

Final Assignment



In this assignment, you will work on a house price prediction project using the "House Prices: Advanced Regression Techniques" dataset from Kaggle. The focus is on data cleaning and preprocessing tasks, including handling missing values, outliers, and categorical features. You will train a regression model to predict house prices and optimize its performance.

Assignment-7

Task 1: Data Understanding:

Begin by exploring the dataset and understanding its structure, including the meaning and type of each feature. (**Dataset attached with your email**)

Task 2: Data Cleaning:

Perform data cleaning tasks to handle missing values, outliers, and inconsistencies in the dataset. Some potential cleaning tasks include:

- Handling missing values: Identify features with missing values and decide on an appropriate strategy to handle them (e.g., imputation or removal).
- Dealing with outliers: Identify outliers in numerical features and decide on an appropriate approach (e.g., removing outliers or transforming the data).
- Handling categorical features: Convert categorical variables into numerical representations using techniques like one-hot encoding or label encoding.

Task 3: Feature Engineering:

Perform feature engineering to enhance the predictive power of the dataset. This may include creating new features, transforming existing features, or selecting relevant features.

Task 4: Data Preprocessing:

Prepare the cleaned dataset for model training. This involves scaling numerical features, encoding categorical variables, and splitting the data into training and testing sets.

Task 5: Model Training and Evaluation:

Choose an appropriate regression model (e.g., linear regression, random forest, or gradient boosting) and train it on the preprocessed dataset. Evaluate the model's performance using suitable metrics like mean squared error (MSE) or root mean squared error (RMSE).

Task 6: Model Optimization:

Fine-tune the hyperparameters of the chosen model to improve its performance. You can use techniques like cross-validation or grid search to find the best parameter values.

Final Assignment



Deploying the machine learning model on an Amazon EC2 instance involves setting up a virtual server, installing the necessary dependencies, transferring the trained model, and exposing an API endpoint for making predictions. This enables the model to be accessible and scalable for real-time prediction requests.

Assignment-8

Task 7: Model Deployment:

Once you have a satisfactory model, deploy it to make predictions on new, unseen data. You can use the trained model to predict house prices for new instances and assess its real-world applicability.

Task 8: LinkedIn Post:

Once you complete all the above tasks, make a linkedin post from your account for the entire Final Assignment completion.

Submission: The Entire assignment should be submitted by the date mentioned in the email, You have to upload all the files you created in this assignment in GitHub with the proper Repo Name & also upload a Screenshot of the working deployed ML project in your browser with the HTTP link of your AWS host link clearly.