Forecasting of Sales Based on Long Short Term Memory Algorithm with Hyperparameter

1st Ali Khumaidi
Faculty of Engineering
Universitas Krisnadwipayana
Jakarta, Indonesia
alikhumaidi@unkris.ac.id

2nd Ika Ayu Nirmala
Faculty of Engineering
Universitas Krisnadwipayana
Jakarta, Indonesia
nirmalaika.renev@gmail.com

3rd Herwanto
Faculty of Engineering
Universitas Krisnadwipayana
Jakarta, Indonesia
herwanto@unkris.ac.id

Abstract— Increasingly competitive business competition requires business people to re-design their business strategies, one of which is by applying forecasting methods. Several forecasting techniques have been used in sales prediction research with fairly good accuracy in a short period, the research will focus on optimizing the hyperparameter LSTM algorithm to improve the performance of the model formed over the next 60 days. The method used in this study is the Cross Industry Standard Process for Data Mining (CRISP-DM), in the preprocessing the data cleaning, labeling, summary, and data transformation. The data understanding stage applies the Exploratory Data Analysis method. The development of the LSTM model uses several parameters, namely data partition, number of hidden layers, dropout scenarios to prevent overfitting, number of neurons, epoch describing the number of training iterations, batch size is the amount of training data that must be considered in each process of updating the weights. The experimental results of the best LSTM model after experimenting with different parameters are hyperparameter batch size 30, epoch 150, 3 hidden layers and 3 dropouts, resulting in RMSE training of 0.0855 and RMSE testing of 0.0846.

Keywords—Forecasting, Hyperparameter, LSTM, Next 60 days, Product Sales

I. INTRODUCTION

Business actors are currently faced with the challenge of increasingly competitive business competition, which requires every business actor to re-design his business strategy in order to meet market demand. The industrial revolution 4.0 and the development of machine learning can be used by business actors for forecasting with the best accuracy. Several business actors have applied forecasting methods to estimate sales [1]. However, the predictions used are often inaccurate and therefore less effective. This has an impact on product accumulation when consumer demand and sales frequency is low, causing storage costs to increase. In addition, when product demand increases but there is a long enough stock out, causing the store to lose sales. To minimize losses due to errors in predicting sales, predictions are needed by utilizing historical sales transaction data using a method to obtain optimal accuracy of prediction results [2].

Previous research related to sales forecasting has been carried out, Long Short Term Memory (LSTM) is able to predict fluctuating demand well and outperform exponential smoothing (ETS), K-nearest neighbor (KNN), autoregressive integrated moving average (ARIMA), support vector machine (SVM) and network artificial nerves (ANN) [3][4]. Forecasting agricultural commodity prices with the LSTM model is superior to the Holt-Winter's Seasonal method and the SARIMA model with an evaluation based on the RMSE value [5]. Forecasting dependence on electricity demand using LSTM shows better results than traditional methods (SARIMA, ARMA and ARMAX) using MAPE and RMSE [6]. LSTM has excellent performance in modeling customer behavior in a fairly complex environment [7]. Prediction of sales at 1,150 stores in Germany, using the comparison of the Extreme Gradient Boosting (XGB) and Random Forest (RFR) algorithms with LSTM. The results showed that LSTM has 12-14% better accuracy than XGB and RFR [8]. The studies that have been carried out show that LSTM algorithm has a fairly good accuracy for making predictions.

Based on previous studies in the field of sales forecasting, quite a lot has been done with the aim of determining the best-selling product using a certain algorithm or comparing certain algorithms. This research will focus on predicting the number of products sold in the future by optimizing the hyperparameter LSTM algorithm to improve the performance of the model formed over the next 60 days..

II. MATERIAL AND METHOD

A. Dataset

The data source used in this study is the report on sales of Ciwo Pet Shop products from January 2018 to March 2020. The data consists of sales data in one marketplace (Tokopedia) and offline sales in stores. Sample data can be seen in Figure 1.

	Date	Customer	Invoice	City	Product	Quantity	Price
0	01/01/2018	Tilar Handayani	INV/20180101/XVIII/I/126179649	Jakarta Timur	MAKANAN KUCING WHISKAS SACHET - WHISKAS BASAH ...	35.0	4350.0
1	01/01/2018	Tilar Handayani	INV/20180101/XVIII/I/126180557	Jakarta Timur	MAKANAN KUCING WHISKAS SACHET - WHISKAS BASAH ...	35.0	4350.0
2	01/01/2018	Ika Kurniawati	INV/20180101/XVIII/I/126163178	Kota Bekasi	PET CARGO KUCING ANJING KELINCI MUSANG LANDAK ...	1.0	160000.0
3	01/01/2018	Agung Darmawan	INV/20180101/XVIII/I/126180935	Jakarta Selatan	PASIR KUCING WANGI GUMPAL 25 LITER	1.0	83000.0
4	01/01/2018	Asti	INV/20180101/XVIII/I/126214321	Kota Bekasi	PASIR KUCING GUMPAL WANGI KAWAN 10L 10 L 10 LITER	1.0	48000.0
...	...	...	...	...	...	...	...
140948	31/03/2020	Duta Wiryanto	INV/20200331/XX/III/512250603	Kota Yogyakarta	PASIR KUCING WANGI GUMPAL 25 LITER	3.0	67000.0
140949	31/03/2020	Teguh Triono	INV/20200331/XX/III/512231140	Kota Bekasi	PASIR KUCING WANGI GUMPAL 25 LITER	1.0	67000.0
140950	31/03/2020	Marcellus Adrian	INV/20200331/XX/III/511823299	Kota Bekasi	PASIR KUCING WANGI GUMPAL 25 LITER	1.0	67000.0
140951	31/03/2020	Tjahjadi Widjaja	INV/20200331/XX/III/511370148	Jakarta Barat	REPACK ORI CAT FOOD ORICAT 1KG SEJENIS BOLT MA...	3.0	17000.0
140952	31/03/2020	Ditto Ardiyanto	INV/20200331/XX/III/511892475	Kab. Tangerang	GUNTING KUKU HEWAN ANJING KUCING SUGAR GLIDER ...	1.0	15000.0

140953 rows × 7 columns

Fig. 1. Sample of sales data transaction

B. Reseach Stages

The method used in this research is ross Industry Standard Process for Data Mining (CRISP-DM) which is a standard process for problem solving in research units in the field of data mining [9]. The purpose of the research methodology is for the implementation of the research to obtain results that are in accordance with the objectives to be achieved. The series of stages in this research include: Data understanding and preparation, modeling and evaluation.

1) Data preparation

The data preparation stage or data preprocessing includes all activities to build the final dataset from raw data to set up data mining. This stage can be done repeatedly to get the appropriate data [10]. This stage includes the following processes:

a) Cleaning Data: Adjusting the existing data format as needed, eliminating missing values and unneeded data.

b) Construct Data: Preparation before completing the dataset, at this stage it can be updated or created a new record or attribute. Because the data in this study is very varied, a labeling process is carried out to give ID to each type of product and the name of the buyer at random.

c) Data Transformation: This stage is a way of normalizing the data to equalize the format in the form of a general scale with a range of 0 to 1. In this study, the Min-MaxScaler method will be used.

2) Modelling and evaluation

The modeling process can be seen in Figure 2. Consists of 7 processes as follows:

a) Input Data: The data input stage is the stage of entering the pre-processed data into the model scenario that will be tested.

b) Input Hyperparameter: This stage is the process of setting the parameter scenario in order to produce the right parameters so that the LSTM algorithm can study the patterns contained in the data properly [11]. The stages in performing the hyperparameter input can be seen in Table 1.

TABLE I. PARAMETER TEST STAGES

Stage	Parameter
1	Data Composition
2	Hidden Layer
3	Dropout Scenario
4	Batch Size
5	Epoch

At the initial stage, it is necessary to test the composition of training and testing that produces the smallest loss. After obtaining the appropriate composition, the results of the experiment are used to find the best hidden layer. If the number of hidden layers has resulted in the smallest loss, the next step is to create a dropout scenario. Dropout serves to prevent data from being overfitted. If the number of dropouts has resulted in a small error, a data partition with the appropriate layer will be obtained. Next is the determination of the best batch size and epoch based on the data composition and layers that have been determined.

c) Training: At this stage, the application of machine learning-based prediction trials, namely the LSTM algorithm, can be implemented on the training data. The training process aims to provide clues to the data through algorithms so that they can learn from the given data patterns and look for correlations.

d) Model, Training Loss and Validation Loss: At this stage the results of the hyperparameter tuning configuration will be displayed with the results in the form of training loss and validation loss values. The results of the training and validation loss are a representation of how well the resulting model is with the test parameter in the form of the Root Mean Square Error (RMSE).

e) Model evaluation: After going through a series of processes, it is necessary to evaluate the performance of the model using the calculation of RMSE is an alternative method used to measure the accuracy of the forecasting technique used [12]. RMSE is the mean absolute difference between the forecasted and actual values presented in the form of a percentage.

f) 60 Days Future Forecasting: The prediction stage is a process to predict a value for the future based on a certain pattern in the data. In this study, predictions are made to estimate sales in the next year. The results of the prediction process will then be recommended to be used as a decision-making tool.

g) Denormalization: After getting the prediction results from the prediction process, then the data is denormalized, that is, the data is converted into real values again [13]. Because the data is still in the form of a range of intervals that have previously been normalized data. The purpose of denormalization is to make it easier to read the resulting output value. The formula for denormalizing: d = d'(max-min) + min.

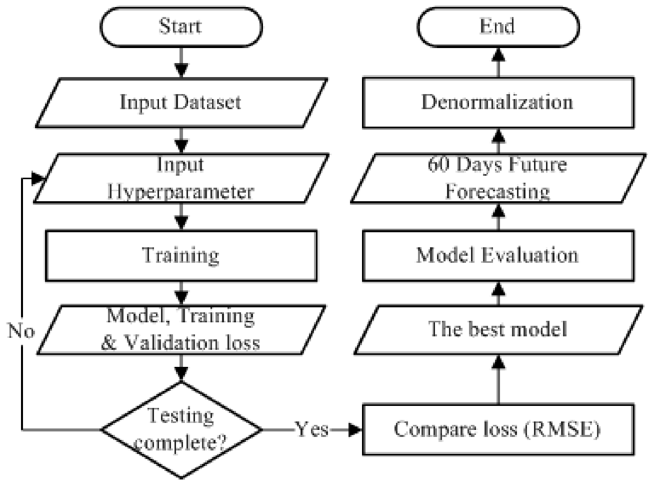

Fig. 2. Flowchart of modelling

III. RESULT AND DISCUSSION

A. Data Preprocessing

Data cleaning is a method to delete unnecessary data and fill in missing data. In this study, the data used were first identified as missing columns in the table. After that, the correlation between the missing tables was checked, which can be seen in Figure 3.

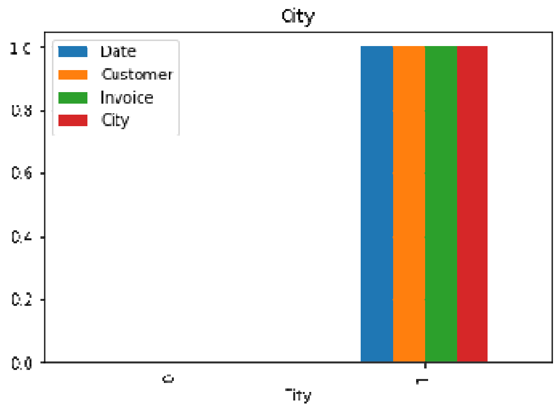

Fig. 3. Missing Value Correlation Chart

Based on the results of the bar plot visualization, it shows that the missing values between columns are related to each other. After reviewing the available datasets, it is necessary to forward fill in the NaN (Not a Number) line and delete transactions without invoice numbers.

Labeling is a way of randomly assigning IDs to product and customer categories to make it easier to define in numeric form. Summary of the number of sales in a daily format to simplify the modeling process so that the algorithms used can learn patterns better. Sumarry results can be seen in Figure 4.

	Date	Price	Quantity	Total
0	2018-01-01	1190200.0	92.0	1781500.0
1	2018-01-02	1830100.0	126.0	2949200.0
2	2018-01-03	4405100.0	165.0	5513300.0
3	2018-01-04	5931450.0	327.0	8161250.0
4	2018-01-05	3536400.0	273.0	5148500.0
...	...	...	...	...
814	2020-03-27	14485200.0	760.0	23278600.0
815	2020-03-28	16073000.0	667.0	25065300.0
816	2020-03-29	15537100.0	577.0	23371700.0
817	2020-03-30	11885683.0	746.0	21249710.0
818	2020-03-31	12882950.0	627.0	22781750.0

819 rows × 4 columns

Fig. 4. Sumarry Results Total Sales Per Day

Data transformation is a way of normalizing data that serves to equalize the format in the form of a general scale (0-1). In this study, the Min-MaxScaler method was used for data modeling [14]. The sample of normalization results can be seen in Table 2.

TABLE II. SAMPLE OF NORMALIZATION RESULTS

Date	Quantity	X'
01/01/2018	92	0,00613079
02/01/2018	126	0,01771117
03/01/2018	165	0,03099455
04/01/2018	327	0,08617166
05/01/2018	273	0,06777929
06/01/2018	488	0,14100817
07/01/2018	372	0,10149863
08/01/2018	575	0,17064032
09/01/2018	426	0,11989109
10/01/2018	327	0,08617166

B. Data understanding

The data understanding stage can provide an analytical foundation for a study by making a summary and identifying potential problems contained in the data before modeling. The data understanding stage applies the Exploratory Data Analysis (EDA) method, namely data exploration techniques using simple arithmetic techniques and graphic techniques in summarizing observational data to facilitate data understanding [15]. The results of the EDA can be seen in Figure 5. The results of the distribution of the data show that there is a difference in the amount between the median (50%) and the mean with a significant difference so that it can be concluded that the number of transactions between buyers is very diverse. Based on the results of exploration, the types of products are very diverse with a total of 1,523 product variations. The results of city distribution exploration based on purchase frequency show that East Jakarta has the highest purchase frequency compared to 266 cities or other regions. Figure 6 explains the results of data exploration showing that sales tend to fluctuate with the assumption that the highest sales are in the cities of Bekasi and East Jakarta and experienced a significant increase in October 2019.

	Quantity	Price	ProductID	CustomerID	Total
count	140953.000000	1.409530e+05	140953.000000	140953.000000	1.409530e+05
mean	4.491980	8.250755e+04	735.295198	13093.044575	1.260052e+05
std	15.581414	1.505750e+05	386.411018	7450.191321	2.350766e+05
min	1.000000	1.000000e+02	0.000000	0.000000	1.000000e+02
25%	1.000000	1.400000e+04	483.000000	6675.000000	2.500000e+04
50%	1.000000	2.500000e+04	731.000000	12959.000000	6.750000e+04
75%	3.000000	7.800000e+04	927.000000	19464.000000	1.360000e+05
max	3000.000000	8.550000e+06	1522.000000	26554.000000	1.800000e+07

Fig. 5. Sumarry Results Total Sales Per Day

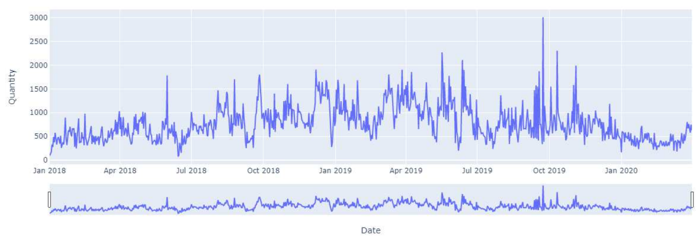

Fig. 6. Sumarry Results Total Sales Per Day

C. Modelling

In develop LSTM model, several parameters are used including data partitioning, number of hidden layers, dropout scenarios to prevent overfitting, number of neurons, epoch describing the number of training iterations, batch size is the amount of training data that must be considered for each process of updating the weights. To get the best results from the LSTM model, the training process will use different parameters to get the best results. From a number of models produced will be compared and analyzed from each loss and RMSE values generated each time the experiment. The provisions of the parameter values used for making the LSTM model can be seen in Table 3. Tests will be carried out on each parameter in Table 1, where the results of parameter testing that have produced a fairly good loss will be used for the next test so that it is expected to produce the best model.

TABLE III. PARAMETER OF LSTM TESTING

Parameter	Quantity	Description
Input	3	Daily Total Sales Data
Hidden Layer	Trial Error	30 - 120 neuron
Dropout	Trial Error	1 - 3 layer
Epoch	Trial Error	50 - 200
Batch Size	Trial Error	10 - 400
Optimizer	1	Adam
Output	1	Number of product sales per day

In the first experiment, experiments were carried out to obtain the optimal composition of the data. The hyperparameters used to find the amount of data composition are 30 Batch Size and 100 Epoch and use the Adam optimizer. Based on the experimental results, the results are shown in Table 4. Based on the experimental results, the losses obtained in tests 1 and 2 are categorized as overfitting because the training loss is smaller than the validation loss. While in experiments 3 to 5 it is categorized as underfitting because the training loss is greater than the validation loss. To get the optimal model from the results of this first experiment, it is taken from the smallest difference between the RMSE Training and RMSE Testing values. Based on the first test, the optimal amount of data composition from the dataset being trained lies in the third experiment with a value of 0.0913 for RMSE Training and 0.0858 for RMSE Testing. As for the amount of data composition that is less than optimal, it is in the fourth experiment with a value of 0.9638 for RMSE Training and 0.0714 for RMSE Testing.

TABLE IV. RESULTS OF DATA PARTITION TESTING

Training	Testing	Loss	Val Loss	RMSE	
		(MSE)	(MSE)	Training	Testing
50	50	0,0078	0,0120	0,0802	0,1094
60	40	0,0077	0,0115	0,0844	0,1070
65	35	0,0092	0,0074	0,0913	0,0858
70	30	0,0093	0,0051	0,9638	0,0714
80	20	0,0106	0,0058	0,1028	0,0760

After getting the results of the first test, stage 2 testing was carried out to get the optimal hidden layer. Phase 2 testing is carried out with Batch size 30, Epoch 100 and data composition of 65:35 which is the result of the first test. The test results to find the hidden layer can be seen in Table 5. Based on the experimental results, the loss value obtained in tests 2 to 5 tends to be overfitting because this value is smaller than the validation loss value. To get the optimal model from the experimental results, it can be seen from the loss and validation loss values which have a small difference, supported by the RMSE Train and RMSE Testing values which also have the smallest difference. Based on the test results, in this study the addition of the number of hidden layers does not really have a significant effect on reducing a test parameter value. This is evidenced by the increase in the loss value generated in the 2nd to 5th experiments in the table above. From the results of this test, the optimal number of hidden layers from the dataset that has been trained is on the hidden 3 neurons with an RMSE Training value of 0.0910 and an RMSE Testing of 0.0889. As for the number of hidden layers that are less than optimal, it is in the 5th experiment with the number of hidden layers 7. After doing the second stage of testing, the third stage of testing is carried out to get the best dropout scenario. Dropout serves to prevent overfitting that is too large. The third stage of testing was carried out using batch size 50, epoch 100, data composition 60:35 based on the results of the first test and the number of hidden layers 3 based on the results of the second test.

TABLE V. RESULTS OF HIDDEN LAYER TESTING

Hidden layer	Loss	Val Loss	RMSE	
	(MSE)	(MSE)	Training	Testing
3	0,0078	0,0120	0,0802	0,1094
4	0,0077	0,0115	0,0844	0,1070
5	0,0092	0,0074	0,0913	0,0858
6	0,0093	0,0051	0,9638	0,0714
7	0,0106	0,0058	0,1028	0,0760

The results of the third experiment can be seen in Table 6. Based on the test results, it shows that as the number of dropouts increases, the resulting value tends to rise and fall, so it can be concluded that the addition of the dropout value does not have a significant effect if it exceeds the predetermined number of hidden layers. In the second experiment the loss training value tends to increase, but the validation value tends to decrease. Meanwhile, in the third test, the training loss value tends to decrease and the validation loss value tends to increase. At this stage to determine the optimal dropout value by comparing the value of the RMSE Training and RMSE Testing results which have the smallest difference in each test. Based on the results at this stage, the optimal dropout value lies in the third test with an RMSE Training value of 0.0906 and an RMSE Testing of 0.0895. After doing the third stage of testing, then the fourth stage of testing is carried out to get the best batch size value. The fourth stage of testing was carried out using batch size 50, epoch 100, data composition 60:35 based on the results of the first test and the number of hidden layers 3 based on the results of the second test and 3 the number of dropouts based on the third test.

TABLE VI. RESULTS OF DROPOUT SCENARIO TESTING

Dropout	Loss	Val Loss	RMSE	
	(MSE)	(MSE)	Training	Testing
1	0,0097	0,0073	0,0857	0,0941
2	0,0102	0,0067	0,1021	0,0821
3	0,0091	0,0080	0,0906	0,0895

The results of the fourth experiment can be seen in Table 7. Based on the test results, the addition of the batch size value tends to produce loss values that tend to go up and down. So it can be concluded that there are no special provisions for the best batch size value. This is because each data has its own treatment in determining the right batch size. At this stage to determine the optimal batch size value by comparing the value of the RMSE Training and RMSE Testing results which have the smallest difference in each test. Based on the results at this stage, the optimal batch size value lies in the fifth test with an RMSE Training value of 0.0914 and an RMSE Testing of 0.0879. After doing the third stage of testing, then the fifth stage of testing is carried out to get the best epoch value. The fifth stage of testing is carried out using the composition of data 60:35 based on the results of the first test and the number of hidden layers 3 based on the results of the second test and 3 the number of dropouts based on the third test and a batch size of 30 based on the results of the fourth stage of testing.

TABLE VII. RESULTS OF HIDDEN LAYER TESTING

Batch Size	Loss	Val Loss	RMSE	
	(MSE)	(MSE)	Training	Testing
10	0,0088	0,0073	0,9031	0,0856
15	0,0086	0,0069	0,0898	0,0829
20	0,0079	0,0064	0,0899	0,0801
25	0,0092	0,0070	0,0911	0,0838
30	0,0097	0,0077	0,0914	0,0879
35	0,0087	0,0068	0,0905	0,0824
40	0,0094	0,0069	0,0926	0,0832

The results of the fifth experiment can be seen in table 8 above. Based on the test results show that the increasing number of epochs, the resulting value tends to fluctuate, so it can be concluded that the determination of the best number of epochs to reduce the amount of loss is based on trials. This is because each data has a different pattern and complexity. At this stage to determine the optimal epoch value by comparing the value of the RMSE Training and RMSE Testing results which have the smallest difference in each test. Based on the results at this stage, the optimal batch size value lies in the third test with an RMSE Training value of 0.0855 and an RMSE Testing of 0.0844.

TABLE VIII. RESULTS OF EPOCH TESTING

Epoch	Loss	Val Loss	RMSE	
	(MSE)	(MSE)	Training	Testing
50	0,0111	0,0077	0,0977	0,0878
80	0,0096	0,0067	0,0997	0,0816
100	0,0087	0,0074	0,0881	0,0860
120	0,0079	0,0089	0,0856	0,0943
150	0,0076	0,0071	0,0855	0,0844
180	0,0087	0,0090	0,0898	0,0949
200	0,0076	0,0095	0,0837	0,0974

Prediction by determining the hyperparameters for the best model that has been obtained in the testing process, namely by using the LSTM algorithm with the modeling architecture in Figure 7 and the results of sales predictions with the LSTM algorithm can be seen in Figure 8. The results are in the form of comparisons between actual data and data from training and testing predictions.

```
Model: "sequential"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
lstm (LSTM)                  (None, 100, 100)          40800
_________________________________________________________________
dropout (Dropout)            (None, 100, 100)          0
_________________________________________________________________
lstm_1 (LSTM)                (None, 100, 100)          80400
_________________________________________________________________
dropout_1 (Dropout)          (None, 100, 100)          0
_________________________________________________________________
lstm_2 (LSTM)                (None, 100)               80400
_________________________________________________________________
dropout_2 (Dropout)          (None, 100)               0
_________________________________________________________________
dense (Dense)                (None, 1)                 101
=================================================================
Total params: 201,701
Trainable params: 201,701
Non-trainable params: 0
```

Fig. 7. LSTM Layer Architecture

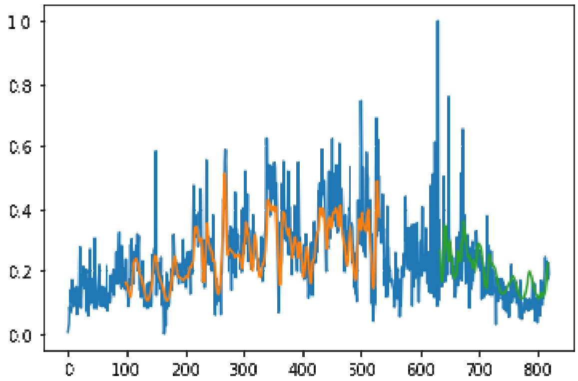

Fig. 8. Chart of Prediction Result

D. Evaluation

After getting the results of the sales prediction, the model accuracy will be calculated using RMSE. The results of the training sample and prediction testing are presented in Tables 9 and 10.

TABLE IX. SAMPLE OF TRAINING DATA RESULTS

y_train	train_ predict	Error	Absolute Value of Error	Square of Error
		Y - Y'	|Y-Y'|	(Y-Y')^2
0,2411	0,1632	0,0779	0,0779	0,00607
0,1685	0,1588	0,0097	0,0097	9,592E-05
0,1433	0,1541	-0,0107	0,0107	0,0001
0,1158	0,1475	-0,0317	0,0317	0,0010
0,0875	0,1393	-0,0517	0,0517	0,0026

TABLE X. SAMPLE OF TRAINING DATA RESULTS

y_test	test_ predict	Error	Absolute Value of Error	Square of Error
		Y - Y'	|Y-Y'|	(Y-Y')^2
0,2731	0,2704	0,0027	0,00275	7,611E-06
0,1927	0,2394	-0,0466	0,0466	0,0021
0,1651	0,2068	-0,0416	0,0416	0,0017
0,1863	0,1796	0,0066	0,0066	4,487E-05
0,1495	0,1654	-0,0158	0,0158	0,0002

By using the hyperparameter composition that has been matched, the researcher uses a 100-day lookback from the testing data to predict the possibility of selling in the next 60 days. The results are presented in graphical form in Figure 9 and 10.

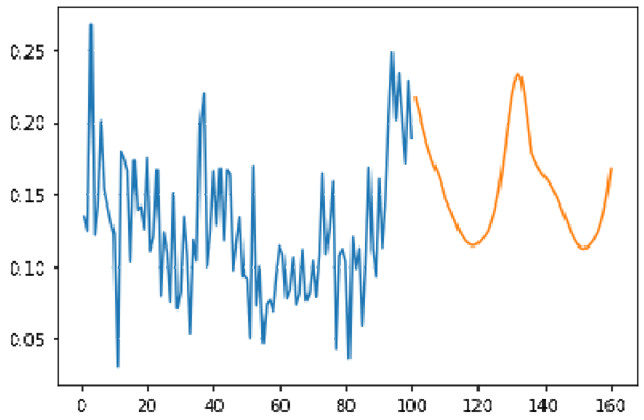

Fig. 9. Chart of Sales Prediction Results for the Next 60 Days

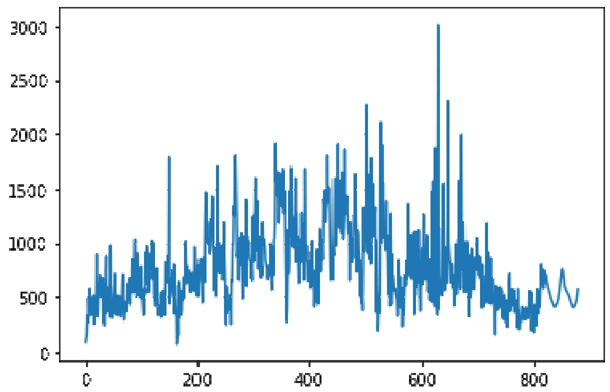

Fig. 10. Chart of Actual Merge and Predicted Results

The diagram above shows a graph starting from day 781 (in blue) until the prediction results are shown by the orange graph line. Prediction results are generated in the form of data with an interval range of 0 to 1. In order to make it easier to read the predicted data, a denormalization process is carried out. The following is a sample of the results of sales predictions for the next 60 days shown in Table 11.

TABLE XI. SAMPLE OF NORMALIZATION RESULTS

Date	Before Denormalization	After Denormalization
01/04/2020	0,216552	709,7967
02/04/2020	0,207718	683,8608
03/04/2020	0,196872	652,0167
04/04/2020	0,187789	625,3489
05/04/2020	0,180030	602,5673
06/04/2020	0,172879	581,5717
07/04/2020	0,168107	567,5609
08/04/2020	0,162540	551,2181
09/04/2020	0,155221	529,7274
10/04/2020	0,146479	504,0633

IV. CONCLUSION

Based on the research that has been done, it can be concluded that the best model obtained from LSTM with hyperparameter batch size 30, epoch 150, 3 hidden layers and 3 dropouts produces RMSE training of 0.0855 and RMSE testing 0.0846. During the training process, it is categorized as overfitting, namely the validation loss value is greater than the training loss. Overfitting occurs because the trained training data is easier to learn than the testing data. In addition, the size of the loss value is strongly influenced by the configuration of hyperparameter tuning such as data partition, hidden layer, batch size and epoch.

REFERENCES

[1] R. M. van Steenbergen and M. R. K. Mes, "Forecasting demand profiles of new products," Decis. Support Syst., vol. 139, p. 113401, Dec. 2020, doi: 10.1016/j.dss.2020.113401.

[2] L. R. Berry, P. Helman, and M. West, "Probabilistic forecasting of heterogeneous consumer transaction–sales time series," Int. J. Forecast., vol. 36, no. 2, pp. 552–569, Apr. 2020, doi: 10.1016/j.ijforecast.2019.07.007.

[3] H. Abbasimehr, M. Shabani, and M. Yousefi, "An optimized model using LSTM network for demand forecasting," Comput. Ind. Eng., vol. 143, p. 106435, May 2020, doi: 10.1016/j.cie.2020.106435.

[4] L. Liang and X. Cai, "Forecasting peer-to-peer platform default rate with LSTM neural network," Electron. Commer. Res. Appl., vol. 43, p. 100997, Sep. 2020, doi: 10.1016/j.elerap.2020.100997.

[5] K. M. Sabu and T. K. M. Kumar, "Predictive analytics in Agriculture: Forecasting prices of Arecanuts in Kerala," Procedia Comput. Sci., vol. 171, pp. 699–708, 2020, doi: 10.1016/j.procs.2020.04.076.

[6] S. Muzaffar and A. Afshari, "Short-Term Load Forecasts Using LSTM Networks," Energy Procedia, vol. 158, pp. 2922–2927, Feb. 2019, doi: 10.1016/j.egypro.2019.01.952.

[7] M. Sarkar and A. De Bruyn, "LSTM Response Models for Direct Marketing Analytics: Replacing Feature Engineering with Deep Learning," J. Interact. Mark., vol. 53, pp. 80–95, Feb. 2021, doi: 10.1016/j.intmar.2020.07.002.

[8] S. Helmini, N. Jihan, M. Jayasinghe, and S. Perera, "Sales forecasting using multivariate long short term memorynetwork models," PeerJ Prepr., vol. 7, pp. 1–16, 2019, doi: https://doi.org/10.7287/peerj.preprints.27712v1.

[9] A. Khumaidi, "Data Mining For Predicting The Amount Of Coffee Production Using CRISP-DM Method," J. Techno Nusa Mandiri, vol. 17, no. 1, pp. 1–8, Feb. 2020, doi: 10.33480/techno.v17i1.1240.

[10] R. Nisbet, G. Miner, and K. Yale, "A Data Preparation Cookbook," in Handbook of Statistical Analysis and Data Mining Applications, Elsevier, 2018, pp. 727–740.

[11] W. Li, B. Wang, J. Liu, G. Zhang, and J. Wang, "IGBT aging monitoring and remaining lifetime prediction based on long short-term memory (LSTM) networks," Microelectron. Reliab., vol. 114, p. 113902, Nov. 2020, doi: 10.1016/j.microrel.2020.113902.

[12] C. Panem, V. R. Gad, and R. S. Gad, "Sensor's data transmission with BPSK using LDPC (Min-Sum) error corrections over MIMO channel: Analysis over RMSE and BER," Mater. Today Proc., vol. 27, pp. 571–575, 2020, doi: 10.1016/j.matpr.2019.12.039.

[13] S. Lightstone, T. Teorey, and T. Nadeau, "Denormalization," in Physical Database Design, Elsevier, 2007, pp. 337–355.

[14] D. Singh and B. Singh, "Investigating the impact of data normalization on classification performance," Appl. Soft Comput., vol. 97, p. 105524, Dec. 2020, doi: 10.1016/j.asoc.2019.105524.

[15] R. Indrakumari, T. Poongodi, and S. R. Jena, "Heart Disease Prediction using Exploratory Data Analysis," Procedia Comput. Sci., vol. 173, pp. 130–139, 2020, doi: 10.1016/j.procs.2020.06.017.