

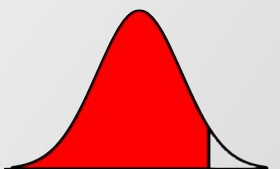


Predicting Rental Prices with Machine Learning: Insights and Applications

Hung-Cheng Chang, 張宏正(Jack)

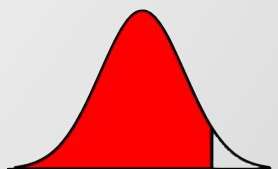
Outline

1. Motivation
2. Flow Chart
3. Data
4. Method
5. Result

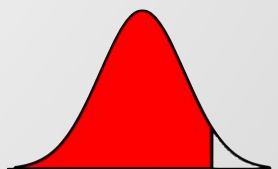
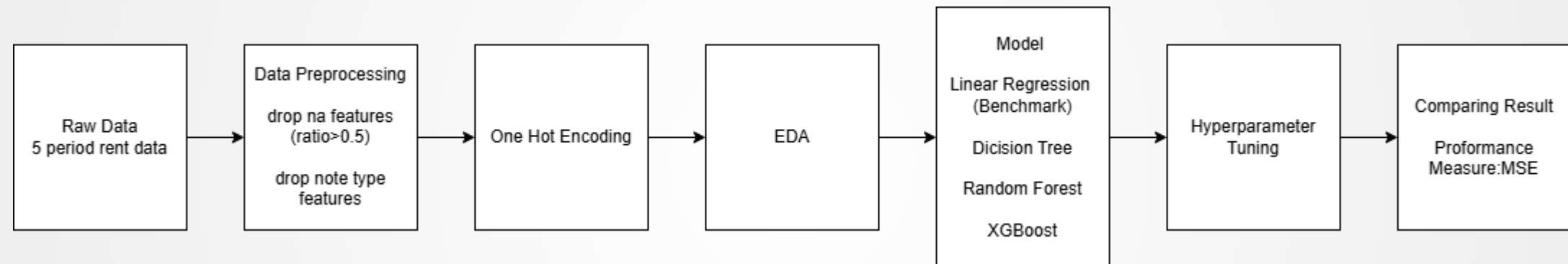


Motivation

- In recent years, the housing issue in Taiwan has become a hot topic, which has led to a sharp rise in housing prices.
- Rent plays an important role in the housing market, directly reflecting the supply and demand relationship.
- It is hoped that the model can serve as a benchmark for both landlords and tenants in deciding rental prices.

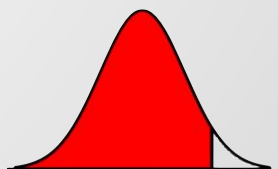


Flow Chart



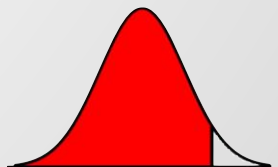
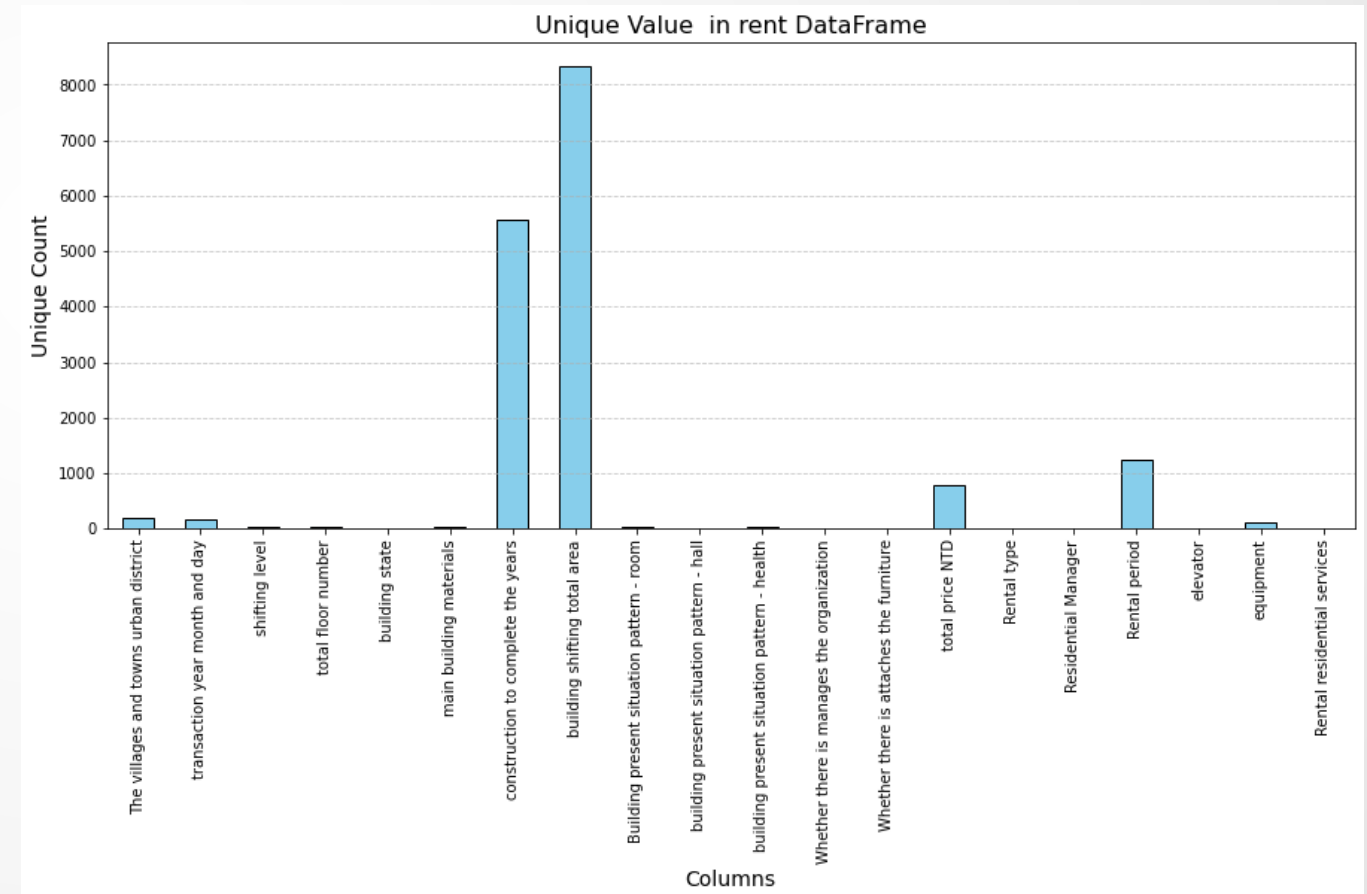
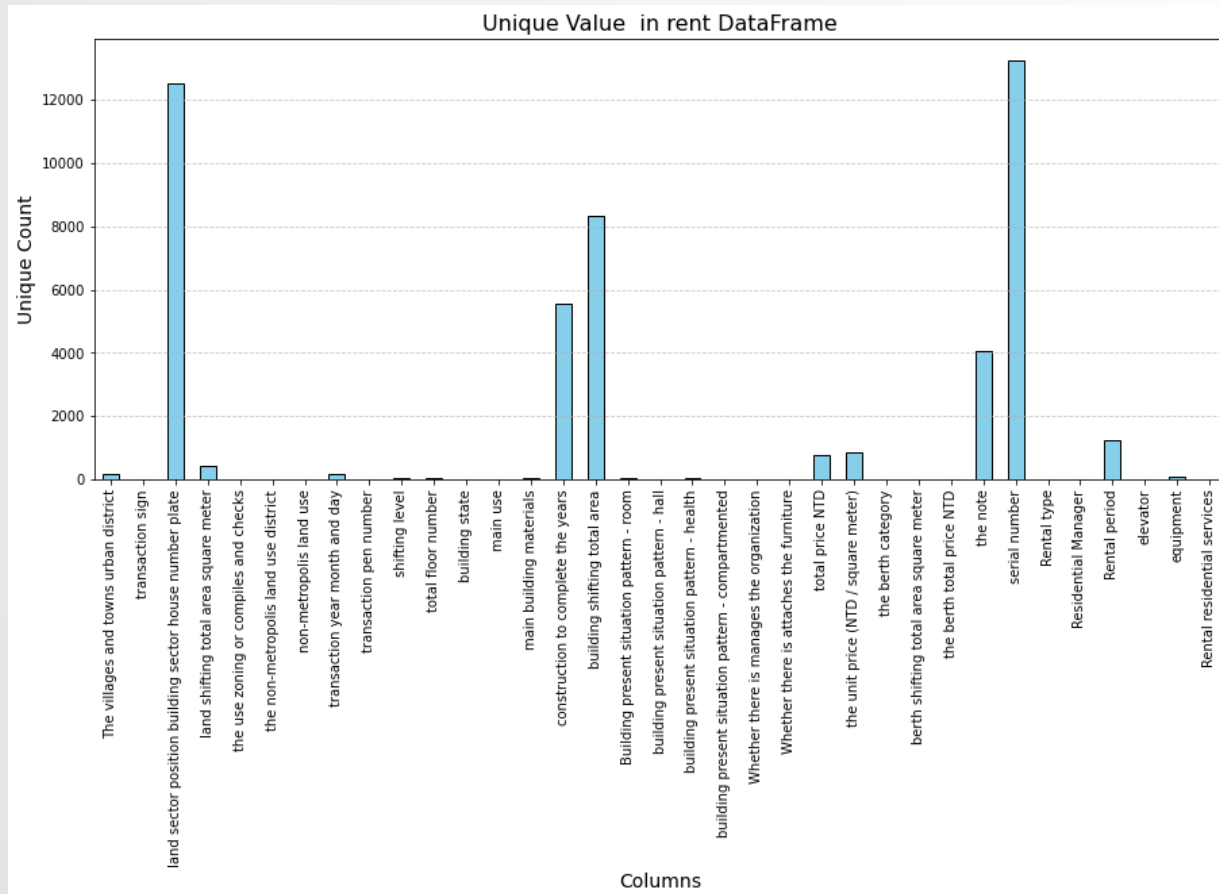
Data

- Source: 不動産成交案件 (Dept of Land Administration M. O. I.)
- Release Date: 240811~240921 (Total 5 periods)
- Size: $n=36498$, $p=35$



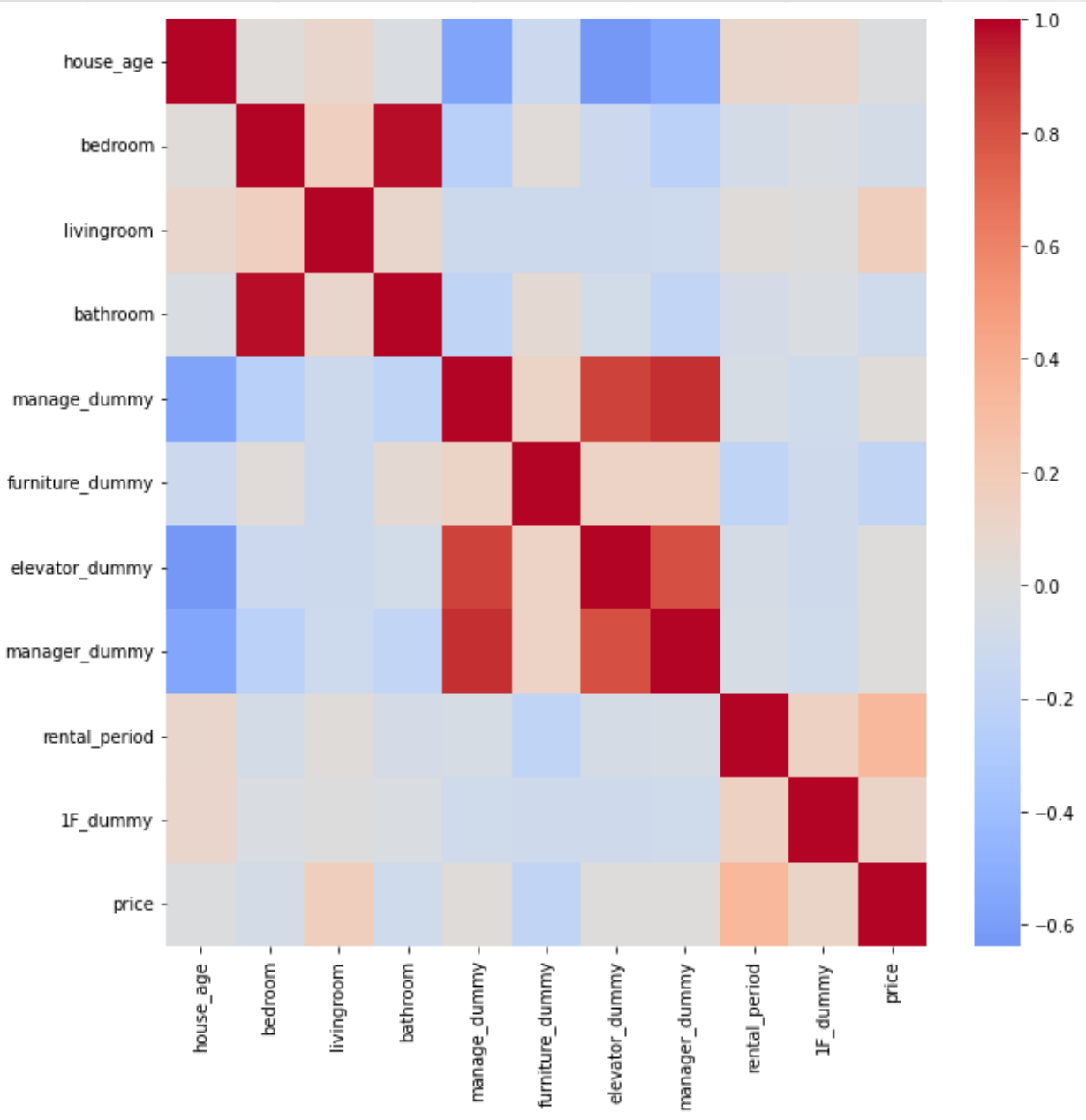
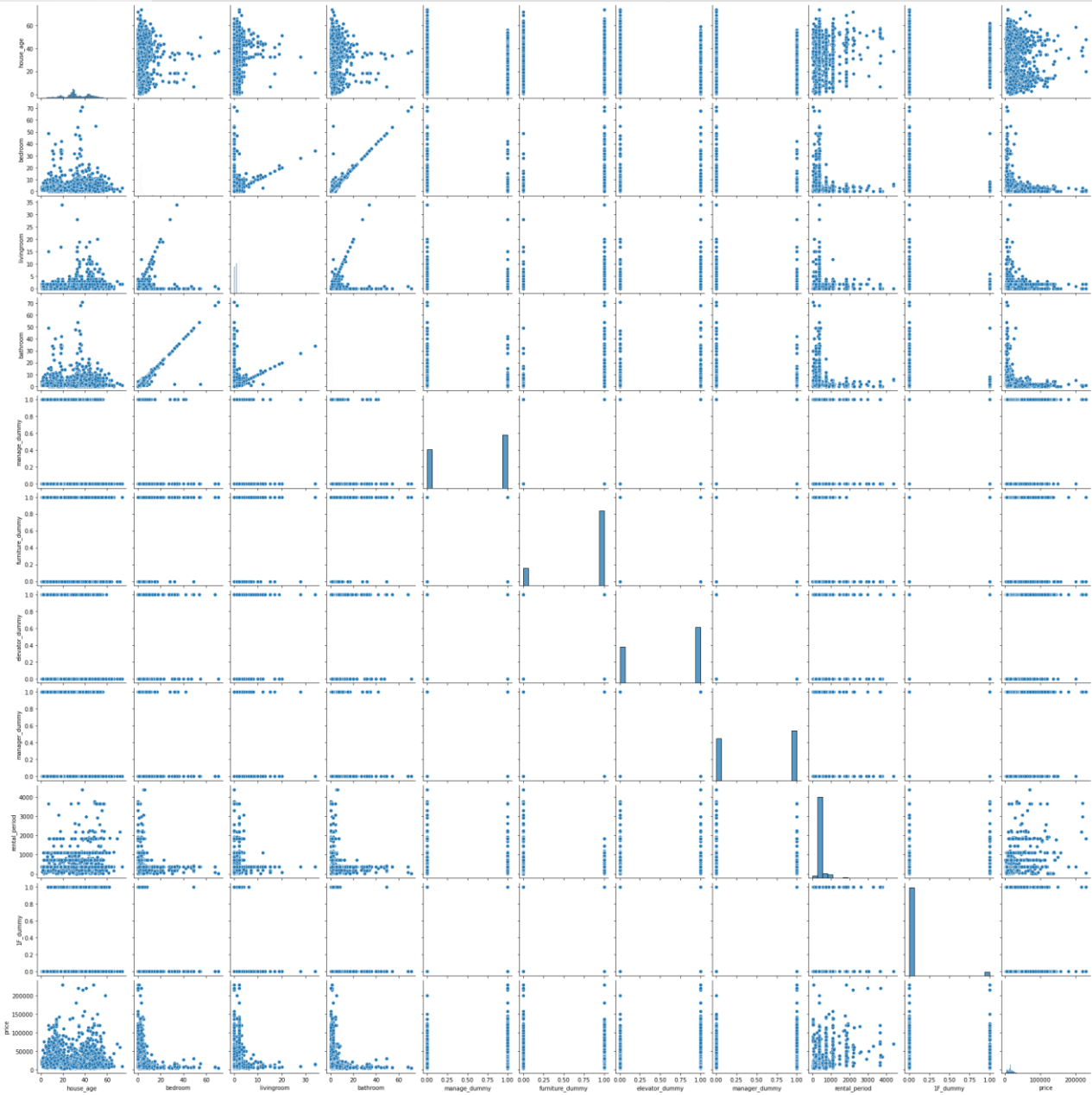
Data

- Drop nan ratio > 0.5 features
- Drop note type features



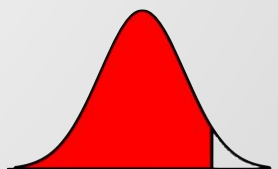
Data

| | house_age | bedroom | livingroom | bathroom | manage_dummy | furniture_dummy | elevator_dummy | manager_dummy | rental_period | 1F_dummy | price |
|-------|-------------|------------|-------------|-------------|--------------|-----------------|----------------|---------------|---------------|-------------|-------------|
| count | 18482 | 18482 | 18482 | 18482 | 18482 | 18482 | 18482 | 18482 | 18482 | 18482 | 18482 |
| mean | 32.05943317 | 3.0226166 | 1.174494102 | 2.231630776 | 0.578130073 | 0.812736717 | 0.607780543 | 0.541662158 | 416.7769722 | 0.041499838 | 16107.43096 |
| std | 11.97784959 | 4.29126347 | 1.275701766 | 4.296573991 | 0.493871328 | 0.390133286 | 0.488258388 | 0.498274722 | 232.8556602 | 0.199448624 | 12317.80348 |
| min | 0.361643836 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2000 |
| 25% | 26.29041096 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 364 | 0 | 9500 |
| 50% | 30.49315068 | 3 | 1 | 1 | 1 | 1 | 1 | 1 | 364 | 0 | 14500 |
| 75% | 42.45890411 | 3 | 2 | 2 | 1 | 1 | 1 | 1 | 364 | 0 | 19000 |
| max | 73.71232877 | 71 | 34 | 71 | 1 | 1 | 1 | 1 | 4382 | 1 | 230000 |



Method

- Linear Regression(benchmark)
- Decision Tree
- Random Forest
- XGBoost



Method

• Linear Regression(benchmark)

No hyperparameter

• Decision Tree

| | | | | |
|-------------------|------|----|----|----|
| Decision Tree | | | | |
| max_depth | None | 10 | 20 | 30 |
| min_samples_leaf | | 1 | 5 | 10 |
| min_samples_split | | 2 | 10 | 20 |

• Random Forest

| | | | | |
|-------------------|------|----|-----|-----|
| Random Forest | | | | |
| n_estimators | | 50 | 100 | 200 |
| max_depth | None | 10 | 20 | 30 |
| min_samples_split | | 2 | 10 | 20 |

• XGBoost

| | | | | |
|---------------|--|------|-----|-----|
| XGBoost | | | | |
| n_estimators | | 50 | 100 | 200 |
| learning_rate | | 0.01 | 0.1 | 0.2 |
| max_depth | | 3 | 5 | 10 |

• Using GridSearchCV scoring:MSE, CV=5



Result

Linear Regression(benchmark)

No hyperparameter

Dicision Tree

| | | | | |
|-------------------|------|----|----|----|
| Dicision Tree | | | | |
| max_depth | None | 10 | 20 | 30 |
| min_samples_leaf | | 1 | 5 | 10 |
| min_samples_split | | 2 | 10 | 20 |

Random Forest

| | | | | |
|-------------------|------|----|-----|-----|
| Random Forest | | | | |
| n_estimators | | 50 | 100 | 200 |
| max_depth | None | 10 | 20 | 30 |
| min_samples_split | | 2 | 10 | 20 |

XGBoost

| | | | | |
|---------------|--|------|-----|-----|
| XGBoost | | | | |
| n_estimators | | 50 | 100 | 200 |
| learning_rate | | 0.01 | 0.1 | 0.2 |
| max_depth | | 3 | 5 | 10 |



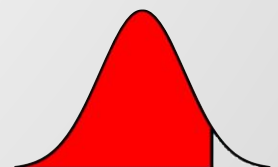
Result

| | Validation | | TEST | |
|-------------------|-------------|------|-------------|------|
| | MSE | R2 | MSE | R2 |
| Linear Regression | 63668686.28 | 0.55 | 71618550.41 | 0.52 |
| Dicision Tree | 57766098.56 | 0.59 | 91632634.07 | 0.38 |
| Random Forest | 30186153.49 | 0.79 | 42628068.83 | 0.71 |
| XGBoost | 31570006.62 | 0.78 | 32161241.09 | 0.78 |



Result

| | Validation | | TEST | |
|-------------------|-------------|------|-------------|------|
| | MSE | R2 | MSE | R2 |
| Linear Regression | 63668686.28 | 0.55 | 71618550.41 | 0.52 |
| Dicision Tree | 57766098.56 | 0.59 | 91632634.07 | 0.38 |
| Random Forest | 30186153.49 | 0.79 | 42628068.83 | 0.71 |
| XGBoost | 31570006.62 | 0.78 | 32161241.09 | 0.78 |



Result

