**Approach**

**Problem Statement:**

To predict Chennai regions house selling price

**Features and Target:**

|  |  |
| --- | --- |
| Features | Numeric/Category |
| PARK\_FACIL | Label |
| INT\_SQFT | Number |
| DATE\_SALE | Number |
| DIST\_MAINROAD | Number |
| N\_BEDROOM | Number |
| N\_BATHROOM | Number |
| N\_ROOM | Number |
| DATE\_BUILD | Number |
| QS\_ROOMS | Number |
| QS\_BATHROOM | Number |
| QS\_BEDROOM | Number |
| QS\_OVERALL | Number |
| REG\_FEE | Number |
| COMMIS | Number |
| AREA | One Hot |
| SALE\_COND | One Hot |
| BUILDTYPE | One Hot |
| UTILITY\_AVAIL | One Hot |
| STREET | One Hot |
| MZZONE | One Hot |
| SALES\_PRICE (Label) | Number |

**Steps:**

* Get the Data
* Create the test set
* Stratified Sampling
* EDA [ Data Visualization such as Univariate and Bivariate]
* Experiment new attributes
* Prepare the Data for ML Algorithm
* Looking for Correlations
* Data Cleaning

1. Feature engineering
2. Missing values imputation by median
3. Encoding
4. Feature Scale

* Pipeline and Transformation
* Feature Importance
* Statistics
* Select and Train Model

1. Linear Regression
2. Decision Tree
3. Random Forest
4. KNN Regression
5. Gradient Boost
6. XG Boost

**Accuracy**

|  |  |  |
| --- | --- | --- |
| Model | Accuracy | Comment |
| Linear Regression | 94% | Good fit |
| Decision Tree | 100% | Over-Fit |
| Random Forest | 99.10% | Over-Fit |
| KNN Regression | 90% | Good Fit |
| Gradient Boost | 91% | Good Fit |
| XG Boost | 96% | Good fit |

**Insight:**

1. Number of Square feet, Number of rooms and Number of bed room are top 3 predictor of Selling price of home
2. T Nagar and Anna Nagar are top two highest selling price of home
3. Maximum Square feet constructed around 2000 square feet in last 3 years
4. Sales Prices of housing is declined in 2009,2010 and 2011
5. Commercial building has more business than homes
6. Home with parking lot has 5% higher price than without parking lot
7. Average age of building is 20 years to 25 years. There is always scope of new upcoming projects in coming years