2021 IEEE 2nd International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE 2021)

2021 IEEE 2nd International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE) | 978-1-6654-1540-8/20/$31.00 ©2021 IEEE | DOI: 10.1109/ICBAIE52039.2021.9389930

Sales Forecasting Using Deep Neural Network And SHAP techniques

Junhang Chen[1]
Wuhan University of Technology
Wuhan, China
chenjunhang@whut.edu.cn

Wang Koju[1],
Nankai University,
Nankai, China,
858374497@qq.com

Shurui Xu[2*]
Xuhai College of China University of Mining and Technology,
XuZhou, China,
924340680@qq.com

Zhe Liu[3]
Central South University
Changsha City, Hunan
2862527658@qq.com

***Abstract*—In today's business world's consumer-centric environment, enterprises seeking good sales performance require a balance between meeting customer demand and controlling the cost of inventory. Good sales forecasts play a crucial role in assisting enterprises to improve. With the development of Artificial Intelligence, increasing methods have been adopted to solve the forecasting problem. In this paper, we propose the Neural Network sale prediction model for predicting Walmart's sales. Moreover, we evaluate our NN model on the datasets provided by the Kaggle platform. Experiments have shown that compared with other machine learning model, our NN model achieves superior performance. Our RMSE metric is 2.92 and 2.58 lower than the Linear Regression algorithm and the SVM algorithm. Furthermore, to mine attributes of different dimensions to make prediction well, we utilize SHAP to interpret our NN model.**

***Index Terms*—Sales Forecast, Neural networks, Machine learning algorithms, SHAP**

I. INTRODUCTION

Nowadays, good forecasts play a vital role in many fields of scientific, industrial, commercial and economic activity [1]. Forecasting of the future demand in sales is key to the planning and activity of trade and business. Forecasting helps business organizations to make improvements, to make changes in business plans and provides a solution related the stock storage[2]. Therefore, this paper's research topic mainly focuses on the problem of estimating the unit sales of Walmart retail goods, which is provided by Kaggle competition.

With the development of artificial intelligence, machine learning algorithms have been applied in various industries to solve practical issues. And the major application is the sale forecasting, which assists a large quantity of company to better distribute their goods, control the cost of inventory and meet the market demand. Thus, we utilize the 1913 day sales of Walmart supermarkets to predict the sales in the next 28 days. Our dataset is about the historical sales data of Walmart from the Kaggle M5 competition. Unlike the previous studies, only depending on extrapolating the statistical trend and requiring a lot of extra information like customer and product analysis. we proposed a Multi-layer feed-forward(MLF) neural networks to predict the sales in the next 28 days, and it preforms better than other machine learning models using Root Mean Square Error(RMSE) as the evaluating metrics. And to get the relationship between each feature of the sample and its results, we use SHAP to evaluate the contribution of the features in our model to the results. In the whole process, Python is the primary tool for data processing, modeling, and analysis.

A. Related Work

To counter the problem more effectively, it is necessary to understand the technologies involved in sales forecasting, which has been studied by a lot of scholars and experts. There is a lot of research work about using machine learning algorithms to predict the sales of products and commodities [4, 5]. Lin et al. [4] predict vehicle sales by apply SVR (Support Vector Regression) algorithm. Gao et al. [5] use a hybrid optimization approach to predict automobile sales.

Over many decades, linear methods were often used to forecast. Because they are convenient to develop and implement, and they relatively simple to understand and interpret. However, experiments have shown that there are many limitation in that linear models in that they are not capable to capture any nonlinear relationship in the data. Using the linear models to solve the complicated nonlinear relationships is not satisfactory.[6]

Currently, owing to the development of deep learning, Artificial Neural Networks (ANN) [7][8] provide a promising tool for forecasting. The inherently nonlinear structure of neural networks is particularly useful for capturing the complex underlying relationship in many real world problems. What's more, the most important progress to using the ANNs for forecasting applications is that they not only could solve the nonlinear structures, but also they can model linear processes.

B. Our contribution

- ✓ We propose a multi-layer neural networks algorithm to forecast the sales in the next 28 days of Walmart.
- ✓ We introduce experimental results and utilize RSSM as the evaluating metrics, and our NN model has a better approximation than the SVM algorithm and linear Regression.

978-1-6654-1540-8/21/$31.00 ©2021 IEEE

✓ To evaluate the contribution of the features in our model to the results, we use SHAP to get the SHAP summary plot.

The rest of this article is organized as follows. Section II introduces the feature engineering. Next, our NN mode is detailed presented in Section III. Then, in Section IV, we show the RSSM of the NN model and compare it with other models. Furthermore, we utilize SHAP to analyze the feature contribution to the result. At last, section V summarizes our research work.

II. FEATURE ENGINEERING

A. Input data introduction

We get data that covers stores in US and includes item level, department, product categories, and store details from Kaggle website. In addition, it has explanatory variables such as price, promotion, day of the week, and special events. The dataset of Walmart retail goods contains three kinds of data, which are calendar, Sell price and Sales data. We segment these data into training dataset containing 1913 sale days data and text dataset containing 28 days data.

1）Calendar contains information about the dates on which products are sold and special events information, like national vacations and religious memorial days.

2）Sell price Contains information about the price of the products sold per store and date, which could indicate the promotion activity. The sell price of each product is marked on a weekly basis.

3）Sales data file contains historical daily unit sales data for each product and store, i.e. 30490 time series data, including the item id, the category of item, the date, the store number, the department number, Weekly Sales (sales in a given department in a given store), etc.

B. Label encoding

We convert the categorical columns into numeric using label encoding. The columns are item_id, dept_id, cat_id, state_id, wday, month, year, event_name_1, event_type_1, event_name_2, event_type_2.

C. Exception handling

About the exception handling[9], we make sales of some goods that were show as Nan in the tabular of supply chain as the median data.

III. NEURAL NETWORK FORECASTING MODEL

In this part, we introduce our NN sales forecasting model. Neural networks are series of algorithms that endeavor to recognize underlying relationships in a set of data through a process that mimics the way the human brain operates. And the neural networks of simple processing elements operating on their local data and communicating with other elements. There exist many types of neural network[10], but the basic principles are similar. Each neuron in the network need to receive input signal, to process them and to send an output signal.

Because of the limited information provided by the dataset, a neural network should be a good choice as it can auto select the hidden features and fit the data accordingly. The common simple multi-layer neural networks should be sufficient enough to output one value for each input feature. We used it to predict the sales in the next 28 days.

The network has four layers, tow layers are hidden layers, the others are output layers and input layers。 The framework of NN model is shown in Figure 1.

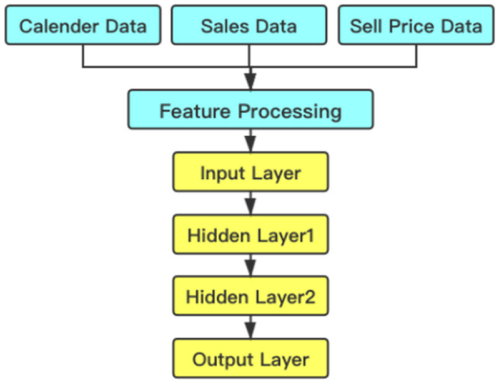

Figure 1: NN model framework

We use Keras and TensorFlow to train our model, and NN model basic setting are as the Table 1 shows:

Table 1: basic setting of NN model

epochs	12
batch_size	28
metrics	mse
Optimizer	adam
activation function	ReLU

IV. EXPERIMENTS

In this part, we applied our neural network model on the dataSet of the Walmart supermarkets sales.

In order to comprehensively evaluate the performance of our neural network model, we compare it with other different classic machine learning models using the same evaluation metric and dataset. In this paper, firstly we select the RMSE metric to evaluate the performance of various models.

The RMSE, which is the abbreviation of Root Mean Square Error[11], unlike the MSE, the RMSE uses same unit of measurement as the parameter of interest, and it is considered an excellent general-purpose error metric for numerical predictions.

The formula of the RMSE is as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(S_i - O_i)}$$

where O_i are the observations, S_i predicted values of a variable, and n(n=28) is the number of observations available for analysis.

The lower the RMSE value, the better the prediction effect of the model.

The experimental results of competing models and our model are shown in Table 2.

Table 2: performance of different models

Models	RMSE
Linear Regression	6.12
SVM	5.78
NN	**3.20**

It is obvious to find that our NN model has the lowest RMSE value. Specifically, the RMSE of our Neural Network model is 2.92 lower than Linear Regression algorithm [12] and 2.58 lower than the SVM algorithm[13]. Experiments show our method using the NN model is more effective in forecasting sales of Walmart.

Furthermore, we require to analyze the importance of each feature, unlike the Algorithms based tree, which facilitates the calculate the feature importance, the Neural Network model is black-box model presenting the prediction as a mystery to the user. It is difficult and complicated to gain an intuition of what happening inside---how certain feature influence its output. Therefore, we use SHAP(SHapley Additive exPlanations)[14]， an effective method that provides a unified approach for interpreting output of machine learning methods. As the Figure 2 shows：

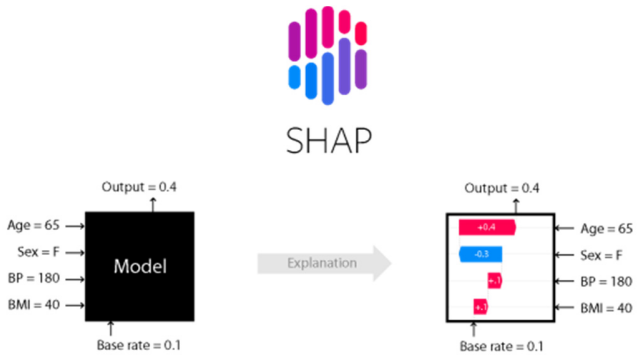

Figure 2: SHAP introduction

Instead of the original feature, SHAP replaces each feature x_i with binary variable z_i that represents whether x_i is present or not:

$$g(z') = \phi_0 + \sum_{i=1}^{M} \phi_i z_i = bias + \sum contribution\ of each\ feature$$

The idea of SHAP to compute Φ_i is from the Shapley value in game theory. The SHAP value will calculate the difference between feature i and the output for all feature combinations except i. The formula is as follows:

$$\phi_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(M-|S|-1)!}{M!}[f_x(S \cup \{i\}) - f_x(S)]$$

Where S is the subset of features from all features N except for feature i, $\frac{|S|!(M-|S|-1)!}{M!}$ is the weight factor counting of permutations of the subset S. $f_x(s)$ is the expected output given the feature subset S

$$f_x(S) = E[f(x)|x_s]$$

which is similar to the marginal average on all other features than the subset S.$f_x(S \cup \{i\}) - f_x(S)$is the difference made by feature.

Based on these, we get the SHAP summary plot, as the Figure 3 shows. The figure contains all samples and every dot is a sample. Color indicates characteristic value (red is high, blue is low). The x-axis indicates that the samples are sorted by Shap value and The y-axis indicates that the features are sorted by Shap value.

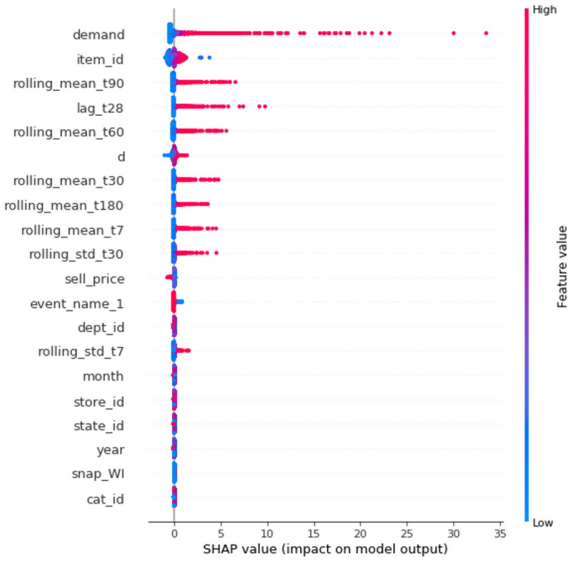

Figure 3: SHAP summary plot

From the Figure 3, it is easy to explore that demand is the most important and the higher demand is, the higher sales the model will output. Demand and model output are roughly positively correlated. Additionally, item_id, rolling_mean_t90 and lag_t28 also have higher importance of the model output.

V. CONCLUSION

In this paper, we propose a NN model to predict Walmart sales on Walmart sales dataset provided by Kaggle website: M5 Forecasting - Accuracy, Estimate the unit sales of Walmart retail goods. We use TensorFlow and Keras to train our model and experiments show our NN model performs better than Linear Regression algorithm and SVM algorithm. Additionally, we utilize SHAP to analyze the feature importance of our NN model. In the future, we will improve our model to improve the prediction effect, and apply our model to other sales problems.

References

[1] Mentzer, J.T.; Moon, M.A. Sales Forecasting Management: A Demand Management Approach; Sage: Thousand Oaks, CA, USA, 2004.

[2] Chatfield, C. Time-Series Forecasting; Chapman and Hall/CRC: Boca Raton, FL, USA, 2000.

[3] Brockwell, P.J.; Davis, R.A.; Calder, M.V. Introduction to Time Series and Forecasting; Sp ringer: Cham, Switzerland, 2002; Volume 2

[4] Lin, K., Lin, Q., Zhou, C., Yao, J.: Time series prediction based on linear regression and SVR. In: IEEE International Conference on Natural Computation (2007).

[5] Gao, J., Xie, Y., Gu, F., Xiao, W., Hu, J., Yu, W.: A hybrid optimization approach to forecast automobile sales of China. Adv. Mech. Eng. 9(8), 1687814017719422 (2017a).

[6] Zhang G P. Business forecasting with artificial neural networks: An overview[J]. Neural networks in business forecasting, 2004: 1-22.

[7] Al-Saba, Tawfiq, and Ibrahim El-Amin. "Artificial neural networks as applied to long-term demand forecasting." Artificial Intelligence in Engineering 13.2 (1999): 189-197.

[8] Rabunal, Juan Ramon, and Julian Dorado, eds. Artificial neural networks in real-life applications. IGI Global, 2006.

[9] Goodenough, John B. "Exception handling: issues and a proposed notation." Communications of the ACM 18.12 (1975): 683-696.

[10] You, L., Kou J., & Wang S.. (2019). Online Retail Sales Prediction with Integrated Framework of K-mean and Neural Network. In Proceedings of the 2019 10th International Conference on E-business, Management and Economics (ICEME 2019). Association for Computing Machinery, New York, NY, USA, 115–118.

[11] Chai, Tianfeng, and Roland R. Draxler. "Root mean square error (RMSE) or mean absolute error (MAE)?–Arguments against avoiding RMSE in the literature." Geoscientific model development 7.3 (2014): 1247-1250.

[12] Saber, Ahmed Yousuf, and AKM Rezaul Alam. "Short term load forecasting using multiple linear regression for big data." 2017 IEEE symposium series on computational intelligence (SSCI). IEEE, 2017.

[13] Huang, Wei, Yoshiteru Nakamori, and Shou-Yang Wang. "Forecasting stock market movement direction with support vector machine." Computers & operations research 32.10 (2005): 2513-2522.

[14] Lundberg, Scott, and Su-In Lee. "A unified approach to interpreting model predictions." arXiv preprint arXiv:1705.07874 (2017).