

# Dimensional Data Unsupervised Learning Using an Analytic Hierarchy Process in Determining Attributes in the Classification Algorithm

Shinta Ayunda Putri<sup>1,\*</sup>, Mustakim<sup>2</sup>

Fakultas Sains dan Teknologi, Sistem Informasi, Universitas Islam Negeri Sultan Syarif Kasim Riau, Pekanbaru, Indonesia

Email: <sup>1</sup>shintaikhtiar@gmail.com, <sup>2</sup>mustakim@uin-suska.ac.id

Email Penulis Korespondensi: shintaikhtiar@gmail.com

Submitted: 23/06/2022; Accepted: 28/06/2022; Published: 30/06/2022

**Abstrak**—Penanganan sistematis diperlukan dalam meningkatkan mutu perguruan tinggi, salah satunya yaitu perlu ditingkatkannya kompetensi lulusan setiap tahunnya. Dalam meningkatkan kelulusan Mahasiswa, maka perlu dilakukannya klasifikasi kelulusan Mahasiswa untuk mengetahui apakah Mahasiswa tersebut dikatakan Tepat Waktu (TW) atau Kemungkinan Tepat Waktu (KTW) dengan menggunakan metode BPNN dan PNN. Data yang digunakan yaitu data Alumni Mahasiswa Sistem Informasi angkatan 2013-2020 dengan 7 kriteria yang digunakan yaitu IPK, Total SKS, Jumlah Mata Kuliah Mengulang, Ambil Mata Kuliah TA disemester 7, Prokrastinasi, Kepercayaan Diri, dan Disiplin. Data yang diperoleh selanjutnya dilakukan proses pembagian data *training* dan data *testing* menggunakan K-Means Clustering dengan tujuan untuk mendapatkan hasil akurasi terbaik. Selanjutnya tahap pengklasifikasian menggunakan BPNN dan PNN menghasilkan hasil akurasi sebesar 98% dan 95,45% dengan *learning rate* 0,125 dan nilai *spread* 0,1.

**Kata Kunci:** BPNN; PNN; K-Means Clustering; dan Klasifikasi Mahasiswa.

**Abstract**—Systematic keyword is needed in improving the quality of higher education, one of which is the needed to increase the competence of graduates every year. In increasing student graduation, it is necessary to classify student graduation to find out whether the student is said to be on time (TW) or possibility on time (KTW) using the BPNN and PNN methods. The data used is the Alumni data of the 2013-2020 Information System study program with 7 criteria use, namely GPA, Total Credits, Number of Repetitive Courses, Taking TA Course in Semester 7, Procrastination, Self-Confidence, and Discipline. The data obtained is then carried out in the process of sharing training data and testing data using K-Means Clustering with the aim; of getting the best accuracy results. Furthermore, the classification stage using BPNN and PNN resulted in an accuracy of 98% and 95% with learning rate of 0.125 and a spread value of 0.1.

**Keywords:** BPNN; PNN; K-Means Clustering; Classification

## 1. PENDAHULUAN

Pendidikan menjadi bagian penting yang mempengaruhi kehidupan manusia. Menurut Undang-Undang Nomor 20 Tahun 2003, Perguruan tinggi turut berkewajiban menyelenggarakan pendidikan, penelitian serta pengabdian masyarakat. Perguruan tinggi dan program studi memiliki hubungan timbal balik yang erat. Kementerian menyebutkan bahwa diperlukan suatu perencanaan penanganan sistematis dalam meningkatkan mutu perguruan tinggi dan program studi, salah satunya yaitu perlu ditingkatkan kompetensi lulusan setiap tahunnya[1]. Pada Program Sistem Informasi tercatat masih banyaknya Mahasiswa belum menyelesaikan perkuliahan lebih dari empat tahun. Melalui hal tersebut penting bagi Manajemen/Administrator Program Studi untuk mencari cara dalam meningkatkan kelulusan Mahasiswa dengan mengetahui faktor apa saja yang mempengaruhi masa belajar Mahasiswa[2]. Penggunaan data mining dapat digunakan sebagai pertimbangan dalam pengambilan keputusan lebih lanjut[3].

*Data mining* adalah proses pencarian hubungan dengan menganalisis pola-pola pengetahuan yang sebelumnya tidak diketahui atau tersembunyi dari sekumpulan data besar dalam database melalui statistik matematika dan mesin belajar[4]. Dalam memperoleh keberhasilan hasil estimasi, penerapan teknik *clustering* dilakukan untuk mengelompokkan dataset menjadi dua bagian yaitu *training* dan *testing* yang dapat mewakili sebaran data[5].

Sebagai salah satu metode dalam *data mining* yang bersifat *unsupervised*[6], *clustering* berperan penting dalam memberikan hasil pembagian data yang efektif untuk menentukan data *training* dalam melakukan klasifikasi. K-Means merupakan salah satu metode yang mempartisi data ke dalam bentuk satu kelompok atau lebih yang memiliki tingkat kemiripan yang tinggi. Penelitian yang dilakukan oleh[5] memperoleh hasil bahwa algoritma K-Means memiliki akurasi yang tinggi dibandingkan dengan K-Fold Cross Validation.

Untuk menunjang hasil estimasi digunakan algoritma Probabilistic Neural Network (PNN). Kelebihan yang didapatkan dari penerapan PNN yaitu pelatihan yang bersifat mudah, bobot tidak terlatih tetapi ditugaskan sehingga tidak akan berganti dan hanya vektor baru yang dimasukkan ke dalam matriks bobot ketika pelatihan [7]. Penelitian dengan studi kasus prediksi kelulusan tepat waktu Mahasiswa menggunakan metode Decision Tree dan Artificial Neural Network (ANN) menghasilkan tingkat akurasi tinggi sebesar 74,51% untuk metode Decision Tree dan akurasi sebesar 79,74% dihasilkan oleh metode ANN[8].

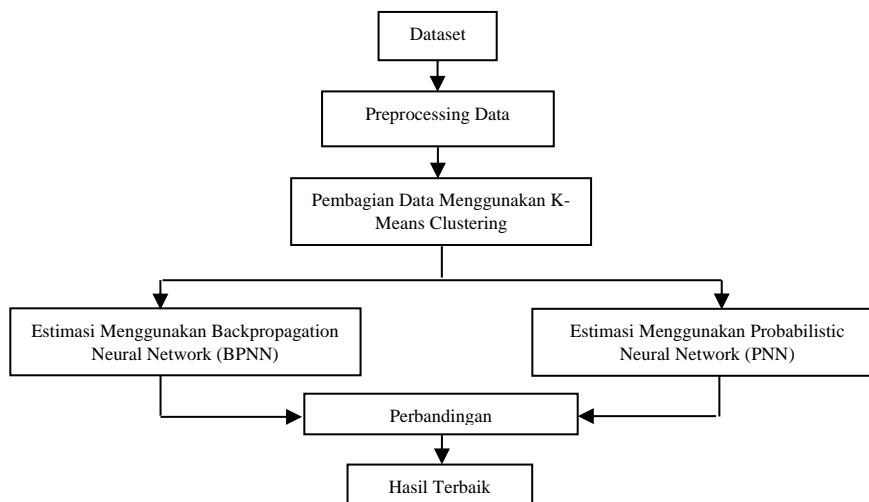
Penelitian terdahulu pernah dilakukan oleh Sari, 2018 mengenai estimasi kelulusan mahasiswa tepat waktu menggunakan tujuh kriteria/atribut yang berpengaruh terhadap kelulusan tepat waktu mahasiswa yaitu Indeks Prestasi Kumulatif (IPK), Total Satuan Kredit Semester (SKS), Ambil Mata Kuliah Tugas Akhir (TA) pada Semester 7, Prokrastinasi, Kepercayaan Diri, Jumlah Mata Kuliah Mengulang dan Disiplin dengan menggunakan algoritma Backpropagation Neural Network (BPNN). Dimana dalam penelitian [9]diperoleh hasil akurasi algoritma BPNN sebesar 87% dengan nilai *error* sebesar 0,3443[9]. Dari ketujuh kriteria yang digunakan, didapatkan 3 kriteria yang

berpengaruh terhadap kelulusan tepat waktu mahasiswa yaitu Prokrastinasi, Total SKS, dan Jumlah Mata Kuliah Mengulang[10]. Berbeda pada penelitian sebelumnya, pada penelitian ini tahap prediksi kelulusan mahasiswa menggunakan algoritma PNN. Hal ini dilakukan dengan tujuan untuk mengetahui algoritma mana yang memiliki tingkat akurasi tertinggi dari algoritma yang dikerjakan sebelumnya.

## 2. METODOLOGI PENELITIAN

### 2.1 Tahapan Penelitian

Pada penelitian ini, data yang digunakan merupakan data Lulusan Mahasiswa atau Alumni pada Prodi Sistem Informasi Tahun Ajaran 2013-2020. Atribut yang digunakan pada penelitian ini sebanyak 7 kriteria yang meliputi IPK, Total SKS, Jumlah Mata Kuliah Mengulang, Mengambil Mata Kuliah TA di Semester 7, Prokrastinasi, Kepercayaan Diri, dan Disiplin. Dataset awal yang terkumpul kemudian dilakukan tahap preprocessing meliputi *cleaning data*, *transformation data* dan *normalization data*. Sebelum dilakukan tahap klasifikasi, data terlebih dahulu dibagi menjadi data training dan data testing menggunakan K-Means Clustering. Selanjutnya proses mengestimasi menggunakan algoritma Backpropagation Neural Network dan Probabilistic Neural Network. Kemudian hasil dari tiap percobaan dibandingkan sehingga didapatkan hasil terbaik berdasarkan nilai uji evaluasi Root Mean Square Error (RMSE) terendah. Alur metodologi penelitian dapat dilihat pada Gambar 1.



Gambar 1. Metodologi Penelitian

#### 2.1.1 K-Means Clustering

K-Means Clustering merupakan salah satu algoritma tanpa pengawasan yang paling banyak digunakan untuk mengelompokkan data dalam sejumlah klaster tertentu[11]. Alur perhitungan pada algoritma k-means dimulai dengan menentukan jumlah cluster atau nilai k yang ingin dibentuk dan menentukan nilai *centroid* awal secara acak. Kemudian melakukan perhitungan jarak antar data ke masing-masing centroid dengan menggunakan persamaan Euclidean Distance dengan rumus:

$$DL_2(x_1 - x_2) = \|x_1 - x_2\|_2 = \sqrt{\sum_{j=1}^n (x_{2j} - x_{1j})^2} \quad (1)$$

Selanjutnya kelompokkan setiap data berdasarkan jarak terdekat atau dengan melihat nilai jarak terkecil antar data dengan *centroid*nya[12].

#### 2.1.2 Backpropagation Neural Network (BPNN)

Backpropagation merupakan salah satu contoh algoritma pembelajaran dalam jaringan syaraf tiruan[13] yang dapat digunakan dalam prediksi dan klasifikasi. Tahapan pada algoritma BPNN yaitu:

- a. Inisialisasi bobot secara acak untuk setiap sinapsis dalam jaringan saraf
- b. Ulangi untuk semua lapisan tersembunyi atau *hidden layer*
- c. Untuk semua *sample training* X lakukan:

1. *Forward Propagation*

Tiap unit *input* menerima sinyal dan meneruskan ke lapisan tersembunyi yang kemudian tiap lapisan tersembunyi menjumlahkan sinyal-sinyal input.

$$Z_{in_j} = V_{0j} + \sum_{i=1}^n X_i V_{ij} \quad (2)$$

Setelah itu sinyal dikirim ke semua *output* dengan rumus:

$$Z_i = f(Z_i n_j) \tag{3}$$

2. Backward Propagation

Menghitung nilai eror disetiap unit *output* dan lapisan tersembunyi menggunakan Persamaan 4 dan Persamaan 5.

$$\sigma_k = (t_k - y_k) f'(y_k) \tag{4}$$

$$\delta_i n_j = \sum_{k=1}^m \delta_k W_{jk} \tag{5}$$

Kemudian bobot pada setiap neuron akan diperbaharui dengan rumus:

$$W_{jk}(\text{baru}) = W_{jk}(\text{lama}) + \Delta W_{jk} \tag{6}$$

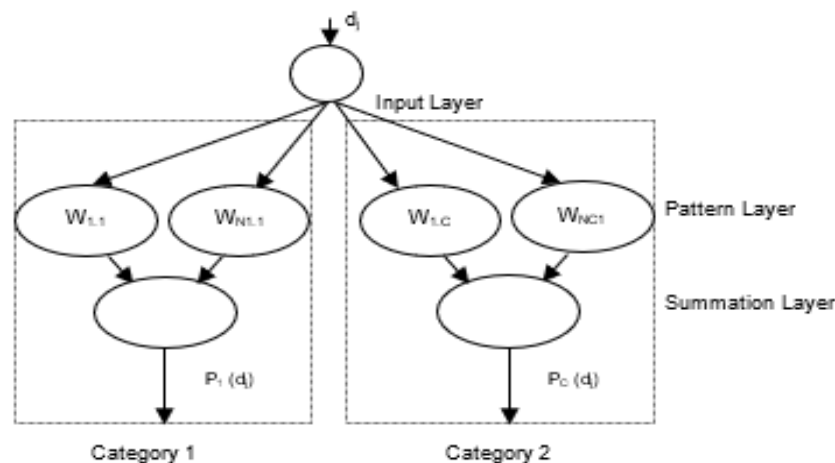
Yang mana nilai  $\Delta W_{jk}$  diperoleh dengan rumus:

$$\Delta W_{jk} = \alpha \delta_j X_i \tag{7}$$

$\alpha$  merupakan learning rate.

2.1.3 Probabilistic Neural Network (PNN)

Jaringan syaraf tiruan *probabilistic* didasarkan pada algoritma statistik yang diterapkan dalam jaringan Bayesian[14]. Algoritma ini termasuk dalam pembelajaran terawasi atau *supervised learning* karena *output* yang diharapkan telah diketahui sebelumnya serta pemodelan yang dibentuk berdasarkan penaksiran peluang yang akan muncul. Arsitektur PNN dapat dilihat pada Gambar 2.



Gambar 2. Arsitektur Algoritma PNN

Output pada algoritma ini akan menghasilkan klasifikasi berdasarkan nilai pada *summation layer* yang memiliki nilai paling besar yang dapat dilihat pada Persamaan 8.

$$P_d(x) = C_k \text{ jika } p(x|C_k) > p(x|C_i) Pr(C_j) \tag{8}$$

2.1.4 Validasi dan Pemodelan Terbaik

Tahapan ini dilakukan untuk mencari dan mengetahui keakuratan hasil peramalan dengan data histori dengan menggunakan Root Mean Square Error (RMSE). Hasil peramalan yang dilakukan akan semakin bagus jika nilai RMSE semakin kecil. Adapaun perhitungan RMSE dapat dilihat pada Persamaan 9.

$$RMSE = \sqrt{\frac{\sum (Y_t - \hat{Y}_t)^2}{n}} \tag{9}$$

3. HASIL DAN PEMBAHASAN

Pada penelitian ini, dataset yang digunakan berasal dari data Lulusan Mahasiswa atau Alumni pada Program Studi Sistem Informasi dari tahun 2013-2020 dengan total keseluruhan data awal sebanyak 210 data dengan penggunaan atribut atau kriteria sebanyak 7 kriteria yang meliputi IPK, Total SKS, Jumlah Mata Kuliah Mengulang, Mengambil Mata Kuliah TA disemester 7, Prokrastinasi, Kepercayaan diri, dan Disiplin. Tahap selanjutnya setelah mengumpulkan dataset adalah melakukan *preprocessing* data yang meliputi *cleaning data*, *transformation data*, dan *normalization data*. Dataset yang digunakan serta hasil *preprocessing* data dapat dilihat pada Tabel 1 dan Tabel 2.

**Tabel 1.** Data Lulusan Mahasiswa Sistem Informasi

No	Kode	IPK	Total SKS	Jumlah MK. Mengulang	Ambil MK. TA Smst.7	Prokrastinasi	Kepercayaan Diri	Disiplin
1	SI001	3	149	3	Tidak	Rendah	Sedang	Tinggi
2	SI002	3.12	140	3	Ya	Sedang	Sedang	Sedang
3	SI003	3.37	149	0	Ya	Rendah	Tinggi	Tinggi
4	SI004	3.3	144	0	Ya	Rendah	Tinggi	Tinggi
5	SI005	3.16	171	3	Tidak	Sedang	Sedang	Sedang
...	...	...	...	...	...	...	...	...
210	SI210	2.55	145	0	Ya	Sedang	Sedang	Sedang

**Tabel 2.** Hasil Preprocessing Data

No	Kode	IPK	Total SKS	Jumlah MK. Mengulang	Ambil MK. TA Smst.7	Prokrastinasi	Kepercayaan Diri	Disiplin
1	SI001	0,669	0,2692	0,333	0	0	0	1
2	SI002	0,718	0,2	0,333	1	0,5	0	0
3	SI003	0,819	0,269	0	1	0	1	1
4	SI004	0,790	0,231	0	1	0	1	1
5	SI005	0,734	0,438	0,333	0	0,5	0	0
...	...	...	...	...	...	...	...	...
145	SI145	0,480	0,238	0	1	0,5	0	0

### 3.1 Penentuan Kelas Target

Terdapat 2 target yang ditetapkan untuk mengklasifikasi kelulusan Mahasiswa yaitu, Tepat Waktu (TW) dan Kemungkinan Tepat Waktu (KTW). Untuk penetapan target merujuk pada penelitian yang dilakukan oleh [10] menggunakan metode Analytic Hierarchy Process dengan diambil 3 nilai *eigen* tertinggi sebagai kriteria yang paling berprioritas yaitu kriteria Prokrastinasi, Total SKS dan Jumlah Mata Kuliah Mengulang dengan nilai *eigen* berturut sebesar 0,301; 0,216; 0,170 yang selanjutnya dalam penentuan kelas target dikondisikan dari 3 kriteria berpengaruh tersebut. Penentuan kelas target dapat dilihat pada Tabel 3. Untuk memudahkan dalam menentukan kelas target, data yang telah di normalisasi terlebih dahulu di denormalisasi seluruhnya sehingga berbentuk data transformasi yang dapat dilihat pada Tabel 4.

**Tabel 3.** Penentuan Kelas Target

No	Jika	Maka
1	Prokrastinasi dalam kategori “Rendah”	
2	Total SKS > 136 pada semester 7	Tepat Waktu (TW)
3	Mata Kuliah Mengulang < 3	
4	Salah satu atau semua kondisi tidak terpenuhi	Kemungkinan Tepat Waktu (KTW)

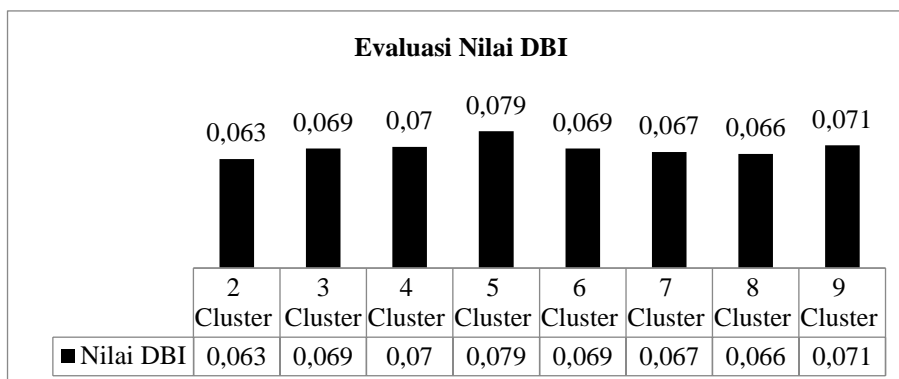
**Tabel 4.** Data Setelah Penetapan Target

No	Kode	IPK	Total SKS	Jumlah MK. Mengulang	Ambil MK. TA Smst.7	Prokrastinasi	Kepercayaan Diri	Disiplin	Kelas Target
1	SI001	3	149	3	Tidak	Rendah	Sedang	Tinggi	KTW
2	SI002	3.12	140	3	Ya	Sedang	Sedang	Sedang	KTW
3	SI003	3.37	149	0	Ya	Rendah	Tinggi	Tinggi	TW
4	SI004	3.3	144	0	Ya	Rendah	Tinggi	Tinggi	TW
5	SI005	3.16	171	3	Tidak	Sedang	Sedang	Sedang	KTW
...	...	...	...	...	...	...	...	...	...
145	SI145	2.53	145	0	Ya	Sedang	Sedang	Sedang	KTW

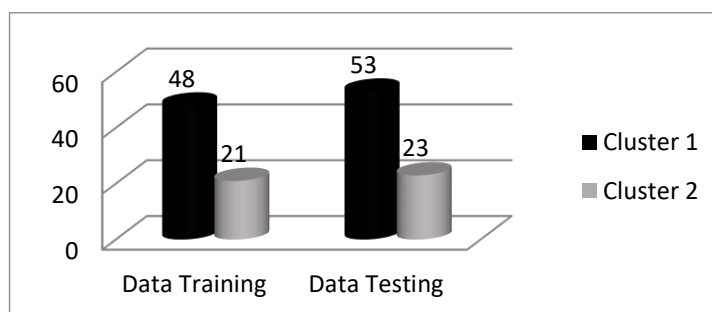
### 3.2 Pembagian Data Menggunakan K-Means Clustering

Data yang telah melalui tahap normalisasi selanjutnya dilakukan proses pembagian data menjadi data *training* dan data *testing* yang pada tahap ini menggunakan algoritma K-Means. Hasil percobaan pembagian data latih dan data uji menggunakan algoritma K-Means dapat dilihat pada Gambar 3. Dari Gambar 3. diketahui bahwa *cluster 2* atau  $k=2$

terpilih sebagai *cluster* terbaik dengan nilai DBI terendah sebesar 0,063. Pada *cluster* 1 dan *cluster* 2 secara berturut memiliki anggota sebanyak 69 *record* dan 76 *record* yang nantinya dilakukan pengelompokkan kembali dengan ketentuan 70%;30% data *training*;data *testing* disetiap *cluster* yang dapat dilihat pada Gambar 4.



Gambar 3. Grafik Evaluasi Nilai Davies Bouldin Index (DBI) K-Means Clustering



Gambar 4. Pembagian Data Training dan Data Testing

### 3.3 Pemodelan Menggunakan BPNN

Pada penelitian ini, learning rate yang diuji sebanyak 10 percobaan dengan rentang nilai 0,001; 0,0025; 0,005; 0,0075; 0,01; 0,025; 0,05; 0,075; 0,1; 0,15 dan dengan 1000 *epoch* atau perulangan maksimal. Hasil percobaan learning rate dapat dilihat pada Tabel 5.

Tabel 5. Hasil percobaan BPNN

Percobaan Ke	Learning Rate	RMSE
1	0,001	0,681
2	0,005	0,022
3	0,0025	0,045
4	0,0075	0,022
Percobaan Ke	Learning Rate	RMSE
5	0,01	0,022
6	0,05	0,022
7	0,025	0,022
8	0,075	0,022
9	0,1	0,022
10	0,15	0,022

Berdasarkan Tabel 4. diketahui bahwa nilai learning rate terbaik diperoleh pada saat percobaan ke 2 dengan nilai learning rate; *error*; akurasi berturut sebesar 0,005; 0,022; 98%.

### 3.4 Pemodelan Menggunakan PNN

Pada penelitian ini, nilai parameter penghalus atau *spread* yang diuji sebanyak 6 percobaan dengan rentang nilai 0,000001; 0,00001; 0,0001; 0,001; 0,01; 0,1. Hasil percobaan pada tahap ini dapat dilihat pada Tabel 6.

Tabel 6. Hasil Percobaan PNN

Percobaan ke	Nilai <i>Spread</i>	RMSE
1	0,1	0,0455
2	0,01	0,0455
3	0,001	0,1136

4	0,0001	0,1818
5	0,00001	0,1818
6	0,000001	0,1818

Berdasarkan Tabel 5. diketahui bahwa nilai dengan akurasi terbaik diperoleh pada percobaan ke 1 dengan nilai spread sebesar 0,1 dengan nilai *error* sebesar 0,0455 dan akurasi sebesar 95,45%.

#### 4. KESIMPULAN

Dari percobaan yang dilakukan dengan menggunakan algoritma K-Means Clustering pada proses pembagian data menjadi data *training* dan data *testing* didapat hasil *cluster* terbaik pada saat  $k=2$  dengan nilai DBI sebesar 0,063. Selanjutnya, pada tahap klasifikasi kelulusan mahasiswa digunakan metode Backpropagation Neural Network dan Probabilistic Neural Network. Dari percobaan yang telah dilakukan diketahui bahwa algoritma BPNN memiliki akurasi yang lebih tinggi dibandingkan PNN dengan nilai akurasi BPNN sebesar 98% dan PNN sebesar 95,45%, sehingga pada kasus dengan menggunakan data lulusan Mahasiswa penggunaan algoritma BPNN memiliki performa yang lebih baik dibanding PNN.

#### REFERENCES

- [1] Akreditasi Program Studi Sarjana Ban-Pt, no. November. 2008.
- [2] J. S. Bassi, E. G. Dada, A. A. Hamidu, and M. D. Elijah, "Students Graduation on Time Prediction Model Using Artificial Neural Network," IOSR J. Comput. Eng., vol. 21, no. 3, pp. 28–35, 2019, doi: 10.9790/0661-2103012835.
- [3] A. Noercholis and M. Zainuddin, "Comparative Analysis of 5 Algorithm Based Particle Swarm Optimization (Pso) for Prediction of Graduate Time Graduation," MATICS J. Ilmu Komput. dan Teknol. Inf., vol. 12, no. 1, pp. 1–9, 2020, doi: 10.18860/mat.v12i1.8216.
- [4] Y. Yin, L. Long, and X. Deng, "Dynamic Data Mining of Sensor Data," IEEE Access, vol. 8, pp. 41637–41648, 2020, doi: 10.1109/ACCESS.2020.2976699.
- [5] Mustakim, "Effectiveness of K-means clustering to distribute training data and testing data on K-nearest neighbor classification," J. Theor. Appl. Inf. Technol., vol. 95, no. 21, pp. 5693–5700, 2017.
- [6] G. E. I. Kambey et al., "Penerapan Clustering pada Aplikasi Pendeteksi Kemiripan Dokumen Teks Bahasa Indonesia," J. Tek. Inform., vol. 15, no. 2, pp. 75–82, 2020.
- [7] C. WU, H. JIANG, and P. WANG, "Education quality detection method based on the probabilistic neural network algorithm," Diagnostyka, vol. 21, no. 4, pp. 79–86, 2020, doi: 10.29354/diag/127194.
- [8] E. P. Rohmawan, "Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode Decision Tree dan Artificial Neural Network," J. Ilm. Matrik, vol. 20, no. 1, pp. 21–30, 2018.
- [9] N. K. Sari, "Estimasi Kelulusan Tepat Waktu Mahasiswa Menggunakan Algoritma Backpropagation Neural Network (BPNN)," 2018.
- [10] Mustakim, N. K. Sari, Jasril, I. Kusumanto, and N. G. I. Reza, "Eigenvalue of analytic hierarchy process as the determinant for class target on classification algorithm," Indones. J. Electr. Eng. Comput. Sci., vol. 12, no. 3, pp. 1257–1264, 2018, doi: 10.11591/ijeecs.v12.i3.pp1257-1264.
- [11] J. M. Haut, M. Paoletti, J. Plaza, and A. Plaza, "Cloud implementation of the K-means algorithm for hyperspectral image analysis," J. Supercomput., vol. 73, no. 1, pp. 514–529, 2017, doi: 10.1007/s11227-016-1896-3.
- [12] R. Gustrianda and D. I. Mulyana, "Penerapan Data Mining Dalam Pemilihan Produk Unggulan dengan Metode Algoritma K-Means Dan K-Medoids," J. Media Inform. Budidarma, vol. 6, no. 1, p. 27, 2022, doi: 10.30865/mib.v6i1.3294.
- [13] S. J. A. Sarosa, F. Utamingrum, and F. A. Bachtiar, "Breast cancer classification using GLCM and BPNN," Int. J. Adv. Soft Comput. its Appl., vol. 11, no. 3, pp. 157–172, 2019.
- [14] M. Alweshah, L. Rababa, M. H. Ryalat, A. Al Momani, and M. F. Ababneh, "African buffalo algorithm: Training the probabilistic neural network to solve classification problems," J. King Saud Univ. - Comput. Inf. Sci., vol. 34, no. 5, pp. 1808–1818, 2022, doi: 10.1016/j.jksuci.2020.07.004.