Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 2, April 2024
DOI : https://doi.org/10.33395/v8i2.13609
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Optimization of Artificial Neural Network Algorithm with Genetic Algorithm in Stroke Prediction

Serin Wulandari[1)*], Yogi Isro'Mukti[2)], Tri Susanti[3)]
[1)2)3)] Informatics Engineering Study program, Pagar Alam Institute of Technology, Indonesia
[1)]Serinwulandari12@email.com, [2)]yogie.isro.mukti@email.com, [3)]trisusantisubagyo8@email.com

Submitted :Mar 7, 2024 | **Accepted** : Mar 31, 2024 | **Published** : Apr 8, 2024

Abstract: This research is motivated by health problems in the community that are less considered so that it causes a disease such as stroke. Factors of lifestyle, poor diet and other factors that can be the cause of stroke. Starting from unhealthy lifestyle patterns such as smoking, drinking alcohol, unhealthy diet, lack of physical activity, psychological stress, and environmental pollution causing various diseases. so that this study aims to optimize Artificial Neural Network with Genetic Algorithm to increase accuracy in predicting stroke. Therefore, where later the data that has been obtained will be processed to see what factors determine the cause of stroke. The data used, namely kaggle and mendeley, will be processed using RapidMiner, with a development method (CRISP-DM) and a testing method using this testing Confusion Matrix to conduct evaluations used to calculate the accuracy of classification data in predicting stroke. The results of this study, stroke disease classification model accuracy kaggle Artificial Neural Network dataset with Genetic Algorithm accuracy 95.13% and AUC 0.667 and mendeley dataset accuracy 98.20% and AUC 0.712. For model evaluation with Artificial Neural Network algorithm with genetic dataset algorithm kaggle using X-fold validation average accuracy of 95.14% and AUC 0.686.7 and mendeley dataset resulted in accuracy of 98.20% and AUC 0.712.5. So as to produce from an algorithm a new attribute from the results of the classification model that has been carried out, namely heart disease, ever married, work type and residence type.

Keywords: Stroke, Artificial Neural Network, Genetic Algoritm.

INTRODUCTION

Health is very important to pay attention to in human life, but sometimes people forget to take care of their health so that diseases suddenly appear. Patients with degenerative diseases or (diseases caused by a decrease in the function of body organs) that cannot be separated from modern lifestyle changes and life demands that cause greater psychological stress. So that it can cause a disease such as stroke caused by several factors with unhealthy lifestyle patterns such as smoking, drinking alcohol, unhealthy diet, lack of physical activity, psychological stress, and environmental pollution That is why with an unhealthy lifestyle can cause various diseases (Isfaizah & Widyaningsih, 2019). (Isfaizah & Widyaningsih, 2019)

Stroke is the leading cause of disability worldwide and the second leading cause of death (WHO, 2022). Stroke is a blood vessel that disrupts blood flow to the brain causing impaired brain function. As a result, blood flow to the brain becomes stiff, numb, or weak, and the disease usually usually occurs on one side of the body (Widyaswara Suwaryo, Widodo, & Setianingsih, 2019). According to World Stroke Organization data, there are new cases reaching 13.7 strokes every year and for the number of death cases around 5.5 million that occur due to stroke, therefore it is necessary to make efforts to predict stroke disease (Ali et al., 2023).

Based on the results of observations and interviews conducted at the Pagar Alam City Health Office about stroke caused by disruption or decrease in blood supply to the brain due to rupture or blockage of blood vessels. Contributing factors such as unhealthy lifestyle, poor diet, hypertension and other factors that cause stroke. Therefore, the Pagar Alam City Health Office has tried to reduce the risk of stroke caused by these factors. However, the results of stroke prevention have not been optimal. To solve the problem, classifying data with algorithms is needed to make accurate predictions of stroke.

Artificial Neural Network is a basic algorithm and simple method used in deep learning approaches. Artificial Neural Network has several advantages compared to other methods, namely Artificial Neural Network can provide

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 2, April 2024
DOI : https://doi.org/10.33395/v8i2.13609
e-ISSN : 2541-2019
p-ISSN : 2541-044X

results that can recognize patterns well and are easily developed into various variations according to existing problems or parameters (Jayadianti, Cahyadi, Amri, & Pitayandanu, 2020) besides that Artificial Neural Network (ANN) is easy to use to solve complex problems so as to shorten data processing time, which computer and information technology practitioners consider it an ideal technique to apply in their projects. This method allows computer programs to mimic and mimic the functions of the human nervous system. This method is usually used with 2 layers. A perceptron is a layer consisting of interconnected points with possibilities based on their weight (Arifin, Haidi, & Dzalhaqi, 2021). Relationship with the author, in this study the author uses Artificial Neural Network in predicting stroke which can later contribute to the treatment of stroke early.

Based on research conducted by (Mutiara, Nurlelah, Ermawati, & Firdaus, 2022) the results of research from the comparison of the Artificial Neural Network method with Particle Swarm Optimization (PSO) resulted in an accuracy of 95.66%, Artificial Neural Network with Genetic Algorithm (GA) produced an accuracy of 96.55% while to use a Neural Network alone without optimization resulted in accuracy reaching 94.51%. The relationship with the author, in this study will predict stroke by improvising between Artificial Neural Network (ANN) and Genetic Algorithm (GA) to get better accuracy. Based on the results of observations, interviews and background above, the author takes the title: "Optimization of Artificial Neural Network Algorithm with Genetic Algorithm in Stroke Prediction".

LITERATURE REVIEW

Optimization

Optimization is the process of carrying out a series of actions that are arranged systematically to achieve a goal and improve performance to the maximum. This optimization role can help improve the accuracy of stroke prediction and speed up the diagnosis process, which in turn can affect overall management of stroke. Thus, optimization plays a key role in improving the effectiveness and accuracy of stroke prediction (Wulandani, Amallia, & Yusra, 2022).

Stroke

Stroke is a disease where when experiencing interference with blood circulation to the brain causes brain tissue to die so that it can cause paralysis or death (Tamburian, Ratag, & Jeini Ester Nelwan, 2020).

Artificial Neural Networks

Artificial Neural Networks (ANN) are intelligent methods in advanced computing that use training and learning to quantitatively analyze information. Due to its highly nonlinearity, large parallel processing capacity, high robustness, and high fault tolerance, the network is suitable for handling complex model internal relationships. Neural networks of the feed forward type flow directly from the input layer to the output layer through several hidden layers in the first type (Hukubun, 2022).

Genetic Algorithm

Genetic algorithms are optimization techniques generally used for practical problems that aim to determine the best parameters. Charles Darwin's theory of natural selection that only populations with high fitness scores can survive is the basis of the mechanism of using genetic algorithms. Genetic algorithms have been used to find the best value solutions to complex problems (Ariadi, 2021).

METHOD

Data Collection Methods

To obtain data for the writing of this research there are several methods that will be needed as follows:

Observation

Observation is a method of data collection carried out through direct review and observation of problems accompanied by detailed recording of the causes and factors that have the potential to cause stroke to be studied (Prawiyogi, Sadiah, Purwanugraha, & Elisa, 2021).

Interview

Interview is one of the data collection techniques that involves direct questioning between data collectors and data sources (Trivaika & Senubekti, 2022). In this data collection technique, the author directly interviewed the head of the NCD sub-coordinator of the Pagar Alam City Health Office.

Documentation

Documentation is a data collection technique by taking pictures or documents to obtain data, both written and electronic documents or others (Apriyanti, Lorita, & Yusuarsono, 2019).

Literature Study

Literature study is a data collection technique carried out studying several research journals and documents that are relevant or related to the research being carried out (Trivaika & Senubekti, 2022).

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 2, April 2024
DOI : https://doi.org/10.33395/v8i2.13609
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Research Location

In this study, researchers conducted research at the Pagar Alam City Health Office which is located at Jl.AIS Nasution, Alun Dua, North Pagar Alam District. Researchers chose to conduct this study at the Pagar Alam City Health Office because it can improve stroke prevention so that it can help people deal with stroke early.

CRISP DM Method

This research uses Cross Industry Standard Process Model For Data Mining (CRISP DM) as a method to solve a problem. There are several stages consisting of business understanding, data understanding, data preparation, modeling, evaluation, and deployment. In this method there are 6 stages that can be explained as follows:

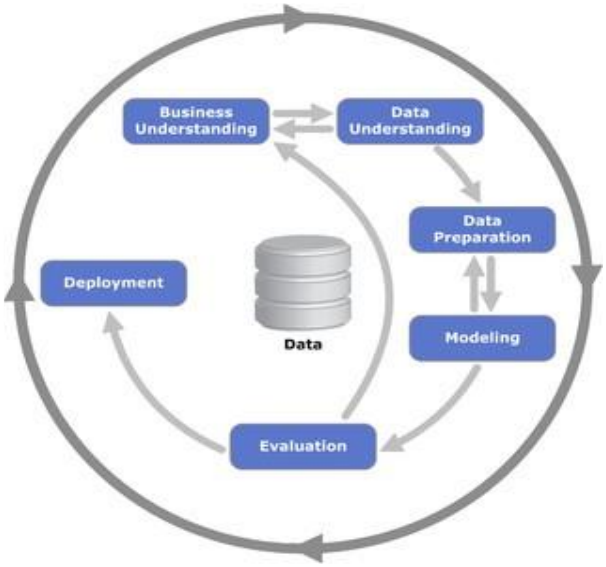


Fig 1. Crisp-Dm groove

Bussines Understanding

For this stage is the first stage that needs to be done such as understanding the purpose needs of a business. So that you can determine plans and strategies in order to achieve these goals.

Data Understanding

This stage includes preparing this writing data using a stroke prediction dataset from the kaggle https://www.kaggle.com/code/rishabh057/healthcare-dataset-stroke-data site and mendeley https://data.mendeley.com/datasets/x8ygrw87jw/1 data.

Data Preparation

At this stage the quality of data for modeling must be improved, such as selecting and sorting data for use in this case the data used through the process of data selection, cleaning and transformation.

Modelling

This stage involves deep learning directly to determine the method used. At this stage build the best modeling scenario by choosing the algorithm to use. The modeling that will be carried out in this study will be carried out using several techniques. Furthermore, all results from modeling will be evaluated using predetermined methods. Here are the techniques that will be used in modeling:

1. Based on the existing problems and looking at the data pattern here, the author determines the completion of the Classification.
2. Using Genetic Algorithm (GA) optimization techniques

Evaluation

At this stage evaluate models that approach the analysis of stroke risk factor data. At this stage, it is done to see the level of performance of the pattern that has been generated from the predetermined algorithm.

To measure the evaluation of model performance based on confusion matrix with accuracy and AUC value. The formula used is as follows (Lestari & Sirodj, 2022):

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+TN+FN} \quad (1)$$

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 2, April 2024
DOI : https://doi.org/10.33395/v8i2.13609
e-ISSN : 2541-2019
p-ISSN : 2541-044X

$$AUC = \frac{1 + TP_{rate} - FP_{rate}}{2} \quad (2)$$

Deployment

This stage is the last stage that will be carried out some attributes that have been processed by the author in the previous stages. So from the pattern above we can see what are the risk factors for stroke.

RESULT

The result of the study was a stroke classification model producing an accuracy rate on the kaggle dataset with an Artificial Neural Network algorithm with a Genetic Algorithm accuracy of 95.13% and AUC of 0.667. As for the dataset mendeley Artificial Neural Network algorithm with Genetic Algorithm accuracy of 98.20% and AUC 0.712.

The author compared how well the accuracy generated from 2 datasets on the Artificial Neural Network algorithm with the Genetic Algorithm in predicting stroke using rapidminer tools. The dataset used was 5,109 data taken from kaggle and 43.3389 data taken from mendeley. For model evaluation using X-fold validation with an average accuracy of 95.14% and AUC of 0.686.7. The mendeley dataset produces an average accuracy of 98.20% and an AUC of 0.712.5

DISCUSSIONS

It can be seen that the stages of the process using CRISP-DM are as follows:

Bussines Understanding

For this stage of business understanding, there are several things that need to be done so that they can determine plans and strategies to achieve these goals. This research requires an understanding of the concept of stroke and risk factors for stroke. Therefore, it is necessary to make efforts in predicting risk factors for stroke. Therefore, researchers will use Artificial Neural Network and Genetic Algorithm to predict stroke to see how well the accuracy of classification of factors that affect the occurrence of stroke so that it can help take stroke prevention measures early.

Data Understanding

At the data understanding stage, there are 48,510 data to be used taken from Kaggle 5,110 data and Mendeley data there are 43,400 data, which have 12 attributes including id, gender, age, hypertension, heart disease, ever married, work type, residence type, avg glucose level, BMI (Body Mass Index), smoking status, and stroke. Data obtained in excel form The data that has been obtained will be carried out to the next stage, namely data selection, data processing / cleaning and data transformation to make it easier to understand, structured and accurate.

Data Preparation

In the data preparation process, data management is carried out which will go through various stages of processing carried out using rapidminer tools, namely:

Data Selection

This stage is done by selecting attributes before starting the analysis stage on the dataset so that the data is easier to use. The data obtained must be ensured relevant data in order to support accuracy. This process is important to ensure that only the data that is important is used in building predictive models. This stage is from 12 attributes obtained from the kaggle and mendeley datasets. Then a filter was carried out using excel where the attributes used were 11 attributes including id, gender, age, hypertension, heart disease, ever married, work type, residence type, avg glucose level, BMI (Body Mass Index), smoking status, and stroke the selected attributes were attributes that influenced the prediction process.

Processing/Cleaning

Furthermore, the data cleaning stage ensures that the data to be managed does not have missing values in Kaggle and mendeley data. Each attribute in the kaggle and mendeley dataset has a missing amount of data so that the data needs processing. From the initial dataset kaggle 5,110 data after cleaning in the gender attribute section so that the data became 5,109 while the mendeley dataset 43,400 data after cleaning became 43,389 data.

Transformation

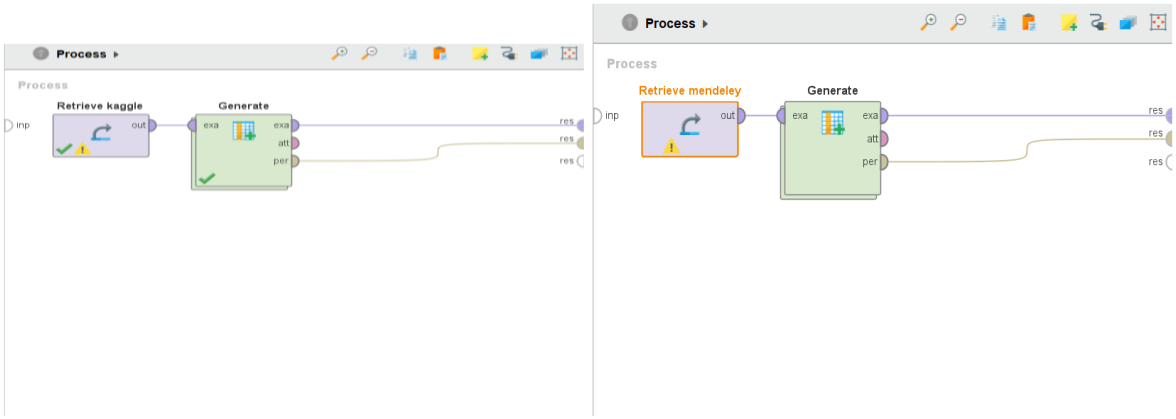
At this stage the kaggle and mendeley dataset is carried out a transformation process to convert the value of several nominal value attributes into numerical such as attributes: gender, age, hypertension, heart disease, ever married, work type, residence type, avg glucose level, BMI (Body Mass Index), smoking status and for stroke attributes will be converted from numeric values to nominal.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 2, April 2024
DOI : https://doi.org/10.33395/v8i2.13609
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Modelling

The process of processing data with genetic algorithm optimization on artificial neural network algorithms using Kaggle and Mendeley datasets. The kaggle dataset is linked to the genetic algorithm using optimize by generation (GGA).



(a) (b)

Fig 1. (a) Kaggle Process (b) Mendeley Process

In the optimize by generation (GGA) operator process, there is a validation process that will be used to conduct testing.

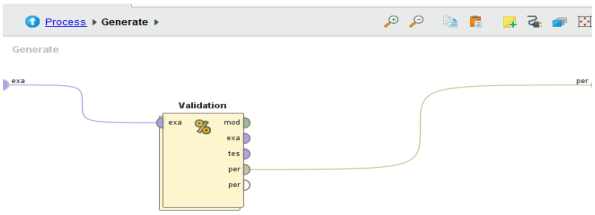


Fig 2. Validation

And the validation process has 2 processes, namely training using neural net while for testing there is apply model and performance to produce information.

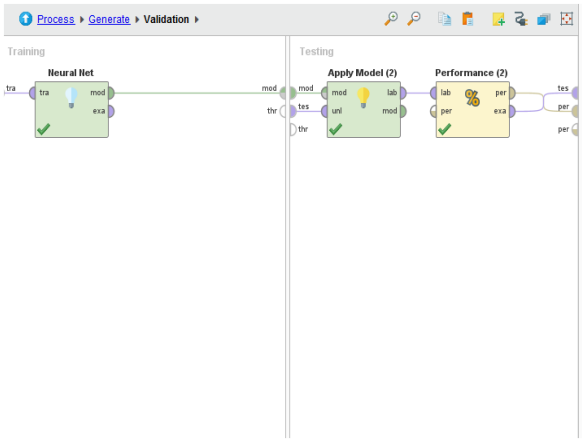


Fig 3. Test model

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 2, April 2024
DOI : https://doi.org/10.33395/v8i2.13609
e-ISSN : 2541-2019
p-ISSN : 2541-044X

The results of model testing conducted by adding optimize by generation (GGA) to the kaggle dataset produce 4 attributes can be seen in the figure below.

Open in Turbo Prep Auto Model Filter (5,109 / 5,109 examples):

Row No.	stroke	heart_disea...	ever_married	work_type	Residence_...
1	Yes	1	1	1	1
2	Yes	0	1	3	0
3	Yes	1	1	1	0
4	Yes	0	1	1	1
5	Yes	0	1	3	0
6	Yes	0	1	1	1
7	Yes	1	1	1	0
8	Yes	0	0	1	1
9	Yes	0	1	1	0
10	Yes	0	1	1	1
11	Yes	0	1	1	0
12	Yes	1	1	4	0
13	Yes	0	1	1	1

Fig 4. Optimize results by generation (GGA)

Furthermore, for the process of artificial neural network algorithms with genetic algorithms on the mendeley dataset. For the artificial neural network algorithm with optimize by generation (GGA) produces 4 attributes can be seen in the figure below.

Open in Turbo Prep Auto Model Filter (43,389 / 43,389 examples):

Row No.	stroke	heart_disea...	ever_married	work_type	Residence_...
1	No	0	0	0	0
2	No	0	1	1	1
3	No	0	0	1	1
4	No	0	1	1	0
5	No	0	0	2	0
6	No	0	1	1	1
7	No	0	1	1	1
8	No	1	1	3	0
9	No	0	1	1	0
10	No	0	1	3	1
11	No	0	1	4	1
12	No	1	1	1	1
13	No	0	1	1	0

Fig 5. Optimize results by generation (GGA)

Evaluation

This stage is the stage of evaluating the model that has been done to see the level of performance generated by the algorithm that has been determined. Based on calculations that have been carried out using the results of Artificial Neural Network testing with Genetic Algorithm on kaggle and mendeley datasets. The results of the tests conducted using X-fold validation to obtain accurate validation results on the kaggle and mendeley datasets using Artificial Neural Network with Genetic Algorithm.

Table 1. Testing

X validasi	ANN+GA(kaggle)		ANN+GA(mendeley)	
	Accuracy	AUC	*Accuracy*	AUC
X1	95,13 %	0.671	98,20%	0.745
X2	95,13 %	0.653	98,20%	0.705
X3	95,13 %	0.676	98,20%	0.705
X4	95,15 %	0.771	98,20%	0.708
X5	95,17 %	0.756	98,20%	0.710
X6	95,13 %	0.658	98,20%	0.706

*name of corresponding author

X7	95,13 %	0.666	98,20%	0.710
X8	95,13 %	0.671	98,20%	0.713
X9	95,13 %	0.667	98,20%	0.712
X10	95,13 %	0.678	98,20%	0.711
Rata-rata	95,14%	0.686,7	98,20%	0.712,5

It can be seen in table 1 above is the result of trials that have been carried out using the Kaggle and Mendeley datasets where tests carried out using x-fold validation as many as 10 experiments to see the accuracy and AUC generated from the 2 datasets so as to produce different average accuracy and AUC. For the kaggle dataset it produces an average accuracy of 95.14% and the AUC produced is 0.686.7 while for the mendeley dataset it produces a higher accuracy of 98.20% and the AUC is 0.712.5 so that it has a difference in the dataset used.

Table 2. Test Results

	Akurasi	AUC
ANN+GA(kaggle)	95.13%	0.667
ANN+GA(mendeley)	98.20%	0.712

And for table 2 can be seen above are the results that have been done at the modeling stage, the results of the classification model can be seen in figures 6 and 7 using artificial neural networks with genetic algorithms. So that the test results that have been done for the kaggle dataset produce an accuracy of 95.13% and the AUC is 0.667 while the mendeley dataset produces an accuracy of 98.20% with an AUC of 0.712 the results of the 2 datasets have the difference that the mendeley dataset produces high accuracy and AUC compared to the kaggle dataset.

Deployment

This stage is the last stage that produces information from several stages that have been done before, namely knowing the classification model in Artificial Neural Network with Genetic Algorithm producing 4 attributes, namely heart disease, ever married, work type and residence type on 2 datasets kaggle and mendeley. From this process, it can be seen that the difference in accuracy in the 2 datasets generated with the Artificial Neural Network algorithm with the Genetic Algorithm has different accuracy.

CONCLUSION

This study used a kaggle dataset of 5,109 data and mendeley 43,389 data on an artificial neural network algorithm with a genetic algorithm. The results of the study were a kaggle dataset stroke classification model on the Artificial Neural Network algorithm with Genetic Algorithm produced an accuracy of 95.13% and AUC 0.667 and for the mendeley dataset produced an accuracy of 98.20% and AUC 0.712. As for the evaluation of the model with the Artificial Neural Network algorithm with Genetic Algorithm on the kaggle dataset using X-fold validation with an average accuracy of 95.14% and AUC of 0.686.7 and for the mendeley dataset which resulted in an average accuracy of 98.20% and AUC of 0.712.5, it can be concluded that the more datasets used, the more accuracy results will increase. As for future suggestions by using other optimization techniques to produce a high level of accuracy.

REFERENCES

Ali, M., BL, A. B., Robbani, F. Y., Hanafi, I., Anugrah, M. R., Ansari, N. V., & Wijaya, S. P. (2023). Pentingnya Pencegahan Dini Stroke, *02*(01), 65–71.

Apriyanti, Y., Lorita, E., & Yusuarsono, Y. (2019). Kualitas Pelayanan Kesehatan Di Pusat Kesehatan Masyarakat Kembang Seri Kecamatan Talang Empat Kabupaten Bengkulu Tengah. *Profesional: Jurnal Komunikasi Dan Administrasi Publik*, *6*(1). https://doi.org/10.37676/professional.v6i1.839

Ariadi, D. (2021). Aplikasi Algoritma Genetika Dalam Mengoptimasi Tuned Mass Damper Untuk Mereduksi Getaran Pada Gedung Akibat Beban Gempa. *Jurnal Kacapuri : Jurnal Keilmuan Teknik Sipil*, *4*(1), 19. https://doi.org/10.31602/jk.v4i1.5125

Arifin, I., Haidi, R. F., & Dzalhaqi, M. (2021). Penerapan Computer Vision Menggunakan Metode Deep Learning pada Perspektif Generasi Ulul Albab. *Jurnal Teknologi Terpadu*, *7*(2), 98–107. https://doi.org/10.54914/jtt.v7i2.436

Hukubun, A. J. M. (2022). Neural Network, (November).

Isfaizah, & Widyaningsih, A. (2019). Menurunkan Tingkat Stres dan Penyakit Degeneratif dengan Pendekatan Focus Grup Discussion di PT Kayu Lapis Indonesia, *1161*(2010), 37–43.

*name of corresponding author

Jayadianti, H., Cahyadi, T. A., Amri, N. A., & Pitayandanu, M. F. (2020). Metode Komparasi Artificial Neural Network Pada Prediksi Curah Hujan - Literature Review. *Jurnal Tekno Insentif*, *14*(2), 48–53. https://doi.org/10.36787/jti.v14i2.150

Lestari, T. S., & Sirodj, D. A. N. (2022). Klasifikasi Penipuan Transaksi Kartu Kredit Menggunakan Metode Random Forest. *Jurnal Riset Statistika*, *1*(2), 160–167. https://doi.org/10.29313/jrs.v1i2.525

Mutiara, E., Nurlelah, E., Ermawati, E., & Firdaus, M. R. (2022). Komparasi metode ann-pso dan ann-ga dalam prediksi penyakit tuberkulosis, *09*(02), 366–381. DOI: http://dx.doi.org/10.20527/klik.v9i2.462

Prawiyogi, A. G., Sadiah, T. L., Purwanugraha, A., & Elisa, P. N. (2021). Penggunaan Media Big Book untuk Menumbuhkan Minat Membaca di Sekolah Dasar. *Jurnal Basicedu*, *5*(1), 446–452. https://doi.org/10.31004/basicedu.v5i1.787

Tamburian, A. G., Ratag, B. T., & Jeini Ester Nelwan. (2020). Hubungan antara hipertensi, diabetes melitus dan hiperkolesterolemia dengan kejadian stroke iskemik. *Journal of Public Health and Community Medicine*, 1*(1), 27–33.* DOI: https://doi.org/10.35801/ijphcm.1.1.2020.27240

Trivaika, E., & Senubekti, M. A. (2022). Perancangan Aplikasi Pengelola Keuangan Pribadi Berbasis Android. *Nuansa Informatika*, *16*(1), 33–40. https://doi.org/10.25134/nuansa.v16i1.4670

WHO, W. H. O. (2022). World Stroke Day 2022. Retrieved from https://www.who.int/srilanka/news/detail/29-10-2022-world-stroke-day-2022

Widyaswara Suwaryo, P. A., Widodo, W. T., & Setianingsih, E. (2019). Faktor Risiko yang Mempengaruhi Kejadian Stroke. *Jurnal Keperawatan, 11*(4), 251–260. https://doi.org/10.32583/keperawatan.v11i4.530

Wulandani, S. A., Amallia, T., & Yusra, Z. N. (2022). Optimalisasi Target dan Realisasi Pajak Pada E-Filling di Kota Bandung, *1*(1). https://doi.org/10.36787/jti.v14i2.150

*name of corresponding author