The 2nd International Conference on Multidisciplinary Engineering (2nd ICOMDEN 2024)

The 2nd International Conference on Multidisciplinary Engineering: 00056 (2024) https://doi.org/10.29103/icomden.v2.xxxx
Volume 2, 2024
eISSN: xxxx-xxxx
Research Original Article

APPLICATION OF DATA MINING USING THE NEURAL NETWORK BACKPROPAGATION METHOD TO DETERMINE THE ELIGIBILITY OF SMART INDONESIA PROGRAM SCHOLARSHIP RECIPIENTS

Oriza Sativa ✉[1] **Bustami**[2] **Yesy Afrillia**[3]
[1] Department of Informatics Engineering, Malikussaleh University, Bukit Indah, Lhokseumawe, 24353, Indonesia, oriza.200170194@mhs.unimal.ac.id
[2] Department of Informatics Engineering, Malikussaleh University, Bukit Indah, Lhokseumawe, 24353, Indonesia
[3] Department of Informatics Engineering, Malikussaleh University, Bukit Indah, Lhokseumawe, 24353, Indonesia, yesy.afrillia@unimal.ac.id
✉ Corresponding Author: oriza.200170194@mhs.unimal.ac.id | **Phone: +6285763057730**

Abstract
The government provides students from impoverished or vulnerable backgrounds with financial assistance, educational opportunities, and expanded access through this scholarship program. Underprivileged students at SD Negeri 04 Lembah Melintang are still selected manually by each homeroom instructor through the collection of student and student parent data. The dataset utilized is comprised of 407 data points, including 326 training data and 81 test data, collected from pupils at SD Negeri 04 Lembah Melintang between 2022 and 2024. The objective of this research is to develop, execute, and evaluate the Neural Network Backpropagation method for the classification of PIP scholarship eligibility determination. The following attributes are included in this study: the status of the father, the status of the mother, the income of the father and the income of the mother, the job of the father and the job of the mother, distance from home, number of dependents, and means of transportation, with the classification results Eligible and Ineligible. This research produces an accuracy rate of 95%, with Recall 90%, Precision 100% and F1-Score of 94%.

Keywords: Data Mining, Classification, Neural Network Backpropagation

Introduction

PIP Scholarship Program Provides financial assistance, expands reach, and educational opportunities from the government to students from underprivileged families. In selecting underprivileged students at SD Negeri 04 Lembah Melintang, manual selection is still carried out by checking student data and parent data by each homeroom teacher. Because the large number of student data causes the results of determining eligibility to be inaccurate, an information system is used with One approach to obtaining information from data is through data mining so that knowledge is obtained which later this knowledge will be very valuable to be used as a basis for decision making [1].

Data mining techniques using the *neural network backpropagation method* are one of the effective solutions in processing and analyzing data. An information system is a set of procedures designed to process data in such a way as to produce valuable information that can be used by all individuals to make the right decisions [2]. By developing a web-based information system that implements the *neural network backpropagation* method , it is hoped that it can speed up the selection process, increase accuracy, and minimize errors in determining students who are considered eligible and ineligible to receive the PIP scholarship.

The use of the same method is also applied in the classification of loan customers [3], book lending [4], with quite satisfactory results. In this study, the author uses the *neural network backpropagation method* in classifying the determination of PIP scholarship eligibility. Classification is made to predict categorical labels, such as "safe" or "risky" for loan applications, "yes" or "no" for marketing data, or "treatment A", "treatment B", and "treatment C" for medical data, in a classification model in data mining. These categories can be presented with values that suit their needs. In cases where the arrangement of these values does not have a particular meaning [5].

The backpropagation algorithm is classified as supervised learning during the learning process. The training strategy using this algorithm can achieve a balance between the network's capacity to identify patterns used during the training

process. Several examples of signatures will be given for the learning or training process through this backpropagation artificial neural network. Furthermore, this ANN will be able to carry out the recognition process after receiving signature input for testing [6].

This research is expected to contribute to the development of a decision support system in the selection of PIP scholarship recipients and increase the accuracy and speed of the selection process, so that this scholarship program can be more effective and efficient.

Literature Review

Data Mining

Data mining is a subset of big data, which has become one of the most widely used languages in the industry. Big data is an abstract concept. which is usually defined as a combination of very complex and large data that is difficult to process using conventional database management tools. Big data is often identified by five main attributes: accuracy, value, diversity, velocity, and quantity [7].

Scholarship

A scholarship is a provision of financial assistance intended to be used to advance a person's education [8]. Every citizen has the right to obtain and participate in basic education and the government is obliged to finance it, as stated in Article 31 of the 1945 Constitution. And in Presidential Instruction Number 7 of 2014, among other things, mandates the Smart Indonesia Program (PIP) to the Ministry of Education and Culture to prepare the Smart Indonesia Card (KIP) and distribute Smart Indonesia Program (PIP) funds to students whose parents are unable to finance their education.

System

A system is a group of individuals who work together according to systematic and structured rules to form a unit that performs a function to achieve a certain goal. The system is characterized by various properties, including system components, system output, system processing, and system targets [9].

The author attaches several previous studies such as: The use of Backpropagation artificial neural networks in MSME classification provides the highest accuracy of up to 98.4294% [10].

Backpropagation neural network parameters affects the accuracy of the results. Retinal fundus image identification can be applied by utilizing backpropagation neural networks as a classification method according to the specified target with an accuracy value reaching 95% [11].

The conjugate gradient algorithm-based backpropagation neural network is able to accurately classify the condition of houses as habitable or uninhabitable, as evidenced by the average classification accuracy of 98.96% during the training process and 97.58% during the analysis process, based on this model [12].

Materials & Methods

Time and Place of Research

This research was conducted by the author by taking the research location at SD Negeri 04 Lembah Melintang, which is planned to start in March 2024 until completion. This research was conducted directly, where the data was obtained from DAPODIK provided by the administrative staff of SD Negeri 04 Lembah Melintang.

System Requirements Analysis

The system requirements analysis stage aims to truly understand the needs of the system to be designed and built. The results of this analysis will be a reference in designing the system to be built.

3.4.1 Hardware Requirements Analysis

Hardware is a device that is very necessary in a computer system, the hardware used in this study is the Asus VivoBook 15 A516JAO, with the following specifications:

a) Intel(R) Core(TM) i3-1005G1
b) 8 GB RAM
c) 512GB SSD

Software Requirements Analysis

Software is also one of the supporting factors for creating and designing a system to be built. The software used in designing this system is:

a. Operating System: Microsoft Windows 11, 64 bit
b. Supporting Applications: Visual Studio Code, Laragon, Google Chrome

Input Needs Analysis

The data inputted to perform processing for the classification of determining PIP scholarship recipients. The data used in this study is data obtained directly from SD Negeri 04 Lembah Melintang.

Process Needs Analysis

The process is carried out by the system after receiving input data from the user. The data is processed using the neural network method with the backpropagation algorithm to obtain results based on certain rules.

Output Needs Analysis

The system will issue the classification output results from the PIP scholarship recipient data that are included in the eligible or ineligible category using the calculation procedure from the neural network method with the backpropagation algorithm so that it is expected to be able to help related agencies.

Research Stages

The diagram below shows the research stages:

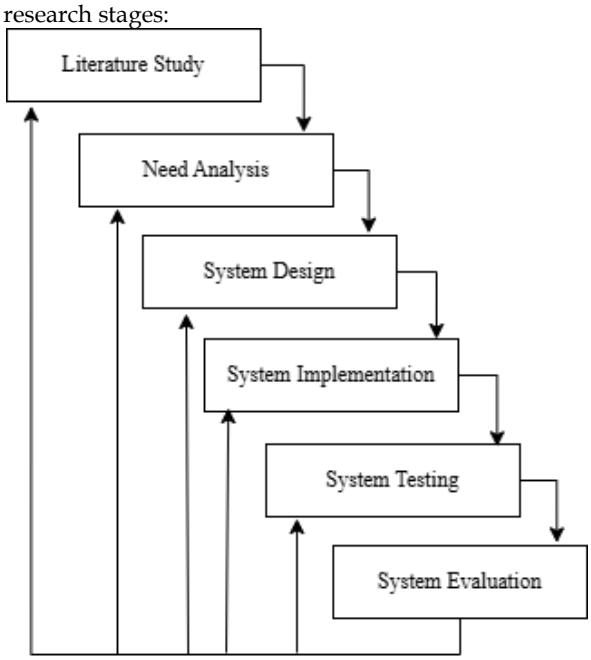

Figure 1. Research Stage

This research procedure was conducted using the waterfall method. Initially, the literature review was conducted through a cascade method system evaluation, which consists of several stages, including literature review and needs analysis, system design, system implementation, system initiation, and system evaluation.

System Scheme

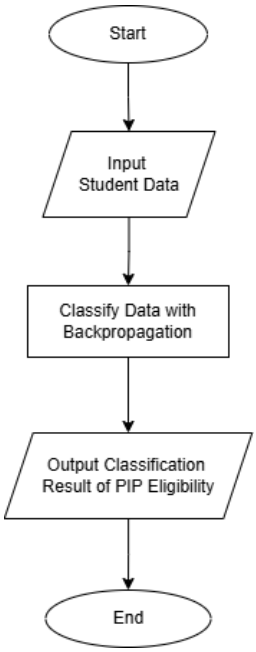

Figure 2. System Schematic

The system scheme starts from student data input, then the data will be processed by the system and perform calculations with *backpropagation* to produce classification results. After the classification is complete, the system will

display the output results of the classification, namely whether or not students are eligible to receive PIP.

Backpropagation Method Scheme

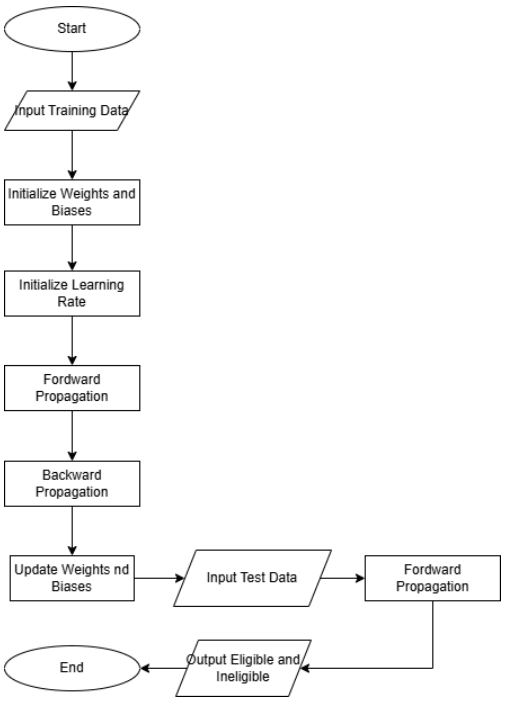

Figure 3. Backpropagation Method Scheme

Begin by inputting training data into the neural network to measure the model's performance, initializing weights a nd biases, and setting the learning rate for the training. Follow this with forward propagation, the process of passing inp ut through the network layers to generate predictions, and backward propagation, which calculates the error gradient an d updates the weights and biases accordingly. Adjust weights and biases to optimize the model's prediction accuracy. Th en, input test data to evaluate the previously trained model with new weights and biases using forward propagation aga in to generate outputs of eligible and ineligible. Done.

Results and Discussion

Research Result

In this study, the author will implement the *backpropagation method* for the classification of eligibility of PIP scholarship recipients at SDN 04 Lembah Melintang. The classified data will later be used as knowledge for SDN 04 Lembah Melintang. This system uses the *python programming language.* The main component of the system is data input, the inputted data is student data that contains several criteria that are used as objects of assessment, namely: father's status, mother's social condition, father's career, father's income, mother's career, mother's income, distance from home, total number of dependents, and means of transportation. The output of this system is the classification of scholarship recipient eligibility, which is classified as eligible or unfit.

Research Discussion

Attribute Conversion:

Table 1. Attribute Conversion

Attribute	Conversion
Means of transportation	x1
Father's Status	x2
Father's occupation	x3
Father's Income	x4
Mother Status	x5

Mother's Job	x6
Mother's Income	x7
Distance from Home	x8
Number of Dependents	x9
Class	y

Each attribute is converted, the attributes consist of means of transportation converted into x1, father's status x2, father's job x3, father's income x4, mother's status x5, mother's job x6, mother's income x7, distance from home x8, number of dependents x9, *Class* is y.

Table 2. Transportation Attribute Value Conversion

Means of transportation	Value
Motorcycle	1
Other	2
Shuttle car/bus	3
Andong/bendi/sado/dokar/delman/becak	4
Bicycle	5
Walk	6

Table 3. Conversion of Father Status Attribute Values

Father's Status	Value
Life	1
Died	2

Table 4. Conversion of Father's Job Attribute Values

Father's occupation	Value
Civil Servants/TNI/Polri	1
Self-employed	2
Private sector employee	3
Wholesaler	4
Small Trader	5
Fisherman	6
Laborer	7
Farmer	8
Other	9
Doesn't work	10
Already dead	11

Table 5. Conversion of Father's Income Attribute Values

Father's Income	Value
5,000,000 - 20,000,000	1
2,000,000 - 4,999,999	2
1,000,000 - 1,999,999	3
500,000 - 999,999	4
Less than 500,000	5
No Income	6

Table 6. Conversion of Mother Status Attribute Values

Mother Status	Value
Life	1

Died	2

Table 7. Conversion of Mother's Job Attribute Values

Mother's Job	Value
Civil Servants/TNI/Polri	1
Self-employed	2
Private sector employee	3
Businessman	4
Small Trader	5
Laborer	6
Farmer	7
Other	8
Doesn't work	9
Already dead	10

Table 8. Conversion of Mother's Income Attribute Values

Mother's Income	Value
2,000,000 - 4,000,000	1
1,000,000 - 1,999,999	2
500,000 - 999,999	3
Less than 500,000	4
No Income	5

Table 9. Conversion of Distance from Home Attribute Values

Distance from Home (Km)	Value
0	1
1	2
2	3
3	4
4	5
5	6

Table 10. Number of Dependents

Number of Dependents
1
2
3
4
5
6
7
8
9

Backpropagation Calculation

Dataset

Table 11. Dataset

No	Name	x1	x2	x3	x4	x5	x6	x7	x8	x9	y

1	Adinda Novita	1	1	8	3	1	2	3	2	1	0
2	Aditya Rahman	1	1	6	2	1	8	3	2	2	0
3	Clarrisa Salsabila Anggiat	1	1	2	3	1	2	3	2	1	0
....											
407	Zul Afnan	5	1	7	3	1	6	3	2	5	1

Step 0: setting weights and biases, and setting the learning rate

w1 = 0.032 w2 = 0.002 w3 = -0.02

w4 = -0.006 w5 = -0.003 w6= -0.008

w7 = -0.006 w8 = 0.003 w9 = 0.023

Bias = 0 Learning rate = 0.01

Step 1: determine the min and max values of the training data

Table 12. Min and Max Values of Data Attributes

	x1	x2	x3	x4	x5	x6	x7	x8	x9
Min	1	1	1	1	1	1	1	1	1
Max	6	2	11	6	2	10	5	6	8

Step 2: normalization of training data

$$Normalization = \frac{original\ data - min}{\max - min} \quad (1)$$

Table 13. Normalization of Training Data

No	Name	x1	x2	x3	x4	x5	x6	x7	x8	x9
1	Adinda Novita	0.0	0.0	0.7	0.4	0.0	0.1	0.5	0.2	0.0
2	Aditya Rahman	0.0	0.0	0.5	0.2	0.0	0.8	0.5	0.2	0.1
3	Clarrisa Salsabila Anggiat	0.0	0.0	0.1	0.4	0.0	0.1	0.5	0.2	0.0
....										
326	Aulia Rahman	1.0	0.0	0.7	0.6	0.0	0.9	1.0	0.2	0.6

Step 3: feedforward phase

1. Calculate all outputs in hidden units layers

$$\sum = b + x1.w1 + x2.w2 + \cdots + x_n.w_n \quad (2)$$

Data 1 = 0 + (0 * 0.032) + (0 * 0.002) + (0.7 * -0.02) + (0.4 * -0.006) + (0* -0.003) + (0.11 * -0.008) + (0.5 * -0.006) + (0.2 * 0.003) + (0 * 0.0023)
= -0.0197

2. Calculating output at *output layer* using binary step activation function (hard limit)

$$y(o) = \begin{cases} 0, if\ o \leq 0 \\ 1, if\ o > 0 \end{cases} \quad (3)$$

Table 14. Output Layer Calculation Results

	Σ	o
1	-0.020	0
2	-0.017	0
3	-0.008	0
....		

326	0.018	1

Step 4: calculate the error in hidden layers

$$e = y - o \qquad (4)$$

Table 15. Hidden Layer Error Calculation Results

	y	o	error
1	0	0	0
2	0	0	0
3	0	0	0
....			
326	1	1	0

Step 5: weight phase update

1. Updating new weights

$$w_{new} = w_{old} + e.X.\eta (5)$$

Data 1
w1 = 0.032 + (0 * 0 * 0.01)
= 0.032
w2 = 0.002 + (0 * 0 * 0.01)
= 0.002
w3 = -0.02 + (0 * 0.7 * 0.01)
= -0.02
w4 = -0.006 + (0 * 0.4 * 0.01)
= -0.006
w5 = -0.003 + (0 * 0 * 0.01)
= -0.003
w6 = -0.008 + (0 * 0.11 * 0.01)
= -0.008
w7 = -0.006 + (0 * 0.5 * 0.01)
= -0.006
w8 = 0.003 + (0 * 0.2 * 0.01)
= 0.003
w9 = 0.023 + (0 * 0 * 0.01)
= 0.023

2. Updating new bias

$$b_{new} = b_{old} + e.\eta \quad (6)$$

Data 1 (b) = 0 + (0 * 0.01)
= 0
Data 2 (b) = 0 + (0 * 0.01)
= 0
Data 3 (b) = 0 + (0 * 0.01)
= 0
Data 4 (b) = 0 + (0 * 0.01)
= 0

New Weights and Biases

w1 = 0.066
w2 = -0.028
w3 = -0.027
w4 = -0.03
w5 = -0.023
w6 = 0.055
w7 = 0.048
w8 = 0.005
w9 = 0.027
Bias = -0.07

Test Data Classification Results

Table 16. Test Data Classification Results

Data	Σ	o	Status
1	-0.0522	0	Ineligible
2	-0.0315	0	Ineligible
3	-0.0305	0	Ineligible
4	0.0689	1	Eligible
5	-0.0436	0	Ineligible
6	-0.0304	0	Ineligible
7	-0.0567	0	Ineligible
8	0.0931	1	Eligible
9	0.0669	1	Eligible
10	-0.0093	0	Ineligible
11	-0.0132	0	Ineligible

12	0.0730	1	Eligible
13	-0.0632	0	Ineligible
14	-0.0408	0	Ineligible
15	0.0792	1	Eligible
16	0.0689	1	Eligible
17	0.0328	1	Eligible
18	-0.0638	0	Ineligible
19	-0.0611	0	Ineligible
20	-0.0329	0	Ineligible
21	-0.0147	0	Ineligible
22	0.0735	1	Eligible
23	0.0669	1	Eligible
24	-0.0567	0	Ineligible
25	0.0689	1	Eligible
26	-0.0341	0	Ineligible
27	-0.0611	0	Ineligible
28	-0.0632	0	Ineligible
29	-0.0258	0	Ineligible
30	-0.0466	0	Ineligible
31	-0.0424	0	Ineligible
32	-0.0857	0	Ineligible
33	0.0700	1	Eligible
34	-0.0586	0	Ineligible
35	-0.0746	0	Ineligible
36	0.0751	1	Eligible
37	0.0604	1	Eligible
38	-0.0776	0	Ineligible
39	0.0444	1	Eligible
40	-0.0563	0	Ineligible
41	-0.0607	0	Ineligible
42	-0.0164	0	Ineligible
43	-0.1148	0	Ineligible
44	-0.0593	0	Ineligible
45	0.0730	1	Eligible
46	0.0270	1	Eligible
47	-0.0607	0	Ineligible
48	-0.0261	0	Ineligible
49	0.0398	1	Eligible
50	0.0751	1	Eligible
51	0.0336	1	Eligible
52	-0.0787	0	Ineligible
53	0.0101	1	Eligible
54	0.0541	1	Eligible
55	-0.0752	0	Ineligible
56	-0.0304	0	Ineligible
57	0.0398	1	Eligible
58	0.0398	1	Eligible
59	0.0371	1	Eligible
60	0.0828	1	Eligible

61	0.0689	1	Eligible
62	-0.0789	0	Ineligible
63	-0.0625	0	Ineligible
64	-0.0810	0	Ineligible
65	0.0316	1	Eligible
66	0.0669	1	Eligible
67	0.0869	1	Eligible
68	-0.0653	0	Ineligible
69	0.0710	1	Eligible
70	0.0648	1	Eligible
71	-0.0004	0	Ineligible
72	0.0378	1	Eligible
73	0.0751	1	Eligible
74	-0.0651	0	Ineligible
75	0.0689	1	Eligible
76	-0.0589	0	Ineligible
77	0.0378	1	Eligible
78	-0.0810	0	Ineligible
79	0.0689	1	Eligible
80	-0.0571	0	Ineligible
81	0.0193	1	Eligible

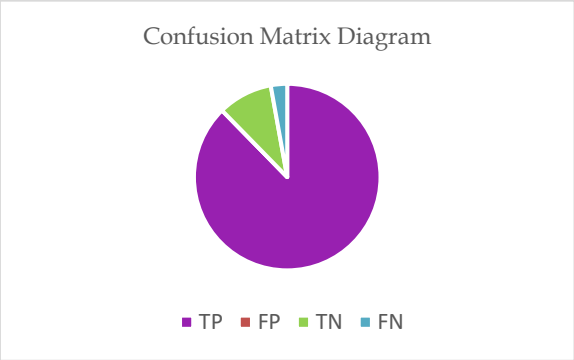

Figure 4. Confusion Matrix Diagram

Table 17. Confusion Matrix

n = 81	Current Eligible (1)	Current Ineligible (0)
Prediction Eligible (1)	TP = 37	FP = 0
Prediction Ineligible (0)	FN = 4	TN = 40
	41	40

Information :
TP: The model predicts that students are eligible, and indeed students are eligible.
TN: The model predicted that the student was ineligible and it was true that the student was ineligible.
FP: The model predicts that students are eligible, but it turns out that students are ineligible.
FN : The model predicted that the student was ineligible, but it turned out that the student was eligible.

Table 18. Evaluation Results

Accuracy	= (TP+TN)/(TP+FP+FN+TN) = (37+40) / (37+0+4+40) = 0.95061	Recall	= (TP)/ TP+FN) = 41 / (41+0) = 0,9024

	= 0,95*100% = 95%		=0,90*100% =90%
Precision	=(TP)/ (TP+FP) = 37/ (37+0) = 1 = 1*100% = 100%	F1-Score	=2*(Precision*Recall)/(Precision+Recall) =2*(1*0,90)/ (1+0,90) = 0,94 = 0,94*100% = 94%

Conclusions

A web-based system using the Python programming language capable of classifying the eligibility of recipients of the Program Indonesia Pintar (PIP) scholarship using the Neural Network Backpropagation method has been successfully designed, with classification labels of Eligible and Ineligible. The implementation of the Neural Network Backpropagation method in this classification system produced values of TP = 37, FP = 0, TN = 40, and FN = 4. This study indicates that using the Neural Network Backpropagation method for determining scholarship eligibility yielded excellent evaluation results with an accuracy of 95%, recall of 90%, precision of 100%, and an F1-score of 94%. It can be concluded that this classification system can be applied to classify the eligibility of PIP scholarship recipients.

Acknowledgments

The authors wish to thank Ms. Yesy Afrillia and Mr. Bustami for their valuable guidance and advice during this research. We also extend our gratitude to Malikussaleh University and SD Negeri 04 Lembah Melintang for their support in providing research facilities..

References

[1] A. Razi, "Klasifikasi Penerima Beasiswa Aceh Carong (Aceh Pintar) Di Universitas Malikussaleh Menggunakan Algoritma Knn (K-Nearest Neighbors)," Jurnal Tika, Vol. 7, No. 1, Pp. 79–84, Apr. 2022, Doi: https://doi.org/10.51179/tika.v7i1.1116.

[2] Y. Afrillia And S. Ramadani, "Sistem Informasi Pencarian Tata Letak Buku Pada Perpustakaan Menggunakan Metode Binary Search," Jurnal Informatika Kaputama (Jik), Vol. 6, No. 1, Pp. 31–35, Jan. 2022, Doi: https://doi.org/10.59697/jik.v6i1.132.

[3] N. Hadianto, H. B. Novitasari, And A. Rahmawati, "Klasifikasi Peminjaman Nasabah Bank Menggunakan Metode Neural Network," Jurnal Pilar Nusa Mandiri, Vol. 15, No. 2, Pp. 163–170, Sep. 2019, Doi: https://doi.org/10.33480/pilar.v15i2.658.

[4] N. Norhikmah And R. Rumini, "Klasifikasi Peminjaman Buku Menggunakan Neural Network Backpropagation," Sistemasi, Vol. 9, No. 1, P. 1, Jan. 2020, Doi: https://doi.org/10.32520/stmsi.v9i1.562.

[5] S. S. M. S. Retno Tri Vulandari, Data Mining: Teori Dan Aplikasi Rapidminer. Penerbit Gaya Media, 2017. [Online]. Available: https://books.google.co.id/books?id=1EDZzwEACAAJ

[6] B. C. Octariadi, "Pengenalan Pola Tanda Tangan Menggunakan Metode Jaringan Syaraf Tiruan Backpropagation," Jurnal Teknoinfo, Vol. 14, No. 1, P. 15, Jan. 2020, Doi: https://doi.org/10.33365/jti.v14i1.462 .

[7] I. Hidayat, E. Darnila, And Y. Afrillia, "Clustering Zonasi Daerah Rawan Bencana Alam Di Kabupaten Mandailing Natal Menggunakan Algoritma K-Means," G-Tech: Jurnal Teknologi Terapan, Vol. 7, No. 3, Pp. 1218–1226, Jul. 2023, Doi: https://doi.org/10.33379/gtech.v7i3.2880.

[8] N. Aulia, N. Suarna, And W. Prihartono, "Klasifikasi Penentuan Penerima Program Indnesia Pintar Di Krwilbidikcam Greged Menggunakan Algoritma C4.5," Jati (Jurnal Mahasiswa Teknik Informatika), Vol. 7, No. 6, Pp. 3913–3919, Feb. 2024, Doi: https://doi.org/10.36040/jati.v7i6.8294.

[9] E. Y. Anggraeni, E. Risanto, Y. Basuki, D. Nofianto, A. A. C, And A. Offset, Pengantar Sistem Informasi. Penerbit Andi, 2017. [Online]. Available: https://books.google.co.id/books?id=8VNLDwAAQBAJ

[10] T. Hardoyo And E. H. P. Eko, "Klasifikasi Usaha Mikro Kecil Menengah Menggunakan Jaringan Syaraf Tiruan Backpropagation," Konstelasi: Konvergensi Teknologi Dan Sistem Informasi, Vol. 2, No. 1, Apr. 2022, Doi: https://doi.org/10.24002/konstelasi.v2i1.5625.

[11] M. F. Syahputra Et Al., "Hypertensive Retinopathy Identification Through Retinal Fundus Image Using Backpropagation Neural Network," J Phys Conf Ser, Vol. 978, P. 012106, Mar. 2018, Doi: https://doi.org/10.1088/1742-6596/978/1/012106.

[12] J. R. Prabowo, R. Santoso, And H. Yasin, "Implementasi Jaringan Syaraf Tiruan Backpropagation Dengan Algoritma Conjugate Gradient Untuk Klasifikasi Kondisi Rumah (Studi Kasus Di Kabupaten Cilacap Tahun 2018)," Jurnal Gaussian, Vol. 9, No. 1, Pp. 41–49, Feb. 2020, Doi: https://doi.org/10.14710/j.gauss.v9i1.27522.