

---

## Comparison of Machine Learning Techniques in the Classification of Parkinson's Disease Sufferers

Titik Khotiah<sup>1)\*</sup>, David Fahmi Abdillah<sup>2)</sup>, Ilham Basri K<sup>3)</sup>, Fery Arianto<sup>4)</sup>, Abdul Rohman<sup>5)</sup>

<sup>1)2)3)4)5)</sup>Institut Teknologi dan Bisnis Ahmad Dahlan Lamongan

<sup>1)</sup>[titik.ahmaddahlan@gmail.com](mailto:titik.ahmaddahlan@gmail.com), <sup>2)</sup>[davidfahmi18@gmail.com](mailto:davidfahmi18@gmail.com), <sup>3)</sup>[ilhambasrik@gmail.com](mailto:ilhambasrik@gmail.com), <sup>4)</sup>[fery.ar31@gmail.com](mailto:fery.ar31@gmail.com),

<sup>5)</sup>[rockyrohman@gmail.com](mailto:rockyrohman@gmail.com)

---

### ABSTRACT

Parkinson's disease is a progressive and relatively common *neurodegenerative* disorder in the central nervous system where sufferers can have difficulty moving. This disease has a high mortality rate in the world of around 9.3 million in 2021. Meanwhile, in Indonesia, it is estimated that as many as 12,980 people die every year due to Parkinson's cases. This increase in cases of death is due to the lack of information about the initial symptoms and dangers of the disease, besides it is important to know how to prevent it early. Early detection of Parkinson's disease can prevent symptoms of a certain age thereby increasing life expectancy. The existence of a computer-based system for diagnosing Parkinson's disease is called a classification system where the system applies the *Machine Learning* method. This study aims to compare the performance of algorithms in the classification system of people with computer diseases. In this study, it used methods in *Machine Learning* such as K-NN, *Multi Layer Perceptron* (MLP), *Linear Regression* and *Support Vector Machine* (SVM). The data set in this study was obtained using the Weka application. The dataset used was *Parkinson's Disease* data totaling 195 rows of data taken from the *UCI Machine Learning Repository Datasets*. The results of the experiment based on the four algorithms showed that the poor performance was the *Multi Layer Perceptron* approach to regression data with an RSME value of 0.459. Meanwhile, the *k-Neural Network* Algorithm is a good classification technique for Parkinson's problem with an RMSE value of 0.1895.

**Keywords:** Parkinson, K-NN, MLP, Linear Regression, SVM

---

### 1. INTRODUCTION

Parkinson's disease (PD) is one of the relatively common *neurodegenerative* disorders with the course of its development classified as *a prodromal*, stage 1, 2, 3 and severe condition. In the medical world it is often difficult to identify the severity of Parkinson's disease and predict the progression of the disease. The disease often occurs in individuals older than 60 years (Mozhdehfarahbakhsh et al., 2021). Symptoms of Parkinson's disease are the body feeling weak or feeling stiffer in part of the body, as well as tremors or subtle shaking in one of the hands. This disease exerts an influence on a small part of the midbrain called substantia nigra which functions as a messenger to various nerves in the spine where the spine has control of the muscles in the body (Oktariza et al., 2019).

One of the *methods of Machine Learning* in the bioinformatics domain is to look for the best possibilities for the selection of classification techniques. From the available *public datasets*, a classification or clustering is carried out that computes the possibility of whether a person suffers from a disease or not. The use of feature selection or weighting is useful for improving performance, but is not done in these experiments. As a description of Parkinson's disease, experiments were carried out using several classification algorithms in *Machine Learning* such as k-NN or *Linear Regression*, *Support Vector Machine*, and *Neural Network* (Artificial Neural Network). A system that can provide information about Parkinson's disease and ways of dealing with it such as early diagnosis and treatment. A computer-based system that can solve these problems there is a classification system, where the system can provide information and perform analysis as done by classification, one of the methods that can be applied in the classification system is K-Nearest Neighbor, this method is very good in carrying out classification based on previous events (Prakoso et al., 2020).

Previous research related to the classification of Parkinson's disease has been widely carried out, namely in the research of Prakoso & Wiriatmadja (2020) stated that the results of testing the K-Nearest Neighbor classification

\* Corresponding author



with Confusion Matrix obtained the best accuracy results of 80% on the jumlah 116 dataset Parkinson Disease (Prakoso et al., 2020). Another study that discusses Parkinson's is the classification method using the K-Nearest Neighbor algorithm, obtained the highest accuracy of 90% in 14 Parkinson Disease datasets (Desinaini et al., 2019). Peneliti compared with the methods present in previous studies. The proposed system has the highest classification accuracy achievement which through cross-validation analysis is 10 times, amounting to 97.47%. This method is proposed as a candidate for the appropriate method for the diagnosis of *People With Parkinson's (PWP)* due to its excellent results.

This disease has a high mortality rate in the world which is around 9.3 million in 2021. Meanwhile, in Indonesia, it is estimated that as many as 12,980 people die every year due to Parkinson's cases (Bismi & Harafani, 2022). This increase in cases of death is due to the lack of information about the initial symptoms and dangers of the disease, besides it is important to know how to prevent it early. Early detection of Parkinson's disease can prevent symptoms of a certain age thereby increasing life expectancy. The existence of a computer-based system for diagnosing Parkinson's disease is called a classification system where the system applies the *Machine Learning* method. This study aims to compare the performance of algorithms in the classification system of people with Parkinson's disease.

**2. LITERATURE REVIEW**

Literature review is one of the references in research and writing in the discussion. The benefits of this literature review also enrich the theory used. Some of the theories included in this study relate to:

**Support Vector Machine (SVM)**

SVM is an algorithm that tries to find a linear separator (hyperplane) between the data points of two classes in multidimensional space. SVM is suitable for dealing with the interaction between redundant features and features (Zuriel & Fahrurrozi, 2021).

**Multi Layer Perceptron (MLP)**

Single layer networks (perceptron) and MLP are algorithms of the Neural Network family. Multilayer networks (MLP) are able to realize non-linear functions, but have weaknesses on the local minima side & high dimensions of weight-space (Azzahra et al., 2022)

**Linear Regression**

LR is a supervised algorithm used to determine the relationship between two numerical or continuous or quantitative variables (X and Y) The formula/formula of Linear Regression (Haryadi et al., 2021) is:

$$y=a+bx \tag{1}$$

**K-Nearest Neighbor**

The k-Nearest Neighbor (k-NN) algorithm is one of the supervised learning algorithms in which the results of the new data classification are based on the majority category of the k-th closest neighbor. In classification, this algorithm uses the majority of votes among the classifications of k objects (Johanes & Santoso, 2020).

The following is a table of previous studies relevant to this study:

**Table 1.**  
**Previous Research**

<b>Do not.</b>	<b>Researcher Name, (Year)</b>	<b>Heading</b>	<b>Equation</b>	<b>Difference</b>
1	(Arrazaq dkk., 2017)	Diagnosis of Parkinson's Disease Through VGRF-Based Walking Pattern Analysis Using Wavelet and	Parkinson's dataset, Alogirtma Support Vector Machine	VGRF with Wavelet

\* Corresponding author



		Support Vector Machine		
2	(Desinaini dkk., 2019)	Application of Fuzzy K-Nearest Neighbor (FKNN) To Detect Parkinson's Disease	Parkinson's Disease Dataset	Fuzzy K-Nearest Neighbor (FKNN)
3	(Prakoso et al., 2020)	Classification System in Parkinson's Disease using the K-Nearest Neighbor Method	Dataset Parkinson Disease, Nearest Neighbor K-Algorithm	116 dataset PD dari UCI
4	(Utami, 2020)	Classification of REM Behaviour Disorder Sleep Disorders Based on EEG Signals Using Machine Learning	Algoritma Support Vector Machine (SVM)	Data RBD
5	(Handayani, 2021)	Application of Support Vector Machine (SVM) Algorithms for Classification Pattern Analysis on Parkinson's Datasets	Parkinson's Dataset, Support Algoritma Vector Engine (SVM)	Pendekatan PPE
6	(Sumarlinda & Lestari, 2022)	K-Nearest Neighbor (KNN) Application for Cardiovascular Disease Classification	Algoritma K-Nearest Neighbor (KNN)	Cardiovascular disease dataset

The previous research that was used as a reference in this study was the results of a classification system using K-Nearest Neighbor in the Parkinson's disease dataset with Confusion Matrix testing obtained the best accuracy results of 80% in the number of 116 datasets (Prakoso et al., 2020).

### 3. METHOD

#### Research Data

The data in this study The dataset used was *Parkinson's Disease* data totaling 195 rows of data taken from the UCI *Machine Learning Repository Datasets*. Dataset in matrix format (excel) taken from 31 samples with 23 of them positive for Parkinson's. It consists of 10 independent attributes and 1 output column with a value of 0 or 1. A value of 0 means healthy and 1 means a sufferer (Hariharan & Sindhu, 2014). The full attributes can be seen in Table 1.

**Table 1. Parkinson's Dataset Attributes**

Do not.	Attribute Name	Definition
1	MDV P:Fo (Hz)	Mean of the base frequency of vowels
2	MDV P: Fhi(Hz)	Maximum base frequency of vowels
3	MDV P: Flo (Hz)	Minimum base frequency of vowels
4	MDV P:Jitte r(%)	Basic frequency variations
5	MDV P:Jitte r(Abs)	Basic frequency variations
6	MDV P:RAP	Basic frequency variations
7	MDV HAL:PPQ	Basic frequency variations
8	Jitte r:DDP	Basic frequency variations
9	MDV P:Berkilauan	Basic frequency variations
10	MDV P:Berkilau(dB)	Wave variations
11	Glitter:APQ3	Wave variations
12	Glitter:APQ5	Wave variations
13	MDV P:APQ	Wave variations
14	Berkilauan:DDA	Wave variations
15	NHR	Noise ratio for tonal components in sound

\* Corresponding author



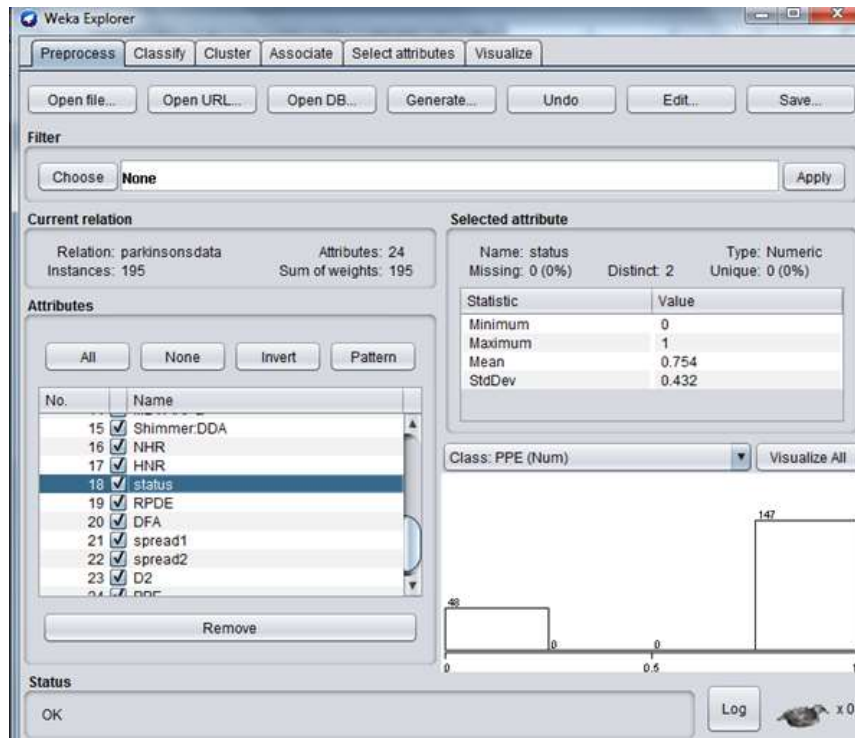
16	HNR	Noise ratio for tonal components in sound
17	Condition	Health status
18	RPDE	Complexity
19	DFA	Fractal signal-level exponents
20	penyebaran1	Base frequency variation (nonlinear)
21	penyebaran2	Base frequency variation (nonlinear)
22	D2	Complexity
23	PPE	Base frequency variation (nonlinear)

**Data Analysis**

In this study, the data above was processed using the help of the Weka application. The calculation process is started by taking data on disease symptoms in dogs and then the data is classified according to the symptoms of the disease. The calculation process of K-Nearest Neighbor begins with calculating the distance Eucliden distance, after obtaining the value of the result of the distance Eucliden distance begins to sort the calculation of the *Eucliden* distance based on the smallest distance. The sorting results are taken by a number of  $k = 1$ , namely the top data (the smallest distance value) to show the recommendations of the results of the Parkinson's diagosa. Furthermore, the next calculation uses methods in *Machine Learning* such as *Support Vector Machine* (SVM), *Multi Layer Perceptron* (MLP) and *Linear Regression*. The final results will be compared to each of these algorithms' performance in classifying Parkinson's disease.

**4. RESULT AND DISCUSSION**

The experiment was conducted using the Weka v3.8 application. The test was performed with 10-fold cross-validation (Tempola et al., 2018).

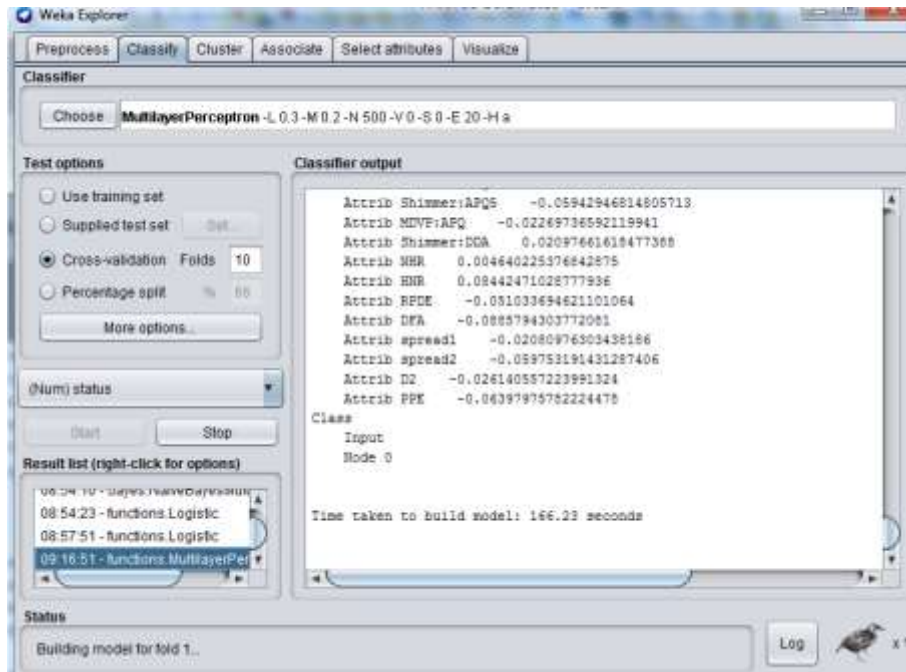


**Figure 1.** Parkinson's data input process with 195 instants, and 23 attributes.

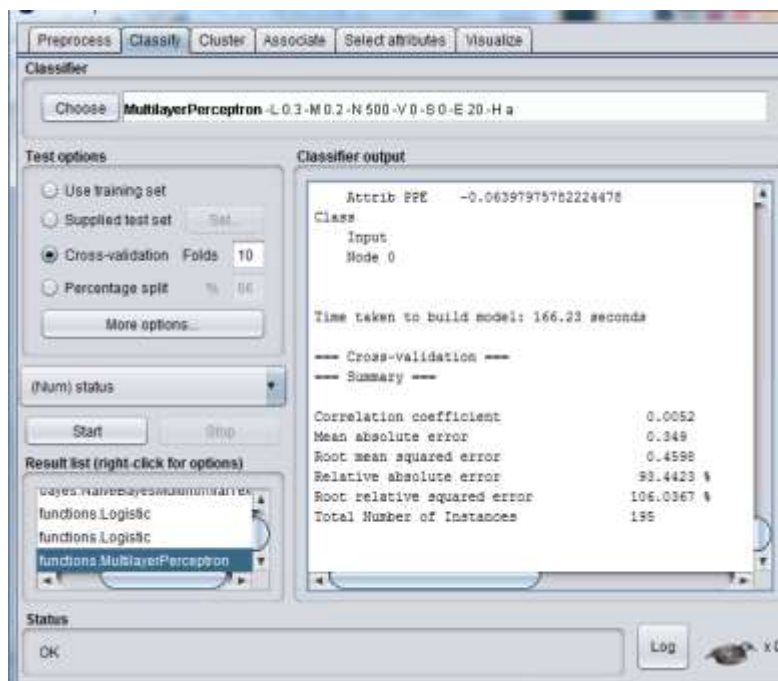
\* Corresponding author



Figure 1. is the first process of entering Parkinson's data into the Weka app. The dataset consists of 195 instant data and 23 attributes. After all the data is inputted into the application, the next process is experimentation with a predetermined algorithm. The first experiment was carried out using the *Multi Layer Perceptron (MLP)* algorithm, with results such as figure 2 below:



**Figure 2a.** The process using *Multi Layer Perceptron*



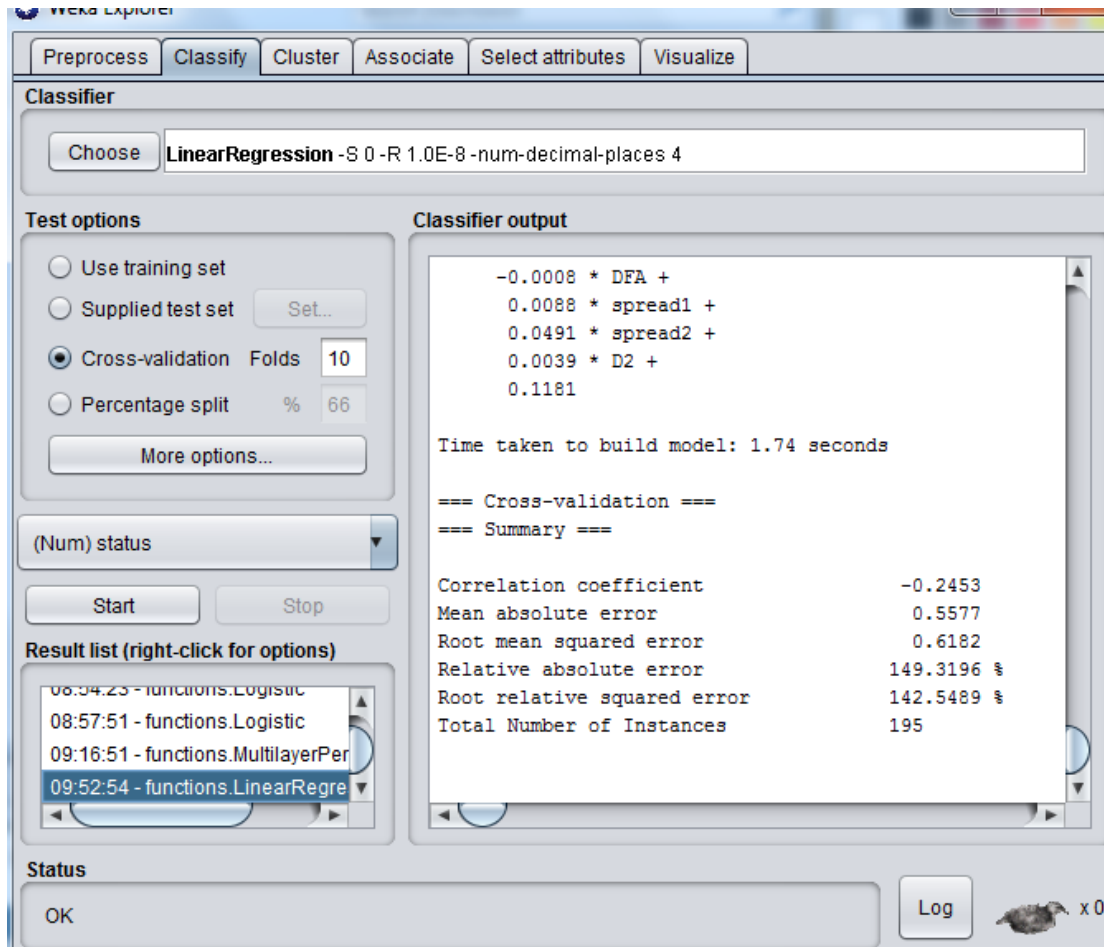
**Figure 2b.** Results using *Multi Layer Perceptron* with RMSE 0.46 result

\* Corresponding author



Figure 2 shown above is a testing process with the MLP algorithm. The epochs process for 10 x folding berlang sung is very long which is 166 seconds with a MAE result of 0.35 and RMSE of 0.46.

The second experiment used *Linear Regression* with the results looking like the following image:



**Figure 3.** Results using *Linear Regression* with RMSE 0.6182 result

The figure above is the result of testing the Parkinson's Disease dataset with *Linear Regression*. The test results were MAE values of 0.557 and RMSE of 0.618. Meanwhile, the time required to execute with the algorithm is 1.74 seconds.

Next is an experiment with the *Support Vector Machine (SVM)* reg approach in the Weka application. The results on testing with SVM can be seen as shown below:

\* Corresponding author



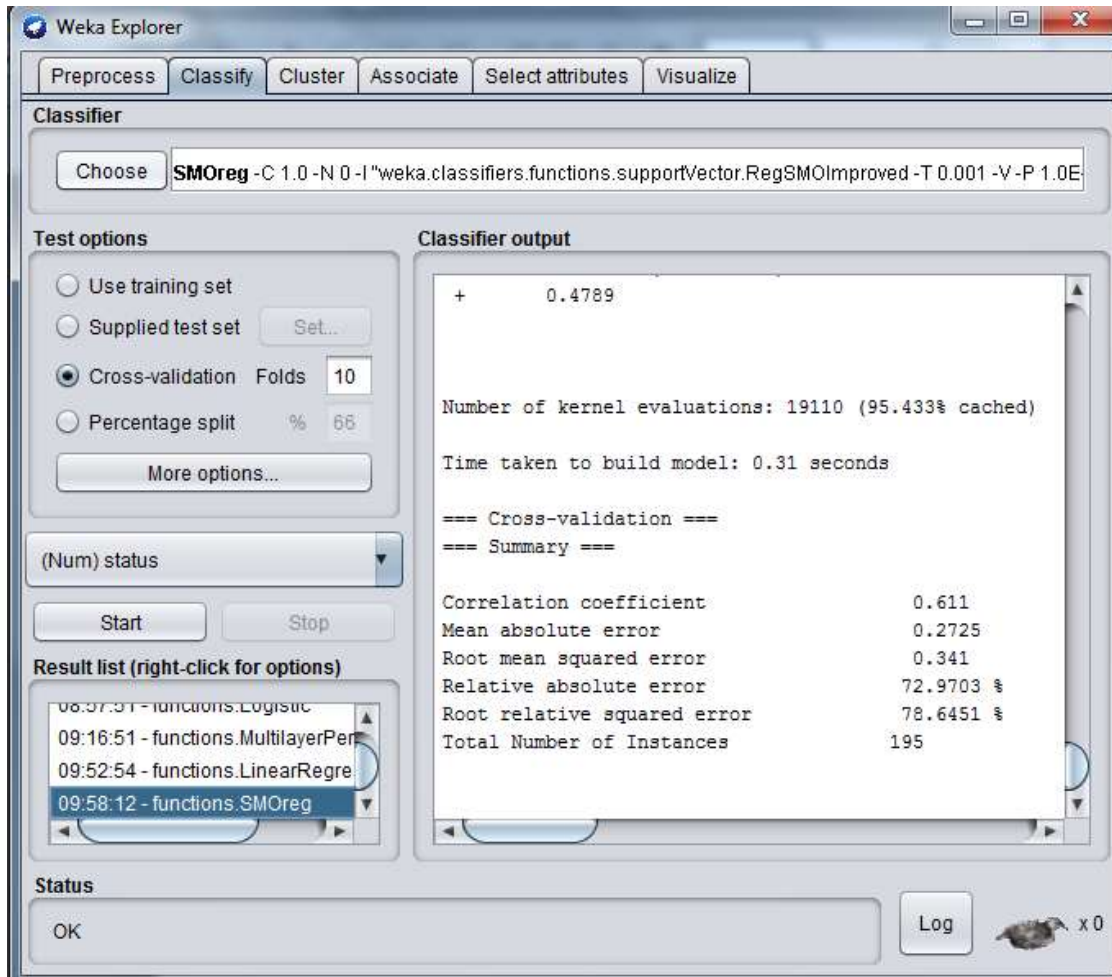


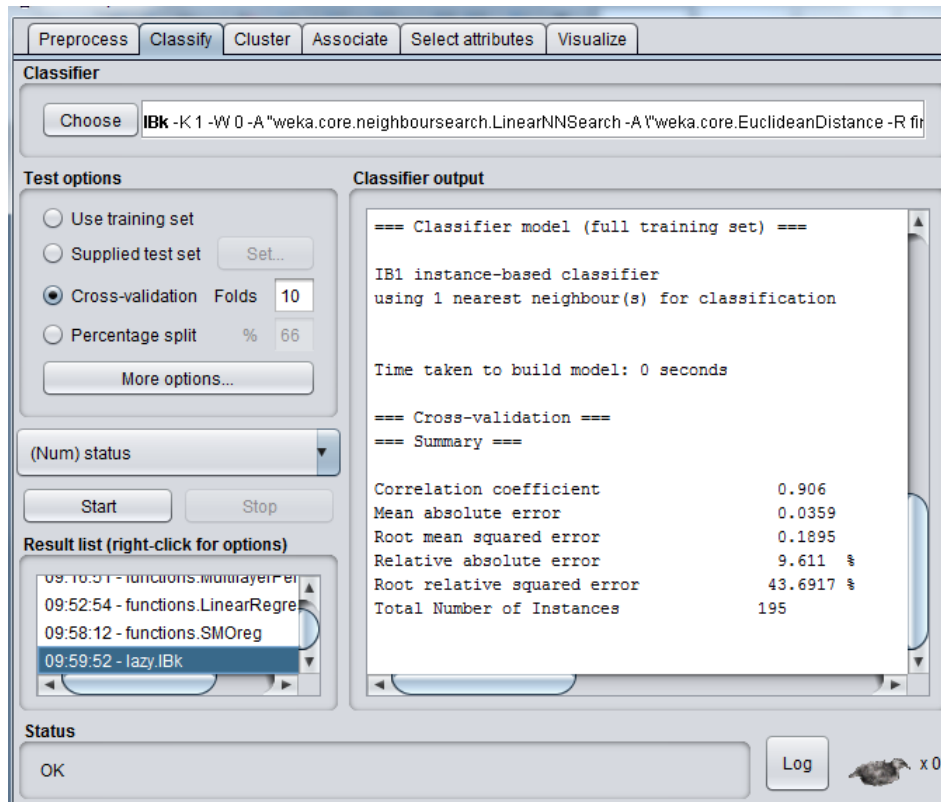
Figure 4. Results using SVM (for regression) with RMSE result 0.341

The above figure is an output output with the SVM algorithm for regression with MAE results of 0.27 and RMSE of 0.341. Meanwhile, the time required to complete the test is 0.31 seconds.

Furthermore, the fourth experiment using the K-Nearest Neighbor (KNN) algorithm. The results of this fourth experiment are very different than using the *Multi Layer Perceptron* (MLP) algorithm. The test results with the K-NN method can be seen as shown below:

\* Corresponding author





**Figure 5.** Results using k-NN with RMSE result 0.1895

The figure above describes the test results of the Parkinson Disease dataset with the K-NN algorithm. This method gives the best results for the error rate (Table 2). MAE results of 0.036 and RMSE 0.1895 were also obtained with the fastest computational time. Based on the test results using the four algorithms, namely MLP, SVM, Linear Regression and KNN, it can be concluded that the comparison of the results of each algorithm can be concluded as in the following table:

Table 2. Comparison of RESULTS

	MLP	LG	SVM (Regresi)	k-NN
MAE	0.35	0.557	0,27	<b>0,036</b>
RMSE	0.459	0,618	0,341	<b>0,1895</b>

### 5. CONCLUSION

The conclusion that can be drawn from this experiment is the poor performance of the *Multi Layer Perceptron* approach to the *Parkinson's Disease* data with an RSME value of 0.459. Meanwhile, the *k-Neural Network* Algorithm is a good classification technique for Parkinson's problem with an RMSE value of 0.1895. This is supported by previous research Zuo (2013) which used PSO optimization on advanced k-NN algorithms with a high degree of accuracy.

\* Corresponding author



## 6. REFERENCES

- Arrazaq, G. A., Magdalena, R., Atmaja, R. D., Prodi, S., Telecommunications, T., Elektro, F. T., & Telkom, U. (2017). Diagnosis of Parkinson's disease through analysis of walking patterns based on Vgrf using *wavelets and Support Vector Machine Diagnosis of Parkinson's Disease Through Gait Analysis based on vgrf using wavelet and support vector machine*. 4(2), 1855–1862.
- Azzahra, J. F., Sumarti, H., & Kusuma, H. H. (2022). Classification of COVID-19 and SARS Cases Based on Texture Traits Using the Multi-Layer Perceptron Method. *Journal of Physics*, 12(1), 16–27. <https://journal.unnes.ac.id/nju/index.php/jf/article/download/35685/13196>
- Bismi, W., & Harafani, H. (2022). *Comparison of Deep Learning Methods in Classifying MRI Scan Images of Brain Diseases*. 12(3), 177–185.
- Desinaini, L. N., Mualimah, A., Novitasari, D. C. R., & Hafiyusholeh, M. (2019). Fuzzy K-Nearest Neighbor (FKNN) application to detect Parkinson's disease. *InPrime: Indonesian Journal of Pure and Applied Mathematics*, 1(1), 8–16. <https://doi.org/10.15408/inprime.v1i1.12827>
- Handayani, P. K. (2021). Application of Support Vector Machine (SVM) algorithms for classification pattern analysis in Parkinson's Dataset. *Indonesian Journal of Technology, Informatics and Science (IJTIS)*, 3(1), 31–35. <https://doi.org/10.24176/ijtis.v3i1.7530>
- Haryadi, D., Marini Umi Atmaja, D., Rahman Hakim, A., & Suwaryo, N. (2021). Identify the risk level of stroke using multiple linear regression algorithms. *Deny Haryadi, SNTM*, 1(November), 1198–1207.
- Irnawati, O., & Nuraeni, N. (2022). Decision Tree Optimization with Particle Swarm Optimization for Gender Prediction Classification of E-Commerce Users. *Journal of Information Engineering and Educational Technology*, 6(1), 37–41. <https://doi.org/10.26740/jieet.v6n1.p37-41>
- Johanes, R. E., & Santoso, E. (2020). *Implementation of the K-Nearest Neighbor Algorithm for Disease Detection Classification in Dogs*. 4(5), 1580–1583.
- Mozhdehfarahbakhsh, A., Chitsazian, S., Chakrabarti, P., Chakrabarti, T., Kateb, B., & Nami, M. (2021). MRI-based Deep Learning Model for Predicting Stages of Parkinson's Disease. *MedRxiv*, February, 2021.02.19.21252081. <https://doi.org/10.1101/2021.02.19.21252081>
- Oktariza, Y., Amalia, L., Sobaryati, S., & Kurniawati, M. Y. (2019). Evaluation of the quality of life of Parkinson's patients based on levodopa-based therapy. *Indonesian Journal of Clinical Pharmacy*, 8(4), 246. <https://doi.org/10.15416/ijcp.2019.8.4.246>
- Prakoso, R. D. Y., Wiriaatmadja, B. S., & Wibowo, F. W. (2020). Classification System in Parkinson's Disease Using the K-Nearest Neighbor Method. *National Seminar on Computer Technology & Science (SAINTEKS)*, 2016, 63–68.
- Sumarlinda, S., & Lestari, W. (2022). K-Nearest Neighbor (KNN) Application for Cardiovascular Disease Classification. *Proceedings of the National Seminar on Technology ...*, 55, 259–262. <http://ojs.udb.ac.id/index.php/Senatib/article/download/1897/1487>
- Tempola, F., Muhammad, M., & Khairan, A. (2018). Comparison of Classifications Between KNN and Naive Bayes on Determination of Volcano Status with K-Fold Cross Validation. *Journal of Information Technology and Computer Science*, 5(5), 577. <https://doi.org/10.25126/jtiik.201855983>
- Utami, A. N. (2020). Classification of REM Behaviour Disorder Sleep Disorders Based on EEG Signals using Machine Learning. *Journal of Intelligent Systems*, 03(02), 216–230.
- Zuriel, H. P. P., & Fahrurozi, A. (2021). Implementation of the Support Vector Machine Classification Algorithm For Twitter User Sentiment Analysis Of Psbb Policy. *Scientific Journal of Computer Informatics*, 26(2), 149–162. <https://doi.org/10.35760/ik.2021.v26i2.4289>

\* Corresponding author



This is an Creative Commons License This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International (CC BY-SA 4.0).