Demand Forecasting in Retail Based on Deep Learning Models

Panupong Wanjantuk
Department of Computer Engineering
Khon Kaen University
Khon Kaen, Thailand
panupong@kku.ac.th

Temsiri Sapsaman
Department of Production and Robotics Engineering
King Mongkut's University of Technology North Bangkok
Bangkok, Thailand
temsiri.s@eng.kmutnb.ac.th

Ruth Banomyong
Department of International Business, Logistics and Transport
Thammasat University
Bangkok, Thailand
ruth@tbs.tu.ac.th

Prarawan Senachai
Department of Marketing
Khon Kaen University
Khon Kaen, Thailand
prarse@kku.ac.th

***Abstract*—Forecasting future demand is essential to making supply chain decisions. Time series forecasting methods use historical demand to produce a forecast. The purpose of this paper is to apply a deep learning model called Long Short-Term Memory (LSTM) using historical sales data from a Thailand retailer for forecasting sales in retail supply chain. The data used in the analysis includes 46 months of actual daily sales of selected consumer goods. Based on this data, LSTM models have been trained and evaluated by using different look-back window sizes and different amounts of time to forecast the future sales. The result of the research is the appropriate deep learning models and parameters for demand forecasting that concludes from the analysis and evaluation of sales forecasting in the retail supply chain based on LSTM models. For each model, the relationships between the performance and the parameters, the look-back window sizes and the number of predicted time points into the future, are presented. Accurate demand forecasting in consumer goods could increase the competitive power of a retailer and improve its performance. The forecasting model can be used by businesses to optimize their inventory level, increase their bargaining power for purchasing, and ensure product availability. A novelty is the use of the LSTM model trained on real transaction data from a retail company that has based its business on the supply chain with suppliers and recipients in Thailand.**

Keywords— Demand Forecasting, Deep Learning, Supply Chain Management, Time Series, LSTM

I. Introduction

A. Retail business and management

Retail business is one of the most dynamically developing market segments [1] and globally speaking, the most prominent [2]. It involves selling consumer goods and services to customers for their daily use as a role at the end of the supply chain [3]. However, retail businesses have had high competition [4] and numerous challenges [3]. The significant problems of the retail trade are related to supply chain inefficiencies, that can be found in both directions: upstream and downstream. For instance, high out-of-stock conditions, long lead times, inefficiencies affecting the accuracy of demand forecasts, and low on-shelf availability result in losses of revenue even if the products are available on site. Meanwhile, overstocking affects non-sustainable actions, perishable products, the blockage of storage space, and needle handling, including planning efforts [5].

Thailand, retailing emerged almost 70 years with a rapid expansion of supermarkets, hypermarkets, and convenience stores in the last two decades [6]. However, department stores significantly impacted Thai retailing and are said to be considering spinning off the supermarkets to become free-standing operations [7]. Indeed, supermarkets were appreciated as a one-stop shopping place and much preferred by many Thai consumers on factors ranging from the atmosphere (air conditioning), car parking availability, hygiene, and price [6]. It is found that socioeconomic and geographic diffusion is already advanced, with most Thai people having access to and utilizing modern retail, while modern retail controls half of the food sales [6]. The success of the retail business in Thailnd is to offering low-price deals, and improved purchasing convenience by using advanced IT [8].

Accordingly, retail management is crucial since it provides guidance for the future usage of services and infrastructure components [9] given that retail management involves all the steps that are required to fulfill the needs of the customers [3] and to enhance the shopping experience by providing good product availability [10]. Retail management also involves deploying accounting and management information systems to control activities, warehouse activities, and distribution, as well as to develop new products, including marketing activities [3]. Finally, proper retail management can save time, ensure customer satisfaction, and maintain business goals [10].

Accordingly, retail management is crucial since it provides guidance for the future usage of services and infrastructure components [9] given that retail management involves all the steps that are required to fulfill the needs of the customers [3] and to enhance the shopping experience by providing good product availability [10]. Retail management also involves deploying accounting and management information systems to control activities, warehouse activities, and distribution, as well as to develop new products, including marketing activities [3]. Finally, proper retail management can save time, ensure customer satisfaction, and maintain business goals [10].

B. The time series method of demand forecasting in Retail SCM

Making a supply chain decision to forecast future demand is necessary [11]. However, demand forecasting truly depends upon the quality of the data sets [5]. There are various factors, which are related to demand forecasting for the customers in retail business, such as seasonality [12], [13], new trends, crisis, the commercial behavior of their competitors in the marketplace [12], marketing factors [14], weather [15], etc. Consequently, better demand forecasting for the retail industry could optimize stocks, minimize costs, reduce inventory problems, and could, as a result, increase revenues and profits for retailers [12].

Recently, Predictive Analytics (PA) in Supply Chain Management (SCM) has received attention from both academicians and researchers and has been recognized as a means to predict demand forecasting in the era of Industry 4.0 [5]. Unfortunately, for different settings and types of data, suitable forecasting methods still remain unclear [16]. Despite this, time series analyses are the most common method that can be used for demand forecasting [5]. Moreover, time series analyses are critical elements in the automation and optimization of business processes and provide an intuitive overview that can assist in solving large-scale forecasting problems and in enabling data-driven decision-making processes [9].

Forecasting time series data is vital in many aspects of business [17]. It can be utilized to investigate both short-term (6-12 months) and long-term levels (over 1 year) [5]. Indeed, the time series method has gained acceptance in various fields. For instance, in retail businesses, data mining has been used, and database researchers have used the time series method to forecast the consumption of individual households and the demand for all products that the large retailers offer to their customers [9]. Various techniques have forecasted the subsequent lag time series data [17]. Long Short-Term Memory (LSTM) is one of the practical techniques that is used in time-series prediction [17]; [18]. In 2017, their usage grew in many disciplines, such as economics, finance, business, and computer science [17].

LSTM has been assimilated into the Recurrent Neural Network (RNN), which has the capability of remembering the values from earlier stages and employing them for future use [17]. It can remember information far back in the time series like a gated cell, which can be used to store information and is similar to computer memory [18] and consists of a forget gate (hidden layer), an input gate, and an output gate [19]. However, it is necessary to determine what a neural network looks like before delving into LSTM [17]. Finally, formulating sales predictions as time-series forecasting is vital for any business. When retailers employ analytics, the data becomes a factor that can be employed to create and/or to increase the added value of their businesses [20].

Consequently, this study approaches sales prediction via the LSTM model, which is expected to handle the non-linearity of sales forecasting with a higher prediction accuracy [18]. The investigation compares the performance of LSTM models with different window of information. The data are the weekly values of the regional wholesale revenue over 4 years for 10 consumer goods placed in household retails.

II. Long Short-term Memory (LSTM)

Time Series Forecasting (TSF) is predicting the future values of variables based on the previous and current values of time series [21]. Describing the relationships between these values, functions or models for time series are traditional models, machine learning models, and deep learning models. Using a window of historical data, deep learning model is created to predict the data point in the future. The deep learning algorithm and the size of the look-back window affects the performance of the model. This study investigates the performance of the Long Short-term Memory (LSTM), which is based on the Recurrent Neural Network (RNN). The RNNs use the neural networks to represent the relationship between the past variables (inputs) and the future variable (output). This neural network consists of hidden layers and weights that are recurrently adjusted using the inputs to minimize the errors between the calculated and the known values. The RNNs can result in inaccuracy due to gradient vanishing problem; that is, the RNN models are the result of data from a small window of time or short-term memory. To address this problem the LSTMs introduced four additional layers: forget layer (1), input layer (2), output layer (3), and addition layer (4) [11]. The four layers manipulate the information used in creating the model and allow learning over longer-term data. While the forget layer decides information to be stored in the cell state, the input layer decides which to be updated. The output layer generates the potential output and then combines with the results from previous layer in the addition layer. Then the next layer selects the cell states (5) and the last layer computes the output (6).

$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] + b_f\right) \tag{1}$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{2}$$

$$g_t = tanh\left(W_g \cdot [h_{t-1}, x_t] + b_g\right) \tag{3}$$

$$C_t = f_t \times C_{t-1} + i_t \times g_t \tag{4}$$

$$O_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{5}$$

$$h_t = O_t \times tanh(C_t) \tag{6}$$

The Stacked Deep LSTM Network [22] is used in this study with architecture shown Fig. 1. The models are built by stacking two 50-node LSTM layers between the input and output.

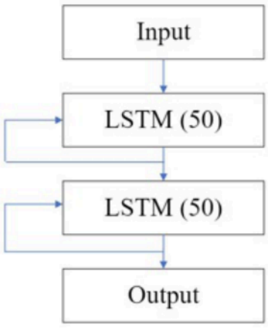

Fig. 1. Stacked Deep LSTM Network

III. 1. DATA AND EXPERIMENTS

A. Dataset

The dataset used in this study is the daily sales of 10 best-seller consumer goods from one retailer in Thailand, which includes data for the four-year time period from July 2018 to May 2022. The 10 consumer goods are 180ml UHT milk, refined sugar 1kg, 200ml UHT milk, canned Fish, rice whisky, energy drink, herbal drink, canned coffee, refined sugar 50kg, and cane sugar. The data was collected daily and the data set has 1,364 rows and 3 columns. The first column is simply an index and was ignored for the analysis. The second column labeled as date and time. The "sale" column is the target variable. The data have been transformed from daily sale to weekly sale for forecasting. The time series for all consumer goods are plotted in Fig. 2. Four consumer goods have less consistency at the end of the time series, i.e., 180ml UHT milk, refined sugar 1kg, 200ml UHT milk, canned Fish, and rice whisky. This indicates unusual demand.

B. Experiments

The experiments were performed with the LSTM stacked model and different values of the look-back window representing the portion of the recent past used as inputs. Window sizes of 1, 2, 3, 4, 5, 6, 7 and 8 weeks were used in the experiment. For the LSTM deep learning models, hyper parameters were set as follows:

- Learning rates = 0.001
- Batch sizes = 32
- Optimizer = Adam
- Epochs = 500

The LSTM model used in the experiment can be summarized as follow:

```
Layer (type)              Output Shape         Param #
=================================================================
lstm_210 (LSTM)           (None, 1, 50)        10400

lstm_211 (LSTM)           (None, 50)           20200

dense_105 (Dense)         (None, 1)            51

=================================================================
Total params: 30,651
Trainable params: 30,651
Non-trainable params: 0
```

C. Measures of Evaluation

To compare the performance of the models, the Mean Squared Error (MSE) was evaluated as the loss function. The training and testing loss were computed as a function of the epochs to detect possible overfitting. The Mean Absolute Error (MAE) (7) and Root Mean Squared Error (RMSE) (8) were computed.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \quad (7)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad (8)$$

IV. RESULTS

TABLE 1 summarizes the results of our experiments. The LSTM models perform the good performance with the low Mean Average Errors (MAE) and Root Mean Square Errors (RSME). Overall, the results indicate the good performance of the LSTM models for demand forecasting with the largest best RMSE as small as 0.508. Smaller windows are likely to lead to more efficient methods. For most cases the smaller-size of look-back window results in better performance. These cases have more consistent time series throughout the entire period, while other cases have significant variations during the end. The best size for the look-back window to predict 1 or 2 weeks. These cases also have less dependency on the size of the look-back window as the MAEs and RMSEs are similar. On the other hand, the four consumer goods with unusual demands seem to need longer look-back window and have more dependency on the size of the window.

V. CONCLUSION

To make well-versed supply chain decisions, results from forecasting future demand is crucial. The demand forecasting using a deep learning algorithm is investigated in this study with the four-year sales of 10 best-seller consumer goods from one retailer in Thailand. The time-series forecasting is performed using the Stacked Deep LSTM Network with various look-back window sizes. The Mean Squared Errors (MSE) and the Root Mean Squared Errors (RMSE) are evaluated to measure the performance of the LSTM models. This investigation has found that the LSTM method is suitable for demand forecasting and the suitable look-back window is 1 or 2 weeks for most cases. Future study can investigate other architectures of the LSTM model or other deep-learning algorithms such as Gated Recurrent Unit (GRU) and Transformer model. Further exploration can be done on data with different pattern from the best seller goods.

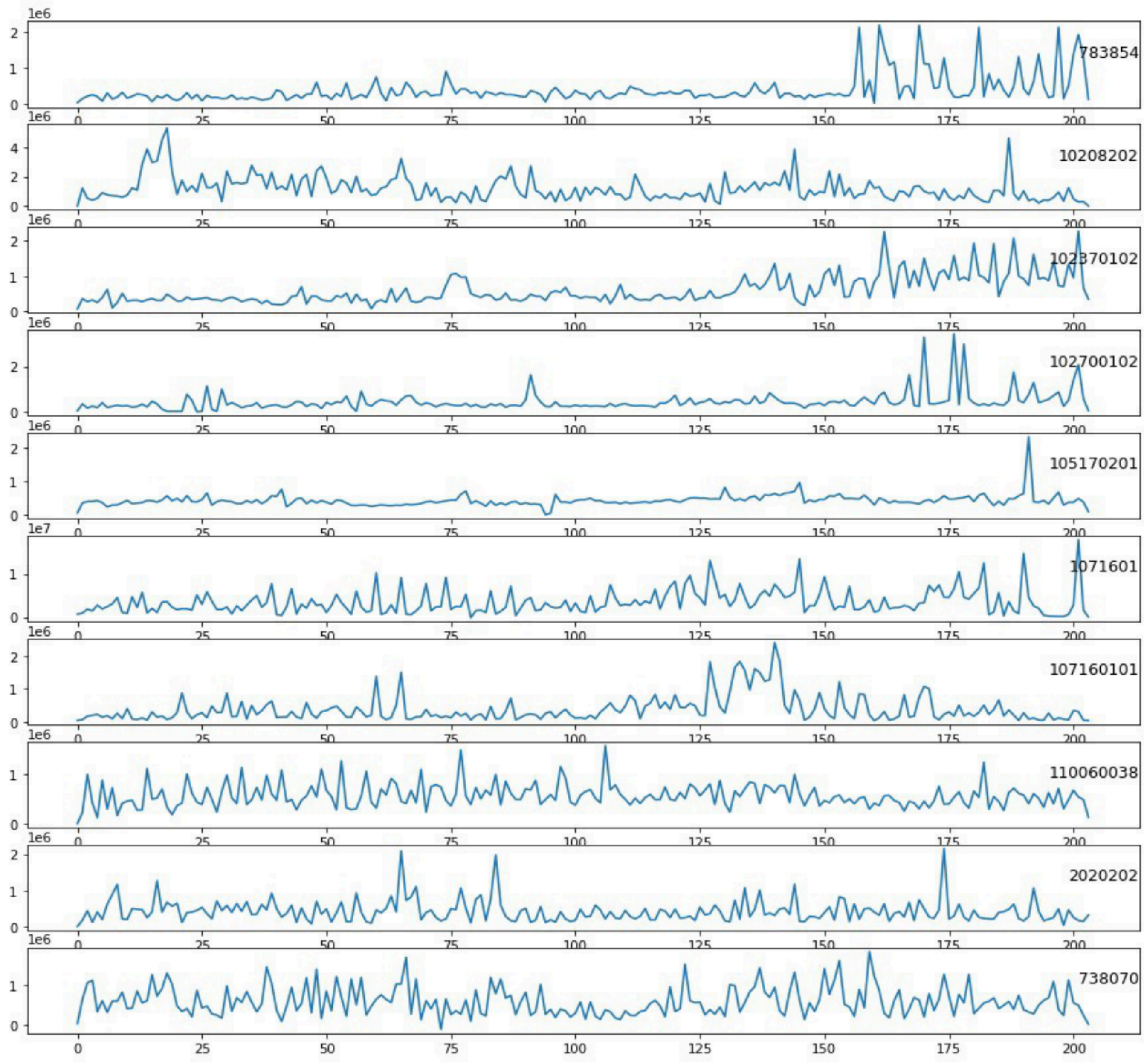

Fig. 2. The time series for 10 consumer goods

TABLE I. MAE AND RMSE OF DEMAND FORECASTING IN 10 CONSUMER GOODS

Look-back Window	0783854		10208202		102370102		102700102		105170201		1071601		107160101		110060038		2020202		738070	
	UHT Milk 180ml		Refined Sugar 1kg		UHT Milk 200ml		Canned Fish		Rice Whisky		Energy Drink		Herbal Drink		Canned Coffee		Refined Sugar 50kg		Cane Sugar	
	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
1 week	0.226	0.321	0.105	0.163	0.201	**0.252**	0.561	1.052	0.217	0.583	**0.202**	**0.292**	**0.094**	**0.112**	**0.114**	**0.140**	**0.105**	**0.167**	**0.118**	**0.151**
2 weeks	**0.204**	0.320	**0.099**	**0.148**	0.309	0.402	**0.303**	0.523	0.202	0.534	0.258	0.374	0.113	0.154	0.118	0.148	0.136	0.215	0.132	0.165
3 weeks	0.208	0.320	0.133	0.205	0.248	0.310	0.324	0.565	0.193	0.417	0.221	0.301	0.095	0.130	0.171	0.213	0.117	0.188	0.203	0.253
4 weeks	0.272	0.379	0.150	0.301	0.262	0.337	0.328	0.559	0.212	0.414	0.255	0.346	0.113	0.154	0.163	0.201	0.136	0.217	0.172	0.210
5 weeks	0.266	0.376	0.151	0.344	0.258	0.342	0.337	0.578	**0.192**	**0.404**	0.243	0.339	0.112	0.148	0.194	0.252	0.149	0.215	0.192	0.214
6 weeks	0.252	**0.304**	0.131	0.186	0.315	0.415	0.403	0.680	0.226	0.421	0.231	0.310	0.106	0.142	0.179	0.242	0.155	0.226	0.146	0.176
7 weeks	0.249	0.343	0.128	0.181	**0.199**	0.262	0.357	0.560	0.366	0.821	0.235	0.344	0.097	0.136	0.199	0.236	0.128	0.211	0.135	0.177
8 weeks	0.250	0.349	0.149	0.242	0.245	0.331	0.306	**0.508**	0.347	0.717	0.272	0.348	0.118	0.149	0.153	0.202	0.129	0.211	0.172	0.206

REFERENCES

[1] N. M. Chernysheva, R. M. Kotov, E. S. Altynbaeva, and Yu. A. Khristoforova, “Current condition of the retail market.,” *Food Process. Tech. Technol.*, vol. 47, no. 4, pp. 145–151, Dec. 2017, doi: 10.21603/2074-9414-2017-4-145-151.

[2] T. Rajesh and P. Asokan, “Effects on consumer sales promotion: A major development in FMCG sector.,” *CLEAR Int. J. Res. Commer. Manag.*, vol. 6, no. 11, 2015.

[3] M. B. R. Vaja, “Retail management.,” *Int. J. Res. Anal. Rev.*, vol. 2, no. 1, 2015.

[4] A. Mittal and D. Jhamb, “Determinants of Shopping Mall Attractiveness: The Indian Context,” *Procedia Econ. Finance*, vol. 37, pp. 386–390, 2016, doi: https://doi.org/10.1016/S2212-5671(16)30141-1.

[5] T. Falatouri, F. Darbanian, P. Brandtner, and C. Udokwu, “Predictive Analytics for Demand Forecasting – A Comparison of SARIMA and LSTM in Retail SCM,” *Procedia Comput. Sci.*, vol. 200, pp. 993–1003, 2022, doi: 10.1016/j.procs.2022.01.298.

[6] M. Kelly, S. Seubsman, C. Banwell, J. Dixon, and A. Sleigh, “Traditional, modern or mixed? Perspectives on social, economic, and health impacts of evolving food retail in Thailand,” *Agric. Hum. Values*, vol. 32, no. 3, pp. 445–460, Sep. 2015, doi: 10.1007/s10460-014-9561-z.

[7] A. Feeny, T. Vongpatanasin, and A. Soonsatham, “Retailing in Thailand,” *Int. J. Retail Distrib. Manag.*, vol. 24, no. 8, pp. 38–44, Sep. 1996, doi: 10.1108/09590559610150375.

[8] E. Gen, *Diversifying retail and distribution in Thailand*. Chiang Mai: Silkworm Books, 2013.

[9] C. Faloutsos, V. Flunkert, J. Gasthaus, T. Januschowski, and Y. Wang, “Forecasting Big Time Series: Theory and Practice,” in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, New York, NY, USA, 2019, pp. 3209–3210. doi: 10.1145/3292500.3332289.

[10] A. Grubor and N. Milicevic, “The Analysis of FMCG Product Availability in Retail Stores,” *Eng. Econ.*, vol. 26, Feb. 2015, doi: 10.5755/j01.ee.26.1.7070.

[11] D. Pliszczuk, P. Lesiak, K. Zuk, and T. Cieplak, “Forecasting Sales in the Supply Chain Based on the LSTM Network: The Case of Furniture Industry,” *Eur. Res. Stud. J.*, vol. XXIV, no. Special Issue 2, pp. 627–636, Jun. 2021, doi: 10.35808/ersj/2291.

[12] A. O. Akyuz, M. Uysal, B. A. Bulbul, and M. O. Uysal, “Ensemble approach for time series analysis in demand forecasting: Ensemble learning,” in *2017 IEEE International Conference on INnovations in Intelligent SysTems and Applications (INISTA)*, Gdynia, Poland, Jul. 2017, pp. 7–12. doi: 10.1109/INISTA.2017.8001123.

[13] C. Chatfield, *The analysis of time series: an introduction*, 6th ed. Boca Raton, FL: Chapman & Hall/CRC, 2004.

[14] N. G. Mankiw, *Study Guide for Mankiw's Principles of Macroeconomics*. Cengage Learning, 2011. [Online]. Available: https://books.google.ht/books?id=UdWmAQAAQBAJ

[15] K. Murray, F. Muro, A. Finn, and P. Popkowski Leszczyc, “The Effect of weather on consumer spending,” *J. Retail. Consum. Serv.*, vol. 17, pp. 512–520, Jul. 2010, doi: 10.1016/j.jretconser.2010.08.006.

[16] F. Petropoulos, S. Makridakis, V. Assimakopoulos, and K. Nikolopoulos, “‘Horses for Courses’ in demand forecasting,” *Eur. J. Oper. Res.*, vol. 237, pp. 152–163, Aug. 2014, doi: 10.1016/j.ejor.2014.02.036.

[17] S. Siami-Namini, N. Tavakoli, and A. Siami Namin, “A Comparison of ARIMA and LSTM in Forecasting Time Series,” in *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*, 2018, pp. 1394–1401. doi: 10.1109/ICMLA.2018.00227.

[18] Q. Yu, K. Wang, J. O. Strandhagen, and Y. Wang, “Application of Long Short-Term Memory Neural Network to Sales Forecasting in Retail—A Case Study,” 2017.

[19] Y. Ensafi, S. H. Amin, G. Zhang, and B. Shah, “Time-series forecasting of seasonal items sales using machine learning – A comparative analysis,” *Int. J. Inf. Manag. Data Insights*, vol. 2, no. 1, p. 100058, 2022, doi: https://doi.org/10.1016/j.jjimei.2022.100058.

[20] E. Hofmann and E. Rutschmann, “Big data analytics and demand forecasting in supply chains: a conceptual analysis,” *Int. J. Logist. Manag.*, vol. 29, no. 2, pp. 739–766, Jan. 2018, doi: 10.1108/IJLM-04-2017-0088.

[21] J. Shi, M. Jain, and G. Narasimhan, “Time Series Forecasting (TSF) Using Various Deep Learning Models.” arXiv, Apr. 23, 2022. Accessed: Oct. 03, 2022. [Online]. Available: http://arxiv.org/abs/2204.11115

[22] U. Singh, S. Chauhan, A. Krishnamachari, and L. Vig, “Ensemble of deep long short term memory networks for labelling origin of replication sequences,” in *2015 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*, Campus des Cordeliers, Paris, France, Oct. 2015, pp. 1–7. doi: 10.1109/DSAA.2015.7344871.