

PREDICT FUTURE SALE

MingZhu Kang

¹ Xi'an Shiyou University, China

Introduction

In deep learning projects, data visualization operations often need to be carried out, including visualization of original image data, visualization of loss and accuracy, etc.

In this study, the data in the data set are collated and analyzed to analyze the impact of commodity sales volume, product category, operating income and the economic scale of the city where the stores are located on the income.

The model is used to train the data set, and the historical sales data is used as the model feature to realize the commodity sales forecast. Effectively guide the store to conduct reasonable inventory management.

Data Sets

- Preprocessing of project data sets
The preprocessing of project data set includes training set, commodity set, commodity data set, commodity category set and test set.
through the data of the training set we determined the parameters of the fitting curve to filter the obvious outliers and checked the abnormal conditions of prices.
- Training set data cleaning
through the data of the training set, we determined the parameters of the fitting curve.
to filter the obvious outliers and checked the abnormal conditions of prices and sales

Structured Data And Analysis

Sale Analysis
According to the data analysis, to the overall sales volume of the store showed a downward trend. and the monthly sales volume was mostly lower than the same period of last year

TurnOver Analysis
Revenue in the 23rd month was up sharply from the 11th month, while sales were down year-on-year. One item's gross revenue was unusually high.

Product Category Influence
The influence of product category on sales volume The first three categories that contribute the most to total sales are the 40th, 30th and 55th categories.

Impact Analysis
The impact of the city in which the store is located on total revenue.
Stores in district 14 cities contributed the most to the total revenue.

Predict Future Trend

By processing closed stores and goods that are not for sale, Keep only stores open for the last six months and items that sell. Combining the features of the store data set and the merchandise data set.

The goal is to use historical sales data to predict future sales. Using the historical sales data as the characteristics of the model, this month's sales results as labels to build a model for regression analysis.

Model Training

LightGBM Model is a fast, distributed, high-performance gradient enhancement framework based on decision tree algorithms. It supports category characteristics.

This project uses lightGBM model for training. select lightGBM model for training and combine the predicted results into the test set. LightGBM supports category characteristics directly and natively by changing the decision rules of the decision tree algorithm, without transformation.

Teams	Pre-ranking algorithm	LightGBM
Segment gain calculation	$O(\text{\#feature\#data})$	$O(\text{\#feature})$
Direct variance mapping	N/A	Double the speed
Direct support category features	N/A	8 times faster on Expo data
Accumulation of statistic	$O(\text{\#feature\#data})$	$O(\text{\#feature\#data})$
Cache optimization	N/A	accelerate 40% on Expo data

Conclusion

Through this project, I have learned a lot, including the effective aspects of problem cutting, code implementation of analysis algorithm, design of analysis process, etc.

which enables me to better grasp the thinking of data analysis on the whole. From data analysis methods to feature engineering and prediction model construction. a lot of time has been spent to study and comb.

In feature processing, there is also a feature of the commodity category not added to the forecasting model.The main reason is that the category is not added to the forecasting model.The main reason is that the category is not added to the forecasting model.

deficiencies in the treatment of price changes, which needs to be strengthened.

In the process of predictive analysis, the feature engineering and model construction is not perfect, which needs to be strengthened.

Acknowledgement
•International Cooperation Project (Y7Z0511101)
of IIE, Chinese Academy of Sciences