

**Name: Hafiz Haider Ali**

**Roll No: BSAI-130-4C**

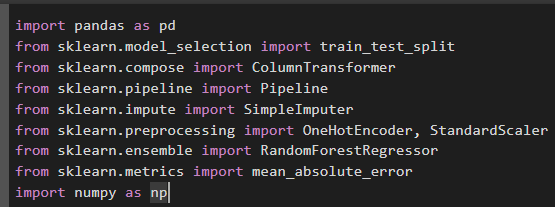
**Date: 27 Feb 2025**

**Subject: PAI LAB**

**House Price Prediction**

This Python program builds a house price prediction model using Random Forest Regression. The workflow follows these key steps:

**1. Importing Necessary Libraries:**



* **pandas**
* **train\_test\_split**
* **ColumnTransformer**
* **Pipeline**
* **SimpleImputer**
* **OneHotEncoder**
* **RandomForestRegressor**
* **mean\_absolute\_error (MAE)**
* **numpy**

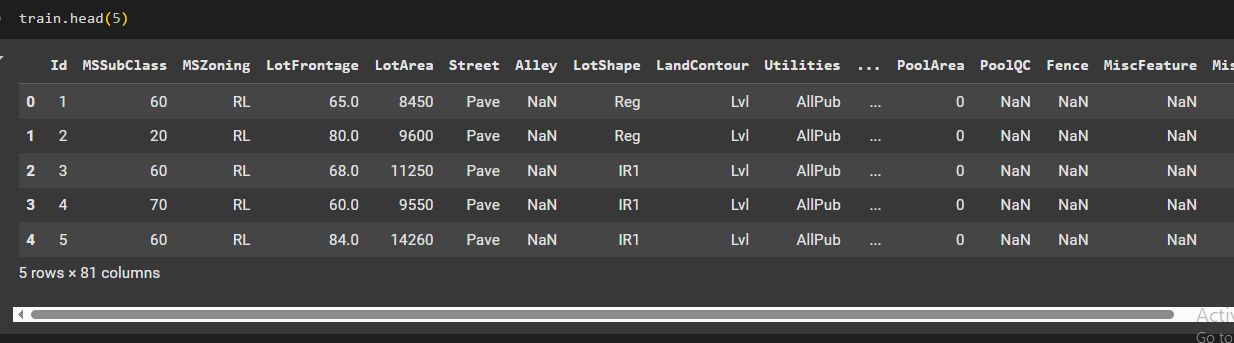
**2. Loading the Data:**

Loads the training and testing datasets from CSV files.



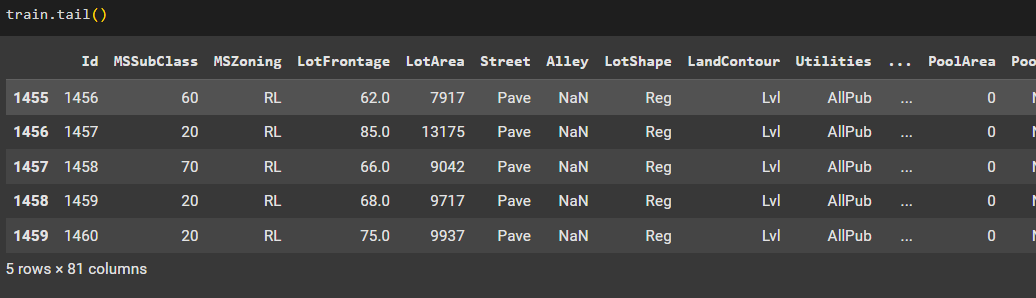
**3. train.head():**

* Helps in getting a quick overview of the structure and values.



**5. train.tail():**

* Displays the last five rows of the training dataset.
* Useful for checking missing values or inconsistencies at the end of the dataset.



**9. Splitting Features and Target Variable:**



* y (target): Stores the house sale prices.
* X (features): Stores the rest of the dataset, excluding the target variable.

**10. Splitting Data into Training and Testing Sets:**



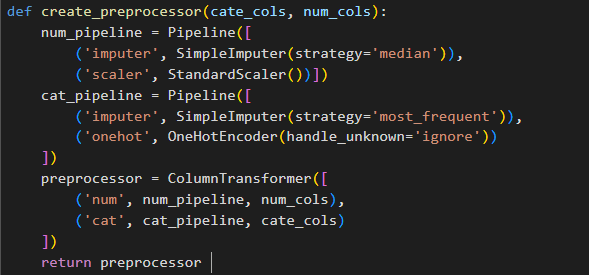
* 80% of the data is used for training (X\_train, y\_train).
* 20% is reserved for testing (X\_test, y\_test).

**11. Identifying Categorical and Numerical Columns:**



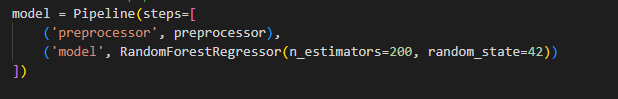
* Categorical columns: Selected if they have a data type of object and contain fewer than 10 unique values.
* Numerical columns: Selected if they are of type int64 or float64.

**12. Data Preprocessing:**



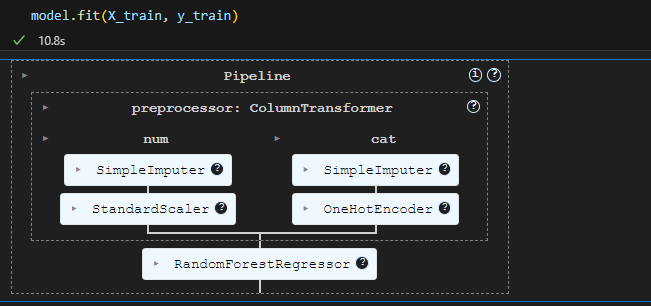
* **Numerical Data:**
* Missing values are replaced with the median using SimpleImputer(strategy='median').

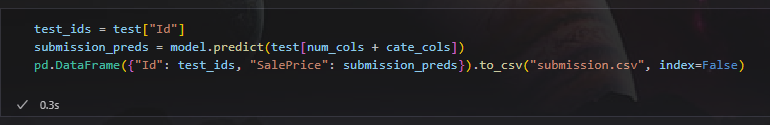
**13. Building the Model Pipeline:**



* The Pipeline ensures that preprocessing and model training happen in sequence.
* RandomForestRegressor is used with

**14. Training the Model:**



****

* The trained model makes predictions on the test dataset.
* Results are saved as submission.csv with Id and SalePrice.

**Accuracy:**

