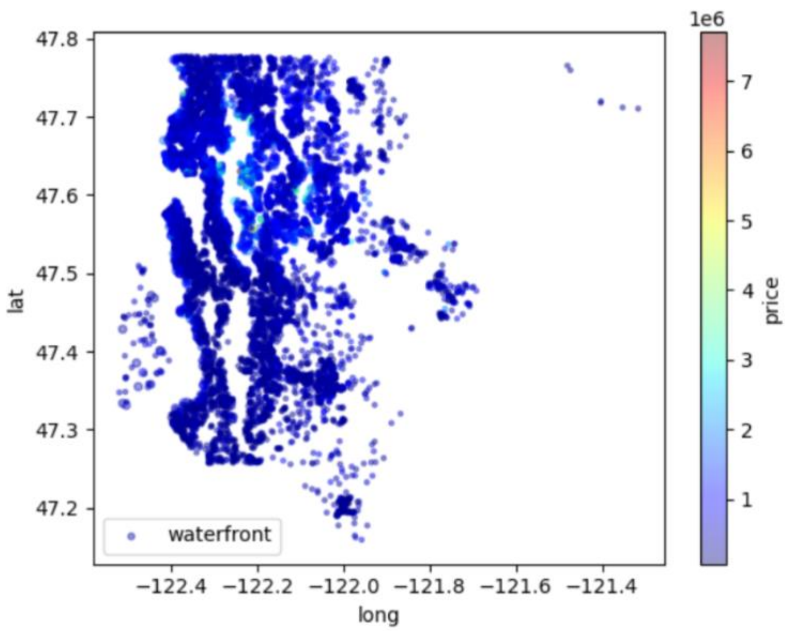
tablel(1.1)

yr_renovated	0.118033
house_age	0.111687
sqft_lot	0.086329
sqft_lot15	0.078120
yr_built	0.056685
long	0.031495
condition	0.028565
year_bought	0.012701
month_bought	-0.013203
id	-0.018426
Zip code	-0.058470

table(1.2)

Correlation between features and price

- Geospatial features show that the location of the house highly affects its price , in the below plot the price is determined by the intensity of the color and the radius of the points indicate the waterfront in that region



fig(1.1) : shows the lat vs long plot with color indicating price and size of dot indicating waterfront at the given location

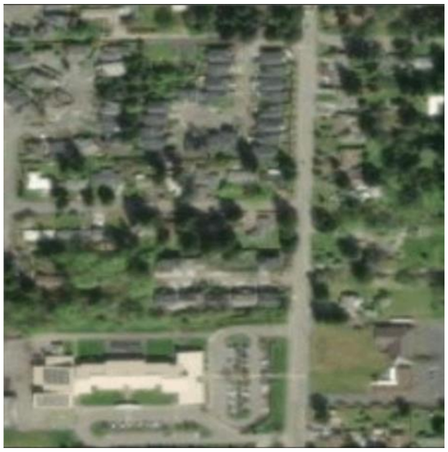
- This shows that the price of a house varies with its location , as we move from north to south the prices drop considerably .

2.Extraction of Satellite imagery

- To incorporate the visual details of the surroundings into our prediction we need satellite images of that location which tells us about the surroundings like the greenery near the house , the urban density (depending upon the number of houses nearby , network of roads etc) , also waterfront if it's near a sea etc.
- To extract the images we use the coordinates provided in the dataset, most of the houses in our data set belong to the seattle or the king county region of the US , hence we used the ESRI world imagery which is a publicly accessible geospatial images data set with high resolution .
- ESRI has images of the US region with high precision up to 30cm per pixel . We extract the images by using the arcgis api . But to ensure that our images capture the surrounding information well, the zoom level matters a lot , that is approximately how much area would you consider around your house to define the locality . Below are satellite view images of a house at different zoom levels .

fig(2.1):0.007(1.5kmX1.5km)

fig(2.2):0.003(600mX600m



fig(2.3): 0.002(400mX400m)

fig(2.4):0.001(200mX200m)

- As we can see the fig(2.1):(1.5kmX1.5km) covers a lot of region but its very clustered , while the fig(2.4): (200mX200m) is very zoomed in , it does not even span for about 1 block on either side of the house whose price we are predicting,which limits our understandings about the surroundings (assuming it to be located at the centre) hence fig(2.2) and fig(2.3) are good choices for learning features from , hence we have used fig(2.2) that is (600mX600m) images that covers 3 blocks on all sides of the house .

- Images of the same zoom level have been extracted for all the training and test data .

- And to access the images easily the images were named as their house_id.png .

3.Analysis of Baseline Models

- Before building a multimodal model that integrates both images and the tabular data for prediction , a few baseline models were trained only on the tabular data , how they were implemented and the RMSE and R2 score obtained by those models is listed below .
- The baseline models tried were :
    1. Linear Regression
    2. DecisionTrees
    3. RandomForest
    4. NeuralNetwork
    5. XgBoost
- All the baseline models were imported from the Sklearn library .
- For Linear regression and Neural network the independent features and the price were scaled first then trained on the scaled features . Linear regression used was a basic straight line and the neural network had 2 hidden layers , with ReLU activation and Adam optimizer .
- For Decision trees , Random forests and Xgboost the data was passed as it is , without any scaling and after experimenting with a few different values of the parameters such as depth of the tree , minimum samples per leaf etc. the following table shows the best results achieved .

BASELINE MODEL	RMSE (in dollars)	R2 SCORE
Linear Regression	207419.27	0.6988
Decision Tree	186146.02	0.7575
Random Forest	160583.81	0.8195
Neural Network	145733.93	0.8513
Xgboost	139585.17	0.8636

table (3.1) : RMSE and R2 scores achieved by diff models

- Hence out of all baseline models tried XgBoost was the best one.Therefore upon predicting prices only based on tabular data could give us a RMSE up to 139585.

4.Architecture and working of Multimodal Regression Model

- Until now we have tried predicting the prices only using tabular data , now we will also integrate the satellite images which we obtained earlier to find the target variable .

  Zipcode embedding
- Most of the tabular features can directly be passed into the neural network of the model after scaling , but zipcode cannot be passed such a way , as zipcode, unlike the other features, is not a feature that can be analysed by its magnitude . Every city has a zip code which tells us about the postal service of that location. We cannot determine which zip code is better by comparing the magnitude of two zip codes; those are just numbers to represent them .
- Hence for such features we use embeddings which help the model to learn the importance of zip code , before we can pass our zipcode to the embedding layer we have to encode them first , that is if you have around 10 zip codes you encode them from 0 to 9 that is how an embedding layer expects input.

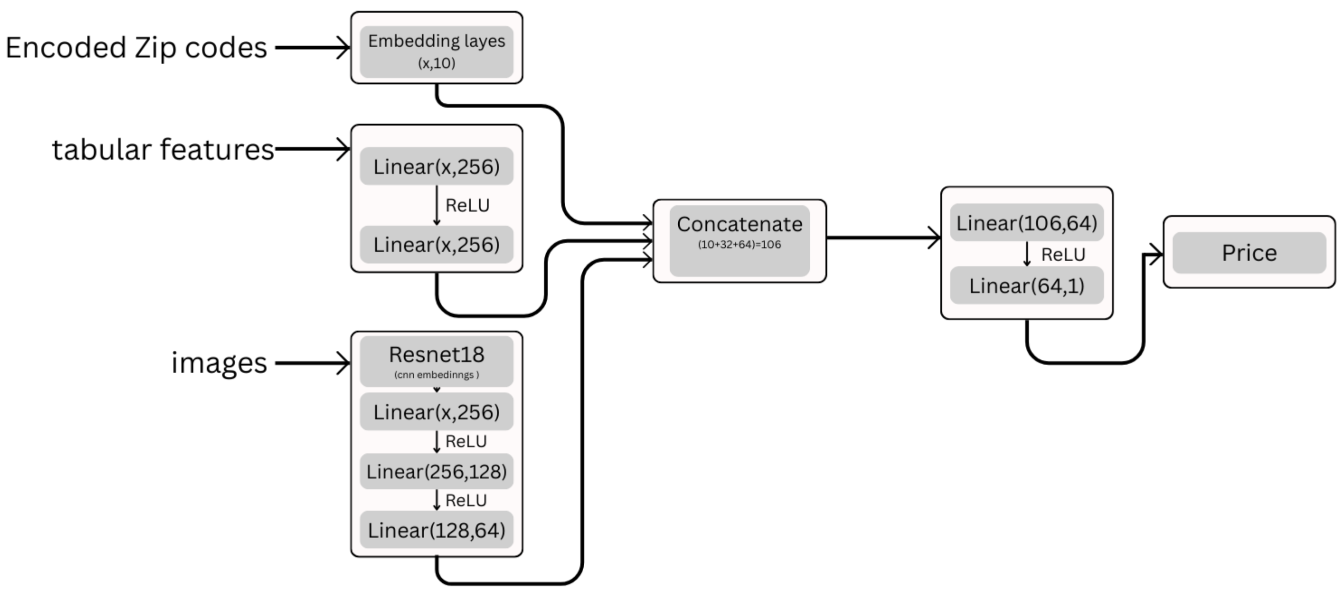
  Resnet18
- In images for better spatial analysis we use pretrained cnn model resnet18,it is a CNN model which is pretrained on the image net dataset which has over 14 million images in it and around 1000 categories , this Resnet18 model is hence robust in spatial analysis of images especially in finding features like green areas , roads etc.
- But before we use the Resnet model we have to turn our images into tensors and normalise them based on Resnets input requirements ,
- The general CNN models can be divided into two categories , feature embedding part and an ANN attached to it , the feature embedding part is the one that learns from the images , the outputs of these layers are used by the ANN to perform the task required like classification or regression .
- The original Resnet models ANN was a classifier,we detach that part and attach our own ANN so that we can use those features for our price prediction , while the feature embedding part is freezed , that is its weights (learnings) cannot be altered when we train our model .

Neural Network for tabular features

- The other features (numerical data) are scaled and sent into a neural network .

Final NN

- After passing our data through the above 3 models accordingly we concatenate the features obtained and pass them through a final nn through which we predict our prices . (*the prices passed in while training were also scaled to avoid overfitting / other possible errors hence the RMSE obtained is also scaled but to obtain it in dollars it is multiplied by the std of the price scaler) .



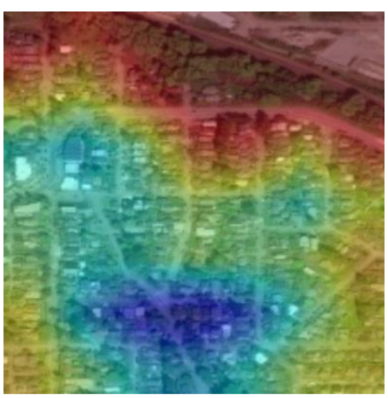
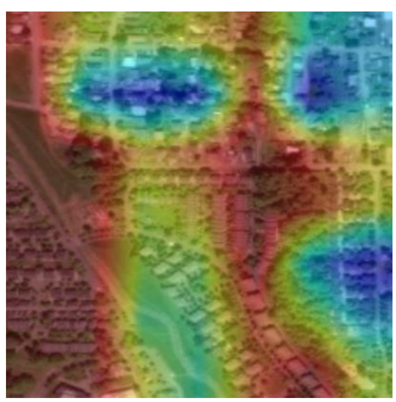
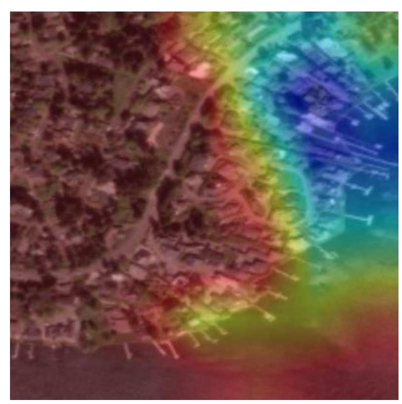
MODEL	RMSE	R2 SCORE
Multimodal regression	123056	0.8792

5.Prediction of prices for test dataset and GradCam analysis

- After the model is trained and tested on validation set its state is saved, i.e, the weights obtained , in the price prediction.ipynb of the Github repo we will reload the model and predict prices on our test data .
- After predicting prices we will apply GradCam on a few images from the test data to see what features in our image are prominent exactly.

GradCam Working

- Gradcam works by analysing how the diff feature maps of the image affect the end result that is price .
- To say how the loss changes wrt the feature maps (which are in the last layer of resnets cnn part , that is resents layer 4) we need the gradient of loss wrt the weights of this layer , hence we unfreeze our resnet before this , its weights will not be affected in the test loop .
- After we get the weights , we average them and assign a score for each feature map present which indicates its prominence , next the feature maps are reconstructed into an image and hence a heatmap is created , the region where the values are high in magnitude will be of darker color (we have used the jet colormap so from blue to red the prominence of that region increases)
- We apply this on a few test cases and the following are the results obtained -

- From these images we can see that areas with a lot of trees are in red colored region , so greenery around the house was an important factor .
- Also areas near the shore , near wide roads compared to narrow roads had more value .

6.ISSUES FACED

While working on this project quite a few problems were encountered such as

- **Images extraction**
  The number of publicly available resources online for extraction of satellite images were low , and among the few available the resolution of a few was very bad , the images extracted from sentinel hubs api were all cloudy
- **Data alignment issues**
  The images were extracted based on their house ids when the number of images extracted and the number of training examples were compared they were not the same later it was realised that the data did not have unique ids but a few houses were resold within a short interval so those purchases were also listed
- **Computational constraints**
  The data set was huge and the use of CNN based image feature extraction resulted in higher computational costs and the limits on availability of gpu on cloud services restricted extensive hyperparameter tuning .
- **Training instability and reproducibility**
  Despite using random seed training results varied upon each run because of the use of drop out and batch ordering so the final model had to be selected carefully

7.CONCLUSION

In this project, we explored the problem of house price prediction by moving beyond traditional approaches that rely only on tabular (numerical) data. While tabular information provides details about the property itself—such as size, number of rooms, and age—it does not capture the characteristics of the surrounding environment. To address this, we incorporated satellite imagery to model neighborhood features such as greenery, road networks, urban density, and proximity to water bodies.

We first started this process by doing exploratory data analysis in which we saw the distributions of all our features and the correlation between them and the price. We also created a few new features from the existing features . baseline models were built only on this tabular data whose performance told us about the limit of using only tabular data and also helped in setting an expectation for our model .

After learning from the baseline models we integrated images to the data and built a multi modal model which takes both of them as inputs , and uses the features learnt from the images along with the tabular data to predict prices . We achieved this using a pretrained resnet18 model to learn the features from the images . The results obtained from this model were slightly better than the ones obtained from the baseline models showing us the importance of surroundings in price prediction .

To understand what features exactly influenced the price we used gradcam which helped us to understand that regions with greenery , wider roads , water availability affect prices a lot .

8.References

- ESRI world imagery
- Pytorch documentation
- CampusX yt channel for pytorch
- Chatgpt (for debugging and conceptual clarifications)