Kumpulan jurnaL Ilmu Komputer (KLIK)
Volume 09 No. 03 2022
ISSN: 2406-7857

BACKPROPAGATION NEURAL NETWORK TO CLASSIFY SUITABILITY BETWEEN ALUMNI'S OCCUPATION AND STUDY PROGRAM

Jaka Permadi, Winda Aprianti, Herfia Rhomadhona
Politeknik Negeri Tanah Laut
Jl. A. Yani Km.06, Ds. Panggung, Pelaihari
jakapermadi.88@politala.ac.id, winda@politala.ac.id, herfia.rhomadhona@politala.ac.id

Abstract

One of the main tasks of Point Career Center (PCC) is to record the suitability of alumni's work with the study program. PCC also provides information about job vacancies. If PCC can predict the suitability of jobs that alumni will get based on academic data and non-academic data, it will help PCC to take policies in an effort to increase the percentage of job suitability of alumni. This research was used BPNN to train dataset with 5 attributes, namely Grade Point Average, high school background, and competency certificate ownership. Non-academic data are parents's occupation. BPNN is applied to 70:30, 75:25, 80:20, and 90:10 ratio with several learning rates and several hidden units. The results of this research are accuracy, precision, and recall in all scenarios is above 70%, and the best performance is the ratio of 80:20 and 90:10 with accuracy = 83.33%, precision = 87.50%, and recall = 83.33%. That indicates BPNN is suitable to classify suitability between alumni's occupation and study program.

***Keywords**: Accuracy, BPNN, Classify, Precision, and Recall*

1. PENDAHULUAN

Point Career Center (PCC) is a career development center at Tanah Laut State Polytechnic. PCC has been tracking studies from 3 (three) departments including Informatics Engineering, Agro-industry and Automotive Engineering since 2012. PCC manages alumni data with admin in each study program. Tracking of tracer studies is carried out regularly every year. Currently, the available tracer study reports are the satisfaction of the industrial and business worlds with the ability of the alumni who work there. In addition, tracer studies also provide information about the suitability of alumni with their study program. This information was obtained from alumni who PCC admins regularly contact. Of course, this is less effective, because it affects the accuracy of the information obtained and will affect the performance of the PCC if the alumni's work is not in accordance with their study program.

Therefore, in order to overcome this problem, this research will classify alumni work with study programs. The dataset in this research is collected from alumni and a unit on campus that functions to manage alumni data called Point Career Center. The dataset consist of 67 data with target class according to the study program and data not according to the study program. The dataset has academic data and non-academic data of alumni. Academic data are Grade Point Average, high school background, and competency certificate ownership. Non-academic data is a parent's occupation. The method used for the classification is the Backpropagation Neural Network (BPNN).

One of the researchers who studied the alumni case was Asroni et al. The research was conducted to predict the waiting period for alumni to get a job using the Naïve Bayes Algorithm [1]. Dataset used is 689 data with 4 (four) attributes, namely gender, architecture, GPA, and year. The accuracy rate generated by the Naïve Bayes Algorithm is 71%. Siahaan and Kardian also conducted a tracer analysis of alumni of the Department of Information Systems and Computer Systems at Gunadarma University in 2013 using the K-Means method [2]. The dataset used is 300 alumni data with 2 (two) attributes, namely field of study and current job. The results of the study formed 2 (two) clusters, namely C1 with 5 alumni jobs and C3 with 3 alumni jobs. In addition, Rezkika et al also conducted a study to predict the waiting time for alumni to get a job using the C45 method [3]. The dataset used is 462 student data with 19 attributes. The test was carried out in 9 test scenarios using k-fold cross-validation with 80.37% accuracy, 79.56% precision, 81.34% recall, 80.44% f-measure, and 0.914 AUC [3].

Some researchers who use neural network algorithms to solve alumni problems such as Sibagariang conduct research on alumni, namely predicting alumni's job prospects using the LVQ method [4]. LVQ method is a neural network algorithm. Sibagariang stated that the accuracy of the LVQ method was 71.25% with 320 lines of training data, 80 lines of test data, 0.8 learning rate, 100 epochs, and 0.1 decreases in learning. Besides LVQ, BPNN is one of the Neural Network methods with three layers, namely the input layer, hidden layer, and output layer, so it is widely applied for pattern detection and classification in the form of sound, text, and images [5]. As Rahmadtulloh et al used BPNN to predict the response of steel frames. The resulting accuracy is 99% with MSE 0.0004 [6]. In line with Pebrianasari et al used BPNN to analyze the Pekalongan batik motif with 5 (five) batik motifs analyzed, where the accuracy of the batik kawung buketan motif was 88.07%, batik burung phoenix motif 87.32%, batik enchim motif 85.68%, batik jawa motif is 90.73% and batik jlamprang motif is 91.31% [7]. The BPNN algorithm can also be used to predict student study length with 4 (four) variables, namely Social Studies semesters 1-4. The resulting accuracy is 99% [8].

Research conducted by Permadi et al compared the K-Nearest Neighbor (KNN) and Backpropagation Neural Network (BPNN) methods to predict the risk of early-stage diabetes [9]. The dataset used in this study is 520 data with 16 attributes. The accuracy result for BPNN is 90% while KNN is 83.75%. Purnama et al stated that BPNN is better than SVM with an accuracy of 83.33% and 83.00% for the analysis of waiting time for alumni to get a job [5]. Purnama stated that the difference in accuracy was not too significant, only 0.33%, while the precision of both methods yielded 100%. Prasetyo and Sari classified the alumni of SMKN 1 Glagah 2018 using the Perceptron Algorithm [10]. This study uses 7 (seven) attributes, namely age, hobbies, place of residence, laptop ownership, industry, education, work, or college. The weighting is repeated so that in the 4th epoch the lowest error is 20%.

2. METODOLOGI PENELITIAN

The stages of this research are shown in Figure 1 and these stages are described further in the next subsection.

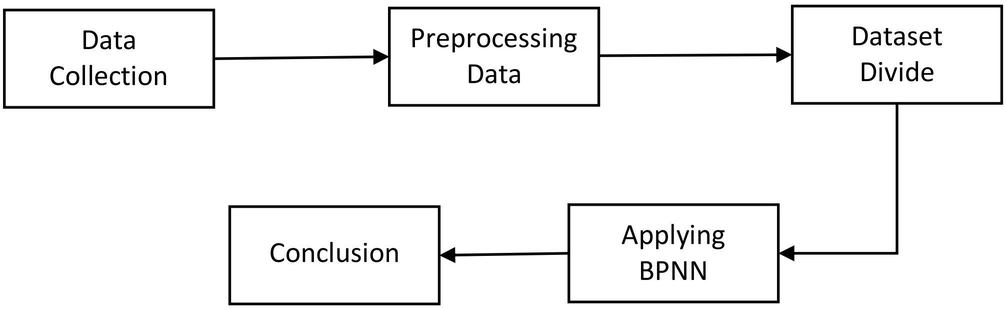

Figure 1. Research Stages

2.1. Data Collection

Dataset in this research is collected from alumni and Point Career Center. The dataset consists of 67 data with target class according to the study program and data not according to the study program. The dataset has academic data and non-academic data of alumni. Academic data are Grade Point Average, high school background, and competency certificate ownership. Non-academic data are parents' occupations. The dataset consists of 67 data which have 5 (five) attributes and 1 (one) target class. The target Class is determined by Point Career Center, which is shown in Table 1.

Table 1. Class of Dataset

Categorical	Numeric
According to the study program	42
Not according to the study program	25

2.2. Preprocessing

The first step in the preprocessing stage is to check the equality of the number between the two classes to ensure that the data held is balanced. If data is imbalanced, there needs to be a balance of data to get the optimal result of applying BPNN. In addition to data balancing, transformation is also required at this stage. High school background, competency certificate ownership, and parental occupation are attributes with categorical data, so these attributes need to be transformed into numerical data. The next step is the normalization of each attribute with min-max normalization.

2.3. Dataset Divide

Dataset divide is the stages of dividing the dataset into training and testing data with several scenarios, namely 70:30, 75:25, 80:20, and 90:10.

2.4. Applying BPNN

In this stage, BPNN will be applied to datasets with various learning rates and various hidden units. The stages of training with BPNN are presented in Figure 2.

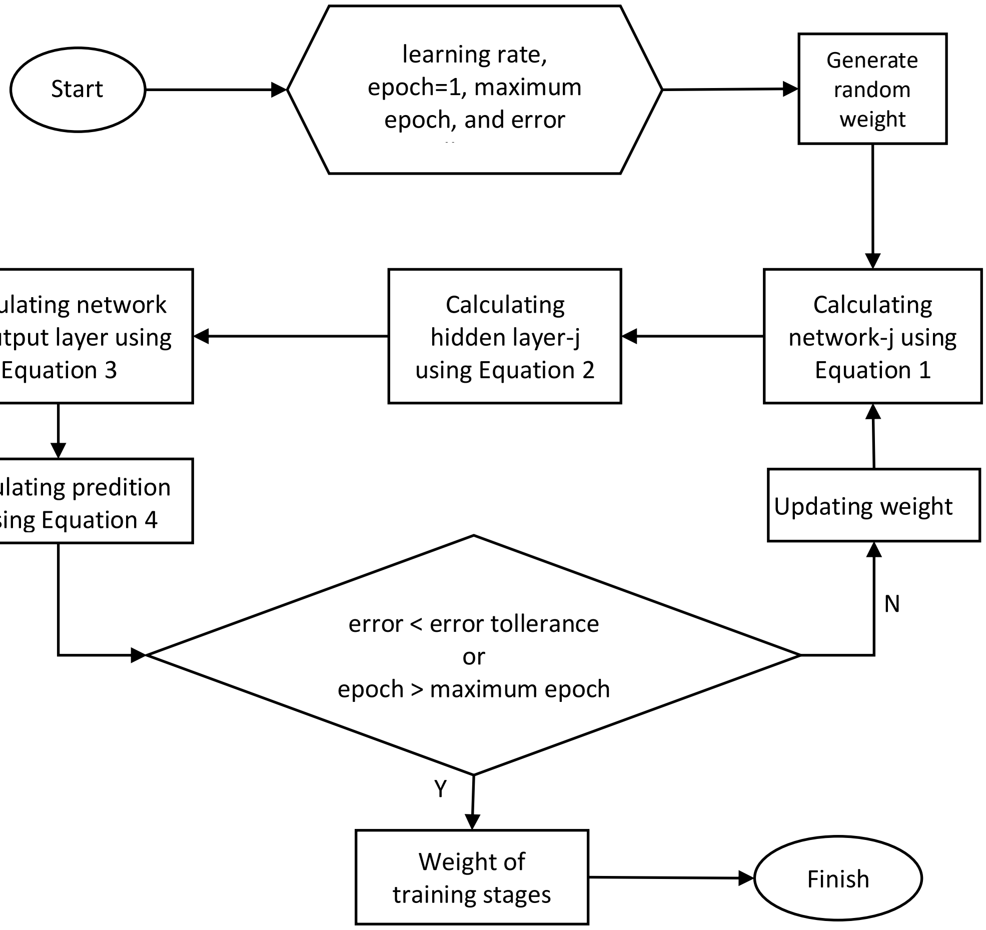

Figure 2. Training Stages

The equations mentioned in Figure 2 are as follows.

$$net_j = \sum_{i=1}^{m} x_i v_{ij} + v_{0j} \quad (1)$$

$$h_j = f(net_j) = \frac{1}{1+e^{-net_j}} \quad (2)$$

$$net_{out} = \sum_{j=1}^{n} h_j w_j + w_0 \quad (3)$$

$$y = f(net_{out}) = \frac{1}{1+e^{-net_{out}}} \quad (4)$$

where net_j is network-*j*, x_i is input unit-*i*, v_{ij} is the weight of network between the input layer and hidden layer, h_j is hidden unit-*j*, net_{out} is network to the output layer, and w_j is the weight between hidden layer and output layer.

Then the weights obtained from the training stages will be used to predict the class of testing data. The prediction result will give a true positive (TP), false positive (FP), true negative (TN), or false negative (FN) value. This value will be used to calculate the values of accuracy, precision, and recall which are presented respectively in Equation 5, Equation 6, and Equation 7.

Accuracy = ((TP + TN) / (TP + FP + FN + TN)) X 100% (5)
Precision = (TP / (TP + FP)) x 100% (6)
Recall = (TP / (TP + FN)) x 100% (7)

2.5 Conclusion

Based on the result of subsection 2.4, conclusion about how good performance of BPNN on a dataset will be gotten.

3. RESULT AND DISCUSSION

Based on class data in Table 1, there is a different number of classes that are categorized according to the study program and classes that are categorized as not according to the study program. So that dataset is balanced to 25 data for each class. The next step is a transformation for categorical attributes.

As for the target class, the class categorized according to the study program was changed to 1 and the class categorized as not according to the study program was changed to 0. Whereas 4 (four) of 5 (five) attributes will be transformed into numeric attributes using Table 2 for high school background, Table 3 for competency certificate ownership, Table 4 for the father's occupation, and Table 5 for the mother's occupation.

Table 2. Transformation for High School Background

Categorical	Numeric
SMK	1
SMA / MAN	2

Table 3. Transformation for Competency Certificate Ownership

Categorical	Numeric
No	0
Yes	1

Table 4. Transformation for Father's Occupation

Categorical	Numeric
Employee	1
Entrepreneur	2
Unemployee	3

Table 5. Transformation for Mother's Occupation

Categorical	Numeric
Employee	1
Unemployee	2
Entrepreneur	3

Then the dataset is normalized using min-max normalization. To test the ability of BPNN, the dataset will be divided into several scenarios, BPNN is applied to the training data as shown in Figure 2, then the accuracy, precision, and recall values are calculated respectively using Equation 5, Equation 6, and Equation 7. The result of these steps is shown in Table 6 until Table 14. The result of applying BPNN to 70:30 ratio at various learning rates is shown in Table 6.

Table 6. The Result of Applying BPNN to 70:30 with Various Learning Rate

Learning Rate	Hidden Node	Accuracy	Precision	Recall
0.1	11	64.29	67.50	64.29
0.2	11	50.00	50.00	50.00
0.3	11	78.57	85.00	78.57
0.4	11	57.14	57.78	57.14
0.5	11	64.29	67.50	64.29
0.6	11	64.29	67.50	64.29
0.7	11	71.43	73.33	71.43
0.8	11	64.29	67.50	64.29
0.9	11	71.43	81.82	71.43

Based on Table 6, the highest value of accuracy, precision, and recall of BPNN is using learning rate 0.3 so then BPNN is applied with learning rate 0.3 with various hidden node numbers as shown in Table 7.

Table 7. The Result of Applying BPNN to 70:30 with Learning Rate = 0.3

Learning Rate	Hidden Node	Accuracy	Precision	Recall
0.3	3	64.29	79.17	64.29
0.3	4	57.14	76.92	57.14
0.3	5	71.43	81.82	71.43
0.3	6	50.00	50.00	50.00
0.3	7	57.14	57.78	57.14
0.3	8	57.14	57.78	57.14
0.3	9	64.29	67.50	64.29
0.3	10	71.43	81.82	71.43

Based on Table 6 and Table 7, the best performance BPNN to 70:30 ratio is BPNN with learning rate = 0.3 and hidden node = 11. The result of applying BPNN to 75:25 ratio at various learning rates is shown in Table 8.

Table 8. The Result of Applying BPNN to 75:25 with Various Learning Rate

Learning Rate	Hidden Node	Accuracy	Precision	Recall
0.1	11	50.00	50.00	50.00
0.2	11	58.33	58.57	58.33
0.3	11	58.33	58.57	58.33
0.4	11	41.67	41.43	41.67
0.5	11	66.67	68.75	66.67
0.6	11	75.00	75.71	75.00
0.7	11	41.67	41.43	41.67
0.8	11	33.33	31.25	33.33
0.9	11	33.33	33.33	33.33

Based on Table 8, the highest value of accuracy, precision, and recall of BPNN is using learning rate 0.6 so then BPNN is applied with learning rate 0.6 with various hidden node numbers as shown in Table 9.

Table 9. The Result of Applying BPNN to 75:25 with Learning Rate = 0.6

Learning Rate	Hidden Node	Accuracy	Precision	Recall
0.6	3	41.67	41.43	41.67
0.6	4	66.67	66.67	66.67
0.6	5	50.00	50.00	50.00
0.6	6	58.33	58.57	58.33
0.6	7	33.33	33.33	33.33
0.6	8	50.00	50.00	50.00
0.6	9	58.33	58.57	58.33
0.6	10	58.33	58.57	58.33

Based on Table 8 and Table 9, the best performance BPNN to 75:25 ratio is BPNN with learning rate = 0.6 and hidden node = 11. The result of applying BPNN to 80:20 ratio at various learning rates is shown in Table 10.

Table 10. The Result of Applying BPNN to 80:20 with Various Learning Rate

Learning Rate	Hidden Node	Accuracy	Precision	Recall
0.1	0	50.00	50.00	50.00
0.2	0	83.33	87.50	83.33
0.3	0	50.00	50.00	50.00
0.4	0	50.00	50.00	50.00
0.5	0	50.00	50.00	50.00
0.6	0	50.00	50.00	50.00
0.7	0	50.00	50.00	50.00
0.8	0	50.00	50.00	50.00
0.9	0	66.67	80.00	66.67

Based on Table 10, the highest value of accuracy, precision, and recall of BPNN is using learning rate 0.2 so then BPNN is applied with learning rate 0.2 with various hidden node numbers as shown in Table 11.

Table 11. The Result of Applying BPNN to 80:20 with Learning Rate = 0.2

Learning Rate	Hidden Node	Accuracy	Precision	Recall
0.2	3	50.00	50.00	50.00
0.2	4	83.33	87.50	83.33
0.2	5	50.00	50.00	50.00
0.2	6	50.00	50.00	50.00
0.2	7	66.67	80.00	66.67
0.2	8	66.67	80.00	66.67
0.2	9	50.00	50.00	50.00
0.2	10	50.00	50.00	50.00

Based on Table 10 and Table 11, the best performance BPNN to 80:20 ratio is BPNN with learning rate = 0.2 and hidden node = 4. The result of applying BPNN to 90:10 ratio at various learning rates is shown in Table 12.

Table 12. The Result of Applying BPNN to 90:10 with Various Learning Rate

Learning Rate	Hidden Node	Accuracy	Precision	Recall
0.1	11	83.33	87.50	83.33
0.2	11	66.67	66.67	66.67
0.3	11	66.67	66.67	66.67
0.4	11	66.67	66.67	66.67
0.5	11	66.67	66.67	66.67
0.6	11	66.67	66.67	66.67
0.7	11	66.67	66.67	66.67
0.8	11	83.33	87.50	83.33
0.9	11	83.33	87.50	83.33

Based on Table 12, the highest value of accuracy, precision, dan recall of BPNN is using learning rate 0.1, 0.8, and 0.9 so then BPNN is applied with learning rate 0.1, 0.8, and 0.9 with various hidden node numbers respectively as shown in Table 13, Table 14, and Table 15.

Table 13. The Result of Applying BPNN to 90:10 with Learning Rate = 0.1

Learning Rate	Hidden Node	Accuracy	Precision	Recall
0.1	3	83.33	87.50	83.33
0.1	4	83.33	87.50	83.33
0.1	5	83.33	87.50	83.33
0.1	6	66.67	66.67	66.67
0.1	7	83.33	87.50	83.33
0.1	8	83.33	87.50	83.33
0.1	9	83.33	87.50	83.33
0.1	10	66.67	80.00	66.67

Table 14. The Result of Applying BPNN to 90:10 with Learning Rate = 0.8

Learning Rate	Hidden Node	Accuracy	Precision	Recall
0.8	3	83.33	87.50	83.33
0.8	4	66.67	66.67	66.67
0.8	5	83.33	87.50	83.33
0.8	6	66.67	66.67	66.67
0.8	7	83.33	87.50	83.33
0.8	8	83.33	87.50	83.33
0.8	9	83.33	87.50	83.33
0.8	10	66.67	66.67	66.67

Table 15. The Result of Applying BPNN to 90:10 with Learning Rate = 0.9

Learning Rate	Hidden Node	Accuracy	Precision	Recall
0.9	3	66.67	66.67	66.67
0.9	4	66.67	66.67	66.67
0.9	5	83.33	87.50	83.33
0.9	6	83.33	87.50	83.33

0.9	7	83.33	87.50	83.33
0.9	8	66.67	66.67	66.67
0.9	9	83.33	87.50	83.33
0.9	10	66.67	66.67	66.67

Based on Table 12, Table 13, Table 14, and Table 15, the best performance BPNN to 90:10 ratio is BPNN with learning rate 0.1, 0.8, and 0.9.

Based on Table 2 until Table 15, the best performance BPNN in various ratios, namely 70:30, 75:25, 80:20, and 90:10 as shown in Figure 3.

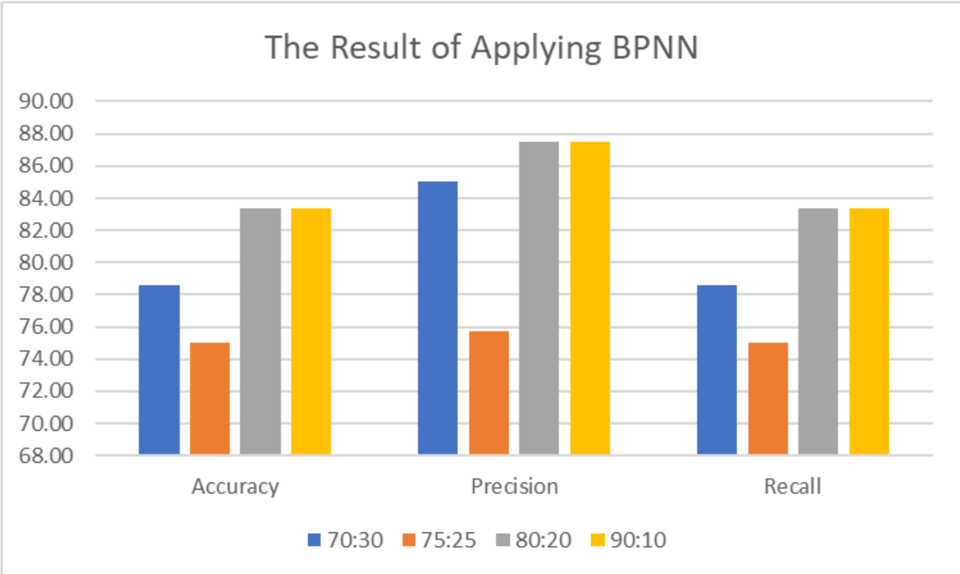

Figure 3. The Result of Applying BPNN

Based on Figure 3, the best performance is applying BPNN to 80:20 and 90:10 ratio with accuracy = 83.33%, precision = 87.50%, and recall = 83.33%. Figure 3 also shows that the performance of applying BPNN to 75:25 is lower than the performance of applying BPNN to 70:30. This is because the determination of the training set is random, so the training data contained in 70:30 is different from 75:25. Suggestions for future research are that a k-fold cross validation method is needed to get the best training set arrangement with the highest testing.

4. CONCLUSION

5 (five) attributes were used in this study, namely the Grade Point Average (GPA), high school background, ownership of competency certificates, and parental occupation. BPNN is applied at 70:30, 75:25, 80:20, and 90:10 with several learning rates and several hidden units. The results of this study are the values of accuracy, precision, and recall are above 70% for all ratio scenarios. The best performance is the ratio of 80:20 and 90:10 with accuracy = 83.33%, precision = 87.50%, and recall = 83.33%. This shows that BPNN can be applied to help PCC classify the suitability between alumni's work and study programs.

ACKNOWLEDGEMENT
This research is a part of a project financed by DIPA Direktorat Akademik Pendidikan Tinggi Vokasi Direktorat Jenderal Pendidikan Vokasi, Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi in Penelitian Dosen Pemula scheme (114/SPK/D4/PPK.01.APTV/VI/2022) fiscal year 2022.

DAFTAR PUSTAKA
[1] A. Asroni, N. M. Ali, and S. Riyadi, "Perkiraan Masa Tunggu Alumni Mendapatkan Pekerjaan Menggunakan Metode Prediksi Data Mining Dengan Algoritma Naive Bayes Classifier," *Semesta Tek.*, vol. 21, no. 2, pp. 189–197, 2018.
[2] V. Berton and A. R. Kardian, "Penerapan Algoritma K-Means Untuk Analisis Tracer Alumni Universitas Gunadarma Jurusan Sistem Informasi dan Sistem Komputer Angkatan 2013," *J. Ilm. KOMPUTASI*, vol. 18, no. 3, pp. 215–228, 2019.
[3] F. Rezkika, B. N. Sari, and A. S. Y. Irawan, "Klasifikasi Masa Tunggu Alumni Untuk Mendapatkan Pekerjaan Berdasarkan Kompetensi Menggunakan Algoritma C4. 5 (Studi Kasus: Fasilkom Unsika)," *Progresif J. Ilm. Komput.*, vol. 17, no. 2, pp. 95–106, 2021.
[4] S. Sibagariang, A. Riyadi, A. Dzikri, F. Suandi, K. T. Sirait, and F. Setiawan, "Prediksi Prospek Kerja Alumni Dengan Algoritma Neural Network," *CESS (Journal Comput. Eng. Syst. Sci.*, vol. 6, no. 1, pp. 91–96, 2021.
[5] D. I. Purnama, R. L. Islami, L. Sari, and P. R. Sihombing, "Analisis Klasifikasi Data Tracer Study Dengan Support Vector Machine Dan Neural Network," *Anal. Klasifikasi Data Tracer Study Dengan Support Vector Mach. Dan Neural Netw.*, vol. 4, no. 2, pp. 46–52, 2021.
[6] A. Widhiantoyo, B. N. Sari, and D. Yusuf, "Penerapan Algoritma Naïve Bayes Dengan Backward Elimination Untuk Prediksi Waktu Tunggu Alumni Mendapatkan Pekerjaan," *JIKO (Jurnal Inform. dan Komputer)*, vol. 4, no. 3, pp. 145–151, 2021.
[7] V. Pebrianasari, E. Mulyanto, and E. Dolphina, "Analisis Pengenalan Motif Batik Pekalongan Menggunakan Algoritma Backpropagation," *Techno. Com*, vol. 14, no. 4, pp. 281–290, 2015.
[8] D. Kartini, "Penerapan Data Mining dengan Algoritma Neural Network (Backpropagation) Untuk Prediksi Lama Studi Mahasiswa," *Pros. SISFOTEK*, vol. 1, no. 1, pp. 235–241, 2017.
[9] J. Permadi, H. Rhomadhona, and W. Aprianti, "Perbandingan K-Nearest Neighbor Dan Backpropagation Neural Network Dalam Prediksi Resiko Diabetes Tahap Awal," *KLIK-Kumpulan J. Ilmu Komput.*, vol. 8, no. 3, pp. 352–365, 2021.
[10] A. Prasetyo and D. A. L. Sari, "Implementasi Jaringan Syaraf Tiruan Model Perceptron Untuk Klasifikasi Karir Alumni SMKN 1 Glagah 2018," *J. ZETROEM*, vol. 4, no. 1, pp. 8–12, 2022.