ISSN 2278-3091
Volume 9, No.3, May - June 2020
International Journal of Advanced Trends in Computer Science and Engineering
Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse31932020.pdf
https://doi.org/10.30534/ijatcse/2020/31932020

Drug Stock Optimization Based on Consumption Patterns for Hospital Formulary Using Deep Learning Approach

Faisal Binsar[1], Tuga Mauritsius[2]
[1]Information Systems Management Department, Binus Graduate Program-Master of Information Systems Management,Bina Nusantara University,Jakarta Indonesia 11480, faisal.binsar@binus.ac.id
[2]Information Systems Management Department, Binus Graduate Program-Master of Information Systems Management,Bina Nusantara University,Jakarta Indonesia 11480, tmauritsus@binus.edu

ABSTRACT

This study uses data on drug expenditure transactions based on prescriptions given by doctors to patients in hospitals based on time series to be able to predict drug needs in the coming period. The calculation of drug needs must be done optimally. Because of the excess drug will be a management problem in the drug and pharmaceutical warehouses and expire, while if the lack of medicine will harm patient health and reduce patient satisfaction and trust that also affects the hospital business. This study developed the Long Short Term Memory (LSTM) network in studying the problem of time series prediction on the use of drugs in prescribing patients in hospitals. The calculation of the prediction of drug needs for the coming period is based on the availability of data on drug expenditure in the previous period. Prediction models are trained and tested with monitoring data from January 2017 to December 2019. The results of research on the predicted value of the model and the actual value show the feasibility and effectiveness of using the LSTM network to predict the optimum drug needs in the coming weeks.

Key words : Deep Learning, Long Short Term Memory, Prediction model, Optimization Model, Drug stock, Formulary

1. INTRODUCTION

Drug management in hospitals is one of the most important components in hospital management and is the largest component that absorbs funding outside services. Drug management aims to ensure that the required drugs can always be available at any time needed in sufficient quantities, the right type, on time, guaranteed quality and used rationally. Inefficient drug management has a major impact on the hospital's financial system. One of the effective drug management processes is to ensure the availability of drugs both in terms of the right type and amount in accordance with needs so as to avoid the lack and excess of the drug.

To maintain the availability of drugs, the team at the hospital that involves various professions must agree and select drugs that will be used and circulated in the hospital. The results of this team's agreement are often referred to as the Hospital Drug Register or Hospital Drug Formulary. If the Hospital Formulary already exists, then writing prescriptions and available drugs must follow the rules as stated in it. Hospital pharmaceutical installations overcome this condition by planning drug procurement as data to be used as a Drug Formulary. The formulary evaluation must be done periodically, updating the formulary is an important factor for optimizing the use of formularies [1].

Planning is preceded by the compilation process of drug use which is a recapitulation of drug use data in the health service unit which is used as a basis for calculating optimum stock [2]. Compiling a formulary for selected drugs requires a large amount of historical drug administration data to patients that can be seen by a certain period. To find out this information can be done through the utilization of patient prescription data in the hospital. These data are a very large and growing set of datasets, known as big data. Big data can be used to determine the formulary of drugs in hospitals [3] by conducting data mining of these data to know the future needs of drugs.

During this time the hospital plans for future drug needs through two analyzes, namely ABC analysis and VEN analysis. Both analyzes offer a value of stock requirements that are not based on history or a good trend of drug expenditure, because they only consider the condition of drug stocks one year before and without considering the optimization aspects. This improper planning results in wasteful budgeting, stagnant and stock out. This study proposes a data mining approach in planning drug needs to be included in hospital formularies using deep learning. The most suitable algorithm for time series data form in the form of drug expenditure in hospitals is Long Short Term Memory (LSTM) [4][5] which is the development of the Recurrent Neural Network (RNN) algorithm.

With a historical drug expenditure dataset derived from the Hospital Information System (HIS), using LSTM can be used as training data based on weekly time to predict the needs of selected drugs in the coming week.

The presentation of this study begins by explaining the need for drug availability in the hospital as well as the prediction of the planned number of needs in section 2, followed by the use of data mining as an expected solution and discussing other related research. Section 3 explains the proposed methodology and model. Next section 4 explains the analysis of the results obtained from the model used. The study concludes with a note of conclusions and suggestions for future research in section 5.

2. LITERATURE REVIEW

2.1 Drugs Availability

Drug shortage is a complex problem that affects all aspects of the health care system. The increasing number of drug shortages harms patient care and has implications for expensive funding [6]. Replacing drugs in the absence of drug stocks will have a negative impact as a result of medication errors [7]. On the other hand, excess drug stock will also cause problems for hospitals. Procurement of drugs that are not based on patient needs, will cause stockpiles of drugs to accumulate. Hospital management must think of larger storage space, the drug will become damaged and expire because it is not used.

To reduce storage costs, regular orders can be made in small quantities. However, it should be noted also that out of stock occurs because purchase costs outside of planning can also be high because of the high value of drugs [8].
Related to the planning of stocks to be included in the hospital formulary, a lot of research has been done. But there are still many who use traditional methods. Improper planning will result in waste in budgeting, stagnant, and stockout [9].

2.2 Data Mining

Data mining is the study of collecting, cleaning, processing, analyzing, and gaining useful insights from data [10]. In other words, data mining is a process of searching for patterns in datasets that involve large amounts of data (big data) to find hidden information or knowledge by extracting and exploring important or interesting patterns from data contained in databases. One of the standards used in data mining is to use the Cross-Industry Standard Process for Data Mining (CRISP-DM). Mariscal, Marba, and Fernandez [11] stated CRISP-DM as a standard defacto for the development of data mining and knowledge discovery projects because it is most widely used in data mining development, analysis of business problems and data [12][13].

2.3 Related Works

Inventory management is a key component of hospital logistics, ensuring medicines and all medical supplies arrive at patients on time. Many hospitals still do it imperfectly, do not utilize the best method, and do not use mathematical or statistical models that can help anticipate changes and periodic needs [14]. Permanasari et al [4] used Long Term Short Memory (LSTM) to predict the need for medicines containing digestive enzymes in hospitals. This method was chosen because it is known to have high accuracy for predicting stationary data. Other researchers raise the issue of spending on drugs, especially for patients who cannot afford expensive health care [5]. Rhemimet et al [15] using LSTM showed a clear increase in predicting cleavage by HIV-1 protease.

LSTM has also been used to predict sales that are accurate and researched in the E-commerce business [16]. In the field of stock calculation, [17] conducted a study to get the level of accuracy in predicting a composite stock price index using Deep Learning, in this case, using the LSTM algorithm. The LSTM to predict water quality was carried out by Liu et al. [18], data on drinking water quality was measured by an automatic water quality monitoring station from the Guangzhou Water Source on the Yangtze River in Yangzhou, Keerthana [19] predicts the concentration of air pollution in the air in the coming years by combining RNN and LSTM outputs. The LSTM network also can study long-term dependencies for power generation load needs [20], and to predict the flow of tourist arrivals [21] [22].

3. METHODOLOGY

The proposed method for obtaining knowledge so that it can be used to estimate drug needs in the coming period is to use a deep learning approach in the form of Recurrent Neural Network (RNN) / Long Short Term Memory (LSTM) algorithm, by previously selecting data that suitable for testing. The framework for this research is as shown in Figure 1.

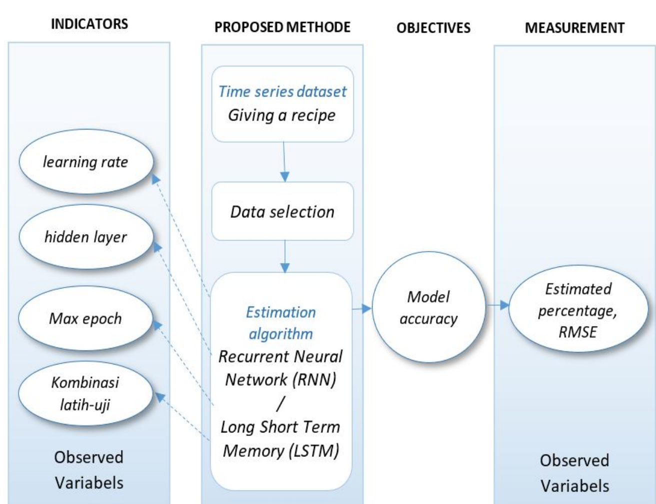

Figure 1: Research Model

The indicators observed were learning rate, number of hidden layers, maximum epoch, and a combination of training data and test data. The objectives in this study are the accuracy of the model estimation of the drug needs in the coming period, the measurement of which is calculated based on the percentage of values, or the estimated amount of the amount of drug use and the value of Root Mean Square Error (RMSE).

This study uses standards from the Cross-Industry Standard Process for Data Mining (CRISP-DM) by carrying out the stages of Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment. While the process of modeling and getting the results of evaluations is done using Python with one of the popular libraries is Keras [23] to build, train, and evaluate various neural networks.

4. RESULT AND DISCUSSION

This study uses the CRISP-DM approach to carrying out the steps. There are six steps or stages, namely: Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment. These six steps are described in each of the following Sub Chapters.

4.1 Data Understanding

The data used in this study is sourced from the Hospital Information System (HIS), one of the data modules is the administration of drugs to patients carried out by the Pharmacy department. Prescriptions are received from doctors who treat patients, both emergency patients, Outpatient and Inpatient Poly. Drug administration data that is already contained in this database is a drug that is actually available and given to patients.

Data obtained from three interrelated tables namely data from the drug reference table, prescription tables, and patient visit tables. After compiling, a table with attributes as in table 1 is obtained.

Table 1: Dataset structure

No	*Field name*	*Data Type*	*Description*
1	Date of visit	Date(10)	Date of a patient visit to the hospital
2	Drug Date	Date(10)	Tanggal resep / pelepasan obat
3	Drug Code	Varchar(10)	Drug code
4	Drug name	Varchar(100)	Drug name
5	Unit	Varchar(30)	Drug expenditure unit
6	Quantity	Int(255)	Quantity of drug released

This study uses a pattern of drug use (Consumption-based). The process of calculating drug needs is based on the previous year's drug use data (time series). From the drug expenditure data given to the patient (prescription), there is information on the number of drugs given along with date information. With data conditions like this, the expenditure of each drug can be calculated for each month.

To get an estimate of the number of drug needs that are planned in the coming period, either the following month or the following year, historical data on the use of drugs that have occurred in previous periods is needed.

4.2 Data Preparation

Data drawn from the Hospital Information System is prepared in steps as shown in Figure 2. Data was collected from 2017 to December 2019 with a total transaction data of 725,547 records. From searching the list of drug names, there are 1,606 drug data records.

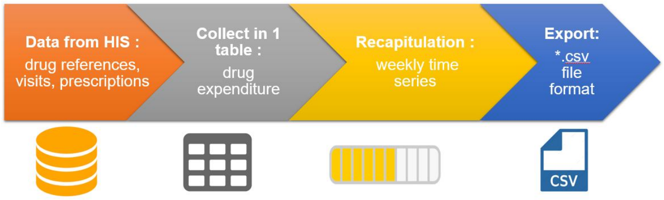

Figure 2: Data preparation phase

This study did not use all drugs but selected 10 types of drugs with criteria appearing regularly every week. With this criterion, the drug chosen is not the drug with the largest amount of expenditure. There are a certain number of weeks of the period found the most drug expenditure, but the drug does not appear routinely weekly so the drug was not selected in the study. After sorting, obtained 10 drugs used in this study are shown in table 2.

Table 2: Drug being studied

No	*Drug name*	*Generic Name*	*Unit of use*
1.	Cefadroxil 500 Mg (Bernofarm)	Cefadroxil	Tablet
2.	Codein 10mg Tablet (KF) 100s	Codeine	Tablet
3.	Infusan Rl (Sanbe)	Infusan	Softbag
4.	Metformin 500mg Tab (Hexapharm)100s	Metformin	Tablet
5.	Metil Prednisolon 4 Mg Tab (Mutifa)100s	Metil Prednisolon	Tablet
6.	Mobafer	Mecobalamin	Tablet

7.	New Diatab (KF)100s	Attapulgite	Tablet
8.	Rhinofed Tab (Dexa)50s	Pseudoefedrine+ Terfenadine	Tablet
9.	Vastigo (Dexa Medica)100s	Betahistin	Tablet
10	Vectrine (Dexa)20s	Erdosteine	Tablet

All transaction records are carried out in the recapitulation process of spending to get a weekly time series for each drug so that the order of time series data is 156 weeks starting from January 1, 2017. Recapitulation is not done in units of days, with consideration to predict the number of drug needs in the future coming will not be planned for the day. The count in days is too narrow and even takes time and energy to make the planner. Planning for medication needs at a hospital is usually carried out for a specific week or month to come.

To formulate the amount of drug expenditure from each day into a week, carried out through the MySql database. Then the data is pulled directly from the MySQL database source and exported into the * .csv file format.

The trend of drug expenditure every week is shown in Figure 3. It can be seen that the trend is very non-linear and it is very difficult to capture trends using this information.

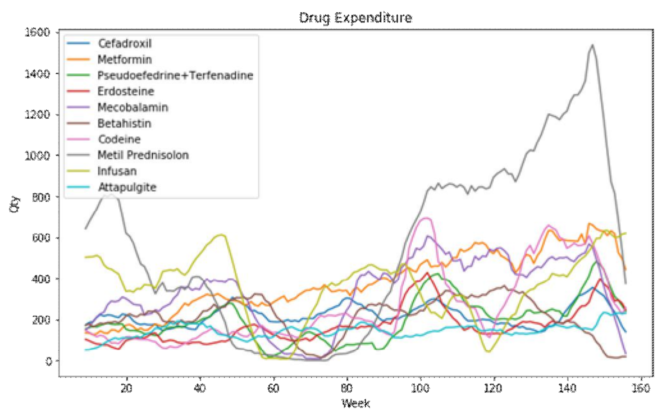

Figure 3: Drug expenditure trends every week

The trend in drug expenditure is influenced by the pattern of disease that develops and the number of patient visits. Metformin has a trend that is always increasing from week to week, this is influenced by the increasing number of patients with diabetes. This drug is given to control high blood sugar levels in patients. While Methyl Prednisolone seems to fluctuate greatly, this is greatly influenced by the number of patients with dyspepsia (ICD code K30), which is experiencing disorders of the digestive tract that can cause various symptoms such as abdominal pain, flatulence, heartburn, nausea, and vomiting. Periods of 60 (2018-02-18) to 80 (2018-07-08) dyspepsia sufferers were found in only 238 cases so that the use of Methyl Prednisolone in this period was at a low point of 406 tablets. Patients with dyspepsia experienced a very significant increase in the 100th week (2018-11-25) to 150 (2019-11-10), namely 2986 cases so that the use of Methyl Prednisolone in this period was at its highest point with the number of 53,124 tablets. LSTM can remember past information while predicting future values with the condition of this past information.

4.3 Modeling

Estimation models are made from datasets using deep learning algorithms, namely Recurrent Neural Network (RNN) / Long Short Term Memory (LSTM). The algorithm will read the attributes in the dataset and make an estimation model for the values of the attributes received.

Modeling is done using Python programming. Python has provided many algorithms for machine learning and especially for deep learning. The RNN / LSTM algorithm in Python is managed in the Keras library. Modeling is done by trying various possible iterations to get the best accuracy. The representation of the proposed model approach algorithm is shown in Figure 4.

```
Import library
Load dataset
Normalize the dataset into values from 0 to 1
Split dataset into train and test sets
Set input units, output units, lstm units and optimizer
for epochs and batch_size do
    Train the LSTM Network
end for
Make predictions
Calculate root mean squared error
```

Figure 4: Pseudo code for prediction of the number of drugs using LSTM

The stages carried out in this study and executed in the Python script are as follows:

- Import the required library
- Convert an array to a dataset matrix
- Read data sets, from CSV files
- Normalization of the dataset
- Determine the percentage of training data and test data (67% training data, the remaining test data)
- Architectural models: 3 gates namely input gate, forget gate and output gate
- Run the LSTM network model
- Make prediction calculations
- Calculate the value of root mean squared error (RMSE)

The study used a number of parameters needed to run the LSTM network model. LSTM networks use input data with specific array structures in the form of batch size, time series, features. The network has a visible layer with 1 input, a hidden layer with 4 LSTM blocks or neurons, and an output layer to produce predictive values. The sigmoid activation function is used as the default in the LSTM block. The network was trained 100 times (epoch) and batch size 1.

Root Mean square error (RMSE) is chosen as a loss function. Adaptive moment estimation (Adam), which has good effects in practical applications, was adopted as an optimization algorithm for the LSTM model, and the weight of the model update and deviation parameters were adjusted. The model results show the x-axis as the original dataset. After being modeled, the data is plotted and shows the original dataset blue, predictions for the training dataset are orange, and predictions on the test dataset are green as shown in Figure 5.

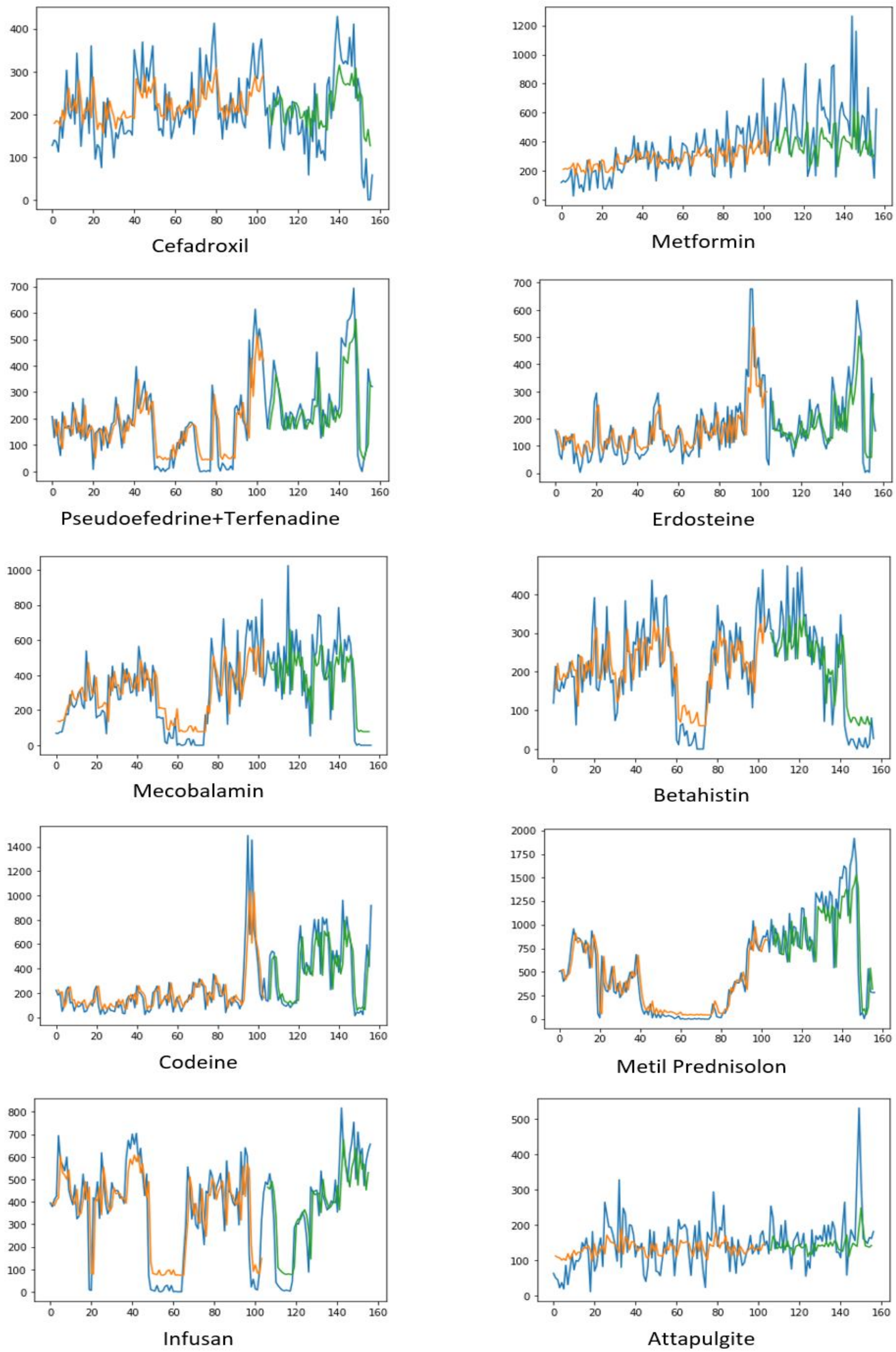

Figure 5: Prediction Model Results

From these results, it can be concluded that this model has a good performance in predicting drug needs in the coming weeks. The results showed the potential application of LSTM and deep learning to optimally predict the number of drugs needed in hospitals based on historical drug expenditure data in the past.

4.4 Evaluation

In predicting the need for this drug, loss function values are obtained at each epoch iteration as shown in Figure 6. In calculating the error of a model during the optimization process, the loss function must be chosen. The loss function [24] determines how many penalties must be set for an instance based on errors in the model's predicted value. The loss function is a method for evaluating how well an algorithm is in modeling the given data. The value of the loss function will continue to decrease when training data in LSTM, which is commonly called the objective function to optimize.

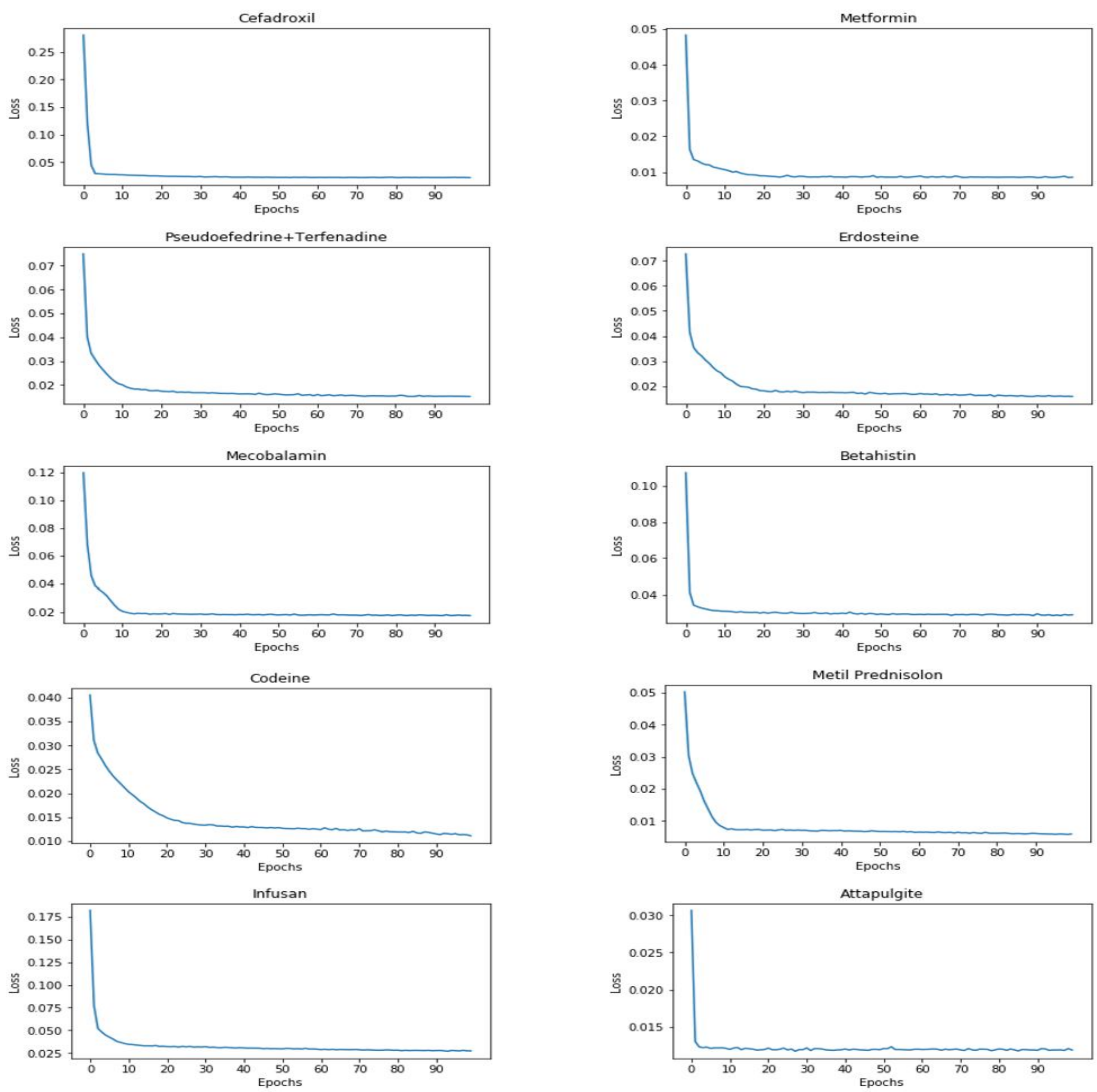

Figure 6: Loss function and epoch iteration

In this research model, RMSE is used to evaluate the degree of adjustment between the predicted value and the actual value. There are two RMSE values namely training RMSE and testing RMSE which are shown in table 3.

The achievement of the value of accuracy of the estimation results of the model built at the modeling stage is evaluated to get a measure of the feasibility of its application in the real world case. By the standard performance measure of a data estimate of a resulting model, validation accuracy is used.

Table 3: RMSE and Loss values

Drug	***RMSE***		***Loss***
	Train	**Test**	
Cefadroxil	64.79	81.60	0.0230
Metformin	126.07	300.11	0.0105
Pseudoefedrine + Terfenadine	86.49	106.24	0.0157
Erdosteine	83.93	105.47	0.0158
Mecobalamin	141.59	206.21	0.0193
Betahistin	82.35	91.00	0.0308
Codeine	155.76	203.23	0.0113
Metil Prednisolon	145.11	319.97	0.0058
Infusan	131.22	120.02	0.0261
Attapulgite	58.84	73.62	0.0131

4.5 Interpretation

Previous studies conducted by Permanasari et al [4] to predict the availability of medicines in hospitals using Long Term Short Memory (LSTM) for the needs of medicines containing digestive enzymes in hospitals, get a value of 12,733 Root Mean Square Error (RMSE). Liu et al. [18] found an MSE value of 0.0017 for LSTM modeling in the measurement of drinking water quality data. While Kaushik et al [5] used two LSTM models, the lowest RMSE was obtained (= 14,617) for a standard LSTM with 7 hidden memory cells and for stacked LSTM the lowest RMSE was obtained (= 13,693) for 4 hidden memory cells in each layer.

There is no standard size for the loss function of certain algorithms in machine learning. There are many factors involved in choosing a loss function for a specific problem such as the type of machine learning algorithm chosen, the ease of calculating derivatives and to some extent the percentage of outliers in the data set.

There are lots of things to do when designing and configuring a deep learning model, such as the number, size, and type of layers in a network. Likewise, the selection of loss functions, activation functions, optimization procedures, and the number of epochs.

5. CONCLUSION

Whether or not the estimated value is indicated by the achievement of drug expenditure figures against the proposed rate is 100% or close to that number. If the expenditure figure is less than the proposed rate, you will get less than 100% of the achievement, or the proposed drug is left in the hospital. Conversely, if the requested drug is not available in the hospital, it means that there is a drug vacancy or the proposed number does not meet the drug needs.

The acquisition of predictive value in this study has shown the achievement of planned drug demand figures in the coming weeks in optimal conditions.

Because deep learning is widely used in problems that have very large datasets, such as the number of tens of thousands or hundreds of thousands of examples, it is recommended to be able to conduct further research for hospitals with higher types with the support of a long history of data backward.

REFERENCES

[1] A. Intang, **The Application Of Drug Formulary System In Public Hospitals Of Makassar City**, *Universitas Hasanuddin Makassar*, 2013.

[2] Kemenkes, ***Pedoman Teknis Pengadaan Obat Publik Dan Perbekalan Kesehatan Untuk Pelayanan Kesehatan Dasar***. *Kementerian Kesehatan RI*, 2008.

[3] D. T. Ma, Carolyn. Smith, Helen wong. Chu, Cherie. Juarez, **Big data in pharmacy practice : current use , challenges , and the future**, *Integr. Pharm. Res. Pract.*, vol. 4, no. February, pp. 91–98, 2016. https://doi.org/10.2147/IPRP.S55862

[4] A. E. Permanasari, A. M. Zaky, S. Fauziati, and I. Fitriana, **Predicting the Amount of Digestive Enzymes Medicine Usage with LSTM**, *Int. Adv. Sci. Eng. Inf. Technol.*, vol. 8, no. 5, pp. 1845–1849, 2018. https://doi.org/10.18517/ijaseit.8.5.6511

[5] S. Kaushik, A. Choudhury, N. Dasgupta, S. Natarajan, L. A. Pickett, and V. Dutt, **Using LSTMs for Predicting Patient ' s Expenditure on Medications**, *Int. Conf. Mach. Learn. Data Sci.*, no. October, pp. 120–127, 2017.

[6] L. D. Sims, B. Stevenson, B. S. Pharm, S. Carolina, and S. Carolina, **Impact of Drug Shortages on Health System Pharmacies in the Southeastern United States**, *Hosp. Pharm.*, vol. 50, no. 4, pp. 279–286, 2015. https://doi.org/10.1310/hpj5004-279

[7] WHO, **Technical Consultation On Preventing And Managing Global Stock Outs Of Medicines**, 2015.

[8] J. D. Quick, **Managing Access to Medicines and Health Technologies**, in *Management Sciences for Health, Inc*, Management Sciences for Health, Inc, 2012.

[9] N. A. Fairuz and T. Yustiawan, **Calculation of Drugs Consumption for Medical Logistics in Rumah Sakit Islam Jemursari Surabaya**, *JAKI*, vol. 5, pp. 155–161, 2017.
[10] C. C. Anggarwal, **Data Mining : Text Book**. *Springer*, 2015.
[11] G. Mariscal, O. Marban, and C. Fernandez, **A survey of data mining and knowledge discovery process models and methodologies**, *Knowl. Eng. Rev. Cambridge Univ. Press*, vol. 25:2, no. 2010, pp. 137–166, 2010.
https://doi.org/10.1017/S0269888910000032
[12] T. Mauritsius and A. S. Braza, **Bank Marketing Data Mining using CRISP-DM Approach**, *Int. J. Adv. Trends Comput. Sci. Eng.*, vol. 8, no. 5, pp. 2322–2329, 2019.
https://doi.org/10.30534/ijatcse/2019/71852019
[13] M. J. Christ, R. Nikolaus, P. Tri, W. Chandra, and T. Mauritsius, **Lending Club Default Prediction using Naïve Bayes and Decision Tree**, *Int. J. Adv. Trends Comput. Sci. Eng.*, vol. 8, no. 5, pp. 2528–2534, 2019.
https://doi.org/10.30534/ijatcse/2019/99852019
[14] C. M. F. Alves, **Demand forecasting in a multi-specialty hospital setting : a comparative study of machine learning and classical statistical methods**, Universidade Do Porto, 2018.
[15] A. Rhemimet, S. Raghay, O. Bencharef, and Y. Chihab, **Long Short-Term Memory Recurrent Neural Network Architectures for Prediction of HIV-1 Protease Cleavage Sites**, *Int. J. Adv. Trends Comput. Sci. Eng.*, vol. 9, no. 1, pp. 194–200, 2020.
https://doi.org/10.30534/ijatcse/2020/29912020
[16] K. Bandara, P. Shi, C. Bergmeir, H. Hewamalage, Q. Tran, and B. Seaman, **Sales Demand Forecast in E-commerce using a Long Short-Term Memory Neural Network Methodology**, *Monash Univ. Melbourne, Aust.*, no. January, 2019.
[17] W. A. Purnomo, **Deep Learning Accuracy Rate In Predicting Composite Stock Price Index (CSPI)**, *Binus University*, 2018.
[18] P. Liu, J. Wang, A. K. Sangaiah, Y. Xie, and X. Yin, **Analysis and Prediction of Water Quality Using LSTM Deep Neural Networks in IoT Environment**, *MDPI J.*, vol. 11, no. 2058, pp. 1–14, 2019.
https://doi.org/10.3390/su11072058
[19] V. S. C. Keerthana R, **Forecasting of the Air Pollution Based on Meteorological Data and Air Pollutants using Deep Learning: A Novel Review**, *Int. J. Adv. Trends Comput. Sci. Eng.*, vol. 9, no. 1, pp. 801–807, 2020.
[20] S. Muzaffar, A. Afshari, and S. Muzaffar, **Short-Term Load Forecasts Using LSTM Networks**, *Energy Procedia*, vol. 158, pp. 2922–2927, 2019.
[21] Y. Li and H. Cao, **Prediction for Tourism Flow based on LSTM Neural Network**, *Procedia Comput. Sci.*, vol. 129, pp. 277–283, 2018.
[22] R. Law, G. Li, D. Ka, C. Fong, and X. Han, **Tourism demand forecasting : A deep learning approach**, *Ann. Tour. Res.*, vol. 75, no. January, pp. 410–423, 2019.
https://doi.org/10.1016/j.annals.2019.01.014
[23] C. Albon, **Machine Learning with Python Cookbook: Practical Solutions from Preprocessing to Deep Learning.** *O'Reilly*, 2018.
[24] T. F. Foster Provost, **Data Science for Business : What you need to know about data mining and data analytic thinking**. *O'Reilly*, 2013.