**CSE 4/587 B**
**Project Phase - 3**

**Team members:**
Yash Dattu Sonkamble (yashdatt)
Sameer Yadav (syadav6)
Anirudh Anilkumar (aanilkum)

**Problem Statement:** The problem at hand is the significant variation in house prices across different neighborhoods in Perth (relevant to other places as well provided the data is available), driven by factors like proximity to amenities, property attributes, and land area.

Our model aims to answer the following key questions:

- How can we develop a near accurate machine learning-based model to predict house prices i.e. are they highly, lowly, or moderately priced in various Perth neighborhoods based on relevant factors?
- What is the significance of accurate house price predictions for both buyers and sellers in making informed real estate transactions?
- How can market transparency be improved through reliable price predictions, and what benefits does this bring to the real estate market?

**Link of the dataset -** https://www.kaggle.com/datasets/syuzai/perth-house-prices

To achieve this, we need the appropriate regression algorithm. From the results we achieved in phase 2, we concluded that Decision Tree Regression with maximum depth equal to 10 gives the most accurate results.

Hence, we choose the Decision Tree Regression algorithm for training our model.

**Implementation:**

1. Our dataset has important features that help predict house prices, like the number of bedrooms, bathrooms, garages, the size of the floor and land area, and the postcode. We've made it easy for users on our web app to predict house prices. They only need to enter details like the number of bedrooms and garages, floor and land area, and postcode.

2. We don't expect users to know or provide details like latitude, longitude, and distances to the nearest school, CBD (Central Business District), or train station. It might be hard for them to find this information.

3. For our training process, we've set up our code to use the user-provided features to make three separate Decision Tree regressor models. These models predict the distances to the nearest school, CBD, and train station. We then use these distance predictions to help in predicting house prices.

4. We chose Decision Tree regressors for these predictions because they showed good results in estimating these three types of distances. Here are some performance metrics::

   CBD Distance:
   Mean Squared Error : 5706701.473170062
   Root Mean Squared Error : 2388.8703341056544
   Mean Absolute Error : 1648.636204050679
   R2 Score : 0.9548664389641164

   Station Distance:
   Mean Squared Error : 2706050.605194796
   Root Mean Squared Error : 1645.0077827155699
   Mean Absolute Error : 1115.5863482833613
   R2 Score : 0.8315004697696524

   School Distance:
   Mean Squared Error : 0.8449407934686658
   Root Mean Squared Error : 0.9192066108708454
   Mean Absolute Error : 0.6186996763669751
   R2 Score : 0.5494227175278654

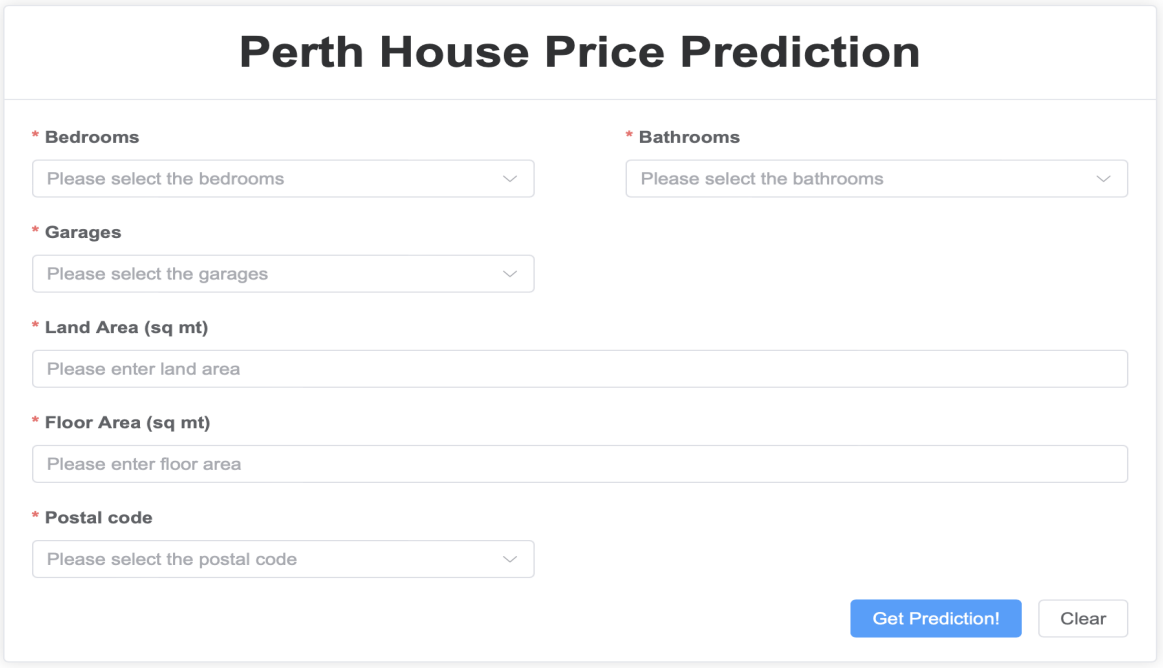
   For graphs and analysis please refer to phase3.ipynb file

5. In addition to the models we mentioned earlier, we also train a Decision Tree regressor model to predict house prices using features like the number of bedrooms, bathrooms, and garages, as well as the floor and land area, postcode, and distances to the CBD, nearest station, and school.
6. We use two more models in our system. One is called 'scaler_price,' which normalizes (adjusts) the user's data and the distance data from the other models when we calculate the house price. The other model is 'scaler_other,' which normalizes the user's data when we calculate the three distances (to the CBD, nearest station, and school). These models help make sure our data is consistent and accurately used for predictions.

**GUI Interface:**

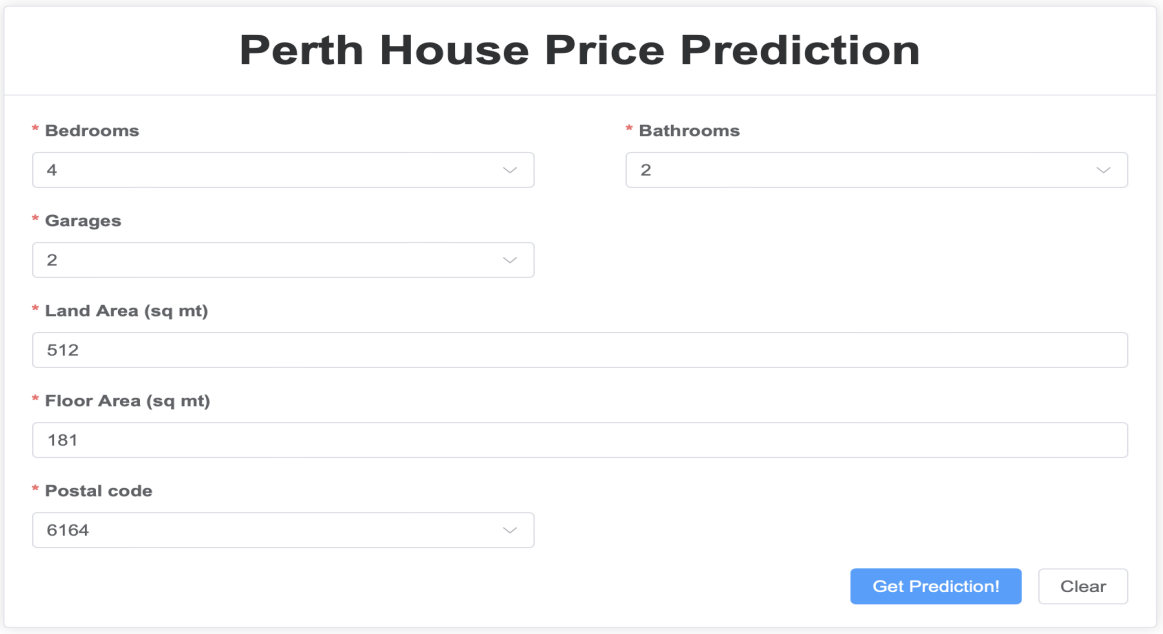
Please refer to readme.txt in the src/phase3/ directory in order to run the web application.

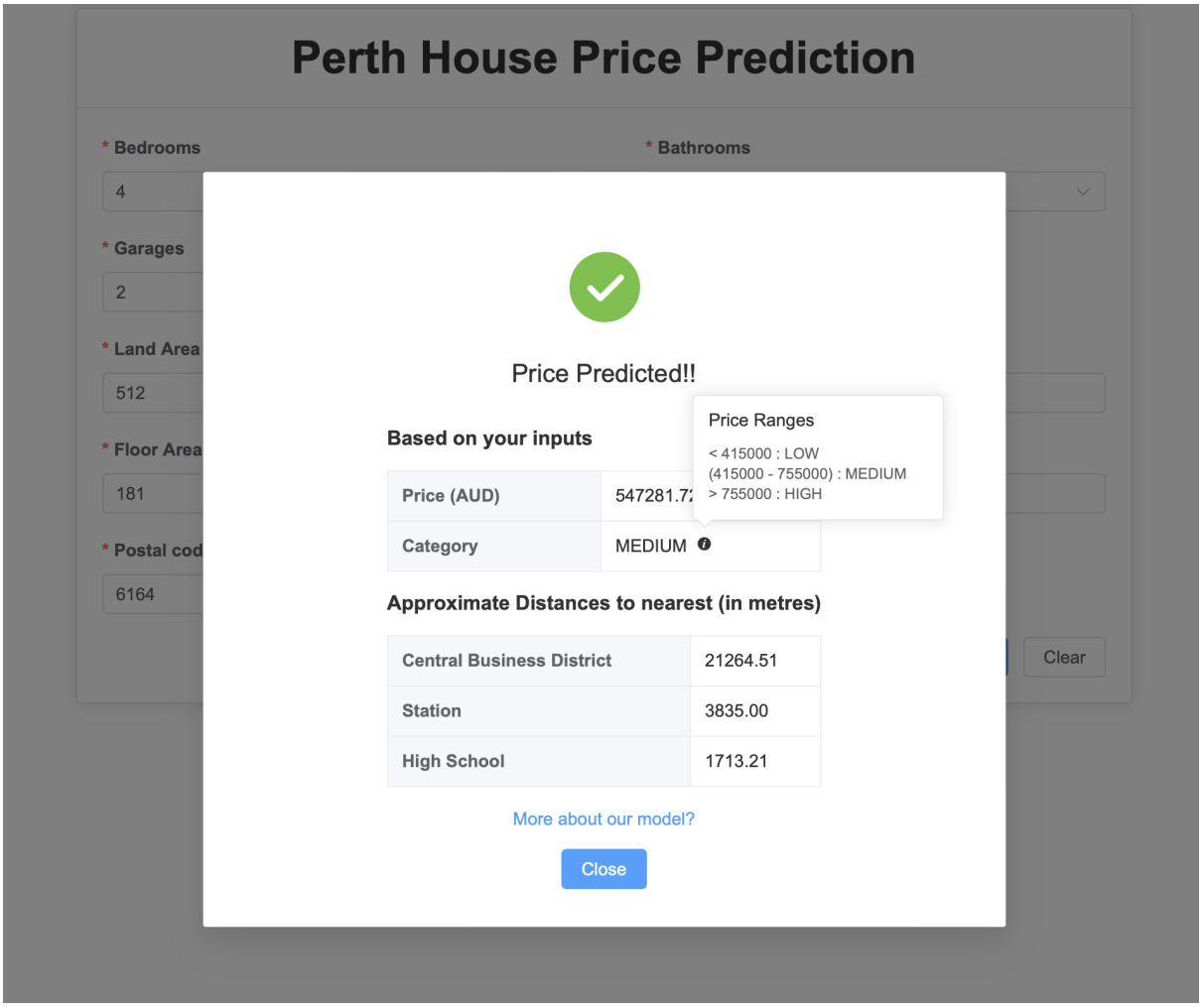
1. After running all the files, user gets the following web application :




After filling out the following entries the user needs to click get "Get Prediction!" button to get the results.

2. Example Usage:

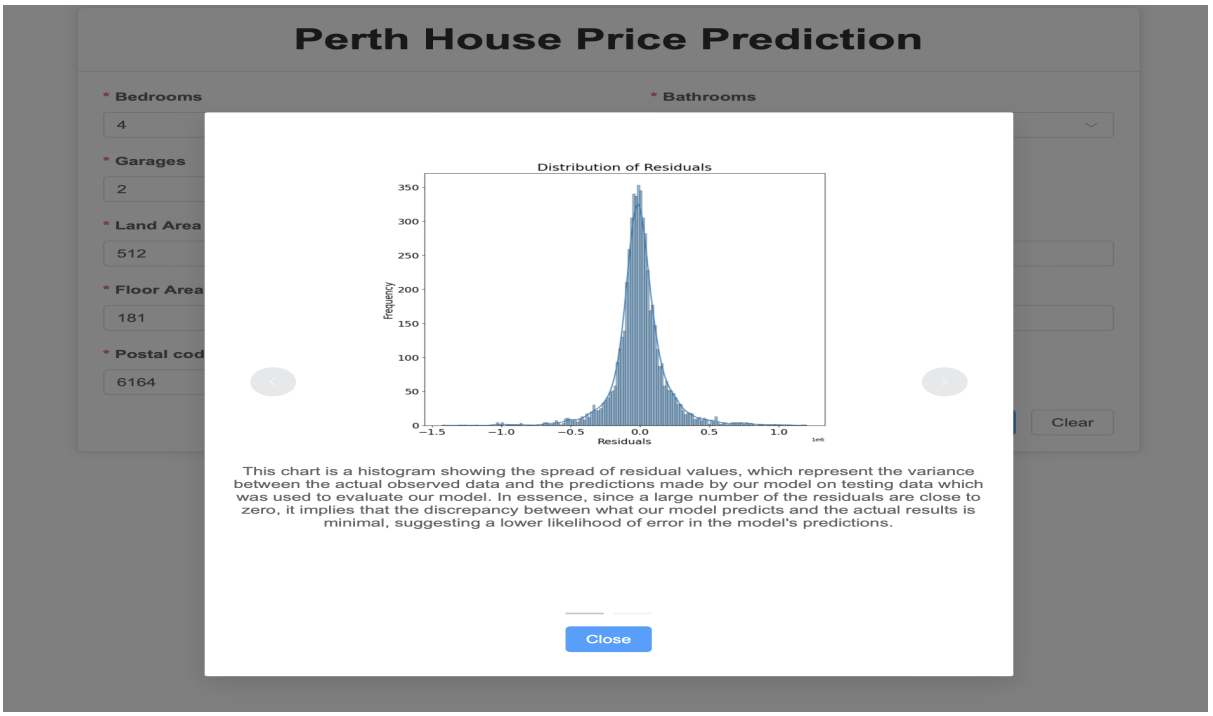


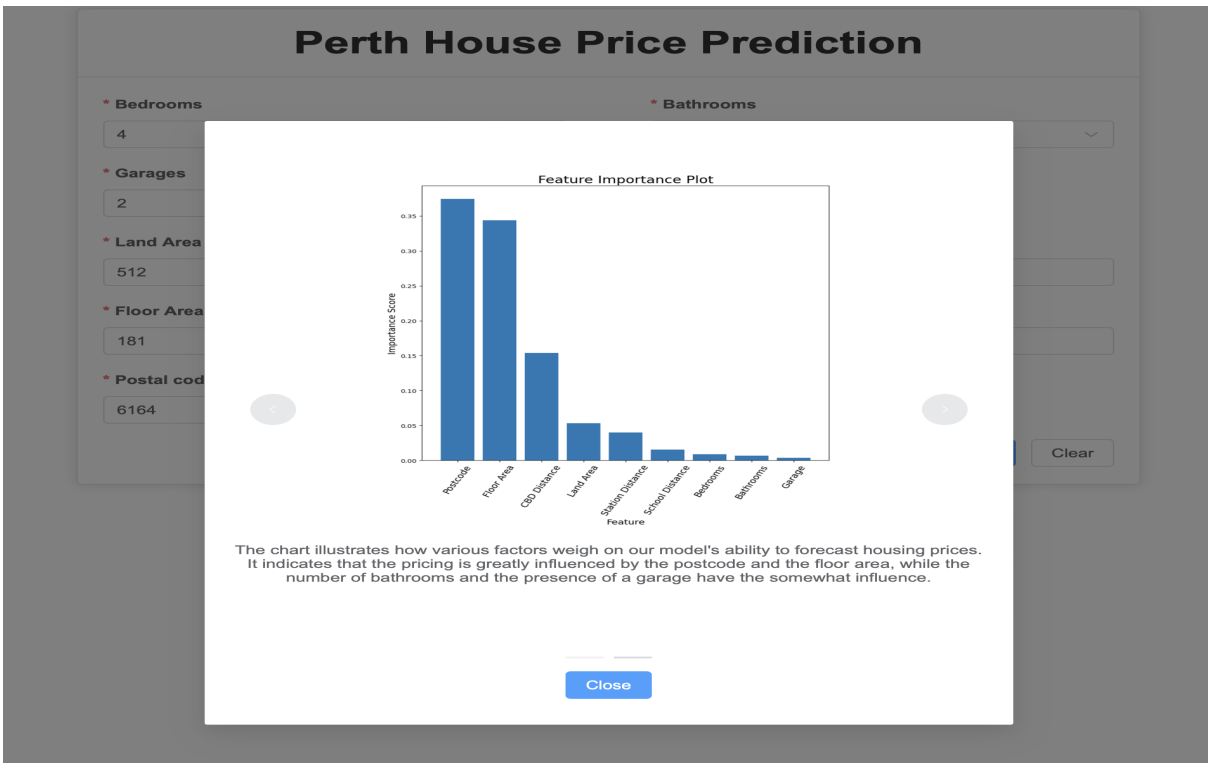



The dialogue provides the user with the following information:

- The price of the house.
- The category of the house.
- A brief tooltip that displays the range for various house categories.
- The distance to the nearest Central Business District (CBD) from the house.
- The distance to the closest train station from the house.
- The distance to the nearest school from the house.

3. To view visualization insights about our model the user needs to click on “More about our model?” link.




The residuals plot and feature importance plot provide insights into the model's performance during training, highlighting the effectiveness through the quality of the residuals plot. Additionally, these plots offer an understanding of which features significantly influence house prices.

Technologies used :

Flask
Vue 3
Element Plus


**Conclusion**

Our final product, a web application, solves this problem by using a machine learning model to predict house prices in Perth's neighborhoods. It considers important factors like location, house features, and land size. This helps both buyers and sellers understand if a house is priced high, low, or moderate. Accurate price predictions make real estate transactions more informed and fair for everyone. Plus, our application increases market transparency, making the real estate market more reliable and beneficial for all involved.

Expanding on our project, We can adapt our model for other cities or regions, not just Perth, by feeding it relevant local data.

# Peer Evaluation Form for Final Group Work

# CSE 487/587B

- Please write the names of your group members.

**Group member 1 : Yash Dattu Sonkamble**

**Group member 2 : Sameer Yadav**

**Group member 3 : Anirudh Anilkumar**

- Rate each groupmate on a scale of 5 on the following points, with 5 being HIGHEST and 1 being LOWEST.

| Evaluation Criteria | Group member 1 | Group member 2 | Group member 3 |
|---|---|---|---|
| How effectively did your group mate work with you? | 5 | 5 | 5 |
| Contribution in writing the report | 5 | 5 | 5 |
| Demonstrates a cooperative and supportive attitude. | 5 | 5 | 5 |
| Contributes significantly to the success of the project . | 5 | 5 | 5 |
| **TOTAL** | 20 | 20 | 20 |

**Also please state the overall contribution of your teammate in percentage below, with total of all the three members accounting for 100% (33.33+33.33+33.33 ~ 100%) :**

**Group member 1 : 33.33%**

**Group member 2 : 33.33%**

**Group member 3 : 33.33%**