**Week Five Progress Report**

|  |  |
| --- | --- |
| **Student:** | C21108612 |
| **Module:** | 21/22-MAT099 Dissertation |
| **Dissertation:** | AI-based Urban Analytics and Planning – Housing Price Prediction |

|  |  |
| --- | --- |
| **Sponsoring organisation:** | N/A |
| **Sponsor Supervisor:** | N/A |
| **University Supervisors:** | Prof. Yipeng Qin |

|  |  |
| --- | --- |
| **Progress:**  **(Summary of work carried out  to-date. Max 250 words)** | 1. Gathered datasets from Kaggle for 6 distinct Indian cities, along with the feature names and descriptions for each data file. 2. Exploratory data analysis and data cleansing completed 3. Handled missing data and duplicates from all the datasets through missing value analysis. 4. The data visualisation process is finished. 5. Based on the interests and heatmap value, I chose the following paired features to do a correlation analysis on. 6. Completed the label encoding from sk-learn to pre-process the data for categorical feature handling 7. Distributed the training and testing data using the Train-Test-Split method. |
| **Any concerns/issues:** | N/A |
| **Methodological approach to be used:**  **(Summary of what methods/approach you will use. Max 250 words)** | 1. I will use **linear regression** models to reduce the residual sum of squares between the targets in the dataset that were observed, and the targets predicted by the linear approximation. 2. The **Random Forest Regressor** utilises averaging to increase predicted accuracy and reduce overfitting by fitting a variety of classification decision trees to different subsamples of the dataset. 3. **Decision Tree Regressor** 4. In order to increase generalizability and robustness over a single estimator, I will be using **ensemble techniques** to integrate the predictions of numerous base estimators created with a given learning process. 5. Neural network Techniques need to be implemented. |
| **Planned Activities:**  (include important milestones) | 1. The project's feature scalability is currently being worked on. 2. Hyper-parameter tuning to look for the best cross validation score in the hyper-parameter space. 3. Evaluation of models and enhancement of model performance 4. To be accomplished: Feature Engineering 5. Final model selection |
| **Agreed Date for Final Presentation :**  (Both University and Company Supervisors to be present) | Friday – 9 December 2022 |