Advices and predictions of China's economy of wholesale and retail trade by machine learning after COVID-19

Liu JinKua Chun-Te Lee
Department of Computer Science, Wenzhou-Kean University, jinkua1@kean.edu

Introduction

This research would conduct an analysis of China's Wholesale and Retail trade data and give a general analysis of it as a background study which found that the retail trade may have the ability to reallocated the resource and encourage the economy growth amount the low economy province. Based on the background and the data from China Statistic Yearbook 2020 and National Bureau of statistic, researcher would implement a mathematical model to predict the retail sales of China's economy. This research would mainly compare and implement three machine learning models for prediction, linear regression, classification and regression tree (CART) and multi-layer perceptron neural network (MLP). And researcher has found that DT-CART would be the most suitable model for the prediction.

Methods and Data

Data is collected from China statistic yearbook 2020 and National Bureau of statistics. China statistic yearbook 2020 collected the China's economy data from 2015 to 2019 and this research acquired the data 15th, Wholesale and Retail Trade. Used document is displayed in table 1 and 2.

Table 1 The main document implement in pre-COVID-19 analysis

Document
Main indicator of enterprises above Designated Size of Wholesale trade by status of Registration and Sector
Main indicators of Enterprises above Designated Size of Wholesale Trade by Region
Main indicators of Enterprises above Designated Size of Retail Trade by Region
Main Indicators of Chain Retail Enterprises by Status of Registration
Main Indicators of Chain Retail Enterprises by Sector and Business Categories

Table 2 The main document implement in Machine Learning Model

Document
The total retail sales of social consumer goods
The retail value of Textiles, Clothing, Shoes and Hats
The retail value of Gold, Jewelry, Jade
The retail value of Furniture
The retail of Cars, Motorcycles and Spare Parts

There's mainly contain two parts of the analysis, one is the pre-COVID-19 data analysis and one is the mathematical model implementation prediction.
In pre-COVID-19 data analysis, Normalization, KDE plot, frequency bar, heatmap, linear regression and data cluster were utilized to investigate the data and determined predictions and advices.
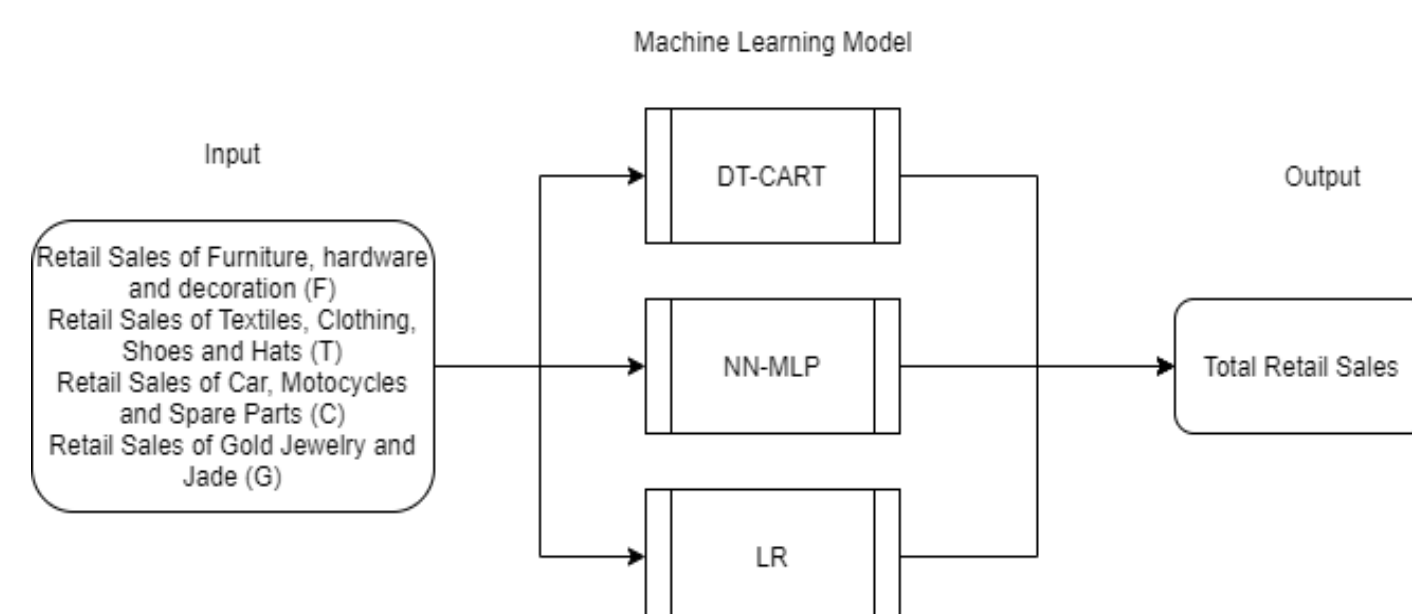In the machine learning model, three method is applied and shown in the figure 1.

Figure 1:The diagram of the Machine Learning Model

Results

Pre-COVID-19 data analysis

1. Total Profit earned by each province from Wholesale and Retail Trade, Figure 3&4
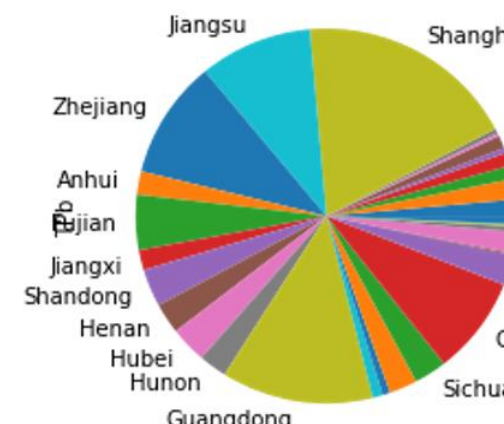

Figure 3: The distribution of total profit of each province in Wholesale data

Figure 4: The distribution of total profit of each province in Retail data

2.The Data Cluster of Wholesale and Retail Trade of each province based on Total profit and Total Assists, Figure 5&6
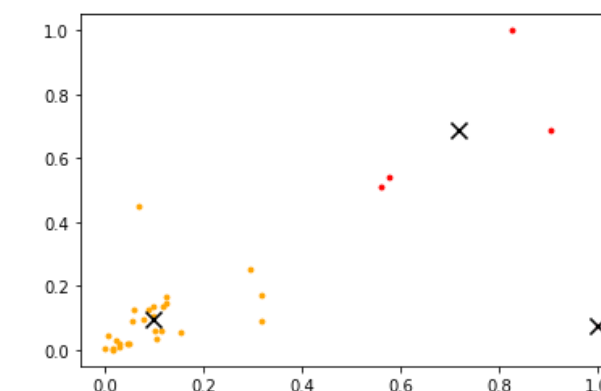

Figure 5: The cluster of each province by total profit and total asset in Wholesale data
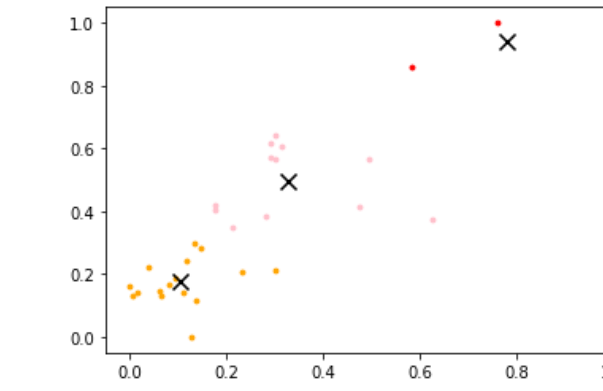

Figure 6: The cluster of each province by total profit and total asset in Retail data

3. In order to have a better analysis of the data, research keep on study the markets of different market in general market, special market and commodity exchange market with 100 million trade turnovers. Moreover, research has defined two variables, Per Booths average profit (PBP) and Per market average profit (PMP).

$$PMP = \frac{Turnover(100M\ yuan)}{NumBooth\ (Unit)}$$

$$PBP = \frac{Turnover(100M\ yuan)}{NumMarket\ (Unit)}$$

Based on the highest dataset above and correspond data provided by the National Bureau of statistics, research has selected four datasets from the special market, Market for Car Motorcycles and Spare Parts (C) and Gold Jewelry and Jade Markets (G) by PBP and Markets for Furniture (F) and Markets for Textiles Clothing Shoes and Hats (T). After that research has study the time correlation of between data. ACF, PACF and CCF diagram is applied to find the time correlation.

Conclusions

In the machine learning model, research has implemented three prediction method and compare the result of each them to find the most suitable method for the prediction. Figure 7& 8 has shown the decision tree and result of DT-CART. Fig 9 has shown the result of the NN-MLP and Fig 10 has shown the result of the LR.
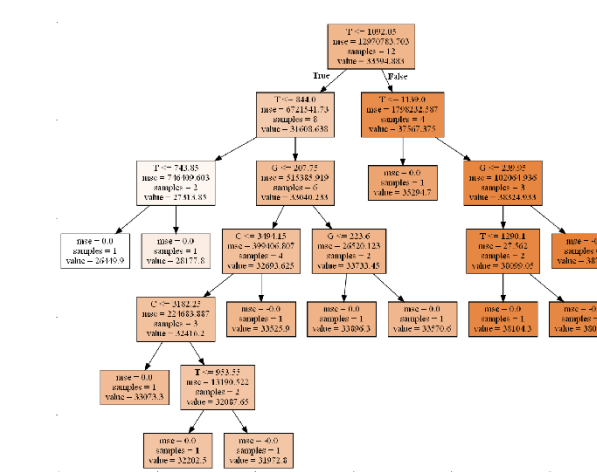

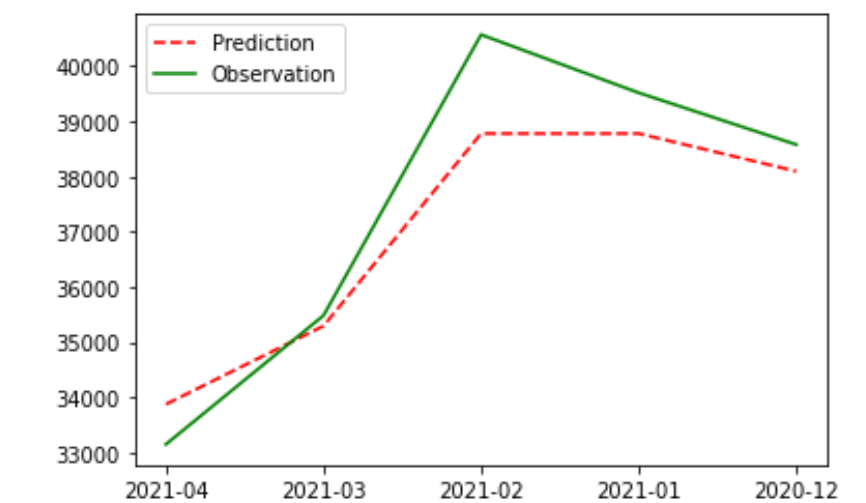Figure 7: DT-CART diagram generates in this research

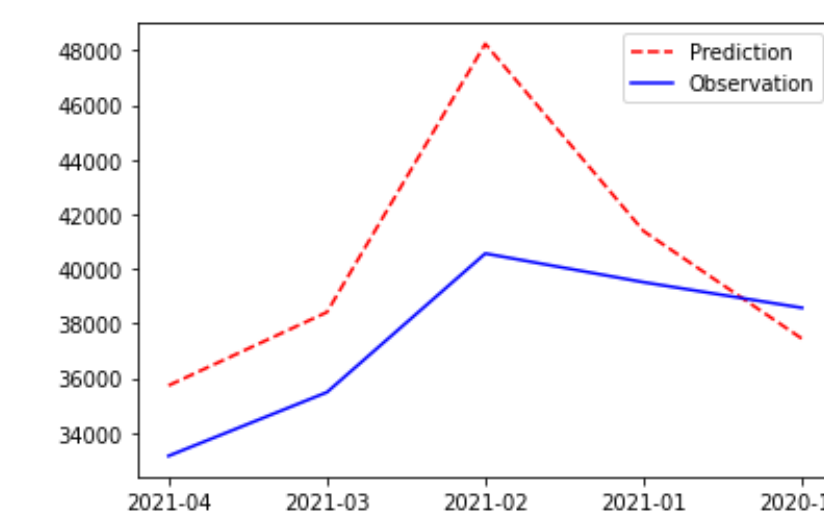Figure 8: Diagram of WRT retail sale prediction for DT-CART

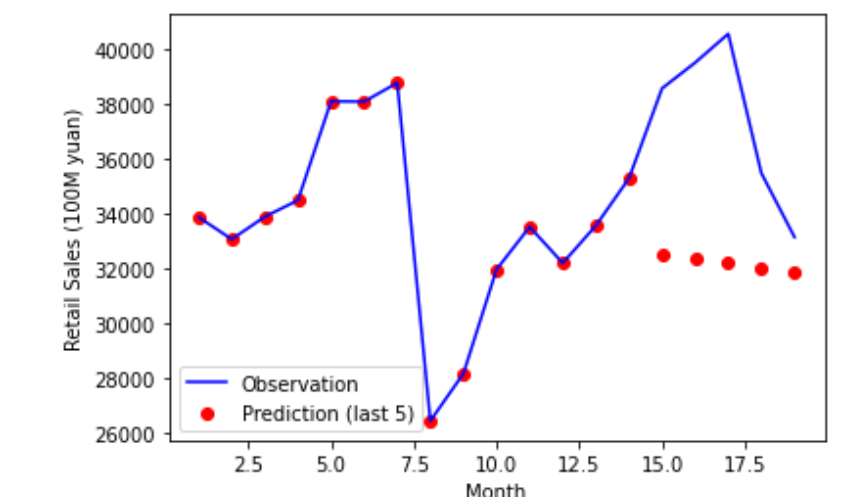Figure 9: Diagram of WRT retail sale prediction for NN MLP

Figure 10: Diagram of WRT retail sale prediction for LR

To analysis the result, researcher has applied the R^2, MAE, MSE and RMSE to the CART-DT and MLP-NN and P-value, R-value and Std-error to the LR shown in table 4 and 5. A better model should have R^2 closed to 1 and smaller MSE, RMSE and Std-Error.

Table 4: The evaluation of the Machine Learning Model (CART & MLP)

	DT-CART	NN-MLP
R^2 of training	1.0	0.03
R^2 of prediction	1.0	1.0
MAE	784.9	3238.4
MSE	908144.5	15796385.5
RMSE	953.0	3974.5

Table 5: The evaluation of the MLM (LR)

	LR
R-value	−0.2
P-value	0.5
Std-Error	235.1

As the graph and table shown, DT-CART would be the best fit machine learning model in this research which has lower MSE and RMSE compared to NN-MLP and a better tendency in the prediction data compared with NN-MLP and the linear regression. Moreover, NN-MLP require more larger data to provide an accurate prediction which is obvious not suitable for post-COVID-19 data. Ultimately, the machine learning model is shown below.
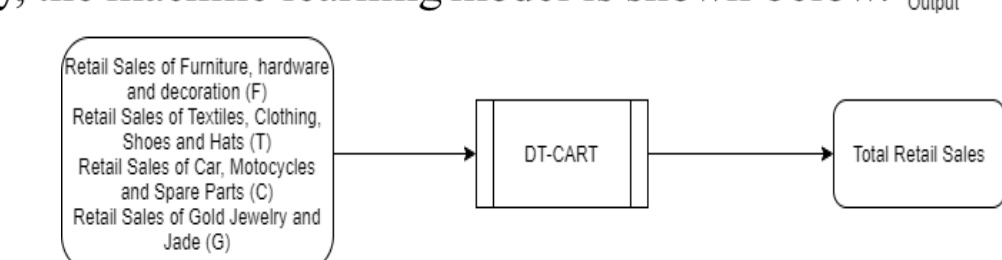
References

National Bureau of statistics of China. (2020). China statistical yearbook 2020. Retrieve from: http://www.stats.gov.cn/tjsj/ndsj/2020/indexeh.htm

Nation Bureau of statistics. Monthly Data. Retrieve from: Naional Data: https://data.stats.gov.cn/easyquery.htm?cn=A01

Roman Timofeev, Dr. Wolfgang Härdle. (2004 Dec月 20日). Classification and Regression Trees (CART) Theory and Applications. Retrieve from: ACADEMIA: https://www.academia.edu/13700196/Classification_and_Regression_Trees_CART_Theory_and_Applications

Spotle.ai. (9/12/2019). Classification and Regression Tree (CART) with Python. Retrieve from: Spotle.ai:https://spotle.ai/feeddetail=/Classification-and-Regression-Tree-CART-with-Python/3644

Zounemat-Kermani Mohammad, Ramezani-Charmahineh Abdollah, Reza, R., Meysam, A., & Ouarda Taha, B. M. J. (2020). Machine learning and water economy: A new approach to predicting dams water sales revenue. Water Resources Management, 34(6), 1893-1911. doi:http://dx.doi.org.kean.idm.oclc.org/10.1007/s11269-020-02529-0

驻卡拉布总领馆经济商务处 、3/26/2020. COVID-19预对对中国经济产生什么影响. Retrieve from: MINISTRY OF COMMERCE PEOPLE'S REPUBLIC OF CHINA: http://www.mofcom.gov.cn/article/i/jyjl/j/202003/20200302947229.shtml