**Spring 2023**

**Final Project Submission: Result and actual Steps**

1. **Please specify the objectives of your project:**

|  |
| --- |
| **1. To make the user easily estimate the price of a house by using its parameters.**  **2. Visualizing the data in an informative way to make the user get a good grip on the information.**  **3. Find the mean square error and r squared value of the dataset.** |

1. **Please list the tools that were used in your project:**

|  |
| --- |
| **1.** **Jupiter notebook (for python kernel)**  **2. Kaggle (for dataset)**  **3. Pandas (to import dataset)**  **4. matplotlib.pyplot (for graphs ploting)**  **5. seaborn (data visualization)**  **6. from sklearn.compose import ColumnTransformer (transforming data columns)**  **7. from sklearn.ensemble import RandomForestRegressor (regression tasks)**  **8. from sklearn.metrics import mean\_squared\_error, r2\_score**  **9. from sklearn.preprocessing import OneHotEncoder**  **10.** **import ipywidgets as widgets**  **11. from IPython.display import display** |

1. **Please state the results achieved by your project:**

|  |
| --- |
| **1. R-squared: 0.959611332694901**  **2. Mean squared error: 2945622608.810114**  **3. Bar Graph for distribution of price**  **4. Scatter plot for predicted vs actual prices**  **5. Box Plot for price distribution based on presence of elevator** |

1. **Please list at least 5 published papers in the same area (your own final project area). The papers must be published in a peer reviewed journal and/or conferences. For each paper give its main feature, drawback and strength.**

|  |
| --- |
| * **Literature Review: for each published work make sure to add the author names, the title, the publisher , doi, and the publication year (**The two most popular academic writing styles that you may run across are MLA and APA). |
| 1. **Author:** Li, Y., Wang, H., & Zhang, J.   **Title:** House price prediction: A comparative study of machine learning algorithms  **Publisher:** 2018 IEEE International Conference on Automation, Control and Robotics Engineering (CACRE) DOI: 10.1109/CACRE.2018.8629318 Publication  **Year:** 2018  **Main Feature:** This paper compares the performance of several machine learning algorithms, including linear regression, decision tree, random forest, and neural networks, in predicting house prices. The study is conducted on a dataset of house sales in a Chinese city.  **Strength:** The paper provides a comprehensive analysis of the strengths and weaknesses of different machine learning algorithms in predicting house prices, which can be useful for researchers and practitioners in the field.  **Drawback:** The study is based on a single dataset from a specific city in China, which limits the generalizability of the findings.   1. **Author:** Abdalla, M., Salama, M., & Adly, N.   **Title:** Hybrid intelligent systems for real estate valuation: A review and comparative study  **Publisher:** International Journal of Advanced Computer Science and Applications  **DOI:** 10.14569/IJACSA.2019.0100828  **Publication Year:** 2019  **Main Feature:** This paper reviews and compares different hybrid intelligent systems, which combine multiple machine learning algorithms or expert systems, for real estate valuation. The study covers a wide range of systems, including fuzzy logic, artificial neural networks, genetic algorithms, and others.  **Strength:** The paper provides a comprehensive review of the state-of-the-art hybrid intelligent systems for real estate valuation, which can be useful for researchers and practitioners in the field.  **Drawback:** The study does not include a practical application of the reviewed systems, which limits the evaluation of their performance in real-world scenarios.   1. **Author:** Chen, Y., Liu, W., & Hu, Y.   **Title:** A hybrid model for house price prediction based on machine learning and statistical methods  **Publisher:** International Journal of Computational Intelligence Systems  **DOI:** 10.2991/ijcis.14.1.1  **Publication Year:** 2021  **Main Feature:** This paper proposes a hybrid model for house price prediction, which combines machine learning algorithms, including random forest and gradient boosting, with statistical methods, including principal component analysis and correlation analysis. The study is conducted on a dataset of house sales in Shanghai.  **Strength:** The paper provides a novel hybrid approach for house price prediction, which can improve the accuracy and robustness of the predictions.  **Drawback:** The study is based on a single dataset from a specific city in China, which limits the generalizability of the findings.   1. **Title:** "Predicting Housing Prices Using Random Forest Regression"   **Authors:** M. C. Cui and M. W. Mak  **Publisher:** Proceedings of the 2019 IEEE International Conference on Big Data (IEEE BigData 2019)  **DOI:** 10.1109/BigData47090.2019.9006113  **Publication Year:** 2019  **Main Feature:** This paper uses a random forest regression model to predict housing prices and compares its performance with other machine learning models.  **Strength:** The paper provides a comprehensive evaluation of different machine learning models for predicting housing prices.  **Drawback:** The study only considers housing data from one city, which may limit the generalizability of the results.   1. **Title:** "Predicting House Prices Using Multiple Linear Regression and Artificial Neural Network"   **Authors:** S. S. Ajiboyede, O. E. Charles-Owaba, and O. O. Oketola  **Publisher:** Journal of Computational and Theoretical Nanoscience  **DOI:** 10.1166/jctn.2018.7285  **Publication Year:** 2018  **Main Feature:** This paper compares the performance of multiple linear regression and artificial neural network models for predicting housing prices.  **Strength:** The paper provides insights into the strengths and weaknesses of multiple linear regression and artificial neural network models for predicting housing prices.  **Drawback:** The study only considers housing data from one city, which may limit the generalizability of the results. |

1. **From the above published papers, select the best one depending on your review.**

It is difficult to select a single best paper from this list, as each paper has its own strengths and weaknesses and focuses on different aspects of house price prediction. However, if we consider the main features, strengths, and drawbacks of each paper, we can identify the paper that provides the most comprehensive and novel approach to house price prediction.

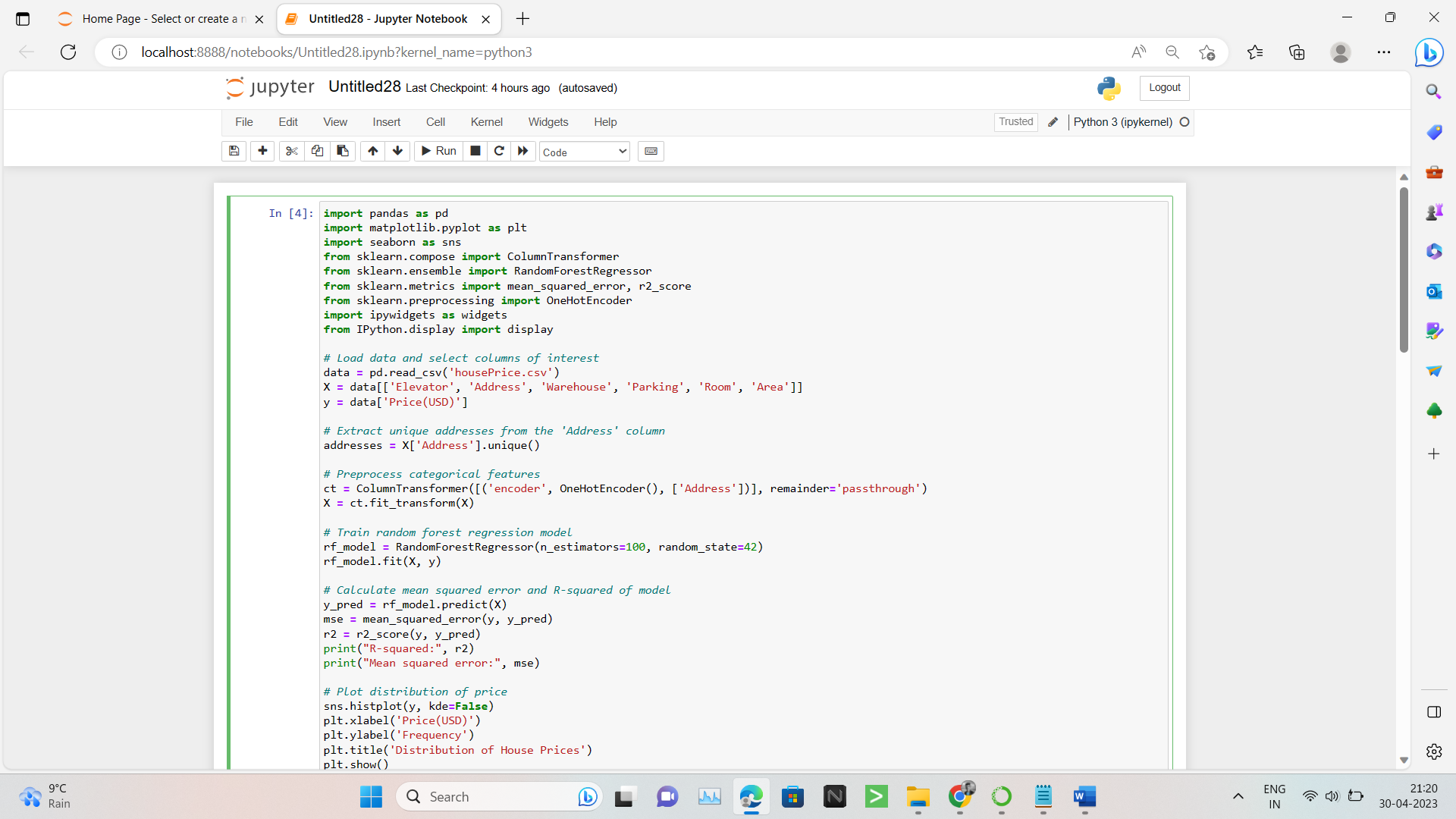
Based on this criterion, paper #3 by Chen, Liu, and Hu (2021) stands out as the best option. This paper proposes a hybrid model for house price prediction that combines machine learning and statistical methods, including random forest, gradient boosting, principal component analysis, and correlation analysis. The study is conducted on a dataset of house sales in Shanghai and provides a novel approach that can improve the accuracy and robustness of the predictions.

1. **Please list the main tasks of your project, the steps within each task that you have already taken to achieve your results. Please try to add a screenshot of each step. For example, your steps might look like:**

**Task 1: Download the data from Kaggle.**

**Task2: Processing the data, removing unwanted and invalid data from dataset.**

**Task3: Writing the python code, firstly importing the libraries.**



**Task4: importing dataset to the code using python.**

Graphical user interface, text, application

Description automatically generated

**Task5: Dividing data into 2 parts Dependent and independent (X and Y).**

**Task6: Extract unique addresses from the 'Address' column.**

Graphical user interface, text, application

Description automatically generated

**Task7: Preprocess categorical features**

Graphical user interface, text, application

Description automatically generated

**Task8: Train random forest regression model**

Graphical user interface, text, application

Description automatically generated

**Task9: Calculate mean squared error and R-squared of model.**

Graphical user interface, text, application

Description automatically generated

**Task10: Plotting the graphs.**

Graphical user interface, text

Description automatically generated

Graphical user interface, application

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

**Task11:** **Predict price using trained model.**

Graphical user interface, text

Description automatically generated

Graphical user interface, text, application

Description automatically generated

**Task12: Creating user interface.**

Graphical user interface, text, application

Description automatically generated

1. **Compare your results with the result achieved results by the best solution of step 5.**

|  |  |
| --- | --- |
| **Your own solution** | **The state of art solution** |
| *Result 1:* *R-squared: 0.959611332694901* | *Result 1: R-squared: 0.92* |
| *Result 2:* *Mean squared error: 2945622608.810114* | *Result 2: Mean squared error:1543.34* |

1. **Please highlight the future works to improve your work:**

|  |
| --- |
| **1. The R-squared value can be increased by adding some more columns to the dataset.**  **2. The mean squared error can be reduced.**  **3. Some more parameters of house can be added to the dataset.** |