Demand Forecasting in Production Planning for Dairy Products Using Machine Learning and Statistical Method

Chayuth Vithitsoontorn
Computer Engineering
Chulalongkorn University
Bangkok, Thailand
6270055421@student.chula.ac.th

Prabhas Chongstitvatana
Computer Engineering
Chulalongkorn University
Bangkok, Thailand
prabhas@chula.ac.th

2022 International Electrical Engineering Congress (iEECON) | 978-1-6654-0206-4/22/$31.00 ©2022 IEEE | DOI: 10.1109/iEECON53204.2022.9741683

***Abstract*— Demand forecasting is an essential task of every industry. Efficient forecasting relieves the excessive stock and out-of-stock problem, reducing revenue loss. This research performs a direct multistep forecast approach of demand forecasting on 8 dairy products of 5 different dairy production plants with 5-year data. Widely used traditional statistical method and the stage of the art deep learning method for sequence problems are picked. ARIMA and LSTM. The models are compared in many aspects, monthly observations against weekly observations, univariate against multivariate, and statistical against deep learning using model error and business metrics. The result shows that both statistical and deep learning method are reliable and are suitable to be used in demand forecasting. There is no single best optimization algorithm. ARIMAs predict the future in an average straight line. It shows the best result on few wavering series, whereas LSTMs predict the future value follow the seasonal of series. It beats ARIMAs on strong trend series. Training the model on monthly observations provide better error score.**

Keywords*— *Demand Forecasting; Deep Learning; Long Short-Term Memory (LSTM); Autoregressive Integrated Moving Average (ARIMA); Time Series Data

I. Introduction

With a rise in the number of competitors, entrepreneurs need to adapt their business to attract new customers and keep their existing guests, to expand or keep their market share, one needs to know the customer demand and handle them wisely, providing decent customer experiences from every aspect of a business. From the manufacturing aspect, the best thing we can do is satisfy the order demand of customers to reduce the churn rate. There are ways to suffice the demand: (1) applying minimum stock limit policy, (2) substituting customer order with similar products, (3) collecting the customer's demands at the lowest chain of the supply chain, (4) cooperating with the lower chain to know the customer demand, and last (5) forecasting the demand which is in the scope of this research. This study uses a real-world dataset for comparative analysis of demand forecasting methods between traditional statistical method and recent method: deep learning. Also, to investigate the feature candidates and provide insight from the analyzed historical and expected to be useful for dairy product forecasting or typical time series forecasting.

A. *ARIMA*

Auto-Regressive Integrated Moving Average (ARIMA) is a class of statistical models for forecasting time series data. The model uses past time series values, lags, and lagged forecast error to predict the future values. The model has three characteristic terms. (1) Auto Regressive (AR) is number of lag observations (2) Integrated (I) is the number of times required to differentiate data to make it stationary. (3) Moving Average (MA) is a size of the moving average or number of lagged forecast errors.

Then a linear regression model is constructed using the above-specified terms. ARIMA can also have more than one regression model by providing additional explanatory variable for multivariate forecasting problem called ARIMAX.

B. *LSTM*

Long Short-Term Memory (LSTM) [1] is a neural network type model. It is an improved version of recurrent neural network (RNN). It is often used in handling long sequence data problems such as NLP and time series data like RNN. LSTM contains additional gates known as memory cells allowing better control of gradient flow in weight adjusting of backpropagation methods, [2] demonstrates the solving of vanishing gradient problem found in RNN.The LSTM model is used in this work as an evidence of dairy sale time series prediction capability.

C. *Encoder Decoder LSTM*

Encoder Decoder LSTM is one type of LSTM model with additional layers for encoding input and decoding to output. The concept behind the architecture is to transform the features into smaller dimensions for better learning of the network.

II. Related Works

Several research studies regression problem on variety industries. In early days, statistical methods are widely used and researched, following the deep learning methods and the develop of hybrid models to gain benefits from both methods.

A. *Single models*

C. P. da Veiga, C. R. P. da Veiga, W. Puchalski, L. dos Santos Coelho and U. Tortato [3] predicts monthly Brazilian retail sales, WNN provides the least error. Fattah, J., et al [4] forecast monthly closing stock price by comparing between ARIMA and LSTM model. LSTM has less error than ARIMA. Chniti, G., H. Bakir, and H. Zaher [5] forecast daily phone prices using SVR and LSTM model. SVR produces the least error for univariate while LSTM has the fewest error for multivariate, more variables increase the model accuracy.

B. *Hybrid models*

Fan, D., et al., [6] compare ARIMA, LSTM and Hybrid ARIMA-LSTM. LSTM performs better than ARIMA if the data has non-linear relation. The hybrid model between ARIMA, and LSTM provides the best result. They show that better feature selection led to better model accuracy. Livieris, I.E., E. Pintelas, and P. Pintelas, [7] and [8] conclude that hybrid models provide the best result and LSTM is a good candidate for a single model.

C. *Other suggestions*

Ren, S., H.-L. Chan, and T. Siqin [9] suggest using both qualitative and quantitative methods for the demand forecasting. Firstly, predict the result with the model then follow with a judgmental forecast by a domain expert. It provides a method to capture fast-changing market information problems. Cadavid, J.P.U., et al., [10] categorize the demand forecasting in machine learning for production planning into 11 activities. Hyndman, R.J. and G. Athanasopoulos [11] suggest using number of days in month to reduce bias in the forecasting model. Hyndman, Rob J., and Anne B. Koehler [12] proposed a new easily interpretable measure MASE.Molnar, Christoph.[13] Explains the importance of model interpretation and interpretation guideline.

III. DATASETS

The diary sale transaction data are collected from the ERP system of the Dairy Farming Promotion Organization of Thailand (DPO) within Oct 2016 – Sep 2021 (5 financial years) recorded by 5 plants. Over 100 products are on sale, but only 25 products require the forecast because they cover 90% of the total revenues. The series with less than 5 missing monthly observations and products produced by all plants is selected. 8 products remain or a total of 40 different series are remained. Sale volume of the series is highly fluctuated ranging from 10 unit/month to over 800 thousand unit/month. The holiday data are collected from the Bank of Thailand.

IV. RESEARCH METHODS

The research method flow is following the standard principles of model prediction task as below.

A. *Data Exploration*

Loess (STL) is used for Seasonal and Trend decomposition to each series where the individual series will be decomposed into a trend, a seasonal, and a remainder then the data is visually inspected to create assumptions and use together with the final experiment result for the conclusion.

B. *Data Preprocessing*

The series are imputed depending on the aggregation of each field. Lastly, we convert zero values of sale field to the smallest positive number to prevent the divide by zero problem. The average is imputed with mode value.

C. *Feature Selection*

Several statistical methods are performed (Pearson‘s correlation, Spearman’s correlation, and low variant) plus recursive feature elimination using Gradient boosting algorithm then follow with judgmental decision to select the appropriate features.

D. *Feature Extraction and Transformation*

The value is aggregated to monthly or weekly level, derive new fields from existing fields. Convert the series to the model’s preferred format by following steps.

1) *Box-Cox Transformation and standardlization*
2) *Time serie format to supervised format*
3) *Feature-scaling to range [0,1]*

E. *Model Training and Validation*

1) *Train-Test Split*

The first 4 years data are reserved as an initial training set and the latter year an as out-of-sample test set. Time series cross-validation [11] is used as an evaluation method. The method reuses tested observations as a trainset. It appends them to the original trainset at the end of each iteration.

2) *Model*

Sixes forecasting algorithms are used for each series, two statistical methods (ARIMA and ARIMAX) and four deep learning methods (Univariate/Multivariate LSTM and theirs Encoder-Decoder version)

3) *Hyperparameter tuning*

Auto ARIMA models with Akaike Information Criteria (AIC) is used for ARIMA model and grid search is used for selecting the parameters for LSTM model. Mean square error (MSE) is picked as a loss function. To deal with the stochastic nature of neural network-based models, the training of LSTM models is repeated three times and record the average score among those iterations.

V. EVALUATION METHODS

Three metrics are evaluated in terms of modeling error scores; [12](RMSE, MAE and MASE) and two business metrics; ([14] fill rate and revenue loss)

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \hat{x}_i)}{N}} \quad (1)$$

$$MAE = \frac{\sum_{i=1}^{N}|x_i - \hat{x}_i|}{\mathrm{N}} \quad (2)$$

$$MASE = \frac{MAE}{MAE_{naive}} \quad (3)$$

$$Fill\ rate = \frac{\sum_{i=1}^{N} \min_{+}\{x_i, \hat{x}_i\}}{\sum_{i=1}^{N} x_i} \quad (4)$$

$$Revenue\ loss = \sum_{i=1}^{N} \max_{+}\{(x_i - \hat{x}_i), 0\} \times price_i \quad (5)$$

The result from these metrics and additional information are used to explain the character of each model in the following aspects.

- *Statistical methods and Deep learning methods*
- *Monthly and Weekly Observations*
- *Univariate and Multivariate models*
- *Typical LSTM and encoder-decoder version.*

TABLE I. SUPERVISED FORMAT

Monthly			Weekly		
Step	***X***	***Y***	***Step***	***X***	***Y***
1	M1…M3	M4 … M6	1	W1…W12	W13 … W24
2	M2…M4	M5 … M7		…	…
3	M3…M5	M6 … M8		…	

Monthly			Weekly		
Step	*X*	*Y*	*Step*	*X*	*Y*
4	M4…M6	M7 … M9	12	W13…W24	W25 … W36
	…	…		…	…

TABLE I. shows supervised format of input data. Each step moves observation by 1. Grey color indicates compare step between monthly and weekly observation.

VI. RESULT

A. Features selections

TABLE II. FEATURES SELECTION

Features	Description	Week	Month
Average price per unit	Average of price per unit	X	X
Average discount percent	Reduction of price in percent		
Giveaway	Giveaway volume		
Return volume	Return volume		
Holiday count	Holiday count (month or week)		
Day count in month	Day count		
Financial year	Financial year	X	X
Month	Month of year		
Week no	Week of month	X	

TABLE II. shows selected features use in weekly and monthly observation series. The selected features have at least 0.25 correlation value and chosen as the best 5 features by recursive feature elimination method.

B. Model metric performance

TABLE III. AVERAGE SCORES OF MONTHLY SERIES

Model	RMSE	MAE	MASE	Fill rate
AR	21501.657	18390.864	0.817	0.821
ARX	21935.72	18843.995	0.83	0.816
U-LSTM	21509.296	18027.518	0.73	0.743
M-LSTM	20693.862	17240.801	0.699	0.763
EU-LSTM	20978.78	17511.478	0.718	0.744
EM-LSTM	21146.189	17835.905	0.706	0.76

TABLE III. show average scores of among each model. ARIMA (AR), ARIMAX (ARX), Univariate LSTM (U-LSTM), Multivariate LSTM (M-LSTM), Encoder decoder Univariate LSTM (EU-LSTM) and Encoder-decoder multivariate LSTM (EM-LSTM). Grey highlights mark the best score of each metric. M-LSTM performed best for all model error metric performance and ARIMA shows the best fill rate score.

C. Observation type comparision

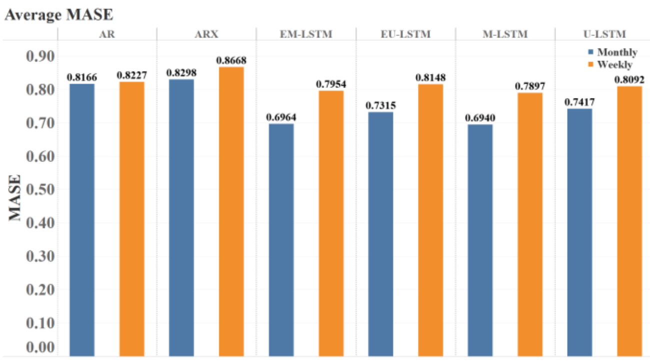

Fig. 1. Average MASE score of montly and weekly observations

Fig. 1 compares MASE score between each model and observation type. The orange bar denotes weekly observation and blue bar denotes monthly observation. All models perform better with monthly observations. LSTMs acquire better benefit from using monthly observations than ARIMAs because of no or less zeroes in data point of the series in which causes the under forecast bias in weekly observations LSTMs.

D. Unsatisfied predicted series

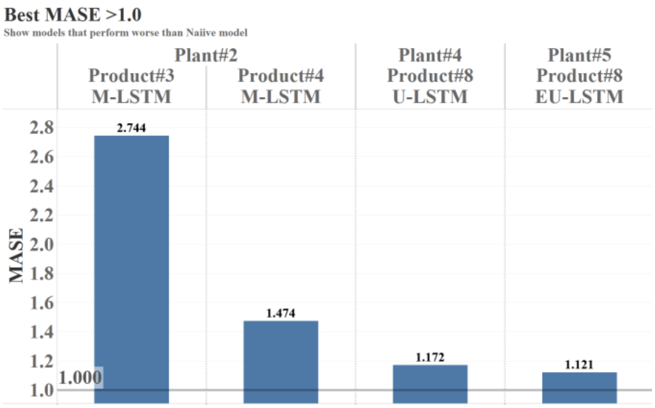

Fig. 2. Series with MASE greater than 1.0 of monthly series

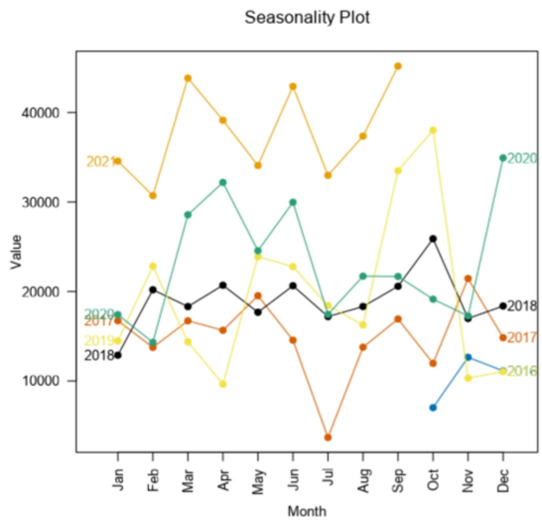

Fig. 3. Seasonality plot between of one of low MASE series

Fig. 2 shows the worst perform series (MASE >1.0) and the least MASE on that series. There are 4 series out of 40 series. From those series, we found an interesting reason and depicts in Fig. 3. There is no overlap between 2021 seasonal line and any other seasonal line at all. It shows that on the concerned series 2021-year value is clearly an outlier compared to other years.

E. Model bias

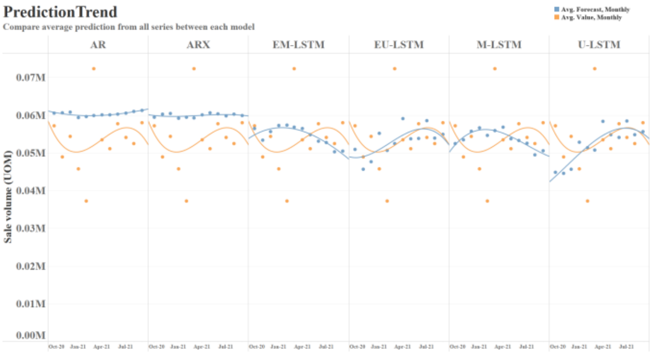

Fig. 4. Actual values trend and predicted values trend of monthly series (Average)

From Fig. 4, orange line and orange dot represent actual trend line and actual values. Blue line and blue dot represent prediction trend and predict values. LSTMs tried to follow the seasonal and trend where ARIMAs predicted in straight line with average from past values. Overall forecast of ARIMAs are above the actual values. It explains why ARIMA models acquired higher fill rate ranks in TABLE III. with lower average MASE scores.

F. Revenue loss by plants

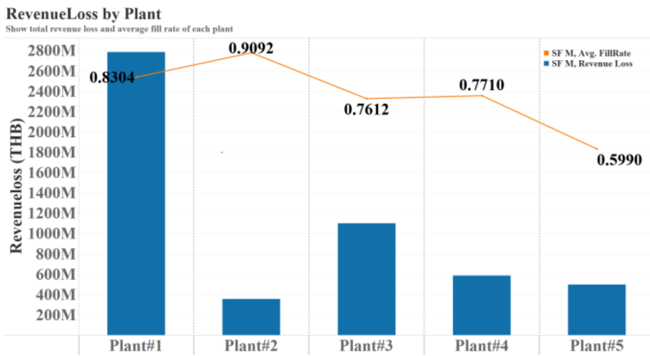

Fig. 5. Total Revenue loss by plant and average fill rate by plant of a monthly series

Fig. 5 shows possibility in improving the demand forecasting to reduce the revenue loss of each plant. Plant#1 has the highest revenue loss (Blue bar) and plant#5 has the least fill rate (orange line). There is a huge room for entrepreneur to investigate the causes of Plant#1 error and find the countermeasures.

VII. CONCLUSION

There is no single best algorithm for all the series. Using the right one for the right situations is important. The ARIMA is a simple model that one can implement with a simple tool such as Excel. It has over forecast behavior, leading to a better fill rate than LSTM because of no penalty on over forecast. It assumes series are stationary, so it works well with unwavering series because of their prediction nature. In contrast, LSTMs predict in a riskier way by following seasonal and trends. This makes LSTM performs better with more complex series, with multivariate models due to a good feature candidate. Encoder-Decoder version of LSTM performs better in univariate LSTM but slightly worse in multivariate LSTM. ARIMA outperforms LSTM on most uptrend series and loses to LSTM on downtrend series or more complex series. Training the model with monthly observations provide a better result because of batch order frequency is greater than week causing the cavities in weekly observations. For future work, applying penalty factor for case of over forecast to the formula of fill rate metric is an advice as it will be more concerned if the product's lifespan is short or there is a limit in storage space. [15] may aids the forecasting in different hierarchy such as product category level. Next, the attention-based mechanism [16] may provide better result intuitively that the model will focus more weight on recent data points. Finally, instead of using a local model for each series, the global model [17] shows equal or better performance compared to the local model and possibly aids outlier in seasonality.

REFERENCES

[1] Hochreiter, S. and J. Schmidhuber, *Long short-term memory.* Neural computation, 1997. **9**(8): p. 1735-1780.

[2] Brownlee, J. Deep Learning with Time Series Forecasting, Machine Learning Mastery. 2018 [cited 2018 November 6th, 2018]; Available from: https://machinelearningmastery.com/machine-learning-with-python/.

[3] C. P. da Veiga, C. R. P. da Veiga, W. Puchalski, L. dos Santos Coelho and U. Tortato

Journal of Retailing and Consumer Services 2016 Vol. 31 Pages 174-181

[4] Fattah, J., et al., Forecasting of demand using ARIMA model. International Journal of Engineering Business Management, 2018. 10: p. 1847979018808673.

[5] Chniti, G., H. Bakir, and H. Zaher. E-commerce time series forecasting using LSTM neural network and support vector regression. in Proceedings of the International Conference on Big Data and Internet of Thing. 2017.

[6] Fan, D., et al., Well production forecasting based on ARIMA-LSTM model considering manual operations. Energy, 2021. 220: p. 119708.

[7] Livieris, I.E., E. Pintelas, and P. Pintelas, A CNN–LSTM model for gold price time-series forecasting. Neural computing and applications, 2020. 32(23): p. 17351-17360.

[8] Aburto, L. and R. Weber, Improved supply chain management based on hybrid demand forecasts. Applied Soft Computing, 2007. 7(1): p. 136-144.

[9] Ren, S., H.-L. Chan, and T. Siqin, Demand forecasting in retail operations for fashionable products: methods, practices, and real case study. Annals of Operations Research, 2020. 291(1): p. 761-777.

[10] Cadavid, J.P.U., et al., Machine learning applied in production planning and control: a state-of-the-art in the era of industry 4.0. Journal of Intelligent Manufacturing, 2020: p. 1-28.

[11] Hyndman, R.J. and G. Athanasopoulos, Forecasting: principles and practice. 2018: OTexts.

[12] Hyndman, Rob J., and Anne B. Koehler. "Another look at measures of forecast accuracy." International journal of forecasting 22.4 (2006): 679-688.

[13] Molnar, Christoph. "Interpretable machine learning. A Guide for Making Black Box Models Explainable", 2019. https://christophm.github.io/interpretable-ml-book/.

[14] Sobel, M.J., Fill rates of single-stage and multistage supply systems. Manufacturing & Service Operations Management, 2004. 6(1): p. 41-52.

[15] Punia, S., S.P. Singh, and J.K. Madaan, A cross-temporal hierarchical framework and deep learning for supply chain forecasting. Computers & Industrial Engineering, 2020. 149: p. 106796.

[16] Qin, Y., et al., A dual-stage attention-based recurrent neural network for time series prediction. arXiv preprint arXiv:1704.02971, 2017.

[17] Montero-Manso, P. and R.J. Hyndman, Principles and algorithms for forecasting groups of time series: Locality and globality. International Journal of Forecasting, 2021.