

PAPER NAME

DATA COMPARISON TO PREDICT STUDY ACCURACY BASED ON FAMILY ATTRIBUTES USING MACHINE LEARNING.pdf

WORD COUNT

3847 Words

CHARACTER COUNT

21361 Characters

PAGE COUNT

6 Pages

FILE SIZE

1.4MB

SUBMISSION DATE

Jan 18, 2024 12:15 PM GMT+7

REPORT DATE

Jan 18, 2024 12:15 PM GMT+7

● **15% Overall Similarity**

The combined total of all matches, including overlapping sources, for each database.

- 12% Internet database
- 4% Publications database
- Crossref database
- Crossref Posted Content database
- 9% Submitted Works database

● **Excluded from Similarity Report**

- Bibliographic material
- Quoted material
- Cited material
- Manually excluded sources
- Manually excluded text blocks

DATA COMPARISON TO PREDICT STUDY ACCURACY BASED ON FAMILY ATTRIBUTES USING MACHINE LEARNING

Wijiyanto^{*1}, Afu Ichsan Pradana², Sopingi³

¹Software Engineering Technology, Faculty of Computer Science, Universitas Duta Bangsa Surakarta, Indonesia

²Computer Engineering, Faculty of Computer Science, Universitas Duta Bangsa Surakarta, Indonesia

³Information Systems, Faculty of Computer Science, Universitas Duta Bangsa Surakarta, Indonesia

Email: ¹wijiyanto@udb.ac.id, ²afu_ichsan@udb.ac.id, ³sopingi@udb.ac.id

Article received: date; Revision: date; published: date)

Abstract

The success of students in pursuing education is to be able to complete in a timely manner. Many factors can influence this success such as non-academic factors using family attribute. The data is taken from the Faculty of Computer Science, Duta Bangsa University, which has 365 records and 11 attributes. The data is processed using machine learning algorithms with naïve Bayes and neural network modeling, which previously split the data into training and test with a percentage ratio of 90:10, 80:20, 70:30, 60:40, and 50:50 to find the best performance based on the accuracy of each data comparison. From the results of modeling and evaluation using a confusion matrix, the best performance results for naïve Bayes with a ratio of 80:20 produce an accuracy of 92%, precision of 0.93, a recall of 0.98, and a F1 score of 0.96. As for neural networks, the best performance results are in the 50:50 ratio, produce an accuracy of 91%, a precision of 0.93, a recall of 0.97, and a F1 score of 0.95. The lowest performance results for naïve Bayes are in the 90:10, and the neural network is in the 80:20. Thus, the Naïve Bayes has better performance than the neural network.

Keywords: Comparison, Naïve Bayes, Neural Network, Machine Learning

PERBANDINGAN DATA UNTUK MEMPREDIKSI KETEPATAN STUDI BERDASARKAN ATRIBUT KELUARGA MENGGUNAKAN MACHINE LEARNING

Abstrak

Keberhasilan mahasiswa dalam menempuh pendidikan adalah dapat menyelesaikan secara tepat waktu. Banyak faktor-faktor yang dapat mempengaruhi keberhasilan tersebut seperti Faktor non akademik menggunakan data beratribut keluarga. Data diambil dari FIKOM-UDB yang memiliki 365 record dan 11 atribut, termasuk 1 atribut sebagai label (*class*). Data diolah menggunakan algoritma *machine learning* dengan pemodelan *naïve bayes* dan *neural network*, yang sebelumnya data dipisah (*split*) dalam data latih (*data train*) dan data tes (*data test*) dengan perbandingan prosentase 90:10, 80:20, 70:30, 60:40 dan 50:50 untuk mencari performa terbaik berdasarkan nilai akurasi dari perbandingan masing masing data. Dari hasil pemodelan dan evaluasi menggunakan confusion matrix didapatkan hasil performa terbaik untuk *naïve bayes* dengan perbandingan 80:20 memiliki nilai akurasi 92%, nilai *precision* 0.93, nilai *recall* 0.98 dan nilai *F1 score* 0.96. Untuk *neural network*, hasil performa terbaik terdapat pada perbandingan 50:50 memiliki nilai akurasi 91%, nilai *precision* 0.93, nilai *recall* 0.97 dan nilai *F1 score* 0.95. Sementara hasil performa terendah untuk *naïve bayes* terdapat pada perbandingan 90:10 dan *neural network* terdapat pada perbandingan 80:20. Dengan demikian algoritma *naïve bayes* memiliki performa yang lebih besar dibandingkan dengan algoritma *neural network*.

Kata kunci: Machine learning, perbandingan, naïve bayes, neural network.

1. PENDAHULUAN

Keberhasilan merupakan tujuan yang harus diraih oleh seorang mahasiswa dalam menempuh pendidikan. Salah satu keberhasilan mahasiswa tersebut dapat dilihat

dari ketepatan waktu dalam menyelesaikan studinya[1]. Keberhasilan studi mahasiswa dalam menjalani proses pendidikan menjadi tanggung jawab semua *stakeholder* yang terlibat didalamnya termasuk pemerintah, orang

tua dan tenaga pengajarnya[2]. Secara normal, seorang mahasiswa jenjang sarjana dapat menempuh pendidikan selama 4 tahun dan jenjang diploma selama 3 tahun, namun demikian banyak mahasiswa yang menyelesaikan pendidikan tidak tepat waktu.

Banyak faktor yang mempengaruhi keberhasilan pendidikan mahasiswa, baik akademik atau non-akademik[3] serta memahami faktor-faktor yang mempengaruhinya tersebut[4] dapat memberikan penilaian tentang ketepatan waktu studi yang merupakan keberhasilan dalam pengelolaan pendidikan[5]. Faktor tersebut merupakan data sumber utama yang dapat dimanfaatkan untuk dikelola, sehingga pola data yang ada dapat dimengerti[6]. Menggunakan *artificial intelligence* adalah salah satu cara untuk menggali data pendidikan yang dapat digunakan untuk memprediksi kinerja mahasiswa dan dapat membantu mahasiswa untuk menghindari kegagalan dalam perkuliahnya[7].

Dalam memprediksi kinerja mahasiswa menggunakan algoritma *machine learning* sudah dilakukan oleh peneliti sebelumnya, seperti pemodelan menggunakan *naïve bayes* yang digunakan untuk klasifikasi data berdasarkan probabilitas dan algoritma *neural Network* atau jaringan saraf tiruan yang merupakan pemrosesan untuk menggali informasi yang terinspirasi oleh sistem saraf otak manusia[8]. Algoritma *naïve bayes* menggunakan atribut umur, kota, pekerjaan ayah, prodi, kelompok kelas, jumlah saudara dalam keluarga dan Indeks Prestasi Kumulatif memiliki tingkat keakuratan pada data test-nya sebesar 57,63%[9], penelitian lain menyebutkan hasil akurasi *naïve bayes* sebesar 84,8052%[10]. Sementara itu implementasi menggunakan *feature forward selection* ditambah dengan SMOTE dapat mempengaruhi keakuratan sebesar 87.13%[11]. Hasil akurasi algoritma *naïve bayes* yang tertinggi juga diperoleh dengan data perbandingan 80:20[12]. Sementara itu juga, hasil terbaik terhadap pembagian data train dan data test didapatkan dengan perbandingan 75:25 pada data ulasan film[13] dan perbandingan 55:45 menggunakan skema *holdout validation*[14]. Penelitian lain yang digunakan untuk meningkatkan penjualan pada toko bangunan menggunakan *naïve bayes* menghasilkan akurasi sebesar 0.71[15].

Untuk penelitian menggunakan *neural network* didapatkan akurasi sebesar 63,63% dengan menggunakan data testing 20%[16]. Hasil prediksi kinerja mahasiswa dapat dimanfaatkan dalam melaksanakan monev (monitoring dan evaluasi) terhadap jalannya proses pembelajaran dalam upaya meningkatkan kualitas pembelajaran[17]. Hasil penelitian ini digunakan perbandingan data untuk menghasilkan performa terbaik dalam memprediksi ketepatan studi mahasiswa berdasarkan atribut keluarga menggunakan algoritma *naïve bayes* dan *neural network*.

2. METODE PENELITIAN

Metode penelitian ini dilakukan dengan beberapa tahapan yang dimulai dari *data collection*, *preprocessing*, *splitting data*, *modelling* dan *evaluation*, dimana data akan diolah menggunakan *Jupyter notebook Python 3*. Adapun tahap-tahap yang digunakan adalah sebagai berikut :



Gambar 1: Metode Penelitian

2.1 Data collecting

Data collecting adalah tahap pertama untuk mengumpulkan data awal yang merupakan tahap sangat penting dalam penelitian ini. Data yang baik adalah data yang didapatkan secara objektif dan telah terverifikasi. Data yang dikumpulkan berasal dari empat prodi (2 prodi jenjang diploma dan 2 prodi jenjang sarjana).

2.2 Preprocessing

Preprocessing merupakan langkah yang sangat penting karena dapat mempengaruhi kualitas data [18] dan pada tahap ini dilakukan seleksi fitur yang memiliki pengaruh besar terhadap prediksi yang akan dilakukan. Pembersihan data dilakukan terhadap data yang bernilai null atau kosong[19], pada tahap ini data yang kosong akan dihapus dan tidak diikutkan dalam tahap berikutnya. Karena datanya mempunyai perbedaan nilai dan beragam, maka dilakukan proses membuat skala data agar data yang diproses lebih cepat.

2.3 Splitting data

Perbandingan rasio *data train* dan *data test* dapat mempengaruhi hasil klasifikasi, karena pemilihan rasio yang tidak tepat dapat menurunkan hasil akurasi model yang akan dibangun[14]. *Splitting data* dilakukan untuk membagi data yang sudah melalui tahap *preprocessing* ke dalam data train dan data test. *Data train* dan *data test* dibagi ke dalam lima kelompok perbandingan prosentase yang berbeda beda untuk mengetahui hasil perbandingan yang terbaik. Data dalam setiap perbandingan akan dimasukkan ke dalam algoritma *machine learning* pada tahap *modelling*.

2.4 Modelling

Setelah *data train* dan *data test* dibagi sesuai dengan persentasenya, selanjutnya setiap perbandingan rasio (5 kali) dilakukan proses pemodelan dengan menggunakan algoritma *naïve bayes* karena dapat digunakan untuk klasifikasi menggunakan data yang terdistribusi secara seragam[7] dan algoritma *neural network*, dimana data latih akan dimasukkan ke dalam tahap pelatihan pada masing masing algoritma *naïve bayes* dan *neural network*. Setelah didapatkan modelnya dari masing masing algoritma yang digunakan, data tes

dimasukkan ke tahap testing menggunakan algoritma yang sama.

2.5 Evaluation

Tahap terakhir adalah evaluasi menggunakan *confusion matrix* yang merupakan matrix untuk menggambarkan hasil model klasifikasi untuk dataset yang akan diuji[20]. *confusion matrix* memberikan penilaian *performance* klasifikasi berdasarkan objek dengan benar atau salah dan menghasilkan nilai *accuracy*, *recall*, *precision*, *F1-score*. karena metode terbaik adalah yang mempunyai nilai akurasi tertinggi[21]

3. HASIL DAN PEMBAHASAN

3.1 Data Collecting

Data mahasiswa dikumpulkan dari Perguruan Tinggi swasta angkatan 2019 pada Fakultas Ilmu Komputer Universitas Duta Bangsa Surakarta yang ditarik dari *sisfo* (sistem informasi akademik) yang terdiri dari prodi Sistem Informasi, prodi Teknik Informatika, prodi Manajemen Informatika dan prodi Teknik Komputer sebanyak 372 baris. Data yang didapatkan berupa data dalam format excel dan dirubah ke dalam bentuk file yang berekstensi csv. Data tersebut telah di komparasikan dengan data lulusan berdasarkan surat keputusan yudisium atau kelulusan.

3.2 Preprocessing

a. Data selection

Dari data yang sudah terkumpul, terdapat fitur yang tidak diikutkan untuk tahap berikutnya seperti fitur yaitu nim, nama, tanggal yudisium dan nomor sk yudisium karena tidak memiliki relevansi dengan fitur yang berhubungan dengan data keluarga. Setelah menghapus fitur tersebut ditambahkan 1 label sebagai keterangan mahasiswa lulus tepat waktu (1) atau terlambat (0). Sehingga data yang di proses pada tahap selanjutnya terdiri dari 10 fitur (jenis kelamin, asal, umur, pendidikan ayah, pekerjaan ayah, id penghasilan ayah, id untuk pendidikan ibu, id untuk pekerjaan ibu, id untuk penghasilan ibu, ipk) dan 1 label (keterangan).

Karena data masih dalam bentuk bukan angka maka dilakukan transformasi data untuk dapat di proses ke data mining. Tabel 1 berikut ini adalah keterangan fitur dan label yang digunakan serta tipe data yang digunakan.

Tabel 1 : Fitur dan tipe data setelah proses seleksi

fitur	Tipe data
jk	int64
asal	int64
umur	int64
pendidikan_ayah	int64
id_pekerjaan_ayah	int64
id_penghasilan_ayah	int64
jenjang_pendidikan_ibu	int64
id_pekerjaan_ibu	int64
id_penghasilan_ibu	int64
ipk	float64
keterangan	int64

dtype: object

Dari tabel 1 terlihat 10 fitur memiliki tipe data integer dan 1 fitur memiliki tipe data *float* yaitu ipk, karena data ipk memiliki nilai pecahan atau desimal. Dengan demikian bahwa semua fitur telah bernilai angka dan dapat dilanjutkan untuk tahap berikutnya

b. Data Cleaning

Pembersihan data dilakukan terhadap data yang kosong atau *null*. Dari 372 data ditemukan 7 data yang kosong, sehingga dilakukan penghapusan data. Setelah proses ini dilakukan pengecekan terhadap data apakah masih ada data yang kosong atau tidak (*missing value*). Dari gambar 3 terlihat bahwa pada setiap dataframe sudah bernilai *false* yang berarti sudah tidak ada data yang kosong (*null*).

Tabel 2 : Hasil pemeriksaan terhadap data missing value

Fitur	Value
jk	False
asal	False
umur	False
pendidikan_ayah	False
id_pekerjaan_ayah	False
id_penghasilan_ayah	False
jenjang_pendidikan_ibu	False
id_pekerjaan_ibu	False
id_penghasilan_ibu	False
ipk	False
keterangan	False

dtype: bool

Setelah proses pembersihan terhadap data yang *null*, data yang digunakan sebanyak 365 baris dan 11 kolom. Tabel 3 merupakan hasil pemeriksaan data yang sudah melalui tahap seleksi dan pembersihan data.

Tabel 3 : Data mahasiswa yang diseleksi dan dibersihkan

	jk	asal	umur	pendidikan_ayah	id_pekerjaan_ayah	id_penghasilan_ayah	id_jenjang_pendidikan_ibu	id_pekerjaan_ibu	id_penghasilan_ibu	ipk	keterangan
0	0	1	24	3	5	2	3	0	0	3.39	0
1	0	0	25	3	7	2	3	7	2	3.26	0
2	0	0	26	5	1	3	3	1	4	3.62	0
3	1	0	27	3	4	2	3	0	5	3.42	0

4	0	0	23	3	5	3	2	0	0	3.50	0
...
360	0	0	22	5	1	3	3	7	2	3.79	1
361	0	1	23	3	7	3	5	0	5	3.29	1
362	0	1	21	3	5	3	2	5	2	3.81	1
363	0	0	22	3	5	2	3	5	2	3.61	1
364	0	0	23	1	3	1	1	0	5	3.68	1

365 rows × 11 columns

c. Data scaling

Karena data pada setiap kolom bervariasi maka dilakukan proses skala menggunakan *standartscaler* dengan tujuan untuk membuat rata-rata 0 dan deviasi standar 1 atau menormalisasi data, sehingga pemodelan dapat belajar lebih cepat dalam rangka untuk meningkatkan hasil akurasi yang lebih baik.

```
x_train
array([[ -0.73645969, -0.43536477, -0.47518494, ..., -1.09247328,
         1.44757875,  1.09841869],
       [ -0.73645969, -0.43536477,  0.7372028 , ...,  0.21275738,
        -0.17537402, -0.4387513  ],
       [ -0.73645969, -0.43536477,  0.7372028 , ..., -0.22231951,
        -0.17537402,  0.3398812  ],
       ...,
       [  1.35784756, -0.43536477, -0.47518494, ...,  0.21275738,
        -0.71635827,  0.3398812  ],
       [ -0.73645969, -0.43536477, -0.47518494, ...,  1.08291115,
        -0.17537402, -0.05599924],
       [ -0.73645969,  2.29692448, -0.47518494, ..., -1.09247328,
        -0.17537402,  0.86753511]])
```

Gambar 2: Hasil standartscaler untuk data latih

3.3 Splitting Data

Setelah data disajikan melalui *preprocessing*, selanjutnya data dipisah untuk data train dan data test. Dalam tahap ini data dipisah sebanyak lima kali untuk dilakukan eksperimen menggunakan *machine learning*. Perbandingan data train dan data test dapat dilihat pada tabel berikut ini:

Tabel 4: Prosentase perbandingan data train dan data test

No	Jumlah Data	Data Train	Data Test	prosentase
1	365	328	27	90:10
2	365	292	73	80:20
3	365	255	110	70:30
4	365	219	146	60:40
5	365	182	183	50:50

Dari tabel 4 terlihat bahwa data sebanyak 365, dipisah antara data train dan data test mulai prosentase 10% berbanding 10%, 80% berbanding 20%, 70% berbanding 30%, 60% berbanding 40% dan 50% berbanding 50%.

3.4 Modelling

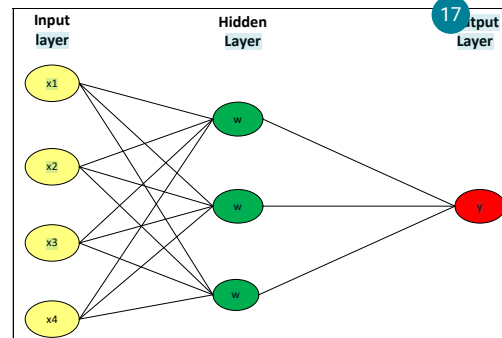
Tahap *modelling* ini dilaksanakan menggunakan algoritma *machine learning* yaitu *naive bayes* dan *neural network* yang diulang sebanyak lima kali sesuai dengan prosentase tabel 4 sesuai perbandingan data latih dan data tes. Dimana pemanggilan fungsi algoritma *naive bayes* telah berhasil. Adapun fungsi algoritma *naive bayes* tersebut dapat mengasumsikan semua fitur dalam model adalah independen[7]. Sehingga dapat digunakan untuk mengklasifikasikan objek

menggunakan data yang terdistribusi secara seragam. Penghitungan tersebut menggunakan rumus :

$$P(c|x) = \frac{P(c)P(x)}{P(x)} \tag{1}$$

$$P(c|x) = P(x_1|x) * \dots * P(x_n|x) * P(c) \tag{2}$$

Untuk pemanggilan fungsi algoritma *neural network* juga berhasil dilakukan dengan parameter *hidden layer size=3*, *learning rate=0.1* dan *max iterasi=100*. Algoritma *neural network* memiliki kinerja untuk klasifikasi yang sangat baik yang bekerja seperti pada *neuron* otak manusia, walaupun membutuhkan waktu yang agak lama karena banyaknya lapisan yang dibentuk[22].



Gambar 3 : Model arsitektur neural network

3.5 Evaluation

Pada tahap ini, evaluasi dilakukan menggunakan *confision matrix* untuk membandingkan hasil prediksi dengan data test. Setelah dibandingkan data tersebut, selanjutnya dilakukan penghitungan untuk mencari akurasi, *precision*, *recall* dan *F1score* dengan hasil pada tabel 5 dan 6.

Tabel 5 : Hasil evaluasi algoritma naive bayes

No	Split	Akurasi	%	Precisior	Recall	F1-Score
1	90:10	0.89	89%	0.91	0.97	0.94
2	80:20	0.92	92%	0.93	0.98	0.96
3	70:30	0.91	91%	0.92	0.98	0.95
4	60:40	0.91	91%	0.93	0.98	0.95
5	50:50	0.91	91%	0.93	0.96	0.95

Dari tabel 5 terlihat semua data telah di evaluasi sesuai memperoleh hasil terbaik terdapat pada perbandingan data 80% data train dan 20% data test memperoleh akurasi sebesar 0.92, sementara hasil

terendah terdapat pada perbandingan data 90% data train dan 10% data test memperoleh akurasi sebesar 0.89.

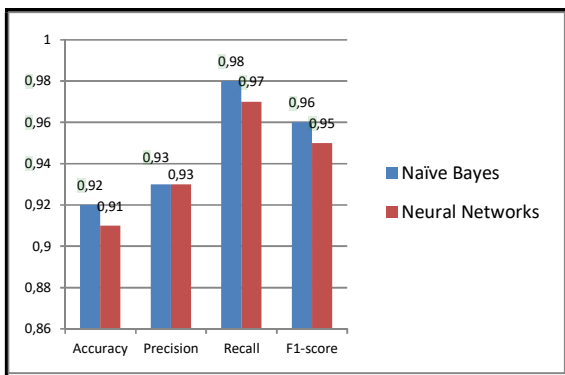
Sementara itu hasil evaluasi untuk algoritma *neural network* yang menggunakan *hidden layer sizes=3, learning rate init=0.1, maxima iterasi=100* adalah sebagai berikut :

Tabel 6 : Hasil evaluasi algoritma neural network

NO	Split	Akurasi	%	Precision	Recall	F1-Score
1	90:10	0.84	84%	0.91	0.91	0.91
2	80:20	0.81	81%	0.91	0.88	0.89
3	70:30	0.86	86%	0.96	0.89	0.92
4	60:40	0.86	86%	0.94	0.91	0.92
5	50:50	0.91	91%	0.93	0.97	0.95

Dari tabel 6 terlihat hasil evaluasi algoritma *neural network* memperoleh performa terbaik terdapat pada perbandingan 50 persen data train dan 50 persen data test dengan akurasi 0.91, sementara untuk hasil akurasi terendah terdapat pada perbandingan data 80 persen data train dan 20% data test sebesar 0.81.

Setelah diketahui hasil terbaik dari masing masing pemodelan *naïve bayes* dan *neural network*, secara lengkap digambarkan hasil evaluasi seperti terlihat pada gambar 5 berikut ini.



Gambar 4 : grafik hasil evaluasi

Dari gambar 4 di atas terlihat bahwa untuk accuracy, *naïve bayes* memiliki nilai yang lebih baik sebesar 0,92. Demikian juga untuk nilai *recall* dan *F1score*, *naïve bayes* memiliki nilai lebih tinggi masing 0,98 dan 0,96. Namun demikian untuk nilai *precision* antara algoritma *naïve bayes* dan *neural network* mempunyai nilai sama 0,93.

4. DISKUSI

Pada penelitian sebelumnya yang sudah dijelaskan pada pendahuluan, kebanyakan menggunakan satu perbandingan data train dan data test. Sementara pada penelitian ini digunakan lima perbandingan data train dan data test menggunakan dua pemodelan machine learning yaitu *naïve bayes* dan *neural network*. Selain itu dataset dan atribut yang digunakan pada penelitian ini berbeda dengan penelitian sebelumnya. Dari kedua pemodelan tersebut masing-masing didapatkan perbandingan data yang terbaik untuk memperoleh nilai akurasi yang tertinggi, sehingga dapat membantu

pengelola pendidikan untuk mengantisipasi yang mengalami kendala khususnya dari lingkungan keluarganya.

5. KESIMPULAN

Berdasarkan hasil penelitian, bahwa atribut keluarga mempunyai pengaruh terhadap keberhasilan dalam memprediksi ketepatan studi mahasiswa di lingkungan fakultas ilmu Komputer Universitas Duta Bangsa Surakarta. Dari data sebanyak 365 yang dibagi untuk perbandingan data latih dan data tes sebanyak lima kali menggunakan algoritma *naïve bayes* dimana hasil tertinggi untuk nilai akurasi sebesar 92% pada perbandingan 292 (80%) data latih dan 73 (20%) data tes dengan nilai *precision* 0.93, nilai *recall* 0.98 dan nilai *F1score* 0.96. Sementara untuk algoritma *neural network* yang menggunakan *hidden layer size 3* dan iterasi 100 kali didapatkan hasil tertinggi untuk nilai akurasi sebesar 91% pada perbandingan 182 (50%) data latih dan 183 (50%) data tes dengan nilai *precision* 0.93, nilai *recall* 0.97 dan nilai *F1score* 0.95. Selain mendapatkan nilai tertinggi, dalam evaluasi ini juga didapatkan hasil terendah untuk algoritma *naïve bayes* dengan nilai akurasi 89% dan algoritma *neural network* dengan nilai akurasi 81%. Dengan demikian algoritma *naïve bayes* mempunyai performa yang lebih baik jika dibandingkan dengan algoritma *neural network* dengan selisih 1%. Pada penelitian selanjutnya dapat dikembangkan dengan algoritma lain dengan menggunakan atribut pendukung tambahan yang lebih tepat seperti atribut akademik dan atribut pribadi mahasiswa dan menggunakan algoritma *deep learning* untuk meningkatkan validitas model dan memperoleh akurasi yang lebih baik.

DAFTAR PUSTAKA

- [1] T. H. Hasibuan and D. Mahdiana, "Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Algoritma C4.5 Pada Uin Syarif Hidayatullah Jakarta," *SKANIKA*, vol. 6, no. 1, pp. 61–74, Jan. 2023, doi: 10.36080/skanika.v6i1.2976.
- [2] L. W. Santoso and Y. Yulia, "Predicting student performance in higher education using multi-regression models," *TELKOMNIKA*, vol. 18, no. 3, p. 1354, Jun. 2020, doi: 10.12928/telkomnika.v18i3.14802.
- [3] I. Issah, O. Appiah, P. Appiahene, and F. Inusah, "A systematic review of the literature on machine learning application of determining the attributes influencing academic performance," *Decision Analytics Journal*, vol. 7, p. 100204, Jun. 2023, doi: 10.1016/j.dajour.2023.100204.
- [4] I. Irawan, M. R. Qisthiano, M. Syahril, and P. M. Jakak, "Optimasi Prediksi Kelulusan Tepat Waktu: Studi Perbandingan Algoritma Random Forest dan Algoritma K-NN Berbasis PSO," *JPSII*, vol. 4, no. 4, pp. 26–35, Oct. 2023, doi: 10.47747/jpsii.v4i4.1374.

- [5] A. Jananto, S. Sulastri, E. Nur Wahyudi, and S. Sunardi, "Data Induk Mahasiswa sebagai Prediktor Ketepatan Waktu Lulus Menggunakan Algoritma CART Klasifikasi Data Mining," *SISFOKOM*, vol. 10, no. 1, pp. 71–78, Feb. 2021, doi: 10.32736/sisfokom.v10i1.991.
- [6] F. Rolansa, Y. Yunita, and S. Suheri, "Sistem prediksi dan evaluasi prestasi akademik mahasiswa di Program Studi Teknik Informatika menggunakan data mining," *JPIS*, vol. 9, no. 1, p. 75, Jun. 2020, doi: 10.31571/saintek.v9i1.1696.
- [7] N. Abuzinadah *et al.*, "Role of convolutional features and machine learning for predicting student academic performance from MOODLE data," *PLoS ONE*, vol. 18, no. 11, p. e0293061, Nov. 2023, doi: 10.1371/journal.pone.0293061.
- [8] Muhamad Ziaul Haq and Nursalim, "Prediksi Indeks Prestasi Kumulatif Mahasiswa berdasarkan Nem dengan Menggunakan Algoritma Neural Network Berbasis Particle Swarm Optimization: Prediction of Student Cumulative Achievement Index Based on NEM Using Particle Swarm Optimization Based Neural Network Algorithm," *J Kol Sai*, vol. 6, no. 2, pp. 147–153, Feb. 2023, doi: 10.56338/jks.v6i2.3303.
- [9] A. Azahari, Y. Yulindawati, D. Rosita, and S. Mallala, "Komparasi Data Mining Naive Bayes dan Neural Network memprediksi Masa Studi Mahasiswa S1," *JTIK*, vol. 7, no. 3, pp. 443–452, May 2020, doi: 10.25126/jtiik.2020732093.
- [10] H. Amalia, A. Puspita, A. F. Lestari, and F. Friyadie, "APPLICATION OF DECISION TREE AND NAIVE BAYES ON STUDENT PERFORMANCE DATASET," *pilar*, vol. 18, no. 1, pp. 53–58, Mar. 2022, doi: 10.33480/pilar.v18i1.2714.
- [11] D. Kurniadi, F. Nuraeni, and S. M. Lestari, "Implementasi Algoritma Naive Bayes Menggunakan Feature Forward Selection dan SMOTE Untuk Memprediksi Ketepatan Masa Studi Mahasiswa Sarjana," *JSC*, vol. 5, no. 2, pp. 63–82, Aug. 2022, doi: 10.37396/jsc.v5i2.215.
- [12] S. Nuralia, H. Harliana, and T. Prabowo, "Implementasi Naive Bayes Classifier Dalam Memprediksi Kelulusan Mahasiswa," *jacis*, vol. 3, no. 1, pp. 63–72, May 2023, doi: 10.47134/jacis.v3i1.57.
- [13] R. M. Awangga and N. H. Khonsa', "Analisis Performa Algoritma Random Forest dan Naive Bayes Multinomial pada Dataset Ulasan Obat dan Ulasan Film," *InComTech*, vol. 12, no. 1, p. 60, Apr. 2022, doi: 10.22441/incomtech.v12i1.14770.
- [14] R. Oktafiani, A. Hermawan, and D. Avianto, "Pengaruh Komposisi Split data Terhadap Performa Klasifikasi Penyakit Kanker Payudara Menggunakan Algoritma Machine Learning," *JSI*, pp. 19–28, Jun. 2023, doi: 10.34128/jsi.v9i1.622.
- [15] F. Valentinus, F. Sujono, I. Ariansyah, and D. A. H. Capah, "Implementation Of Data Mining With Classification And Forecasting Method Use Model Gaussian Naive Bayes For Building Store (STUDI CASE: TB SINAR JAYA)," *J. Tek. Inform. (JUTIF)*, vol. 4, no. 2, pp. 413–420, Mar. 2023, doi: 10.52436/1.jutif.2023.4.2.701.
- [16] U. R. Habibah and A. Solichin, "Prediksi Kelulusan Mahasiswa Dengan Metode Naive Bayes dan Artificial Neural Network: Studi Kasus Fakultas Teknik UNIS Tangerang," *Faktor Exacta*, vol. 15, no. 1, p. 73–83, 2022, doi: 10.30998/faktorexacta.v15i1.11816.
- [17] Muhammad Haris Diponegoro, Sri Suning Kusumawardani, and Indriana Hidayah, "Tinjauan Pustaka Sistematis: Implementasi Metode Deep Learning pada Prediksi Kinerja Murid," *JNTETI*, vol. 10, no. 2, pp. 131–138, May 2021, doi: 10.22146/jnteti.v10i2.1417.
- [18] U. Suriani, "Penerapan Data Mining untuk Memprediksi Tingkat Kelulusan Mahasiswa Menggunakan Algoritma Decision Tree C4.5," vol. 3, no. 2, 2023.
- [19] E. Haryatmi and S. Pramita Hervianti, "Penerapan Algoritma Support Vector Machine Untuk Model Prediksi Kelulusan Mahasiswa Tepat Waktu," *RESTI*, vol. 5, no. 2, pp. 386–392, Apr. 2021, doi: 10.29207/resti.v5i2.3007.
- [20] E. Sutoyo and A. Almaarif, "Educational Data Mining untuk Prediksi Kelulusan Mahasiswa Menggunakan Algoritme Naive Bayes Classifier," *Vol.*, no. 1, 2020.
- [21] M. L. Suliztia and A. Fauzan, "Comparing Naive Bayes, K-Nearest Neighbor, And Neural Network Classification Methods Of Seat Load Factor In Lombok Outbound Flights," *JMSK*, vol. 16, no. 2, p. 187, Dec. 2019, doi: 10.20956/jmsk.v16i2.7864.
- [22] A. Wibowo Putra, W. Widiyono, A. Saifudin, A. Darmawan Soma, and E. Budihartono, "Naive Bayes, Neural Network dan K-Nearest Neighbor untuk Klasifikasi Topik Tugas Akhir," *smartcomp*, vol. 11, no. 4, Oct. 2022, doi: 10.30591/smartcomp.v11i4.4251.

● **15% Overall Similarity**

Top sources found in the following databases:

- 12% Internet database
- Crossref database
- 9% Submitted Works database
- 4% Publications database
- Crossref Posted Content database

TOP SOURCES

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

1	dsitd.ipb.ac.id Internet	1%
2	repository.uin-suska.ac.id Internet	1%
3	doku.pub Internet	<1%
4	jurnal.ustjogja.ac.id Internet	<1%
5	Sriwijaya University on 2020-08-12 Submitted works	<1%
6	Universitas Negeri Jakarta on 2020-07-06 Submitted works	<1%
7	Forum Komunikasi Perpustakaan Perguruan Tinggi Kristen Indonesia (...) Submitted works	<1%
8	Universitas Negeri Medan on 2021-09-05 Submitted works	<1%

9	e-journal.poltekkesjogja.ac.id	Internet	<1%
10	garuda.kemdikbud.go.id	Internet	<1%
11	journal.unhas.ac.id	Internet	<1%
12	repository.its.ac.id	Internet	<1%
13	text-id.123dok.com	Internet	<1%
14	jutif.if.unsoed.ac.id	Internet	<1%
15	journal2.um.ac.id	Internet	<1%
16	123dok.com	Internet	<1%
17	Imperial College of Science, Technology and Medicine on 2020-09-11	Submitted works	<1%
18	Ni Made Meriliana Candra Devi, I Putu Agung Bayupati, Ni Kadek Ayu ...	Crossref	<1%
19	Universitas Amikom on 2023-07-26	Submitted works	<1%
20	ejurnal.stmik-budidarma.ac.id	Internet	<1%

21	pt.scribd.com	Internet	<1%
22	stikomtb.ac.id	Internet	<1%
23	scribd.com	Internet	<1%
24	Lusiana Efrizoni, Sarjon Defit, Muhammad Tajuddin, Anthony Anggraw...	Crossref	<1%
25	Politeknik Negeri Bandung on 2020-02-13	Submitted works	<1%
26	University of Technology, Sydney on 2021-04-15	Submitted works	<1%
27	Wahyudin Hasyim, Alter Lasarudin. "O OPTIMALISASI NEURAL NETWO...	Crossref	<1%
28	eprints.undip.ac.id	Internet	<1%
29	Forum Komunikasi Perpustakaan Perguruan Tinggi Kristen Indonesia (...	Submitted works	<1%
30	President University on 2020-01-23	Submitted works	<1%
31	UIN Sultan Syarif Kasim Riau on 2021-02-15	Submitted works	<1%
32	Universitas Dian Nuswantoro on 2015-08-14	Submitted works	<1%

33	fdocumenti.com Internet	<1%
34	grafiati.com Internet	<1%
35	researchgate.net Internet	<1%
36	medium.com Internet	<1%
37	Mahdiawan Nurkholifah, Jasmarizal, Yusran Umar, Rahmaddeni. "AN... Crossref	<1%
38	Universitas Budi Luhur on 2021-03-16 Submitted works	<1%