Machine-learning algorithm for demand forecasting problem

DT Wiyanti¹, I Kharisudin¹, AB Setiawan², AK Nugroho³
¹Mathematics Department, Faculty of Mathematics and Natural Science, Universitas Negeri Semarang, Indonesia
²Economics Science Department, Faculty of Economics, Universitas Negeri Semarang, Indonesia
³Informatics Engineering Department, Universitas Jenderal Soedirman, Indonesia

*Corresponding author: diantriwiyanti@mail.unnes.ac.id

Abstract. Numerous top algorithms for time series prediction issues have been proposed. Time-series prediction has stirred up wide attention in many research fields because it is an important direction of dynamic data analysis and processing. The prediction is used in a variety of practical situations, especially for the pressing demand for forecasting future data trends based on historical information. While demand forecasting remains a challenge, developments in machine-learning have provided dramatic improvements. In this article, we investigate the feasibility and comparative analysis of Deep Learning approaches to forecasting the demand problem with implementation to a public dataset. We use comparison with RMSE performance metrics to analyze the Deep Learning performance better than other model techniques, including Random Forest, Gradient Boosted Trees, and Support Vector Machine. However, the forecasting problem is a vital need for business decision making.

1. Introduction
Outstanding business planning relies on how much appeal there will be for outputs and treatments that can be recognized by leaders. The more insight into demand from product design to production, logistics to sales, the better prepared the organization will be. On the contrary, if the forecasts were unsatisfied, it will result in the risk of over or under production, weakness in service, or a mere selling of the faulty commodity. At least to achieve maximum income, market opportunities with the greatest potential demand should be pursued. Apart from that getting the right quantity of product at the right time at the right place (production), planning the right production volume and inventory (inventory), and optimizing the supply chain for delivery (supply chain). Besides, finding valuable prospects using long-term customer profiles to identify and get prospects with similar characteristics is extremely important to do. This is where forecasting should begin to be carried out by an organization. Valuable outlook will seize authentic patterns and correlations that be found in historical data but not renew past events that will not occur again, whereas forecasting methods be able to be as common as using last review s such as forecasting (called naive methods), or high complexity, such as neural networks and econometric system of simultaneous equations [1]. Having rigorous demand forecasting and stocktaking management systems for vulnerable operational management is essential because of rivalry and financial constraints in the retail industry, as well as cost-oriented supply chain operations, and retailers entail to optimize their provision to degrade the financial risk [2]. Even though circumstances such as demand uncertainty could be defying, a long-term plan might allow an opportunity for demand planning.
supported by divination and arrangement that put aside time, turn downs costs, and sharpens outcomes [3]. Supply chain analysis is needed in a company because it can help in the process of its development, which includes the process of selecting suppliers, planning logistics, and distributing supplies. The existence of supply chain analysis will have an important effect and have a positive impact on the company, such as getting satisfaction from consumers. If customers are satisfied, then they become loyal customers who use your product for a long time. Besides that, there is also another positive impact, namely increased profit. If there are more loyal customers, the company's profits will also increase because the products offered are sold out. Therefore, business people are expected to be proficient in analyzing data based on the purchase history and profitability of each client, because the profitability is possible to build up over time [4]. Predicted time series is one of the most preferred but throw down the most gauntlet and has played an important role in all kinds of fields including Statistics, Machine Learning, Data Mining, Operations Research for decades [5].

Data Mining and Machine Learning have been successfully devoted to the time series computation of numerous amounts of demand-related data and features, and to forethought coming necessity and prototype using particular learning theorem [2]. The challenges with big data are well established and have spawned a whole new industry around cloud computing services that enable easy access and relatively inexpensive computing solutions [6]. Demand preparation is more about forecasting and expertise to expect the requirement for diverse goods at various points in the stock line which allows an organization to better plan their inventory, ensure product availability according to market needs, and monitor the difference between actual and predicted sales to optimize their production.

All kinds of industries and scenarios where Machine Learning demand forecasting accomplishments have proven effective include novel output introductions, products with short life cycles, and alike [7]. The primary thought of Machine Learning is to develop a formula, which in turn could form models which able to create outlooks based on the data or invent decisions concerning data classification using a training data set to establish a model that can then be used for forecasts [8]. Demand forecasting problems are not a new thing, but machine learning models provide an unprecedented level of sophistication to the task. What Machine Learning can do is identify the relationship between data variables such as historical data and company resource planning systems, point of sale systems, and social media marketing programs, raw material prices, supplier problems, and weather disturbances, with specific factors that drive demand. Uniquely, the model will continue to be updated based on new data, allowing it to adapt quickly. Machine Learning is actionable intelligence derived from data that combine different types of algorithms, from linear discriminant analysis to neural networks to clustering. It is a dependency of Artificial Intelligence (AI) focused on creating computer programs that can learn from experience, and thus adjust their decision-making ability over time technically. AI programs can be used to forecast demand, especially useful for fast-moving consumer goods. This algorithm is potential if it could be harnessed for the business because machine learning will reshape the manufacturing industry in the upcoming year. Among the numerous machine learning methods, deep learning (DL) methods have become greatly leading and are enforced in countless fields such as image and speech recognition, natural language processing, machine translation, and prediction [2].

Deep Learning is a certain domain of Machine Learning theorem that uses complex neural networks, where Machine Learning itself is still a part of artificial intelligence that includes everything that allows computers to behave like humans. Machine Learning is concerned with the extraction of patterns from data sets, where machines able to discover rules for optimal behavior but even could conform to a changing world. Deep Learning would be more and more successful immediately as it requires less hand engineering, so we might easily take out the benefit of the augmentative amount of computation and data available with new deep neural network learning algorithms and architectures that will accelerate this progress [9]. In this study, the Deep Learning algorithm was used to develop the dynamic demand forecast model for the dataset of supply chains used by the company DataCo Global with Rapidminer software. The purpose of this study is to prove that a Deep Learning algorithm is a dominant solution to Machine Learning problems for forecasting demand, especially when decision-makers are faced with complex constraints such as consumer purchasing items and their changing preferences over time.
2. Methods
The forecasting problem is not a new thing. In the world of science, engineering, even business, forecasting has a prominent influence on the organization's decision framing process. Demand forecasting and stock market prediction are particular adoption of forecasting techniques. To speculate the sample \( y(t_{n+1}) \) at a future time \( t_{n+1} \), a set of \( n \) samples \( \{ y(t_1), y(t_2), ..., y(t_n) \} \) in a time sequence, \( t_1, t_2, ..., t_n \), is given [10]. The Artificial Neural Network (ANN) was used to predict sales per customer. It was compared with the commonly used modeling techniques for forecasting problems, which are Random Forest, Gradient Boosted Trees, and Support Vector Machine. As a Deep Learning approach, the ANN can better capture the non-linear relationship in time series data than traditional models [12, 13, 14]. ANN is a prominent Machine Learning formula that can take on human neuron cells to duplicate a non-linear relationship between input and target vectors [10]. Neuron cells obtain input vectors from other cells and generate output using the activation function. The network consists of diverse hidden layers, with each layer containing many neuron cells [15]. The principal of a neuron model is that input, \( x \), together with a bias, \( b \) is weighted by, \( w \), and then summarized. The bias, \( b \), is a scalar value whereas the input \( x \) and the weights \( w \) are vector-valued, i.e., \( x \in \mathbb{R}^n \) and \( w \in \mathbb{R}^n \) with \( n \in \mathbb{N} \) corresponding to the dimension of the input. In this case, the term bias is not permanently present but sometimes it can be omitted. The sum of these terms, i.e., \( z = w^T x + b \) forms then the argument of an activation function, \( \phi \), resulting in the output of the neuron model [16],
\[
y = \phi(z) = \phi(w^T x + b) \tag{1}
\]
Generally, the neuron model is represented more ergonomically by limiting the focus to its main elements, or it is represented by highlighting only part of the input, as shown in Figures 1, a, and b.

![Figure 1. Illustration of an Artificial Neural Network model with (a) activation function \( \phi \), and (b) simplified by the key elements, i.e., the input, the output, and the weight, are depicted [16].](image)

For business demand prediction, we put forward the model construction in four steps merely. The first stage, namely data collection using the monthly in DataCo supply chain dataset from Kaggle public repository with areas of key recorded activities are provision, production, sales, and commercial distribution. Correlation of structured data with unstructured data to generate knowledge is also possible. The collected training data contains 125,000 rows and 28 columns/factors. The deep net architecture applied in the next step requires feature selection, however various irrelevant factors that have a low association and quality issues from the data can be automatically removed. In this part, we chose a column containing numerical data, that is sales per customer which total sales per customer made per customer.

Second, Deep Learning model training with a collection of dominant factors selected automatically. In view of Rapidminer software technology has the built-in mechanism of feature engineering, so that it could be done automatically to retrench the time building cultivation more effectively. The third stage is how the model could comprehend. The trained Deep Learning model comprehensively represents the temporal relationship between various demand forecasting factors and sales per customer. The weights of the neuron links and the attention scores could be applied to determine which original factors have the most influential lag orders. Especially, no manual feature selection or extraction is required in this framework, and all feature engineering tasks are automatically performed by the Deep Learning
model. The fourth step is to assess the model errors. Ideally the optimum model was elected based on the nethermost errors acquired. The following standard way were used to evaluate the forecasting performance of the proposed Deep Learning and other models. The Root Mean Square Error (RMSE) is given by

\[ RMSE = \sqrt{\frac{\sum_{t=1}^{T}(L_a - L_p)^2}{T}} \]  

where \( L_a \) denotes the actual demand, \( T \) represents the total number of forecasting points and \( L_p \) is the corresponding forecasting demand [15]. The Mean Absolute Error (MAE) can be expressed as

\[ MAE = \frac{\sum_{t=1}^{T}|L_p - L_a|}{T} \]  

The MAE of predictions where larger errors contribute proportionally to the error calculations [15]. This measure is less sensitive against outliers. The lower the error, the better the predictive model. The last performance evaluation is model building time, which is the entire time finished building the predictive model.

3. Results and Discussion

The proposed Deep Learning model was used to forecast the load demand of a supply chain. Its performance was also compared to that of the Random Forest, Gradient Boosted Trees, and Support Vector Machine. These algorithms were running with Rapidminer software. Figure 2 (a) shows the predicted values chart of demand data from January 2015 to September 2017. All predictions are equal to the values and lie on the ideal diagonal line which is acceptable because the closer they are to it, the better the predictions are. While Figure 2 (b) shows the histogram frequency of each error value, which is for valuable models, the smaller error values are the most frequent, indicate with closer to zero.

![Figure 2](image)

**Figure 2.** (a) the forecasting result in comparison with actual values, and (b) the histogram of error frequency.

The primary task of developing a forecasting framework is to combine the corresponding factors to reduce forecasting errors as measured by several performance indicators such as root mean square error (RMSE) and mean absolute error (MAE). RMSE predicts where a larger error has a disproportionately large effect, that is, more outliers will be penalized. The predictive model is valid when the error is low. The results as shown in Figure 3 show that the Deep Learning result value is the best of all other algorithms. The Deep Learning model produces an average absolute error (MAE) of 0.371. This means if we predict Sales per customer in 973.74 as a value, so the real value is likely between 973.369 and 974.111. This is seen as a good value in general.
Figure 3. (a) Model building time in second, (b) The prediction and its important factors.

Deep Learning Model process used 4 layers and 10 iterations with the total time spent building the predictive model as shown in Figure 3 (a) is the process that begins to execute until it is finished shows the Deep Learning is the best as well. While Figure 3 (b) is the model that uses the data in each row to calculate a prediction where there are still many values or factors supporting product supply compared to opposite values. This visualization may not be apparent often in a real-life but it will help decision-makers in breakdown the data and variables, make the prediction simulation, and sort out which supporting and opposing values. Some values of the data support the prediction, meaning that they are consistent with the prediction obtained by the model after combining all the data, and they help to explain it. The degree to which the value supports the prediction is indicated by the number of green bars. A few values of the data may support a different prediction than the one obtained after combining all the data. The final prediction is obtained despite these opposing values because the opposing values carry less weight. The degree to which the value opposes the prediction is indicated by the number of red bars. Also, the chart shows confidences which is a confidence value to its prediction by the model assigned, similar to a probability. When summed up, the confidence values for all possible prediction outcomes add up to 1.

4. Conclusion
From the results and discussion obtained, it is concluded that the Deep Learning model has the minimum running time to achieve results compared to other algorithms such as Random Forest, Gradient Boosted Trees, and Support Vector Machine. This fact will be an effective solution for busy decision-makers to find useful solutions to problems. It is predicted that demand forecasting problems and how to respond to them will still be a challenging analysis trend in the coming year. This is due to how heterogeneous consumer desires are that even the best manufacturers will not be successful if they do not produce the product that consumers want. Therefore, the machine learning model with its various algorithms to estimate consumer demand and respond to it is still very relevant to be discussed in future research, mainly for fast-moving consumer products, such as food or fashion, that require real-time forecasting.
References

[1] Hyndman RJ and Athanasopoulos G 2018 *Forecasting: Principles and Practice, 2nd edition* (Melbourne, Australia: OTexts)

[2] Kilimci ZH 2019 *Hindawi* 2019 1

[3] Van derLaan E, van Dalen J., Rohrmoser M, and Simpson R 2016 *J. Oper. Manag.* 45 114

[4] Chen D, Guo K, and Ubakanma G 2015 *Int. J. Bus. Forecast. Mark. Intell.* 2(1) 1

[5] Shi Q et al. 2020 *Proc. AAAI Conf. Artif. Intell.* 4 5758

[6] Bradlow ET, Gangwar M, Kopalle P and Voleti S 2017 *J. Retail.* 93 (1) 79

[7] Blue Pi 2020 Why You Need Demand Forecasting Solutions Using Machine Learning?,” *BluePi Blog*. 2020. https://medium.com/bluepi-blog/why-you-need-demand-forecasting-solutions-using-machine-learning-eb38ba49dc0f (accessed May 19, 2020).

[8] Johansson C, Bergkvist M, Geysen D, De Somer O, Lavesson N, and Vanhoudt D 2017 *Energy Procedia* 116 208

[9] Lecun Y, Bengio Y, and Hinton G 2015 *Nature* 521(7553) 436

[10] Jain AK, Mao J, and Mohiuddin KM 1996 *Computer (Long. Beach. Calif).* 29(3) 31

[11] Ma X, Gildin E and Plaksina T J. *Unconv. Oil Gas Resour.* 9 1

[12] Ma X, Tao Z, Wang Y, Yu H, and Wang Y 2015 *Transp. Res. Part C Emerg. Technol.* 54 187

[13] Ma X, Dai Z, He Z, Ma J, Wang Y, and Wang Y 2017 *Sensors (Switzerland)* 17(4)

[14] Haykin S 1999 *Neural Networks: A Comprehensive Foundation* (International:Prentice-Hall)

[15] Wen L, Zhou K, and Yang S 2020 *Electr. Power Syst. Res.* 179 106073

[16] Emmert-Streib F, Yang Z, Feng H, Tripathi S, and Dehmer M 2020 *Front. Artif. Intell.* 3 1

[17] Pouyanfar S et al. 2018 *ACM Comput. Surv.* 51(5) 1

[18] Bao W, Yue J, and Rao Y 2017 *PLoS One* 12 (7) 1