The implementation of two stages clustering (k-means clustering and adaptive neuro fuzzy inference system) for prediction of medicine need based on medical data

A M Husein1, M Harahap1, S Aisyah2, W Purba3 and A Muhazir3

1 Computer Science Departement, Faculty of Technology and Computer Science, Universitas Prima Indonesia, Medan, Sumatera Utara 20111, Indonesia
2 Information System Departement, Faculty of Technology and Computer Science, Universitas Prima Indonesia, Medan, Sumatera Utara 20111, Indonesia,
3 Computer Science Departement, Faculty of Industry Technology, Institut Teknologi Medan, Medan, Sumatera Utara 20217, Indonesia

Email: amirmahmud@unprimdn.ac.id, mawaddah@unprimdn.ac.id, sitiaisyah@unprimdn.ac.id, windania@unprimdn.ac.id, muhazir@itm.ac.id

Abstract. Medication planning aim to get types, amount of medicine according to needs, and avoid the emptiness medicine based on patterns of disease. In making the medicine planning is still rely on ability and leadership experience, this is due to take a long time, skill, difficult to obtain a definite disease data, need a good record keeping and reporting, and the dependence of the budget resulted in planning is not going well, and lead to frequent lack and excess of medicines. In this research, we propose Adaptive Neuro Fuzzy Inference System (ANFIS) method to predict medication needs in 2016 and 2017 based on medical data in 2015 and 2016 from two source of hospital. The framework of analysis using two approaches. The first phase is implementing ANFIS to a data source, while the second approach we keep using ANFIS, but after the process of clustering from K-Means algorithm, both approaches are calculated values of Root Mean Square Error (RMSE) for training and testing. From the testing result, the proposed method with better prediction rates based on the evaluation analysis of quantitative and qualitative compared with existing systems, however the implementation of K-Means Algorithm against ANFIS have an effect on the timing of the training process and provide a classification accuracy significantly better without clustering.

1. Introduction
Medicines is one of the items needed by the sick community[31]. The sale of medicines in the hospital 50%-60% affects the budget and expenditure of the hospital, so it is needed to predict the optimal medicine needs and because of an error in planning will affect the medical and economical treatment [31]. Medicine planning aims to get the type of the number of medicines in accordance with the requirements and avoid the void based on the pattern of disease and the frequency of the number of patients visit [31]. In general, morbidity method is used for planning of medicines because the expected needs approach the truth and support the improvement of the use of medicines appropriate pattern standard treatment [31]. But it needs a long time and skilled labor, it is difficult to obtain definite disease data and need good recording and reporting [31].
Easy consumption method used does not need data of disease and lack of or excess medicines very small, but it is difficult to get medicine data and the number of patients contacts, lack, excess and loss of medicines are not reliable [31]. Composite method is a method of improvement of the weakness of the method morbidity and consumption, uncertainty number of visits and the frequency of illnesses make this method is less accurate used for the prediction of medicine needs.

Fuzzy Inferences System (FIS) can resolve the problem of uncertainty factors [19], but membership function dependence on the FIS affect the result of the decision-making [32]. Divination with a combination of data mining such as Genetic Algorithm with BP Neural, SOP with Neural Network, SVR with RBFNN outlined on [30] and ARIMA with Neural Network [7]. Neural Network (NN) is a method used by many researchers for divination or prediction [35], NN is better and accurate for divination gold prices compared with ARIMA [13], the ability of NN generalize, identify non-liner relationship and implementation in various applications are very useful [9]. But NN has the weakness of the need for large training data and requires a long computing time to identify the pattern on the data in high dimension[27].

Adaptive Neuro Fuzzy Inference System (ANFIS) is one of the algorithm that combines fuzzy system with artificial nerve network [1]. ANFIS method can be used for divination or prediction [25][1][3][4][6][18], ANFIS method is more accurate in predicting compared ARIMA method and NN [33][24]. But the characteristics of the data used for training will affect the structure of the ANFIS then it affects the accuracy prediction [14].

Clustering is one of the techniques for the classification of objects or patterns in one association so each cluster possess cluster similarity of one another [20]. There are several algorithms in clustering such as K-means, Fuzzy C means (FCM) and Possibilistic c-means (PCM) [26]. FCM is one of the cluster algorithm optimal enough in the classification, but it needs a long computing time [16]. PCM is very strong against noise [17], but it’s too sensitive to the cluster initialization [22]. The K Means Clustering is the algorithm that efficient in grouping a large data set and fast in the computing [16], but limited to numeric data [10].

From the results of several research, combination of K-means Clustering and ANFIS Algorithm is proposed for the prediction of medicine needs. We know with good prediction results will greatly affect the decision-making to improve the medical and medicine need budget more optimal. Data analysis framework needs prediction of medicine based on ANFIS method. The Data is derived from the hospital database. The first stage we use the year 2015-2016 medical data sources to predict the need of medicines with standard ANFIS structure and then calculated the value of the Root Mean Square Error (RMSE) for training and testing. For the second approach, we still use ANFIS algorithm, but after the process of grouping with the algorithm K-Means. The second is the end result of this approach will we compare to know the impact of the classification of accuracy ANFIS prediction. This paper is organized as follows: Part 2 related research. In section 3 the methods used. Section 4 explains the model used, the results and discussion are presented in section 5 and section 6 conclusion.

2. Related Research
Research of Zubir Haider Khan [23], used Artificial Neural Network (ANN) for stock price prediction using two models, one model is for training by using backpropagation and second model is multilayer Feedforward network for testing. Fauzi Yudhi Septiawan [15], applied the Genetic Algorithm to predict the movement of the currency Forex which circulates in Indonesia, USDJPY USDCHF, GBPUSD and EURUSD. Imam Baihaqi Siregar et all [8], performed the comparison method of single ICE, double exponential smoothing holt, triple exponential smoothing based on Exponential Smoothing for palm oil production forecasts and analyzes the accuracy of the model and characteristics of production data forecasts, result of divination single ICE lower based on the value of the RMSE. David Ijegwa Acheme [2], developed a share price prediction using Fuzzy Logic method for decision-making sale and purchase of shares based on the market trends with four indicators are moving average convergence/divergence (MACD), relative strength index (RSI), stochastic oscillator (SO) and on-balance volume (OBV).

Mustain Billah et all [8], did research prediction of share price closing with ANFIS method and ANN on the case of a Securities Exchange Dhaka (BED). Paul Sunjay [28] using ANFIS approach for
optimal supply forecasts. Research of Abbasi [1] for stock price prediction on Iran Khodro Corporation on long-term period by using the triangle membership and four independent variables namely dividends per share (FIXED), price to earning ratio (P/E), closing price and the stock price are used as the independent variables short term. Research of Hien Nguyen Nhu [25], applied the combination of Firefly and ANFIS algorithm. Network model ANFIS trained using the Firefly algorithm and assigned to predict the share price in Vietnam Stock Market with the purpose of comparing the performance among ANFIS with Hybrid Algorithm, Back Propagation and Particle Swarm Optimization (PSO). Research of Osman assured Hegazy had [18] proposed the combination of Genetic Algorithm (GA) with Adaptive Neuro Fuzzy Inference System (ANFIS) for stock price prediction for optimizations. The combination of the previous stock prices as input model for the best stock trend prediction on the next day to see the results of the mean square error (MSE). GA optimize fuzzy parameter to overcome constraints statistics method. Research of Georgia Makridou[24] used ANFIS method for predicting the price of gold and compares with the method 'Buy and Hold' (B&H) strategy. The result of the test accuracy of ANFIS is superior performance (B&H). Ehsan Lotfi [21] proposed ANFIS method can be applied to the oil price forecasts earth by considering input parameters on ANFIS network.

3. The method used

3.1 Adaptive Neuro-Fuzzy Inference System (ANFIS)

ANFIS is Neuro-Fuzzy approach which was introduced by the outskirts (Outskirts et all 1997,1993) which explores fuzzy system Takagai and Sugeno model (1985) and can be used to control application, prediction or divination. ANFIS architecture is similar with neural network, radial function and a little certain restrictions. The process of learning neural network is used to optimize the parameter values of fuzzification. In Neuro-Fuzzy consists of five layers of process. The first layer is the parameter of the membership function fuzzy that his nature is nonlinier association of Exodus system. The learning process on the parameters us error Back-Propagation (EBP) to renew the value of parameters. On the fourth layer, the parameter is a linier parameter against the exodus of the system that make up the basis of fuzzy rules. On this layer is using Least-Squares Estimotor method (LSE).

3.2 K-Means Clustering

The K means clustering is one of the most popular clustering algorithm used because it has simple algorithm, easy to implement and efficient in complexity [5]. Algorithm K-Means determines the cluster of the object based on the attribut/features from the object into the cluster (K). These groups differentiated by its center. This algorithm is very sensitive to the placement of the initial values cluster center. The determination of the partition number of cluster (k) is very important in this algorithm, but there is no terms applicable to determine the amount of cluster(K) formed [34]. The steps in the algorithm K-means are as follows:

1. Prepare the training data.
2. Set the value of the K cluster.
3. Set the initial values centroids.
4. Count the distance between the data and using the formula (Euclidean centroid Distance).
5. Data partition based on the minimum value.
6. And then perform iteration during the data partition is still moving (no longer moving objects to other partitions), when it continues then to points 3.
7. If a group of current data is similar with the previous one, then stop iteration.
8. The data has been partitioned according to the value of the final centroid.

3.3 Dataset

Data need for training and testing ANFIS originates from the previous year data obtained from actual data 2015 and 2016 on one of the Hospital in the City of Medan. Input Data 2015 will be used for the prediction of 2016 and compare the results of ANFIS prediction with the actual data 2016 and so on until the comparison 2017 prediction data with the actual one 2017.
3.4 Accuracy Calculation

ANFIS method prediction is based on the training and testing using the Root Mean Square Error (RMSE). RMSE is an alternative method for evaluation techniques of divination. If the value of the RMSE is getting smaller, and the estimation of the model or variable is more valid.

\[
RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (Y_i - \hat{Y}_i)^2}
\]  

(1)

4. The proposed Model

The proposed model for the prediction of medicine needs at the hospital is based on the results of these observations from two hospitals. In the figure 1 illustrates the framework of research. The first step is to make observations on the private and government hospital, the results of the interview with the head of IFRS, pharmacy and director become as input material, there are three main categories which are used as indicator parameter namely patients, diagnosis and inventory data.

The purpose of this research is to see the influence of the classification of the accuracy of the prediction of ANFIS method sugeno model order one, therefore it uses two different approach. The first stage approach ANFIS method directly applied to perform the training and testing data based medical data. For the second approach, data are grouped with the algorithm k-means as input parameters. The second result of this approach will be made as the material inputs in planning medicine IFRS needs.

5. The results and Discussion

Prediction of medicine needs in IFRS uses the data result of observation at the hospital in 2015 and 2016. The determination of the medical data variable consists of gender, age and ICD patients.

5.1 Prediction with ANFIS without grouping

The training and testing on the first approach directly applies ANFIS method on the data sources with five indicator inputs namely gender, age patients, diagnosis code, long hospitalized, stock, and one of the exodus. There are 34 rules used on the model of the ANFIS Sugeno Order one. In the first stage is to seek the learning rate for optimal learning. This testing uses the amount of data trained as much as 50% of the total data.

RMSE produced by the least is 0.004. The value of the smallest RMSE will be used for the test RMSE and accuracy with a combination of data trained and test data. For the test RMSE and accuracy on the training used the amount of data as much as 85% of the total amount of data. To test the data used 70% of the total amount of data. Each combined data is done by the repetition of ten times to take the value of the RMSE and the accuracy of the test results. The results of the training and testing can be seen in figure 2 and 3.
In the figure 3 is the result of the training data 80% of the total data, with the value of the error 0.004, epoch 500, the results of the training will be made as a test for the prediction of medicine needs with error 0.004, epoch 500 test results seen in fig. 4 for predicting the needs of 2016 and 2017. In the fig. 4 the test result 70% data from the total amount of data for the prediction of medicine needs on 2016 using data 2015, while 2017 uses data 2016. Overall results of the prediction of medicine needs 2016 and 2017, accurate ANFIS method based on the comparison of the predictions with the actual data, but there is still value RMSE data above numbers 0, the value of the average RMSE 0.0669 for 2016 prediction and 0.0594 for prediction of 2017. Graph prediction is shown in figure 4.

5.2 Prediction with K-Means ANFIS (K-ANFIS)
The second approach to prediction of medicine needs is done with the grouping of data. The Data is grouped with the algorithm k-means and then applied to the ANFIS. In table 5 is the result of a cluster of medical data with three groups, namely gender, ICD and age. The process of classification is used in two stages, the first phase of the medical data grouping is 2015 and the second grouping is 2016. Cluster 1 is the most dominant group in the determination of the medicine used, it is very reasonable for an indication of the highest determination of the medicine is the result of disease diagnosis from the doctor. Cluster 3 is the lowest domain because there are no rules of the determination of the medicines for gender. While in cluster 2 is the classification based on age with domain number 2 is the most dominant, this is due to the fact that there are different types of medicines for each of the age of the patient.

The process of computing time on the classification of this data is still too high, this can be seen the result of time which is used for the classification of a domain takes 0.09 seconds while the lowest 0.02 seconds with average 0.05 seconds for one-time iteration from 60 percent of the total data. The number of the cluster of the international Classification of Diseases (ICD) produced 139 clusters isn’t defined and 23 isn’t detected for data 2015, while 2016 produced 145 clusters and isn’t known 12 clusters. On the table 1 explains the distance and the ratio between the cluster.
Table 1. Medical cluster and their number of instances of.

| # Of instances of | Ratio % |
|-------------------|---------|
| Cluster1          | 31.66666667 | 24.1 |
| Cluster2          | 46.47058824 | 46.5 |
| Cluster3          | 55.29411765 | 77.8 |
| Cluster1          | 31.66666667 | 24.1 |

Based on the results of the medical data classification (gender, age, ICD) will be made as ANFIS input parameters for training and testing the predictions of medicine need. The test result can be seen in figure 6.

![Figure 5. Prediction K-ANFIS year 2016 and 2017](image)

In the figure 6 shows the prediction results with applying k-means algorithm for data grouping, significant prediction is more accurate and more quickly in the training process and testing. In the figure 7 and 8 the results of prediction of the comparison ANFIS with K-ANFIS for prediction of medicine needs 2016 and 2017.

![Figure 6. ANFIS and K-ANFIS Prediction 2016](image)

6. Conclusion
Based on the results of the test with the use of medical data of patients 2015 and 2016, for prediction of medicine needs 2016 and 2017, overall ANFIS method is accurate in the prediction, this is seen on the value of the average RMSE, both of the approach is under 0 and comparison of the actual data with
the results of prediction is accurate enough. The second approach is by applying the algorithm k-means of ANFIS which is having an effect on the time of the training process and it provides the accuracy of classification which is significantly better than the first approach without grouping, but the computing time used the algorithm k-means is still too high with an average of 0.05 seconds for each iteration so that it needs further research for computing time optimization by applying the other classification algorithm.

Acknowledgments
This paper is funded by Research Institutes and Community Service (LPPM) of Universitas Prima Indonesia, and supported by one of the private hospital in Medan city and the government hospital at one of district Sumatera Utara for decision making in medicine planning.

References
[1] Abbasi E and Abouec A 2008 Stock Price Forecast by Using Neuro-Fuzzy Inference System Int. J. of Social, Behavioral, Educational, Economic, Business and Industrial Engineering 2.
[2] Ijegwa A D, Rebecca V O, Olusegun F and Isaac O O 2014 A Predictive Stock Market Technical Analysis Using Fuzzy Logic Computer and Information Science 7.
[3] Agrawal S, Jindal M and Pillai G N 2010 Momentum Analysis based Stock Market Prediction using Adaptive Neuro-Fuzzy Inference System (ANFIS) Proc. of the Int. Multi Conf. of Engineers and Computer Scientists IMECS I.
[4] Agrawal A T and Ashtankar P S 2013 An Expert System for Home Health Monitoring: The ANFIS Approach Int. J. of Scientific and Research Publications 3.
[5] Aggarwal N and Aggarwal K 2012 Comparative Analysis of kmeans and Enhanced K-means clustering algorithm for data mining Int. J. of Scientific & Engineering Research 3(3).
[6] Ayat S and Khosravanian A 2014 Identification and Classification of Coronary Artery Disease Patients using Neuro-Fuzzy Inference Systems J. of Mathematics and Computer Science 13 pp 136-141.
[7] Babu C N and Reddy B E 2014 A Moving-Average Filter Based Hybrid ARIMA – ANN Model for Forecasting Time Series Data Applied Soft Computing J. 23 pp 27–38.
[8] Siregar B, Butar-Butar I A, Rahmat RF, Andayani U and Fahmi F 2016 Comparison of Exponential Smoothing Methods in Forecasting Palm Oil Real Production Int. Conf. on Computing and Applied Informatics 801.
[9] Bennett C, Stewart R and Lu J 2014 Forecasting Low Voltage Distribution Network Demand Profiles Using a Pattern Recognition Based Expert System Energy 67.
[10] Bai L, Liang J and Dang C 2011 An Initialization Method to Simultaneously Find Initial Cluster Centers and the Number of Clusters for Clustering Categorical Data Knowledge-Based Systems 246 pp 785-795.
[11] Billah M and Waheed S 2015 Predicting Closing Stock Price using Artificial Neural Network and Adaptive Neuro Fuzzy Inference System ANFIS: The Case of the Dhaka Stock Exchange Int. J. of Computer Applications 129.
[12] Budiono S, Suryawati S and Sulanto SD 1999 Manajemen Obat Rumah Sakit Magister Manajemen Rumah Sakit (Yogyakarta: Fakultas Kedokteran UGM).
[13] Chen H H, Chen M and Chiu C C 2013 The Integration of Artificial Neural Networks and Text Mining to Forecast Gold Futures Prices Communications in Statistics-Simulation and Computation.
[14] Damayanti A, Cahyadi R 2016 Algoritma Genetika Untuk Optimasi Struktur Anfis Seminar Riset Teknologi Informasi SRITI.
[15] Septiawan F Y, Dewa C K and Afiahayati 2017 Prediction Of Currency Exchange Rate In Forex Trading System Using Genetic Algorithm Int. Interdisciplinary Conf. on Sc. Technology Engineering Management Pharmacy and Humanities Held on 22nd – 23rd.
[16] Ghosh S, Dubey SK 2013 Comparative Analysis of K-Means and Fuzzy C Means Algorithms IJACSA Int. J. of Advanced Computer Science and Applications 4.
[17] Hamasuna Y, Endo Y and Miyamoto S 2009 On Tolerant Fuzzy C-Means Clustering and Tolerant Possibilistic Clustering Soft Computing 145 pp 487–494.
[18] Hegazy O, Soliman O S and Toony A A 2014 Hybrid Of Neuro-Fuzzy Inference System and Quantum Genetic Algorithm for Prediction in Stock Market Issues in Business Management and Economics 2 pp 094-102.
[19] Husein A M, Mawengkang H, Budiarthi E 2010 Teknik Fuzzy Inferences System Pada Manajemen Persediaan. (Medan: Magister Teknik Informatika Universitas Sumatera Utara)
[20] Izakian H and Abraham A 2011 Fuzzy C-Means And Fuzzy Swarm for Fuzzy Clustering Problem Expert Systems With Applications 383 pp 1835–1838
[21] Lotfi E and Karimi M R 2014 OPEC Oil Price Prediction Using ANFIS J. of Mathematics and Comp. Sc. 10 pp 286-296.
[22] Ji Z, Sun Q and Xia D 2011 A Modified Possibilistic Fuzzy C-Means Clustering Algorithm for Bias Field Estimation and Segmentation of Brain MR Image Comp. Medical Imaging and Graphics 35 pp 383–397.
[23] Khan Z H, Alin T S, Hussain M A 2011 Price Prediction of Share Market using Artificial Neural Network ANN J. of Computer Applications 22.
[24] Makridou G, Atsalakis GS and Zopounidis C 2013 Gold Price Forecasting With a Neuro-Fuzzy-Based Inference System Int J Financial Engineering and Risk Management 1.
[25] Nhu H N, Nitsuwat S and Sodanil M 2013 Prediction of Stock Price Using an Adaptive Neuro Fuzzy Inference System Trained by Firefly Algorithm Int. Computer Science and Engineering Conf. ICSEC.
[26] De Oliveira, J V and Pedrycz W 2007 Advances in Fuzzy Clustering and its Applications J. Valente de Oliveira and W Pedrycz, Eds Chichester (UK: John Wiley and Sons, Ltd)
[27] Pan S, Iplikci S, Warwick K and Aziz T Z 2012 Parkinson’s Disease tremor classification – A comparison between Support Vector Machines and neural networks Expert Systems with Applications 39 pp 10764–10771.
[28] Paul, S K, Azeem A and Ghosh A K 2015 Application Of Adaptive Neuro-Fuzzy Inference System And Artificial Neural Network In Inventory Level Forecasting Int. J. Business Information Systems 18.
[29] Quick JP, Rankin JR, Laing RO, O’Conor RW 2012 Managing Medicine Supply, The Selection, Procurement, Distribution and Use of Pharmaceutical, Third Edition (USA: Kuminar Press, Conecitcus).
[30] Suryani I, Wahono R S 2015 Penerapan Exponential Smoothing untuk Transformasi Data dalam Meningkatkan Akurasi Neural Network pada Prediksi Harga Emas J. of Intelligent Systems 1.
[31] Satibi 2014 Manajemen Obat di Rumah Sakit (Yogyakarta: Fakultas Farmasi Universitas Gadjah Mada)
[32] Siringoringo R and Situmorang Z 2014 Optimasi Fungsi Keanggotaan Fuzzy Berbasis Algoritma Modified Particle Swarm Optimization Jurnal Ilmiah Komputer dan Informatika KOMPUTA 3.
[33] Yazdani-Chamzini A, Yakhchali S H, Volungevičienė D and Zavadskas E K 2012 Forecasting Gold Price Changes by Using Adaptive Network Fuzzy Inference System J. of Economics and Management 135 pp 994-1010.
[34] Zhang C and Fang Z 2013 An Improved K-means Clustering Algorithm Traditional K-mean Algorithm J. of Information and Computational Science 1 pp 193–199.
[35] Zhang G P 2004 Neural Networks in Business Forecasting (USA: IGI Global).