**Swapna Emmadishetty**

**semmadishetty@regis.edu**

**Accommodation fare Prediction**

**Project Description:**

The objective of this project is to create a machine learning model that can precisely predict the fare of accommodations based on various characteristics, including the number of bedrooms, bathrooms, area, locality, furnishings, tenants, and facing. The property data dataset, which includes training, testing, and validation datasets, will be used to train the model.

**Methodology:**

The following approach will be taken to develop the machine learning model for fare prediction:

1. **Data Cleaning:** The data requires cleaning and pre-processing steps such as string extraction, imputation, categorical variables, variable encoding, feature engineering, etc. This will be done to ensure the data is ready for model training and prediction.
2. **Feature Selection:** After cleaning the data, feature selection techniques such as correlation matrix, PCA, and feature importance will be used to identify the most important features for model training and prediction.
3. **Model Selection:** Various machine learning models such as Linear Regression, Random Forest, and XGBoost will be trained and evaluated on the training and testing datasets. The model with the best performance will be selected for final prediction on the validation dataset.
4. **Hyperparameter Tuning:** The selected model will undergo hyperparameter tuning to optimize the model's performance and reduce overfitting.
5. **Prediction:** The final trained model will be used to predict the rent of houses in Hyderabad based on the features of the validation dataset.

**Conclusion:**

The objective of this project is to create a machine-learning model that can predict the fare of accommodation with accuracy using a variety of features. The model's performance will be enhanced to achieve the lowest MSE by implementing data cleaning, feature selection, model training and selection, hyperparameter tuning, and prediction.

**Source of data:**<https://www.kaggle.com/datasets/saivenkat09/hyderabad-properties>

**Data Brief:**

Three datasets (train, test, and validate) that correspond to the property data are contained in the property data folder. Bedrooms, Bathroom, Facing, Furnishings, Tenants, Area, Price, Locality, and Location are the columns that make up the data. The Price column, which is not shared with the validate dataset, serves as the Y variable for the prediction. The pin code details folder also contains two additional files (Location.txt & Pincode.txt) that contain location and pin code mapping.

**Timeline:**

* Week 3 – Data exploration and cleaning
* Week 4 – Data visualisation
* Week 5 – Feature Selection and Model training
* Week 6 – Model selection
* Week 7 – Hyper parameter-tuning and Finalisation
* Week 8 \_ Presentation