Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 2, April 2023
DOI : https://doi.org/10.33395/sinkron.v8i2.12228

e-ISSN : 2541-2019
p-ISSN : 2541-044X

Classification of Stroke Opportunities with Neural Network and K-Nearest Neighbor Approaches

Nurul Afifah Arifuddin [1)*], I Wayan Rangga Pinastawa [2)] Nurhajar Anugraha [3)]
Musthofa Galih Pradana [4)]
[1, 2,4)] Univeristy Pembangunan Nasional Veteran Jakarta, Indonesia, [3)] Polytechnic Sriwijaya
[1)]nurulafifaharifuddin@upnvj.ac.id, [2)]rangga@upnvj.ac.id, [3)]nurhajar.anugraha@polsri.ac.id,
[4)]musthofagalihpradana@upnvj.ac.id

Submitted : Feb 28, 2023 | **Accepted** : Mar 8, 2023 | **Published** : Apr 1, 2023

Abstract: Stroke is one of the deadly diseases. This is illustrated in stroke deaths in Indonesia which reached a death rate of 131.8 cases. Some of the things that cause a stroke to become a disease with the highest mortality rate are related to transitions in human life in 4 aspects, namely epidemiology, demography, technology, and economics, socio-culture. Of the many influencing aspects, one of the transition points of human life in the technological aspect can be an alternative solution and prevention. Aspects of technology with the utilization of data can be used as a preventive measure for stroke. One approach is to use data mining techniques, which can provide an initial picture regarding the chances of getting a stroke so that it can be used as an early warning for patients. With so many techniques in data mining, this study used a classification or grouping approach using 2 algorithms, namely K-Nearest Neighbor and one of the Neural Network groups, namely Multi-Layer Perceptron. This research will focus on finding the accuracy and best results of the two algorithms in classifying. The final result of this study is that the K-Nearest Neighbor algorithm has a better accuracy of 95% compared to the Multi-Layer Perceptron which produces an accuracy of 88%.

Keywords: Classification; K-Nearest Neighbor; Multi-Layer Perceptron, Stroke

INTRODUCTION

Referring to data from WHO in 2023, there are 10 diseases as the highest cause of death in Indonesia, in the first place is stroke with a total of 131.8 cases of death per 100,000 population (World Health Organization (WHO), 2023). There are many things that cause a stroke to be the highest cause of death and occur a lot, one of which is related to the transition in human life in 4 aspects namely epidemiology, demography, technology, and economics, socio-culture. Judging from these four aspects, one aspect that can actually be used as a solution to the problems that arise is the technological aspect. With sophistication and up-to-date data and the vital role of data, studies can be carried out and become preventive measures based on data. Referring to these data, can be used as a reference for the prevention and early detection of stroke in order to reduce mortality caused by stroke. To be able to make an early identification, of course, it is necessary to know the symptoms and causes of stroke. The distribution of data is one the important things and can be used as a reference or basis for determining the risk of disease. The approach that can be taken is with a classification model to facilitate class predictions based on the data owned so that it can be used as an early reminder for classes that are labeled and at risk of stroke.

*name of corresponding author

In this study, a classification approach was carried out using the Neural Network algorithm, namely Multi-Layer Perceptron and K-Nearest Neighbor. The approach with these two classification algorithms has characteristics and differences in carrying out the grouping or classification process. This research will make a comparison of the two algorithms with the same dataset, from the results of the classification of the two algorithms, the results of success in carrying out the classification will be obtained on the basis of the comparison reference to get results with the most optimal algorithm in carrying out the classification.

LITERATURE REVIEW

Reference research from Adithiyaa applies KNN in predicting optimal process parameters in mixing metal matrix composite castings. The results of this study can be used to predict a more optimal process in mixing metal matrix composite castings (T. Adithiyaaa, D. Chandramohan, 2020). The results of the research on the KNN method are simple and effective non-parametric techniques for solving classification problems. However, the performance of KNN depends on the distance measure used (Arslan & Arslan, 2021). The application of other algorithms in classifying such as the MLP algorithm in this study can make an efficient diagnosis of various cardiovascular diseases with 88.7% accuracy for MLP and 83.5% for CNN (Savalia & Emamian, 2018). The application of MLP in classifying underwater targets using sonar results in predictions with better results (Qiao et al., 2021). Classification can also be done on media images in the MLP algorithm, the results obtained by models based on neural networks to combine different feature groups achieve the best accuracy of 90.2%(Lai & Deng, 2018). Positive results are also shown in the MLP algorithm the fact that the MLP structure is very effective in classification accuracy (Khishe et al., 2018). Disease prediction can also be made easier with classification algorithms, one of which is Neural Network and Naïve Bayes with a final accuracy of 97.06% for Neural Network, and 91.18% accuracy for Naive Bayes. (A.Vincent, 2022). The KNN and Naïve Bayes classification algorithms have been compared to find accuracy values with various scenarios to predict customer churn. The end result is a method that produces better accuracy is the K-Nearest Neighbor method with a value of K=5 (Kaharudin, Musthofa Galih Pradana, 2019) this indicates that the results of Naïve Bayes are still less effective when compared to the Support Vector Machine algorithm (Musthofa Galih Pradana, Azriel Christian Nurcahyo, 2020). However, quite different results were shown in studies that classified diabetes mellitus with Naïve Bayes results that had better accuracy than KNN. (Putry, 2022), diabetes mellitus is one of the things that besides being able to be classified can also be implemented in the form of an expert system (Musthofa Galih Pradana, Bondan Wahyu Pamekas, 2018). KNN was carried out in two experiments, namely by applying Forward Selection and without using the Forward Selection process for predicting thoracic surgery to produce the best accuracy value in KNN by applying the Forwad Selection process (Sanjaya & Fitriyani, 2019). The process of evaluating academic performance can be simulated using classification techniques in Rinna Rachmatika's research with the results of the random forest model being the best algorithm compared to the Decision Tree, KNN, Naïve Bayes, Logistic Regression, Neural Network, Multi-layer Perceptron, and Support Vector Machine models. (Rachmatika & Bisri, 2020). Different results are shown in the research on classifying student performance scores in taking exams with algorithms compared to Logistic Regression, Multi-Layer Perceptron, and Random Forest where the best results are in the Logistic Regression algorithm. (Galih Pradana et al., 2023). The application of KNN in predicting disease risk for patients was carried out in Shahadat Uddin's research with the results of the ensemble approach and the general average distance can be selected as the most suitable KNN variant for disease prediction according to its high level of accuracy, precision, and memory, the type of variant approaches the limit. different from classic KNN and outperforms others in overall performance (Uddin et al., 2022).

METHOD

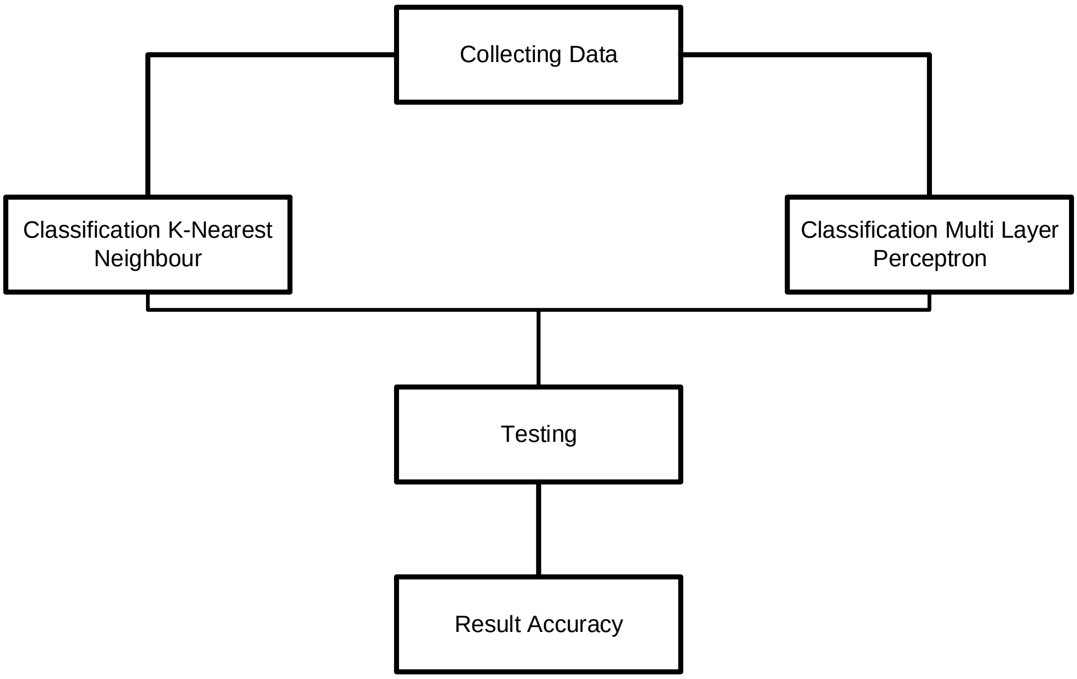
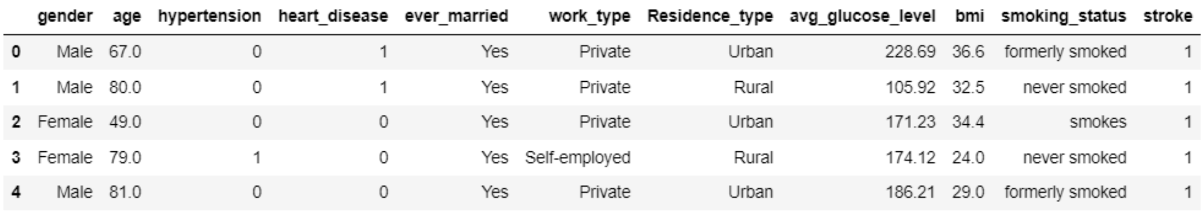

Collecting Data
Classification K-Nearest Neighbour
Classification Multi Layer Perceptron
Testing
Result Accuracy

Figure 1. Flow Research

The stages in the research start from collecting data as a dataset that will be carried out for testing the two algorithms, then proceed with the classification process on the two algorithms, the two algorithms will classify based on classes that are already available with the dataset they have. The data used is a collection of open access data that can be used for research purposes. In carrying out the classification of the two algorithms, it will produce an accuracy value with how much the algorithm is successful in determining the label class according to the original class.

RESULT

The data used in this study are as follows



	gender	age	hypertension	heart_disease	ever_married	work_type	Residence_type	avg_glucose_level	bmi	smoking_status	stroke
0	Male	67.0	0	1	Yes	Private	Urban	228.69	36.6	formerly smoked	1
1	Male	80.0	0	1	Yes	Private	Rural	105.92	32.5	never smoked	1
2	Female	49.0	0	0	Yes	Private	Urban	171.23	34.4	smokes	1
3	Female	79.0	1	0	Yes	Self-employed	Rural	174.12	24.0	never smoked	1
4	Male	81.0	0	0	Yes	Private	Urban	186.21	29.0	formerly smoked	1

Figure 2. Dataset

The dataset used is processed using two algorithms namely K-Nearest Neighbor and Multi-Layer Perceptron. The results obtained by K-Nearest Neighbor are as follows

*name of corresponding author

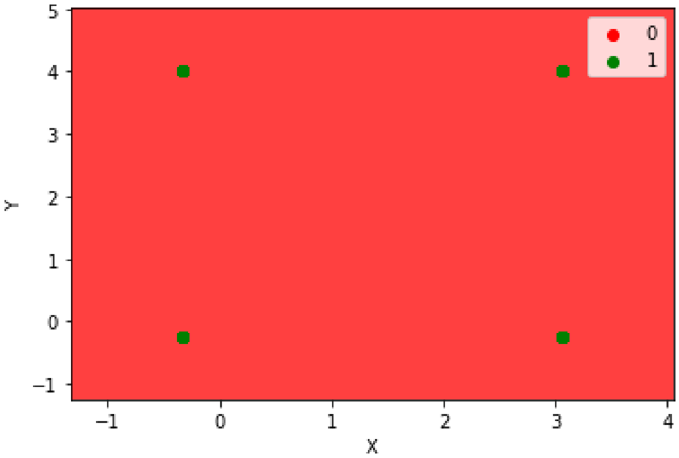


Figure 3. Plot KNN

The plot diagram above shows the distribution of the dataset with the division of training and testing data in green. The test results obtained in detail are shown in Table 1.

Table 1. Result K-Nearest Neighbor

Class	Precision	Recall	F1-Score
0	0.95	1.00	0.98
1	0.00	0.00	0.00

Based on the test results, the precision, recall and F1-Score values are obtained with each number listed in the table. High measured Precision values with True Positive values with lots of positive predicted data.

The results of the MLP algorithm are visualized in the graph shown below in Figure 4.

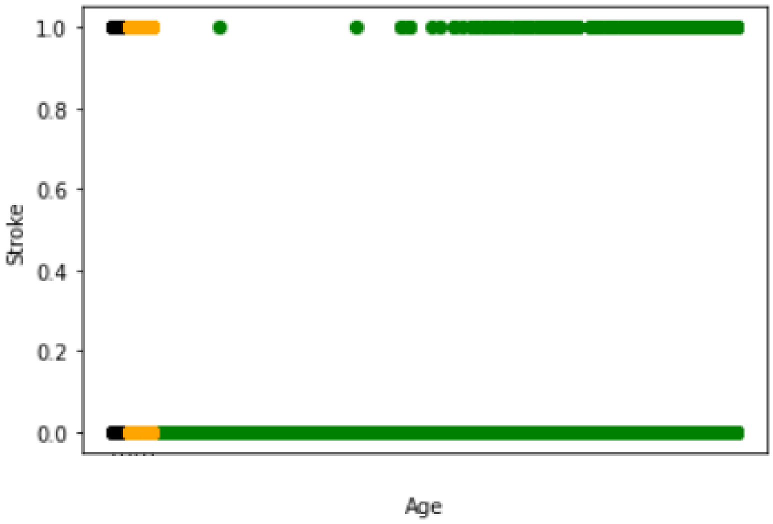


Figure 4. Plot Multi-Layer Perceptron

From the MLP plot it is illustrated that the higher the age, the higher the risk of having a stroke is shown on the green graph. Tests were obtained from details in Table 2.

*name of corresponding author

Table 2. Result Multi-Layer Perceptron

Class	Precision	Recall	F1-Score
0	0.87	0.912	0.93
1	0.89	0.813	0.90

DISCUSSIONS

Based on the results of the tests that have been carried out, testing is carried out to find the accuracy value of both algorithms, the results are shown in Figure 5.

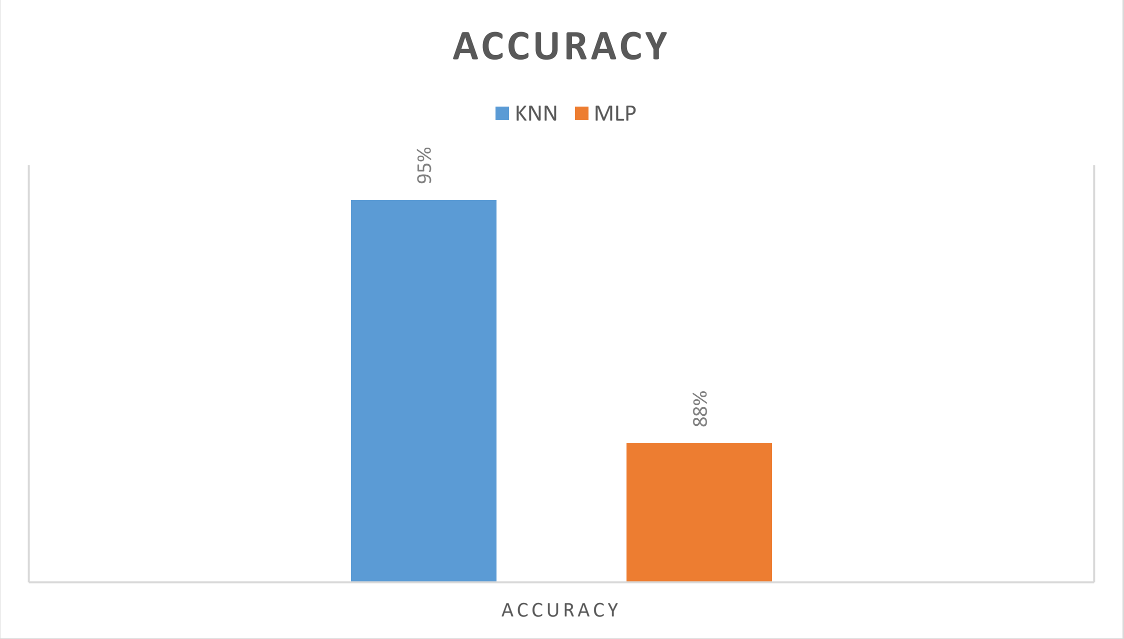


Figure 5. Comparison Accuracy

The results show that the level of accuracy of the two algorithms has a difference that is not too significant with a distance of 7% between 95% on the K-Nearest Neighbor algorithm and 88% on the Multi-Layer Perceptron algorithm.

CONCLUSION

Based on the results of the research conducted, the accuracy produced by the K-Nearest Neighbor algorithm is 95% and the accuracy produced by the Multi-Layer Perceptron algorithm is 88%. With these results, on the characteristics of the data used in this study, the most optimal algorithm is the K-Nearest Neighbor algorithm. For suggestions for future research, it can be enriched with a more diverse dataset or an increase in the performance of the algorithm by modifying and adding elements.

REFERENCES

A.Vincent, J. P. J. F. (2022). Komparasi Tingkat Akurasi Random Forest Dan Knn Untuk Mendiagnosis Penyakit Kanker Payudara. *Universitas Pelita Harapan PSDKU Medan Jurusan Sistem Informasi*, *7*(1), 49–61.

Arslan, H., & Arslan, H. (2021). A New COVID-19 Detection Method from Human Genome Sequences Using CPG Island Features And KNN Classifier. *Engineering Science and Technology, an International Journal*, *24*(4), 839–847. https://doi.org/10.1016/j.jestch.2020.12.026

Galih Pradana, M., Palilingan, K., Vanli Akay, Y., Puspasari Wijaya, D., & Hari Saputro, P. (2023).

*name of corresponding author

Comparison of Multi Layer Perceptron, Random Forest & Logistic Regression on Students Performance Test. 462–466. https://doi.org/10.1109/icimcis56303.2022.10017501

Kaharudin, Musthofa Galih Pradana, K. (2019). Prediksi Customer Churn Perusahaan Telekomunikasi Menggunakan Naïve Bayes dan K-Nearest Neighbor. *Informasi Interaktif*, *4*(3).

Khishe, M., Mosavi, M. R., & Moridi, A. (2018). Chaotic Fractal Walk Trainer For Sonar Data Set Classification Using Multi-Layer Perceptron Neural Network and Its Hardware Implementation. *Applied Acoustics*, *137*(July 2017), 121–139. https://doi.org/10.1016/j.apacoust.2018.03.012

Lai, Z., & Deng, H. (2018). Medical Image Classification Based On Deep Features Extracted By Deep Model And Statistic Feature Fusion With Multilayer Perceptron. *Computational Intelligence and Neuroscience*, *2018*. https://doi.org/10.1155/2018/2061516

Musthofa Galih Pradana, Azriel Christian Nurcahyo, P. H. S. (2020). Pengaruh Sentimen Di Sosial Media Dengan Harga Saham Perusahaan. *Jurnal Ilmiah Edutic*, *6*(2).

Musthofa Galih Pradana, Bondan Wahyu Pamekas, K. (2018). Penyakit Diabetes Mellitus Menggunakan Metode Certainty Factor Design Expert System For Diagnosing Diabetes. *CCIT Journal*, *11*(2), 182–191.

Putry, N. M. (2022). Komparasi Algoritma Knn Dan Naïve Bayes Untuk Klasifikasi Diagnosis Penyakit Diabetes Mellitus. *EVOLUSI : Jurnal Sains Dan Manajemen*, *10*(1). https://doi.org/10.31294/evolusi.v10i1.12514

Qiao, W., Khishe, M., & Ravakhah, S. (2021). Underwater targets classification using local wavelet acoustic pattern and Multi-Layer Perceptron neural network optimized by modified Whale Optimization Algorithm. *Ocean Engineering*, *219*(June 2020), 108415. https://doi.org/10.1016/j.oceaneng.2020.108415

Rachmatika, R., & Bisri, A. (2020). Perbandingan Model Klasifikasi untuk Evaluasi Kinerja Akademik Mahasiswa. *Jurnal Edukasi Dan Penelitian Informatika (JEPIN)*, *6*(3), 417. https://doi.org/10.26418/jp.v6i3.43097

Sanjaya, R., & Fitriyani, F. (2019). Prediksi Bedah Toraks Menggunakan Seleksi Fitur Forward Selection dan K-Nearest Neighbor. *Jurnal Edukasi Dan Penelitian Informatika (JEPIN)*, *5*(3), 316. https://doi.org/10.26418/jp.v5i3.35324

Savalia, S., & Emamian, V. (2018). Cardiac Arrhythmia Classification By Multi-Layer Perceptron And Convolution Neural Networks. *Bioengineering*, *5*(2). https://doi.org/10.3390/bioengineering5020035

T. Adithiyaaa, D. Chandramohan, T. S. (2020). Optimal prediction of process parameters by GWO-KNN in stirring- squeeze casting of AA2219 reinforced metal matrix composites. *Materials Today: Proceedings*, *20*, 329–334. https://doi.org/10.1016/j.matpr.2019.10.051

Uddin, S., Haque, I., Lu, H., Moni, M. A., & Gide, E. (2022). Comparative Performance Analysis Of K-Nearest Neighbour (KNN) Algorithm And Its Different Variants For Disease Prediction. *Scientific Reports*, *12*(1), 1–11. https://doi.org/10.1038/s41598-022-10358-x

World Health Organization (WHO). (2023). *Diseases as the Highest Cause of Death in Indonesia*.