Alcohol Sale Forecast Using Long Short-Term Memory Network

Zufar Bagas Prasetyo (5022201209), Zulfiqar Fauzul Akbar (05111940000101)

Department of Electrical Engineering, Faculty of Intelligent Electrical and Informatics, Sepuluh Nopember Institute of Technology

Department of Informatics, Faculty of Intelligent Electrical and Informatics, Sepuluh Nopember Institute of Technology

Abstract

In this time, data play significant role in any major. Therefore, forecasting is important to process the data so that it can be an insight to do planning or decision making. One of the powerful methods of forecasting is by using artificial neural networks (ANN). For time series analysis, the best neural network architecture is recurrent neural network. However, for standard RNN architecture, any given inputs on the hidden layer would cause either exponential decay or blow up on the output when cycling around the recurrent connections. To resolve that, a method called long short-term memory (LSTM) is introduced. The dataset used in this research from U.S. Bureau Sensus is a monthly alcoholic beverage sale from 1992 to 2018 in millions of dollars. The dataset is then cleaned and fed to the LSTM network to create forecast for certain period ahead. The LSTM model gives great performance with MAPE 4.8%.

Keywords: Sales, LSTM, Neural Network, Forecasting

1. Introduction

In this time, data play significant role in any major. In industry and business, data are compulsory aspects used for planning because the lead time needed for decision making is ranged from several years to a few seconds. Therefore, forecasting is important to process the data so that it can be an insight to do planning or decision making. Many forecasting methods are available to use, from the simplest methods, such as naïve methods, to the highly complex methods, such as neural networks. The development of technology makes forecasting can be done easier (Makridakis et al., 2000).

One of the powerful methods of forecasting is by using artificial neural networks (ANN). ANN are machine learning models inspired by biological neurons. It can find the best functional form characterizing the data (Hill et al., 1994). For time series analysis, the best neural network architecture is recurrent neural network. However, for standard RNN architecture, any given inputs on the hidden layer would cause either exponential decay or blow up on the output when cycling around the recurrent connections. To resolve that, a method called long short-term memory (LSTM) is introduced by changing the structure of the hidden neurons of standard RNN (Hua et al., 2019). Therefore, we want to use this LSTM network to forecast a business sale from the previous year's sales data. Here, we use alcohol sales data consisting monthly sold unit of alcohol for the past 27 years from 1992 to 2019 because it is a business sales data, and it has trend and seasonal data pattern.

2. Review of Literature

2.1. Dataset "Alcohol_Sale"

Dataset is a set of structured data in a relation as an entity (Renear et al., 2010). The dataset used in this research from U.S. Bureau Sensus is a monthly alcoholic beverage sale from 1992 to 2018 in millions of dollars. It has trend and annual seasonal data pattern.

2.2. Forecasting

Forecasting is an activity of predicting what may happen in the future based on current or past data. It is necessary to determine when an event will occur so that appropriate actions can be taken. It becomes an important aid in effective and efficient planning. In organizational and managerial context, a decision-making process can be done with the insight coming from forecasting so that it is important to select the suitable forecasting method for any dataset (Makridakis et al., 2000).

2.3. Artificial Neural Network

Artificial Neural Network is a representation of human brain's neural network. ANN has some neurons that carry and store the certain information. It relates to each other by weights. Generally, there are three layers of ANN. The first is input layer which consists of the neurons that receive input and send it to the next layer. The second is hidden layer which consists of the neurons that receive the information from input layer and process them until they produce outputs sent to the next layer. The third is output layer which takes output from the hidden layer and give response to the user. The structure is shown in figure 2.1 (Sharma & Mehra, 2018).

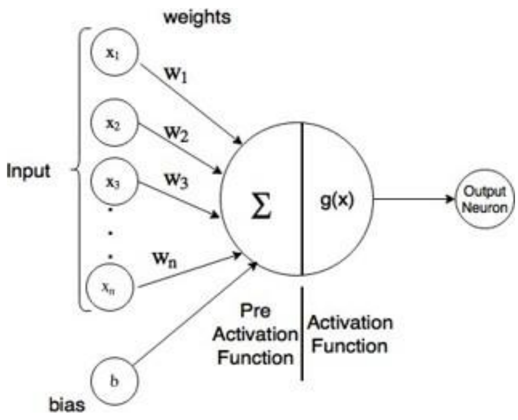


Figure 2.1 Structure of Artifical Neural Network

2.4. Long Short-Term Memory Network

One of the deep learning networks called Recurrent Neural Networks (RNNs) gives a great performance with time series data by recurrent neural connections. However, for standard RNN architecture, the hidden layer's input and output may cause the network either decaying or blowing up exponentially when cycling around recurrent connections. To solve that problem, Long Short-Term Memory (LSTM) was introduced by changing the structure of hidden neurons in traditional RNN. Figure 2.2 highlights the details of LSTM's working principle. Through the cooperation between the memory cell and the gates, LSTM possess powerful ability to predict time series with long-term dependences (Hua et al., 2019).

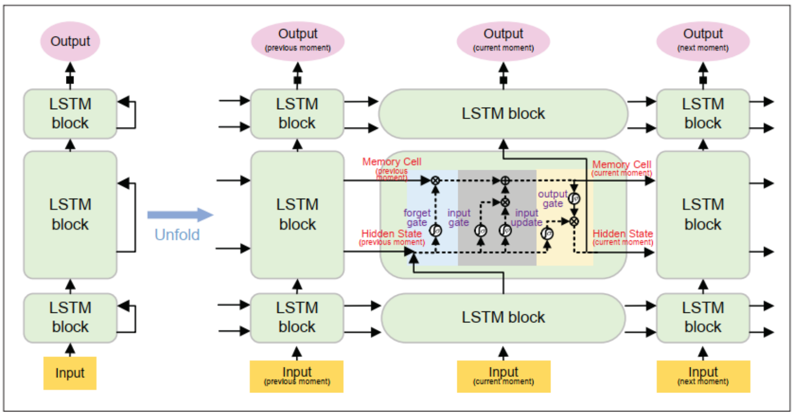


Figure 2.2 Illustration of one-three-layer LSTM network

3. Research Method

Here is the research method used for this research (Larose, 2005):

Figure 3.1 Research Method

3.1. Data Gathering

Data gathering is a process of collecting data and information about the research object. The output of the data gathering is the “alcohol_sales” dataset with csv format as the research object that will be prepared for the next step.

3.2. Data Exploration

Data exploration is a process of getting the insight of data. It can also be defined as a step to understand the data that will help us to determine the analysis method before it is preprocessed. For forecasting, it is usually necessary to see the data pattern so that a correct forecasting method can be determined.

3.3. Data Pre-Processing

Data pre-processing is a step to manipulate or drop some component or all component of the data to enhance the quality and performance before it’s being analyzed. In this step, there are some processes done like data profiling, data cleansing, data reduction, data transformation, data enrichment, and data validation.

3.4. Data Modelling

Data modelling is a step to create an appropriate model used to predict the output of certain variables given in any input in the datasets. For this dataset, a Long Short-Term Memory (LSTM) model is used because in the literature explained in section 2.4, LSTM gives great performance to forecast time series dataset.

3.5. Analyzing and Testing Model

Analyzing and testing model is a step to feed the dataset into a model to result the best accuracy. For Neural Networks model, the dataset is used to train the model so that the weights of the neurons of neural networks is adjusted for certain epochs. The metric used to determine the model's accuracy is MAPE and MAE.

4. Results and Discussion

4.1. Data Gathering

The dataset used for this research is "alcohol_sale" got from U.S. Bureau Sensus. The dataset consists of two columns. The first is "DATE" column which contains the date from 1992-01-01 to 2019-01-01 with monthly difference. The second is "S4248SM144NCEN" (refers to "Sales") column which contains the alcohol sale in millions of dollars unit with total count of 325.

4.2. Data Exploration

In data exploration, it is necessary to see the pattern and calculate some statistical variables of the data. The mean value of the sales is 7886 with standard deviation of 2914.

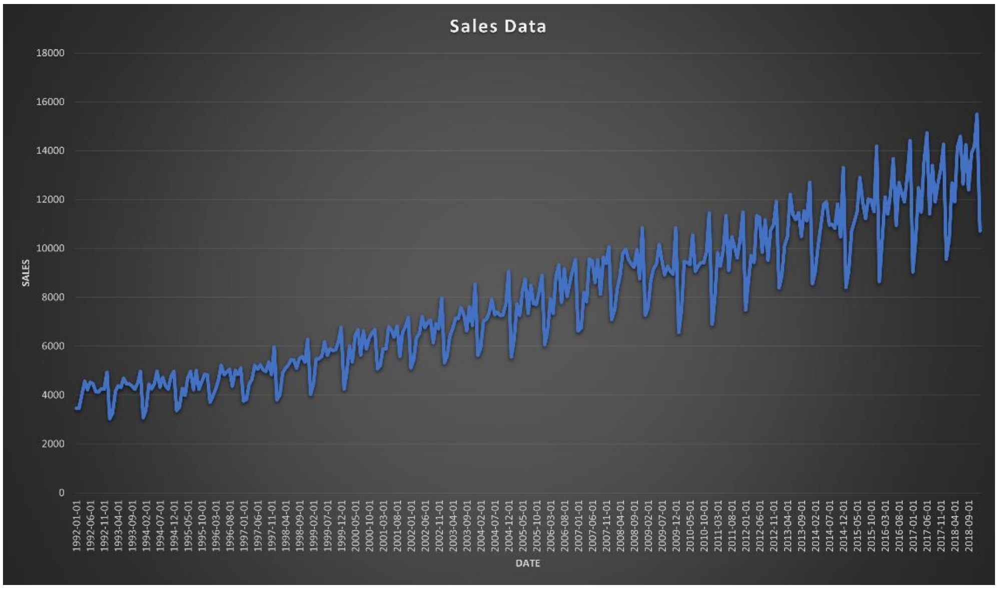


Figure 4.1 Graph of the Sales Data Vs Date

From figure 4.1 can be seen that the sales data has positive multiplicative trend and annual seasonal data.

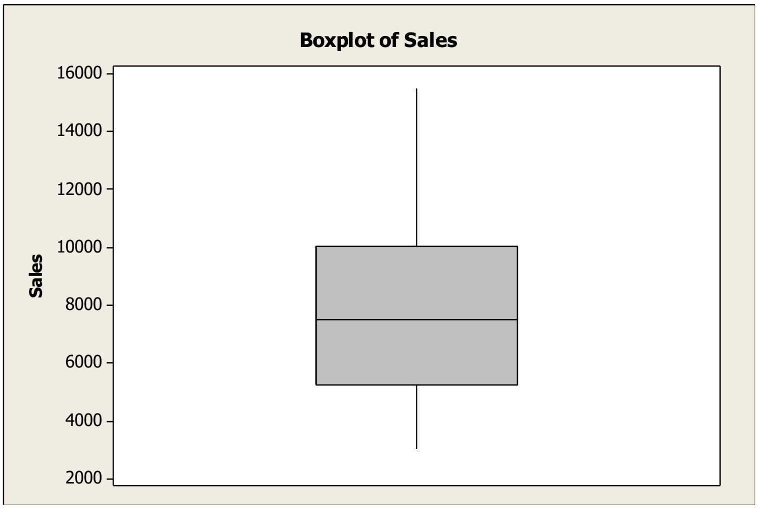


Figure 4.2 Boxplot of Sales Data

Figure 4.2 shows the boxplot of sales data that gives information such as the first quartile value (5225), median value (7481), the third quartile value (10016), minimum value (3031), and maximum value (15504). The boxplot also shows that the whisker and half-box are on the lower side of the median and there is no outlier in the sales data.

4.3. Data Pre-Processing

In Data Pre-processing, there are some steps to do. The first is removing the last data of Sales and DATE column so that all the Sales data is seasonal from the first month to the twelfth month. The second is decompose the data to get the trend and seasonal component of the Sales data using multiplicative decomposition. The LSTM models requires the non-trend component of the input data to perform well so that only the seasonal component of the data will be used as the input of LSTM models.

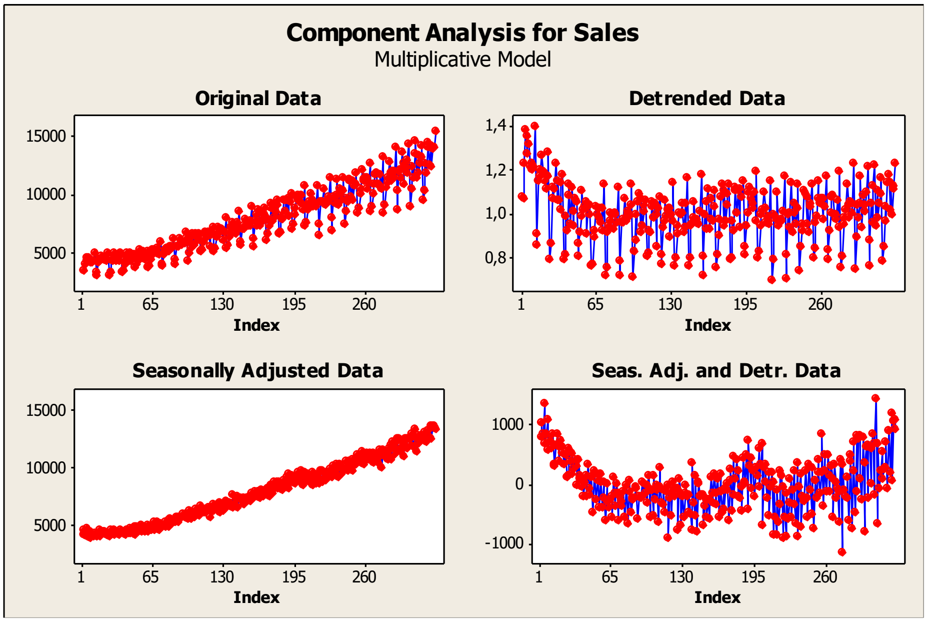


Figure 4.3 Component Analysis for Sales Data

Figure 4.3 shows the seasonal component or the detrended data which will be used for further analysis. The equation of the trend component is obtained as $Yt = 3168.1 + 28.9t$ with Y_t as the Sales and t as the period.

4.4. Data Modelling

In data modelling, the first step is split the data into training set and validation set of detrended Sales data for prediction. The training set consists of 86% data of the dataset and the validation set consists of 14% data of the dataset. The LSTM model consists of 1 input layer, 4 bidirectional hidden layers having 1024, 512, 256, and 128 neurons respectively and 1 output layer.

4.5. Analyzing and Testing Model

In the analyzing and testing model, the first step is feed the training data into the LSTM model. The parameters used in the model are

a) Learning rate: 0.001
b) Metrics: mae
c) Epoch: 53

From the data training process can be shown the mean absolute error and the loss in every epoch in this figure 4.4.

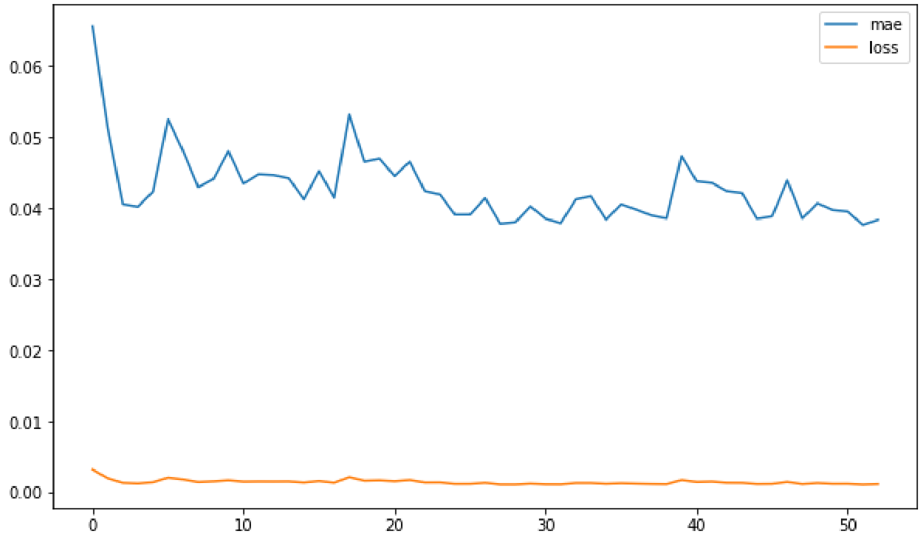


Figure 4.4 Graph of MAE and Loss Vs Epoch

The MAE is decreasing as the epoch increasing, meaning that the model works well in decreasing the MAE. After the data training process is completed, the prediction is made and compared to the validation set to see the performance of the model when handling the new data input.

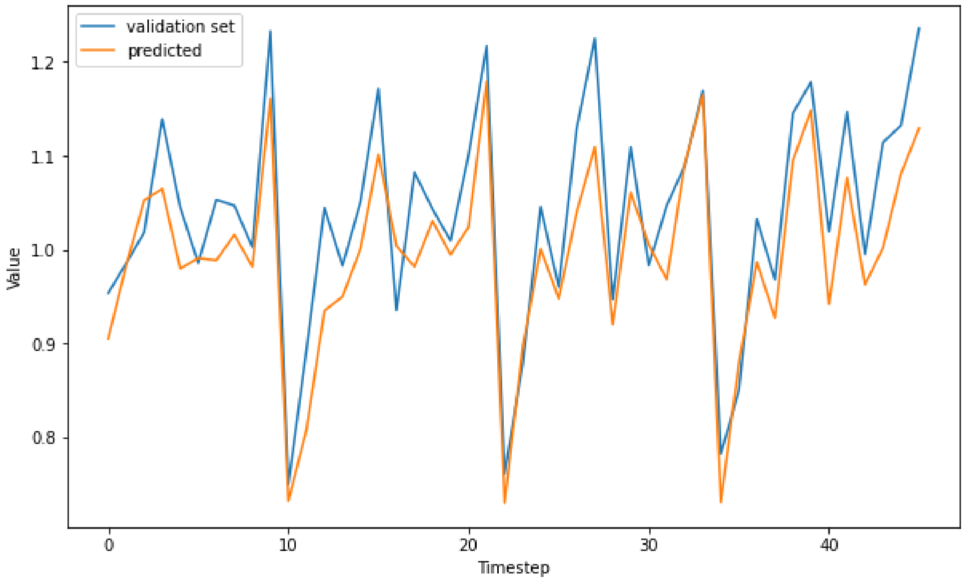


Figure 4.5 Comparation of Detrended Predicted and Validation Data

From figure 4.5, the model prediction fits the validation data well with mean absolute error of 0.05 and mean absolute percentage error of 4.73%.

The predicted data are the detrended data of the Sales data. Therefore, the detrended data should be multiplied with the trend component of the data to get the actual predicted data. By using the equation of the trended component, the equation of actual Sales data is $Ft = (3168.1 + 28.9t) * F_{dt}$ with F_t as the actual predicted data, t as the period, F_{dt} as the detrended predicted data. By using the same technique to the validation set, the comparison of the actual value of both is shown in figure 4.6 below.

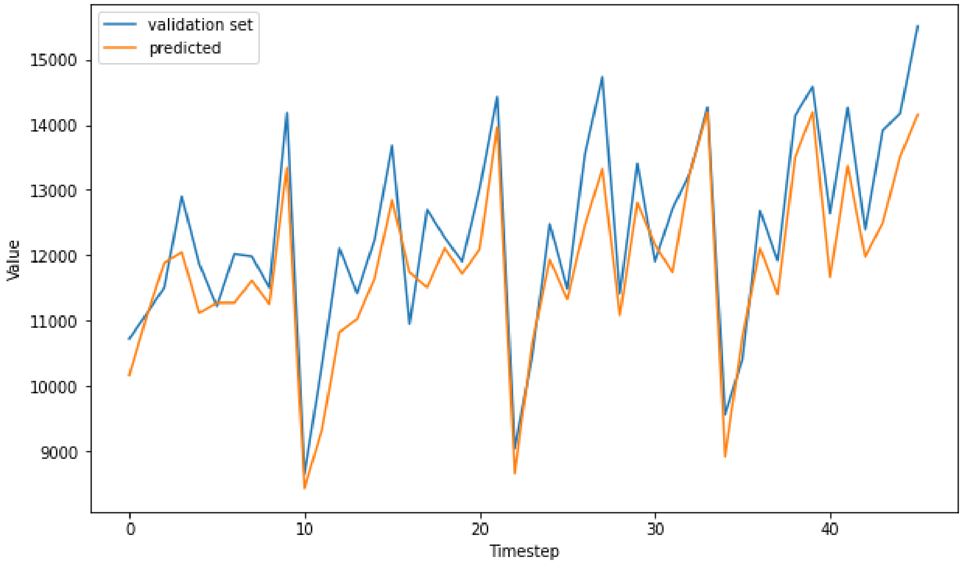
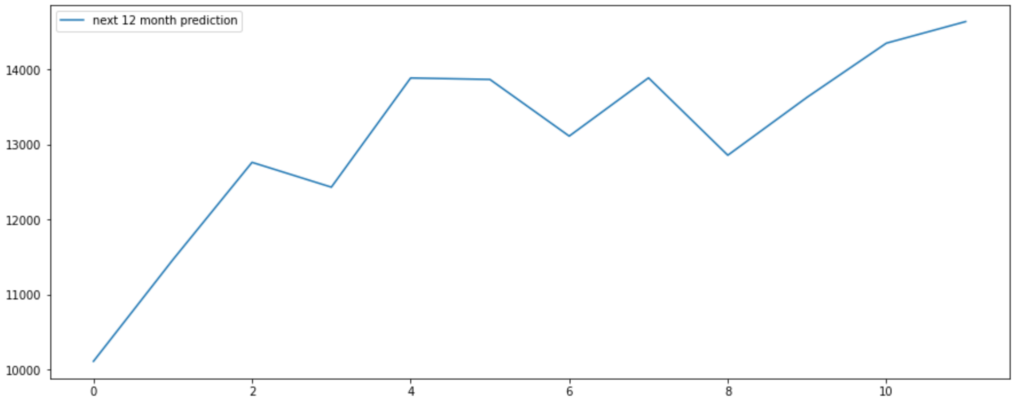


Figure 4.6 Comparation of Actual Predicted and Validation Data

From that figure, the mean absolute error is obtained 602.07 and the mean absolute percentage error is 4.8%. The forecast of Sales data for one year ahead is shown in figure 4.7 below.



4.7 Sales Forecast for One Year Ahead

Having MAPE below 5%, this model has a great performance for forecasting the dataset. Here is the performance comparison of LSTM model with other models.

Table 4.1 Performance comparison of Forecasting Models

Model	**MAPE**
LSTM Network	4.8 %
Moving Average 12 Window Width	10 %
Winter's	4.04 %
ARIMA (1,0,0)	3.45 %
Multiplicative Decomposition	5 %
Additive Decomposition	6 %

5. Conclusion

From the result of forecasting the Sales data using LSTM can be concluded that LSTM model performs well on time series data which has trend and seasonal pattern. However, some traditional forecasting models, like ARIMA, gives better performance so that the hyperparameter of the LSTM network needs to be tuned to increase its performance.

6. Reference

Hill, T., Marquezb, L., O'connor, M., & Remusa, W. (1994). Artificial neural network models for forecasting and decision making. In *International Journal of Forecasting* (Vol. 10).

Hua, Y., Zhao, Z., Li, R., Chen, X., Liu, Z., & Zhang, H. (2019). Deep Learning with Long Short-Term Memory for Time Series Prediction. *IEEE Communications Magazine*, *57*(6), 114–119. https://doi.org/10.1109/MCOM.2019.1800155

Larose, D. T. (2005). Discovering Knowledge in Data: An Introduction to Data Mining. In *Discovering Knowledge in Data: An Introduction to Data Mining*. https://doi.org/10.1002/0471687545

Makridakis, S., Wheelwright C, S., & McGee, V. E. (2000). Metode dan Aplikasi Peramalan Edisi Kedua. In *Erlangga*.

Renear, A. H., Sacchi, S., & Wickett, K. M. (2010). Definitions of dataset in the scientific and technical literature. *Proceedings of the ASIST Annual Meeting*, *47*. https://doi.org/10.1002/meet.14504701240

Sharma, S., & Mehra, V. (2018). Default Payment Analysis of Credit Card Clients. In *Researchgate* (Issue July).