LightGBM Model For Predict Future Sales

Kaggle Competition

Table of Contents

Data Pre-processing

Applying EDA (Exploratory Data Analysis).

Feature Engineering

- Mean encoding
- Lag features

Mean encoding

Calculate mean for groupby features.

Example:

mean(groupby(shop_id))

→ return number of products/month of that shop

Lag features

Use **past** values as features for predicting **future** values.

Model - LightGBM

A gradient boosting framework that uses tree-based learning algorithms

Model - LightGBM’s key features

- Optimization in Speed and Memory Usage
- Optimization in Accuracy
- Support for parallel and GPU learning
- Optimal for categorical features
- Handle large-scale data

Model - How LightGBM's work

- **Gradient Boosting**: uses gradient boosting framework.
- **Decision Trees**: Builds decision trees one at a time, where each new tree corrects errors made by the previously trained tree.
- **Leaf-wise Growth:** Grows tree leaf-wise rather than level-wise.

Model - Hyperparameters

- Learning rate
- Number of trees
- Max depth
- Feature fraction
- Bagging fraction

Data post-processing

Reference:

https://www.kaggle.com/code/abubakar624/first-place-solution-kaggle-predict-future-sales