



Data Preprocessing

- Feature selection
- Investigating missing data
- Remove variable(more than 50% data)
- Remove ambiguity attributes & add attributes
- Set minimum values for attributes
- Check for outliers use Inter-Quartile Range (IQR)

Data Analysis

- Discover the implicit patterns of the data
- By the patterns, observe correlation between attributes

Model Selection

- 3 Type machine learning
 1. **Randon forest**
 2. **XGBoost**
 3. **Light GBM**
- 2 Type techniques
 1. **Hybrid Regression** - 65% Lasso and 35% XGBoost
 2. **Stacked Generalization Regression** - 2-level stacking architecture (first stacking level consisted of Random Forest and LightGBM and the second stacking level was XGBoost)

Model Validate

- Check overfitting, accuracy, RMSLE