

ANALYSIS OF NEURAL NETWORK CLASSIFICATION ALGORITHM TO KNOW THE SUCCESS LEVEL OF IMMUNOTHERAPY

Agung Fazriansyah¹; Mochammad Abdul Azis²; Yudhistira³

^{1,2} Computer Science

STMIK Nusa Mandiri, Jakarta, Indonesia

www.nusamandiri.ac.id

¹agungfazriansyah99@gmail.com, ²mochammad.mmz@nusamandiri.ac.id

³ Accounting information system

Universitas Bina Sarana Informatika, Jakarta, Indonesia

www.bsi.ac.id

³yudhistira.yht@bsi.ac.id

Abstract— Cancer is a disease that is feared by humans at this stage, the genetic term of most diseases that have the characteristics of abnormal cell growth and beyond the normal cell limits so that they can attack cells that cover and are able to spread to other organs. For cancer recovery therapy is immunization therapy. Of course in this alternative treatment still needs to be done research to determine the level of success with existing conditions and parameters. Increasingly sophisticated, developing technology that helps human work. The neural network algorithm is used to analyze large datasets, the purpose of this study is to find the accuracy and immunotherapy methods of the dataset using a neural network learning machine with 200 data training cycles, 0.9 momentum and 0.01 learning levels that produce quite high accuracy 80 % and AUC value of 0.738

Keywords: Neural Network, Cancer, Immunotherapy

Abstrak— Kanker adalah salah satu penyakit yang ditakuti oleh manusia pada dekade ini, istilah genetik dari sebagian besar penyakit yang memiliki karakteristik pertumbuhan sel yang abnormal dan melewati batas sel yang normal sehingga dapat menyerang sel sekelilingnya dan mampu menyebar ke organ lainnya. Untuk terapi penyembuhan kanker diantaranya adalah imunoterapi atau terapi peningkatan kekebalan daya tahan tubuh. Tentu dalam alternatif pengobatan ini masih perlu dilakukan penelitian untuk mengetahui tingkat keberhasilannya dengan kondisi dan parameter yang ada. Teknologi yang semakin canggih, berkembang pesat membantu memudahkan pekerjaan manusia. Algoritma neural network digunakan untuk menganalisis dataset besar, maka tujuan penelitian ini adalah untuk menemukan tingkat accuracy dan keberhasilan metode pengobatan immunotherapy dari dataset

menggunakan machine learning neural network dengan data training cycles 200, momentum 0,9 dan learning rate 0.01 yang menghasilkan accuracy cukup tinggi 80% dan nilai AUC sebesar 0,738

Kata Kunci: Neural Network, Kanker, Immunotherapy.

INTRODUCTION

Cancer is a disease that is feared by humans in this decade. Smoking habits, the culture of consuming foods such as fast food, preservatives, and high-fat foods, as well as increased pollution due to industrialization and urbanization, are causes of cancer that are more easily afflicted by humans. (Widowati & Mudahar, 2009). Cancer is a disease that arises as a result of abnormal growth of tissue cells in the body that develop into cancer cells. This disease is one of the leading causes of death worldwide (Sulaiman et al., 2019) There are many ways to prevent the human body from getting cancer, one of which is immunotherapy. Immunotherapy is cancer therapy by strengthening the body's ability to fight cancer cells by increasing the effectiveness of the immune system (Sulaiman et al., 2019)

In this study, researchers sought to analyze the success rate of immunotherapy therapeutic methods in preventing humans from attacking cancer. Researchers use neural network algorithms that are included in the data mining category. Data mining algorithms are able to analyze data to get unknown patterns in large databases from several industries such as insurance, banking, health and marketing which apply to reduce costs, increase sales. Use and improve research quality. Data mining algorithms especially in the medical world are usually applied in diagnosing various diseases. Predictions about health develop very rapidly by using computer science and information

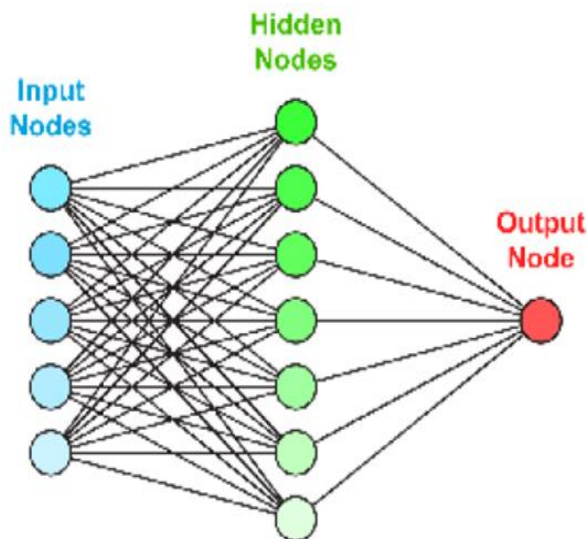
technology in processing health data. (Supriyatna & Mustika, 2018)

MATERIAL AND METHODS

The principle of neural network training using Backpropagation, to illustrate this algorithm that there are three layers of Neural Network algorithms consisting of inputs and outputs, where the input consists of two and only one output as shown in Figure 1 below (Sucipto, 2015).

Weight Initialization

The following figure 1 below is a Layer Neural Network



Sources: (Sucipto, 2015)
Figure 1. Neural Network Layer

Each input (x_{ji} $i = 1,2,3, \dots, n$) receives a signal from input x_1 and then all the hidden layer units will be distributed.

Each *hidden layer* (Z_{ji} $j = 1,2,3, \dots, p$) the sum of the input signal weights

$$Z_{in,j} = W_{bj} + \sum_{i=1}^n x_i w_{ji} \dots\dots\dots(1)$$

Addition of the weight of the incoming signal to the output unit, using the activation function in calculating the output signal

$$y_m = w_{bo} + \sum_j w_{jo} z_j \dots\dots\dots(2)$$

Calculate information from errors
 $\delta = (t - y) * f'(y_m) \dots\dots\dots(3)$

Calculate the corrected weight (new w_{jo})
 $\Delta w_{ja} = \eta \delta z_i \dots\dots\dots(4)$

Calculates the corrected bias
 $\Delta w_{ha} = \eta \delta \dots\dots\dots(5)$

Each unit of *Hidden* ($Z_{\pi j} = 1,2, \dots, p$) adds up the input delta
 $\delta_{inj} = \delta w_{jo} \dots\dots\dots(6)$

To calculate the error multiplied by the activation derivative function
 $\delta_j = \delta_{inj} f'(z_{inj}) \dots\dots\dots(7)$

Weight correction
 $W_{ji} = \delta_j x_i \dots\dots\dots(8)$

Refractive correction
 $f w_{jb} = \delta_j \dots\dots\dots(9)$

Update weights and biases
 $w_{jo}(new) = w_{jo}(old) + \Delta w_{jo} \dots\dots\dots(10)$

each hidden unit *hidden* ($Z_{j1} = 1,2,3, \dots, p$) to update weights and biases
 $w_{ji}(new) = w_{ji}(old) + \Delta w_{ji} \dots\dots\dots(11)$

Iteration will be repeated until it gets the desired value.

Table 1. Data Set Neural Network

	X1	X2	y/z
1	0	0	0
2	1	0	1
3	0	0	1
4	1	1	1

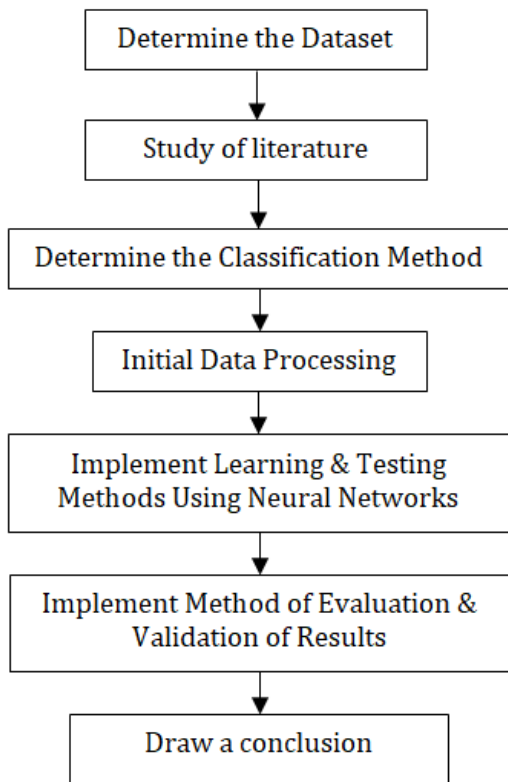
Sources: (Sucipto, 2015)

The table 1 above illustrates a set of data that can be used to teach neural network algorithms. The data set consists of input signals (x_1 and x_2) that can be processed to achieve the desired target (output) z , the neural network is an iterative process (1, 2, 3, 4) wherein each iteration the weight coefficient is modified using new data from the training data set.

Where the data in table 1 above consists of one input layer, with 2 neurons (namely: X_1 and X_2) and has a hidden layer, by having 4 neurons (Z_1, Z_2, Z_3 and Z_4).

$$z = \frac{1}{1+e^{-z_{in}}} \dots\dots\dots(12)$$

Sigmoid activation for 1 output layer with neurons
 $y = \frac{1}{1+e^{y_{in}}} \dots\dots\dots(13)$



Sources: (Rahayu et al., 2019)
Figure 2. Research Steps

The stages of the experiment based on Figure 2 above, which was carried out by the author in conducting this research, were carried out with various activities below:

1. Determine the Dataset

In this study using secondary immunotherapy data obtained from the Repository UCI (California California Invene) with the URL address: <http://archive.ics.uci.edu/ml/>. Filed contained in this dataset in the form of data consisting of age, sex, time, number of warts, area induration diameter, type and attributes of treatment results. With a total dataset of 90 data.

2. Literature Study

A literature study is carried out after obtaining the required dataset, by looking for previous research literature that is related or that has used the same dataset, then study it, and make it as a benchmark for the research conducted.

3. Determine the Classification Method

Based on the existing problems, and the form of the dataset which is supervised (has a label), as well as the study of literature, studied about algorithms that have high accuracy, a suitable method used in this study is Neural Network because it is one of the prediction methods with high accuracy and when added data sets with more accuracy will increase.

4. Initial Data Processing (*Pre Processing*)

Data set of 90 data from cryotherapy data obtained from the UCI Repository will be transformed to obtain data that is truly by the Neural Network algorithm input format with the data input criteria for rapid miner 9.0.

5. Implement Learning and Testing Methods Using Neural Networks

The first process of this method is the process of learning or learning by calculating the predictive value of data in the past. Before the learning process, the immunotherapy value will be processed first to find out whether there is a nominal value in the data or not, if there is, change the nominal value to numeric because there is a boost that cannot read the nominal value. At this stage, the data processing between the training using Neural Network and learning using Neural Network (backpropagation) is then tested to measure the accuracy of the prediction trend.

6. Implement Method of Evaluation and Validation of Results

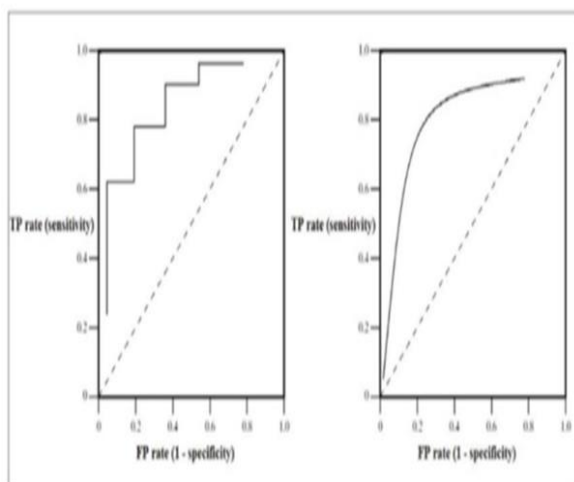
Evaluation is used to observe and analyze the work of the Neural Network on Rapid Miner. Validation is done to measure the predicted results.

a. *Cross Validation*

It is a standardized test carried out to predict the error rate. Each class in the dataset must be represented in the right proportion between training and testing data. Data are randomly divided into each class in the same comparison. To reduce the bias caused by a particular sample, the entire training and testing process is repeated several times with different samples. The error rate at different iterations will be calculated on average to produce an overall error rate (Hastuti, 2012)

b. Measurement using ROC Curve

The *ROC Curve* is a ROC curve that is widely used by researchers to assess predictive results. The ROC curve describes the classification performance regardless of class distribution or error, the vertical end represents a positive value (TP) and the horizontal end represents a negative value (FP). (Sucipto, 2015). See the following figure 3:



Sources: (Sucipto, 2015)
Figure 3. Example of a ROC Graph

Some important things that need to be noted in Figure 3., at the lower left point (0,0) is a point that does not issue a positive classification, (1,1) is represented by the upper right point to explain the positive classification value. Point (0.1) is a classification value for which there are no FN and FP values. Whereas the diagonal line that divides the ROC space represents the space above the diagonal line indicating the good classification and the space below the diagonal line indicates bad classification, while random guesses are found along the diagonal line starting from the lower left to the upper right. (Sucipto, 2015)

c. Confusion Matrix Measurement

Confusion matrix is the number of tables that have provided the amount of test data and are classified correctly and the number of test data that has been classified incorrectly. (Indriani & Nbc, 2014)

An example of a confusion matrix for binary classification is shown in the following table 2:

Table 2. Confusion Matrix for binary classification

		Prediction Class	
		1	0
Actual Class	1	TP	FN
	0	FP	TN

Sources: (Indriani & Nbc, 2014)

The information for Table 2. is stated as follows:

- True Positive (TP), i.e. the number of documents from class 1 that is correct and classified as class 1.
- True Negative (TN), i.e. the number of documents from class 0 that are properly classified as class 0.

- False Positive (FP), i.e. the number of documents from class 0 incorrectly classified as class 1.
- False Negative (FN), i.e. the number of documents from class 1 incorrectly classified as class 0.

Calculation of accuracy is expressed in equations:

$$Accuracy = \frac{(TN+TP)}{(TN+FN+TP+FP)} \times 100 \% \dots\dots\dots(14)$$

7. Draw conclusions

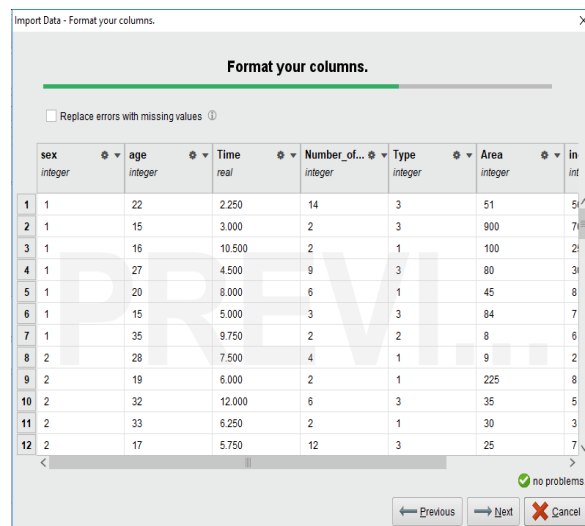
After obtaining the results of the quantitative processing of the dataset, the authors can then conclude whether the dataset has a classification of 1 successful and 0 to fail, and the method of immunotherapy treatment can be predicted by taking into account the parameters.

RESULTS AND DISCUSSION

The Training, Learning, and Testing Process uses Neural Network Algorithms

The Training, Learning, and Testing process uses Neural Network algorithm with a dataset of 90 data as follows:

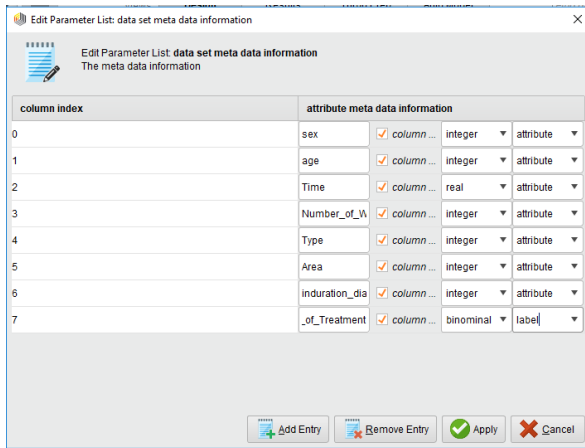
Primary data that are interesting in reading Excel with its attributes are age, number of warts, sex, type, area, time, treatment results and induration diameter as shown in Figure 4 below.



Sources: (Fazriansyah et al., 2019)

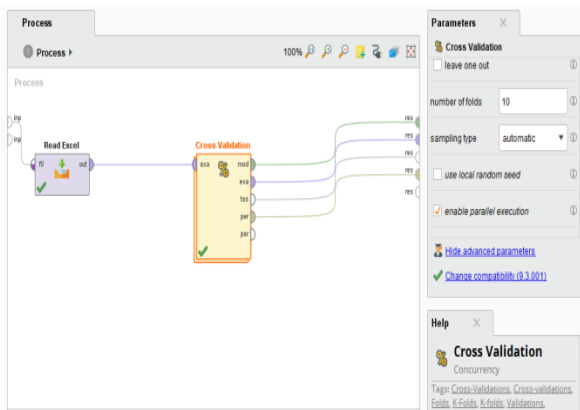
Figure 4. Dataset Input

The next step is setting the attribute type according to the Neural Network algorithm. In this study, only class attributes were changed into binomials, as figure 5 below.



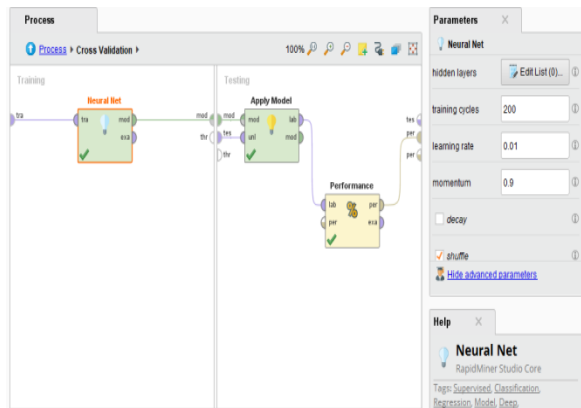
Sources: (Fazriansyah et al., 2019)
Figure 5. Setting attributes

Insert and set validation with some folds 10, as figure 6 below.



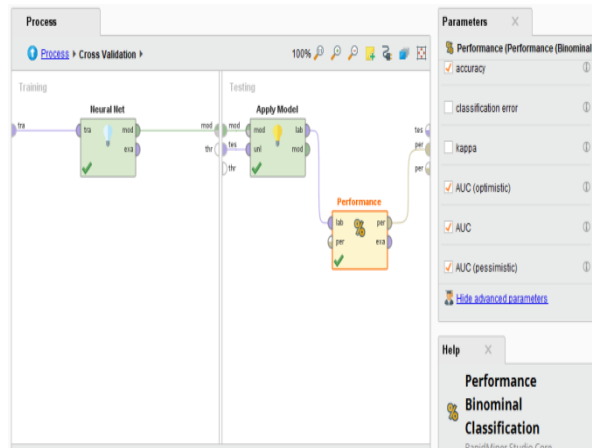
Sources: (Fazriansyah et al., 2019)
Figure 6. Insert and Validation

The next process is the process of inserting the model to be trained, namely by entering the Neural Network algorithm model using 90 data in 200 training cycle datasets, 0.01 from the learning level and 0.9 from the momentum, as in Figure 7 below.



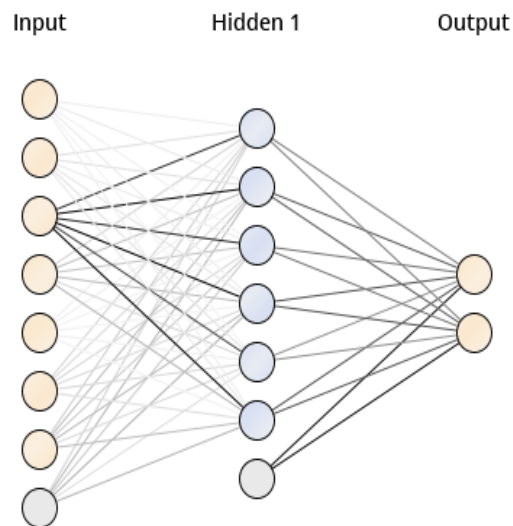
Sources: (Fazriansyah et al., 2019)
Figure 7. Insert the *Neural Network* model

The final process is testing by inserting apply performance models with main criterion accuracy and AUC, as figure 8 below.



Sources: (Fazriansyah et al., 2019)
Gamba Figure 8. Testing with insert apply model and performance

The test results from immunotherapy data show that the accuracy of performance is obtained using the Neural Network algorithm is 80.00% and the AUC value is 0.738. After the testing process by applying the model and performance the Neural Network test results are obtained, The result is like Figure 9 below.



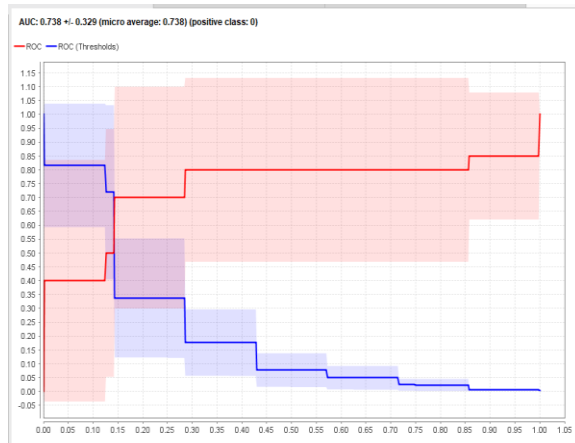
Sources: (Fazriansyah et al., 2019)
Figure 9. Neural Testing Model Results Network

Evaluation and Validation

This study aims to test the accuracy of the immunotherapy success rate analysis using the Neural Network algorithm with data obtained from the UCI Repository assisted by the Rapid Miner

application so that high accuracy is obtained using the neural network algorithm..

1. ROC Curve Test



Sources: (Fazriansyah et al., 2019)

Figure 10. AUC eural Network Curve

Figure 10. Represents the AUC curve resulting from Neural Network validation with an AUC result of 0.738. This shows that the accuracy results obtained fall into the average category.

2. Test the Confusion Matrix

accuracy: 80.00% +/- 12.61% (micro average: 80.00%)

	true 1	true 0	class precision
pred. 1	66	13	83.54%
pred. 0	5	6	54.55%
class recall	92.96%	31.58%	

Sources: (Fazriansyah et al., 2019)

Figure 11. Accuracy Value of Neural Network

Based on Figure 11. it can be predicted that the accuracy using the Neural Network classification method of 80.00% shows that the accuracy results obtained are included in either category.

CONCLUSION

Based on experiments conducted in this study, the immunotherapy dataset was tested using the Neural Network classification algorithm with 200 cycles training, From the resulting learning level of 0.01 and the monetum result 0.9 using the rapid miner application shows high accuracy results of 80.00% and AUC value of 0.738 which means that this dataset is included in the average classification category and can be used as a benchmark for predict success in immunotherapy treatment methods.

REFERENCE

- Fazriansyah, A., Azis, M. A., & Yudhistira, Y. (2019). *Final Report of Independent Research: Analysis Of Neural Network Classification Algorithm To Know The Success Level Of Immunotherapy.*
- Hastuti, K. (2012). ANALISIS KOMPARASI ALGORITMA KLASIFIKASI DATA MINING UNTUK PREDIKSI MAHASISWA NON AKTIF. *SEMANTIK*, 2(1), 241-249. <http://publikasi.dinus.ac.id/index.php/semantik/article/view/132>
- Indriani, A., & Nbc, D. (2014). *Klasifikasi Data Forum dengan menggunakan Metode Naive Bayes Classifier.* 5-10.
- Rahayu, S., Nugraha, F. S., & Shidiq, M. J. (2019). Analisa tingkat keberhasilan cryoterapy menggunakan neural network. *Jurnal Pilar Nusa Mandiri*, 15(2), 1-7. <https://doi.org/10.33480/pilar.v15i2.599>
- Sucipto, A. (2015). CREDIT PREDICTION WITH NEURAL NETWORK ALGORITHM Ir . Adi Sucipto , M . Kom . Sains and Technology Faculty Universitas Islam Nahdlatul Ulama Jepara. *PROSIDING SEMINAR NASIONAL MULTI DISIPLIN ILMU & CALL FOR PAPERS UNISBANK (SENDI_U)*, 15, 978-979. https://www.unisbank.ac.id/ojs/index.php/sendi_u/article/view/3343/
- Sulaiman, F. H., Yulianti, K., & Husain, H. S. (2019). MODEL MATEMATIKA TERAPI KANKER MENGGUNAKAN KEMOTERAPI, IMUNOTERAPI DAN BIOCHEMOTHERAPY. *Jurnal EurekaMatika*, 7(1), 1-10. <https://ejournal.upi.edu/index.php/JEM/article/view/17883>
- Supriyatna, A., & Mustika, W. P. (2018). Komparasi Algoritma Naive bayes dan SVM Untuk Memprediksi Keberhasilan Imunoterapi Pada Penyakit Kutil. *J-SAKTI (Jurnal Sains Komputer Dan Informatika)*, 2(2), 152. <https://doi.org/10.30645/j-sakti.v2i2.78>
- Widowati, L., & Mudahar, H. (2009). Uji Aktivitas Ekstrak Etanol 50°/o Umbi Keladi Tikus (Typhonium Flagelliforme (Lood) Bl) Terhadap Sel Kanker Payudara Mcf-7 in Vitro. *Media Penelitian Dan Pengembangan Kesehatan*, 19(1), 6-14. <http://repository.litbang.kemkes.go.id/1368/1/882-1872-1-PB.pdf>