

# PREDICT FUTURE SALES

MINGZHU KANG

ABSTRACT. Under the vigorous development of new retail trade, accurate forecast of merchandise sales can effectively guide stores to carry out reasonable inventory management. Sales forecasting technology is the key to market competition in the convenience industry. The wide distribution of convenience store chains in the region, the variety of goods, the market demand is difficult to predict, all these factors increase the difficulty to forecast the sales volume of goods.

In order to improve the accuracy of commodity sales forecast, this project is based on machine learning and deep learning theories, takes the historical sales data set of convenience stores as the research object, conducts data preprocessing, feature extraction and other operations, and uses the model to train the data set to predict the future sales of stores.

Keywords: Deep learning, store, sale

## Contents

---

*Date:* 2020-10-16.

*1991 Mathematics Subject Classification.* Artificial Intelligence.

*Key words and phrases.* Deep learning, store, sale.

## 1. INTRODUCTION

In deep learning projects, data visualization operations often required include visualization of original image data, visualization of loss and accuracy, etc. Visualization abstracts Numbers into graphs that are more convenient for us to observe and feel. Through data visualization, data can be represented more intuitively.

In this study, the data in the data set are organized and analyzed to analyze the impact of commodity sales volume, product category, operating revenue and the economic scale of the city where the store is located on the revenue. The model was used to train the data set, the historical sales data was used as the characteristics of the model, and the sales results of this month were used as labels to build the model for regression analysis.

This paper This paper makes use of the existing methods to achieve the forecast of merchandise sales and effectively guide stores to carry out reasonable inventory management. It has good practical significance and application prospect.

## 2. PRELIMINARIES

The project aims to find out the relevant factors affecting the sales volume of goods, and predict the sales volume of different goods in the company's stores in the coming month.

Step 1:

Preprocessing of project data sets

Step 2:

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

Step 3:

Structured data and analysis

This session includes sales analysis, profit analysis, view the overall sales trend. analysis commodity sales products operating revenue and store city economic scale effect on revenue benefit.

Step 4:

Characteristics of the processing Fuses the characteristics of the store data set and the merchandise data set.

Add sales characteristics and use the historical sales data as the characteristics of the model.

Step 5:

Model Adopted

This project uses lightGBM model for training.

## 3. METHOD

In order to accurately predict the future sales of goods, this project uses LightGBM Model.

Before using the Model to train the data set, we need to conduct feature processing on the data set to get the expected results.

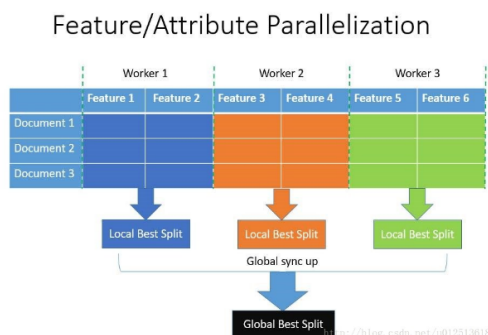
### 3.1 Characteristics of the processing

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.

### 3.2 Model Adopted

This project uses lightGBM model for training. LightGBM is a fast, distributed, high-performance gradient enhancement framework based on decision tree algorithms. It supports category characteristics.

LightGBM supports category characteristics directly and natively by changing the decision rules of the decision tree algorithm, without transformation.



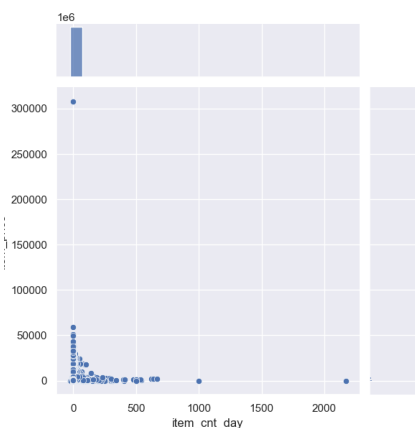
## 4. EXPERIMENT AND ANALYSIS

### 4.1 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

### 4.2 Training set data cleaning

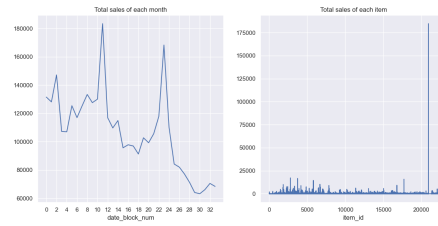
The data from the training set allows us to determine the parameters of the fitting curve. Filter abnormal data and use scatter plots to observe the distribution of commodity prices and daily sales.



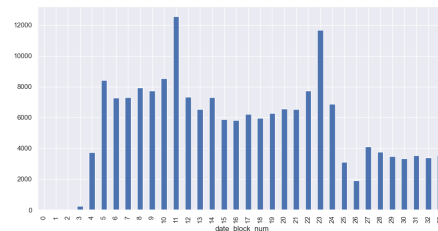
### 4.3 Sales analysis

(1) Shows the store's total monthly sales and total sales per item.

Overall sales were down, and monthly sales were mostly down year on year. One item sold exceptionally well.



(2) Monthly sales of the highest selling item.



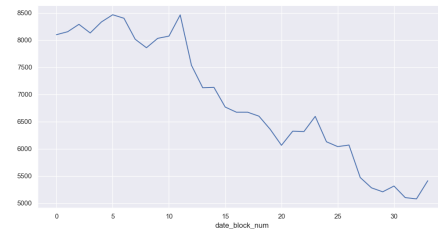
(3) Shows the number of items sold each month.

The number of goods on sale shows an overall trend of continuous decline.

There are two possible reasons for the overall decline in monthly sales:

One is: the heat or life cycle of the goods in the store, and other internal factors related to the goods,

so that the sales volume of goods decreased. Second, the goods are no longer sold due to external reasons such as withdrawal from shelves or shortage of goods, thus affecting the sales volume of goods.

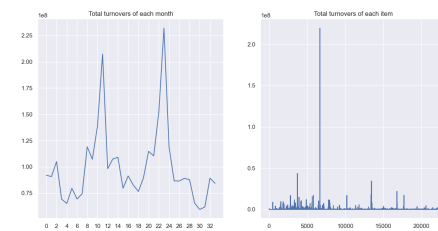


#### 4.4 Revenue Analysis

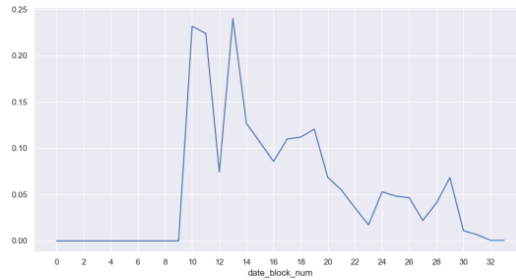
(1) Shows the company's total monthly revenue and total revenue per product.

Revenue in the 23rd month was up sharply from the 11th month, while sales were down year-on-year.

The e total revenue for one item was unusually high.

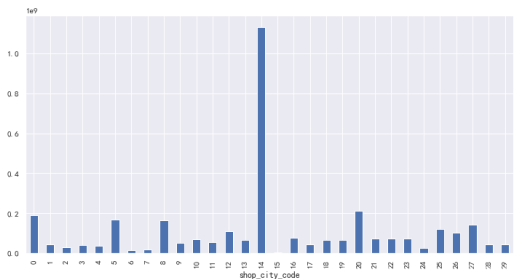


(2) Total revenue ranked no. 1 in merchandise sales per month.

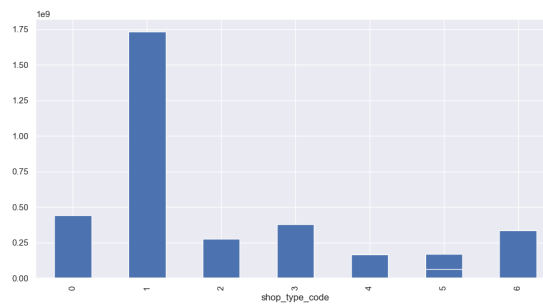


#### 4.5 Impact Analysis

(1) 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.



(2) The impact of store type on total revenue  
Category 1 stores contributed the most to total revenue.



## 5. CONCLUSIONS

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. In the process of predictive analysis, the theoretical and data support for feature analysis and model construction is not concise and powerful enough, which needs to be strengthened.

List of Todos

(A. 1) SCHOOL OF COMPUTER SCIENCE,, XI'AN SHIYOU UNIVERSITY, SHAANXI 710065, CHINA

*Email address, A. 1:* `xxx@tulip.academy`