APPLICATION OF LONG SHORT TERM MEMORY RNN FOR MINIMARKET SALES TRANSACTION PREDICTION

Patrick Josealdo[1], Fatma Indriani[2], Muhammad Itqan Mazdadi[3], Irwan Budiman[4], Andi Farmadi[5]
[1,2,3,4,5]FMIPA ULM Computer Science Study Program
Jl. A. Yani Km 36 Banjarbaru, South Kalimantan
Email: Patrick.ilkom@gmail.com

Abstract

This study aims to determine whether it can build a prediction of sales of goods at the Lapan-Lapan Mart by using the Long Short Term Memory Recurrent Neural Network method that can be used to predict the sale of goods. In this study, the data was taken from the Lapan-Lapan Mart, together with data on 10 different items sold every day. The data is then compiled for the level of sales to be weekly and a total of 52 data is obtained for each item so that the total data is amounted to 520. To get the weight in the LSTM calculation, there are two processes, namely forward and backward . the weight will be used to make predictions using the basic formula of the LSTM.Based on the research that has been done, it is known that the highest accuracy of using MAD (Mean Absolute Deviation) is 91 gr (11.61803507) indomie goods and 1.8kg of lemon daia (2.077000464) for the lowest MAD

Keywords: Long Short Term Memory Recurrent Neural Network, Mean Absolute Deviation, Prediction of sales of goods on the Lapan-Lapan Mart.

1. PRELIMINARY

Buying and selling transactions is an activity that aims to find, influence and give instructions to buyers to be able to adjust their needs to the products offered and enter into agreements regarding prices that are beneficial to both parties. Minimarket is a place for selling goods that are often used by people to buy their daily needs. At the Lapan-Lapan Mart each

The day there are many sales transactions, so the data is stored on *the database* very large. This transaction data can be processed into profitable information for the minimarket itself. A lot of data can also be used as useful information for minimarket owners in policy making.

Data Mining is a process that uses statistical techniques, mathematics, artificial intelligence, machine learning to extract and identify useful information and related knowledge from various large databases. Data mining has an important function to help get useful information and increase knowledge for users. One of the basic functions used is prediction. Prediction is a process of finding patterns from data by using several variables to predict other variables of unknown type or value. Basically, predictions have similarities with classification, but the data are classified based on behavior or values predicted in the future.Yusuf [8].

Time Series Model is a prediction of future values based on the past values of a variable and past mistakes. The time series model is usually used more often for forecasting / prediction. Time series are basically used to do data analysis that considers the influence of time. Data collected periodically based on the order of time, can be in hours, days, weeks, months, quarters and years. In addition, time series analysis can be used for forecasting data over several periods in the future, which is very helpful in planning future planning. Time series data can be found in various fields, for example in the field of economy such as sales data every day, company profits in each year and total export value in each month

According to Sabrina [7] in his research said that Artifiical Neural Network has been very well known for study time series patterns so that they can predict the situation in the next time. *Artificial Neural Network* has several models and the most frequently used are Feed forward Neural Network (FNN) and Recurrent Neural Network (RNN). RNN is often referred to as a modification of FNN and has a more optimal performance in learning time series data. According to Ching-wu [3] in his study said the use *Artificial Neural Network* to be one of the good methods for predicting retail goods sales

According to Hamdani [6] in his research it was concluded that the application of data *Mining* by method *time series* provide future sales forecasting information to the owner *mini Market* An idol that results from excavation *the database* the sale. The level of accuracy is determined by the smallest MAD value. *Mean Absolute Deviation* (MAD) is used to measure the accuracy of a value or forecasting error. The smaller the MAD value the smaller the error value.

Based on the description above, the authors are interested in conducting research by applying *Long Short Term Memory Recurrent Neural Network* to predict the sales transaction at the minimarket in order to get the right accuracy results on the sales transaction

2. RESEARCH METHODOLOGY

The following is a description of this research flow:

a. Data obtained from direct collection in Lapan-Mart, as many as 10 different types of goods data are recapitulated into 52 data per type of goods sold. The data to be used is the Detailed Sales Report of the Lapan-Lapan Mart.

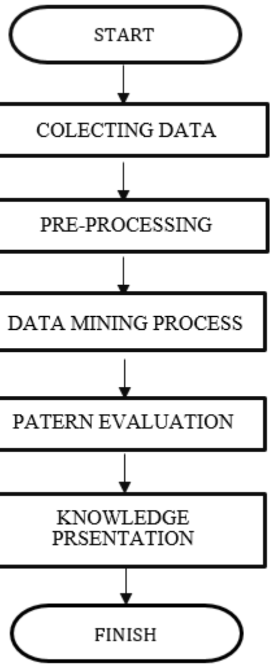
b. Data Selection This variable is done to select the data that matches the variables needed in the study. Data requirements are focused on variables that are relevant to the goal of knowing predictive results.

c. Data Transformation This variable is carried out to process / transform data into a form that is ready to be processed in data mining. Data will be categorized first and then ready to be calculated for the data mining process

d. Distribution of Training Data and Data Testing
dividing data into two groups namely training data and testing data, the ratio of sharing between training data and testing data is 80:20 (40 training data: 12 testing data)..

e. The training data is calculated using LSTM RNN to get results in the form of weights that will be used in the testing data.

f. Validity test is done by testing the level of Accuracy using the MAD method (Mean Absolute Deviation) and MAPE (*Mean Absolute Percent Error)*



Fgure 1. Research Procedure

3. RESULTS AND DISCUSSION

3.1 Results

3.1.1 Data Collection and Literature Search

The data used are sales data of Lapan-Lapan Mart in the form of data on 10 different types of goods, which are then recaptulated into data weekly to obtain 52 data.

3.1.2 Preprocessing

Following are the Preprocessing steps:

a. *Data Cleaning*

Stages of selecting relevant variables needed for research purposes. At this stage the reduction of variables is aimed at reducing unnecessary information and only the variables selected will then be processed in the data mining process.

Table 1. Variables in the data along with an explanation.

No	Variable Name	Information
1	Date	The date the item was sold or refilled
2	Information	Transaction Struck Number
3	Debit	Number of replenishment of goods
4	Credit	Number of items sold
5	Balance	There is nothing
6	Score	Number of items left
7	Price	Prices on goods

b. *Data Transformation and Normalization*

Transformation is done to group data. The stages of the data transformation process are carried out so that the selected data becomes a form that is ready for processing in data mining. Data that are not categorized will first be categorized and then ready to be calculated for the data mining process.

Table 2. Processes before and after transformation

No	Variable	Before Transformation	After Transformation
1	Date	Date per day	Date of Week

After the distribution is done, the next step is to normalize the data of 10 items. The purpose of normalization is for data to be used in the calculation process. The results of normalization in the coca-cola 1.5l data can be seen as follows.

Table 3. Normalization Data

Coca-cola 1,5l		
No	**Input (xt)**	**Input (xt) normalization**
1	**7**	**0.024221**
.	**.**	**.**
51	**19**	**0.065744**

3.1.3 Data Mining Process

The method used in this research is the Long Short Term Memory Recurrent Neural Network method to predict the sale of goods at the Lapan-lapan Mart and the data of goods used is coca-cola 1.5l. The stages carried out in this data mining process are:

a. RNN LSTM Training

The first process is done by calculating using calculations *LSTM* which aims to get the initial weight in the forward process. The data used in the calculation is training data, amounting to 40.

1) Forward process

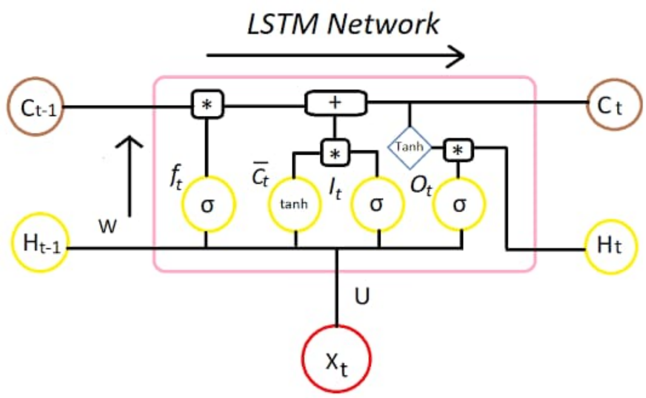


Figure 1. Network of LSTM

Xt = Input
Ht-1 = previous cell output
Ct-1 = previous cell memory
Ht = output
Ct = cell memory results
**, +* = Additions and changes to elements

After seeing from the LSTM plot drawn, it was found that the formula used was as follows.

W, U = is a bobo that is used on each gate
Ft = σ *(xt * Uf + Ht-1 * wf)*
Ĉt = *tanh (xt * Uc + Ht-1 * wc)*
It = σ *(xt * Ui + Ht-1 * wi)*
Ot = σ *(xt * Uo + Ht-1 * wo)*
Then there is the formula used to get the output of the cell
Ct = *(Ft * Ct-1 + It * Ĉt)*
Ht = *Ot * tanh (Ct)*
Then determine the weights of U and W at each gate. In determining the initial weight in the forward process, the weight used is random

Table 4. U and W weights at each gate

Wa	Wi	Wf	Wo	Ua	Ui	Uf	Uo
0.93	0.5	0.04	0.6	0.67	0.64	0.85	0.95

The next process is entered into the calculation to calculate using a predetermined formula using weights and normalized data
Ft = σ *(xt * Uf + Ht-1 * wf)*
The sigmoid function that is used is

$$\sigma = \frac{x(\,xt*Uf+Ht-1*wf)}{1+x(xt*Uf+Ht-1*wf)}$$

The process of this calculation will be used on all data, so the following results are obtained.

Table 5. Results from the calculation of Ft

No	xt (normalization)	*Ft*
1	0.024221	0.500242
	.	.
40	0.051903	0.506262

By using the formula Ĉt we get the following results.
*Ĉt = tanh (xt * Ui + Ht-1 * wc)*

Table 6. Results from the calculation of Ĉt

No	xt (normalization)	*Ĉt*
1	0.024221	0.022522
	.	.
40	0.051903	0.066281

By using the It formula and using the sigmoid function, the following results are obtained.

It = σ *(xt * Ui + Ht-1 * wi)*

Table 7. Results from the calculation of It

No	xt (normalization)	*It*
1	0.024221	0.503028
	.	.
40	0.051903	0.510811

By using the Ot formula and using the sigmoid function, the following results are obtained.

*Ot = σ (xt * Uo + Ht-1 * wo)*

Table 8. Results from Ot calculations

No	x (normalization)	*Ot*
1	0.024221	0.503633
	.	.
40	0.051903	0.514201

After doing calculations on each gate, the next process is to calculate the results of Cell output and Cell memory using the formula Ht and Ct, the following results are obtained.

*Ct = (Ft * Ct-1 + It * Ĉt)*

Table 9. Results from the calculation of Ct

No	x (normalization)	*Ft*	*Ĉt*	*It*	*CT (state)*
1	0.024221	0.500242	0.022522	0.503028	0.011329
	.	.	.	.	.
40	0.051903	0.506262	0.066281	0.510811	0.060634

*Ht = Ot * tanh (Ct)*

Table 10. Results from the calculation of Ht

No	X	x (normalization)	*Ot*	*Ht*
1	7	0.024221	0.503633	0.005706
	.	.	.	.
40	15	0.051903	0.514201	0.03114

2) Backward Process

The next process is calculating the weight to be used in the backward process. The first thing to do is to determine the Label on the data. Labels are real data that will be predicted.

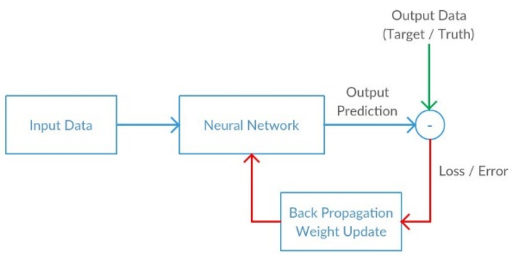


Figure 2. Backward flow

Table 11. Results from determining labels

No	x (normalization)	Label
1	0.024221	0.010381
	.	.
40	0.051903	0

In the backward process there is a formula E, then with that formula the calculation is done using existing data and the following results are obtained.

$E = (x, \dot{x}) = \frac{(X-\dot{x})^2}{2}$

Table 12. Results from determining E

No	x (normalization)	label	*E*
40	0.051903	0	0.001347
.	.	.	.
1	0.024221	0.010381	0.001347

Then next calculate the delta (Δ) and delta out using the formula the following results are obtained.

$\Delta = x - \dot{x}$

Table 13. Results of determining Delta Δ

No	x (normalization)	label	Delta Δ
40	0.051903	0	0.051903
.	.	.	.
1	0.024221	0.010381	0.013841

Furthermore, the process of calculating the sig state and sig out, using the formula obtained results.

Sig out =Δ + Δ out

Table 14. Results from determining Sig out

No	Delta Δ	Delta Δ out	Sig out
40	0.051903	0	0.051903
.	.	.	.
1	0.013841	0.001979	0.01582

*Sig state = Δout * Ot (1-tanh (Ct) Ct1 * ft1)*^{2}

Table 15. Results of determining Sig state

No	DeltaΔ out	*Ot*	*CT (state)*	*CT1*	*Ft1*	*Sig state*
40	0	0.514201	0	0	0	0.026591
.	.	.	.	.	.	.
1	0.001979	0.503633	0.011329	0.011329	0	0.014207

After getting the results from the sig state and sig out, the next process is to calculate each sig of each gate with the data that has been obtained from the forward process with the function obtained by the following results.

*Sig Ĉt = Sig state * It *(1-)*Ĉ*t*^{2}

Table 19. Results for determining Sig Ĉt

No	*Sig state*	*It*	*Ĉt*	*Sig Ĉt*
40	0.026591	0.510811	0.066281	0.013523
.	.	.	.	.
1	0.014207	0.503028	0.022522	0.007143

*Sig Ft = Sig state * Ct-1 * Ft * (1-Ft)*

Table 16. Results from determining Sig Ft

No	*Sig state*	*Ct-1*	*Ft*	*Sig Ft*
40	0.026591	0.052892	0.506262	0,000352
.	.	.	.	.
1	0.014207	0	0.500242	0

*Sig It = Sig state * Ĉt * It * (1-It)*

Table 17. Results from determining Sig It

No	*Sig state*	*Ĉt*	*It*	*Sig It*

40	0.026591	0.066281	0.510811	0,00044
.	.	.	.	.
1	0.014207	0.022522	0.503028	8E-05

*Sig Ot = Ht * tanh (Ct) * Ot * (1-Ot)*

Table 18. Results from determining Sig Ot

No	*Ht*	*Ct*	*Ot*	*Sig Ot*
40	0.03114	0.060634	0.514201	0,000471
.	.	.	.	.
1	0.005706	0.011329	0.503633	1.62E-05

After counting all the sig in each gate, the next step is to calculate the sig x with the formula obtained as follows.

$Sig\ x = W^T * Sig\ gate$

Table 19. Results from determining Sig X

No	Wa	Wi	Wf	Wo	*Sig-Ĉt*	sig-it	sig-ft	sig-ot	*Sig x*
40	0.93	0.5	0.04	0.6	0.013523	0,00044	0,000352	0,000471	0.01309 3
.	.	.	.	.	.	.	.	.	.
1	0.93	0.5	0.04	0.6	0.007143	8E-05	0	1.62E-05	0.006692

The last calculation in the backward process is to calculate the weights W and U using calculations that have been previously done, with the formula can be seen the following results.

Sig W = sig gates * Xt

Table 20. Results from determining Sig W

Xt	Sig W			
0.051903	0,000702	2,28587E-05	1.82E-05	2.44504E-05
.	.	.	.	.
0.024221	0,000173	1,93741E-06	0	3,91379E-07

*Sig U = sig gates +1 * Ht*

Table 21. Results of determining Sig U

Ht	Sig U			
0.03114	0	0	0	0
0.027028	0,000365501	1.19033E-05	9.50163E-06	1.27322E-05
.	.	.	.	.
0.005706	1,64739E-05	1.10529E-07	9.29239E-08	1.11127E-07

After getting the results from sig W and sig U in the data then the results are averaged to get W and U weights.

Table 22. Sig W results

Sig W			
0.009768	0.005834688	0.002718	0.002173402

Table 23. Sig U Results

Sig U

0.006608998	0.001704798	0.002075689	0.001480388

The final step in order to get a new weight is to multiply the weight and use the function of the results can be seen as follows.
Lamda (ʎ) = 0.1

Wnew = W old - ʎ* sig W old

Unew = U old - ʎ* sig U old

Table 24. Sig U and Sig W, W and U old

Sig W	Sig U	W old	U old
0.009768	0.006609	0.93	0.67
0.005835	0.001705	0.5	0.64
0.002718	0.002076	0.04	0.85
0.002173	0.00148	0.6	0.95

Table 25. Weight of new W and U

W new	U new
0.929023	0.669339
0.499417	0.63983
0.039728	0.849792
0.599783	0.949852

After getting the new W and U weights. The weight will be used again in the calculation in the forward process, then the new weight will be sought again through the backward process and the process will continue to be adventurous to get a weight that can be used. In the forward and backward iteration process is 10x.

b. RNN LSTM Testing

The new W and U weights for the training data will be used in the testing data without having to go backward again. The testing data used amounted to 11 from the coca-cola 1.5l data item

Table 26. Coca-cola testing data 1.5l

Coca-cola 1,5l			
No	input (xt)	Normalization (xt)	output (real)
1	0	0	c
.	.	.	.
11	19	0.065744	30

The process of this calculation testing data will be used to calculate Ft, Ot, It, and Ĉt uses the new W and U weights after repeated forward and backward processes, so that the following results are obtained.

Table 27. W and U weights that have been processed

Wa	Wi	Wf	Wo	Ua	Ui	Uf	Uo
0.92021 3	0.49419 6	0.03730 2	0.59784 1	0.66347 6	0.63831 9	0.84796 1	0.94854 3

Table 28. Results from calculations Ft, Ot, It, Ĉt

No	xt (normalization)	*Ft*	*Ot*	*It*	*Ĉt*
1	0	0.5	0.5	0.5	0
.	.	.	.	.	.
11	0.065744	0.503258	0.512782	0.510112	0.068669

After calculating Ft, Ot, It, and Ĉt then calculate the results of Ht output and Ct cell state so that they can get prediction results using LSTM calculations.

Table 29. Results of Ht and Ct calculations

No	xt (normalization)	*Ht / Output*	*Ct / Cell State*
1	0	0	0
.	.	.	.
11	0.065744	0.024312	0.047447

c. Result

The final result of the LSTM formula lies in the forward process which is the initial formula of the LSTM. The next process is to calculate the accuracy of the predicted results using the LSTM calculation. To calculate the accuracy of the results used MAD (Mean Absolute Deviation). As well as using MAPE (*Mean Absolute Percent Error)* accuracy results obtained from the average totals obtained are 6.746636499 and for MAD the training data is 4,551990528, and the MAPE obtained is 52.38% The LSTM calculation process will also be used on 9 other data. These results can be seen from the table below.

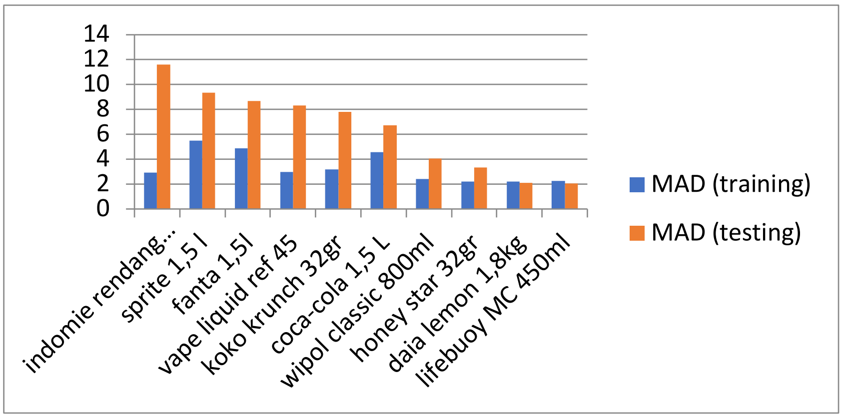
Table 30. Results of all goods

no	Inventory	MAD (training)	MAD (testing)	MAPE
1	coca-cola 1.5 L	4.55	6.74	52.38%
2	Lifebuoy MC 450ml	2.24	2.07	68.79%
3	koko krunch 32gr	3.21	7.81	50.98%
4	Daia Lemon 1.8kg	2.21	2.08	36.29%
5	fanta 1.5l	4.86	8.68	49.30%
6	honey star 32gr	2.23	3.32	59.98%
7	spite 1,5 l	5.49	9.36	30.32%
8	Indomie rendang 91 gr	2.95	11.61	41.10%
9	vape liquid ref 45	2.96	8:30	45.53%
10	800ml Classic Wipol	2.42	4.06	55.38%

3.2 Discussion

This study aims to determine whethercan build a prediction of sales of goods on the Lapan-Lapan Mart by using the Long Short Term Memory Recurrent Neural Network method that can be used to predict the sale of goods. In this study, the data was taken from the Lapan-Lapan Mart, together with data on 10 different items sold every day. The data is then compiled for the level of sales to be weekly and obtained a total of 52 data. Before doing the calculation the data is divided into 2 namely training data and testing data. The training data is used to find out the weights used in the LSTM calculation.

After getting a new weight, the weight will be used to make predictions using the basic formula of *LSTM* the. The results of the prediction are then recalculated using accuracy *MAD (Mean Absolute Deviation).* The accuracy obtained is 6.746636499 by using Coca-cola 1.5l goods data and the results obtained using MAPE (Mean Absolute Percent Error) is 52.38%. After further investigation, there are several things that affect the level of sales of food items that also influence in making predictions using LSTM. Sales time is an important factor that influences the level of sales. In this study it was found that the highest sales occur at the beginning of the fasting month or when there is a big day, which affects the prediction.



Graph 1. Bar Graph based on the highest MAD

4. CONCLUSION

Based on the results of research and discussion that has been done, it can be concluded that, known accuracy results using *MAD (Mean Absolute Deviation)* the highest is indomie rending goods 91 gr (11.61803507) and daisy lemon 1.8kg (2.077000464) for the lowest MAD. If using MAPE the results obtained are MC 450ml (68.79%) lifebuoy the highest and the lowest is 1.5l spite (30.32%).

BIBLIOGRAPHY

[1] Aan, Sgus Jiwa Permana. 2014. Appropriate Eveluation System for Intern Students Using Elman Recurrent Neural Network. Department of Informatics Management, Undiksha University Singaraja: IJCCS, Vol.8, No.1, January 2014, ISSN: 1978-1520

[2] Aprilla C, Dennis, et al. 2013. Learning Data mining with Rapid Miner. Jakarta.

[3] Ching-wu, chu. 2003. A Comparative Study of Linear and Nonlinear Models for Aggregate Retail Sales Forecasting.Int. J. Production Economics 86

[4] Fernando, TGI 2017. *A Recurrent Neural Network Approach in Predicting Daily Stock Prices.* ISSN: 978-1-5386-1676-5 / 17

[5] Gurney, Kevin. (1997). An Introduction to Neural Networks. London: UCL Press.

[6] Hamdani, Widyatmoko et al. 2012. Data Mining Implementation To Predict Sales at Idola Minimarket Jl Pati-Tambakromo KM 2 Karangmuly Village RT 08 RW 1 Using Time Series Method. Informatics Engineering Study Program, Yogyakarta STTA University. Vol 1, No.2, November 2012

[7] Sabrina, Nanggala, et al. 2016*AReasoning and Implementation of Elman Recurrent Neural Network for Prediction of Agricultural Commodity*

Prices.Computational Science Study Program Telkom University. E-Proceeding of Engineering: Vol.3, No.1, April 2016, ISSN: 2355-9365

[8] Yusuf. 2014. Data Mining prediction / forecasting function. http; // yusufmakkaraka.b logspot.co.id/2014/02/data-mining-fungsi-p predict-.html (accessed on tan gal 21 May 2018)

[9] Minami, Shotaro. 2018. Predicting Equity Price with Corporate Action Events Using LSTM-RNN. Journal of Mathematical Finance, 2018, 8, 58-63. ISSN: 2162-2442.