

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

Wijiyanto¹⁾, Afu Ichsan Pradana²⁾, Sopingi³⁾

¹ Program Studi Teknologi Rekayasa Perangkat Lunak Fakultas Ilmu Komputer, Universitas Duta Bangsa Surakarta

² Program Studi Teknik Komputer Fakultas Ilmu Komputer, Universitas Duta Bangsa Surakarta

³ Program Studi Sistem Informasi Fakultas Ilmu Komputer, Universitas Duta Bangsa Surakarta

Jl. Bhayangkara 55-57 Surakarta, Jawa Tengah

Co Responden Email: wijiyanto@udb.ac.id

Abstract

Article history

Received 24 Jan 2024

Revised 14 Feb 2024

Accepted 07 Mar 2024

Available online 30 Apr 2024

Keywords

Machine Learning,

Comparison,

Naïve Bayes,

Neural Network

Students' success in completing their education on time is an important goal. Various factors can influence this success, including non-academic factors such as family data. The data used comes from FIKOM-UDB with 365 records and 11 attributes. The data was processed using machine learning algorithms using naïve bayes modelling and neural networks. The data was divided into training and test data with different percentage comparisons, namely 90:10, 80:20, 70:30, 60:40, and 50:50, to find the best performance based on accuracy value. Evaluation using confusion matrix resulted in the best performance for naïve bayes with a ratio of 80:20, with 92% accuracy, precision 0.93, recall 0.98, and F1-score 0.96. For neural network, the best performance is found in the 50:50 ratio with 91% accuracy, precision 0.93, recall 0.97, and F1-score 0.95. The results show that the lowest performance for naïve bayes occurs at a ratio of 90:10, while for neural networks it occurs at a ratio of 80:20. Thus, the naïve bayes algorithm shows better performance than the neural network so that, the Faculty can apply the naïve bayes model in predicting students in order to anticipate and overcome problems that arise related to student graduation in a timely manner.

Abstrak

Riwayat

Diterima 24 Jan 2024

Revisi 14 Feb 2024

Disetujui 07 Mar 2024

Terbit Online 30 Apr 2024

Kata Kunci

Machine Learning,

Perbandingan,

Naïve Bayes,

Neural Network

Keberhasilan mahasiswa dalam menyelesaikan pendidikan tepat waktu merupakan tujuan yang penting. Berbagai faktor dapat memengaruhi keberhasilan ini, termasuk faktor non-akademik seperti data keluarga. Data yang digunakan berasal dari FIKOM-UDB dengan 365 record dan 11 atribut, di antaranya satu atribut berperan sebagai label (*class*). Data tersebut diproses menggunakan algoritma *machine learning* menggunakan pemodelan *naïve bayes* dan *neural network*. Sebelumnya, data dibagi menjadi data latih dan data uji dengan perbandingan prosentase yang berbeda, yaitu 90:10, 80:20, 70:30, 60:40, dan 50:50, untuk mencari kinerja terbaik berdasarkan nilai akurasi. Evaluasi menggunakan *confusion matrix* menghasilkan performa terbaik untuk *naïve bayes* dengan perbandingan 80:20, mencapai nilai akurasi sebesar 92%, *precision* 0.93, *recall* 0.98, dan *F1-score* 0.96. Sementara untuk *neural network*, performa terbaik terdapat pada perbandingan 50:50 dengan nilai akurasi sebesar 91%, *precision* 0.93, *recall* 0.97, dan *F1-score* 0.95. Hasil menunjukkan bahwa performa terendah untuk *naïve bayes* terjadi pada perbandingan 90:10, sementara untuk *neural network* terjadi pada perbandingan 80:20. Dengan demikian, algoritma *naïve bayes* menunjukkan performa yang lebih baik dibandingkan *neural network* sehingga, Fakultas dapat menerapkan model *naïve bayes* dalam memprediksi mahasiswa dalam rangka untuk mengantisipasi dan mengatasi permasalahan yang timbul terkait kelulusan mahasiswa dengan tepat waktu.

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 (Hasibuan and Mahdiana 2023). Keberhasilan studi mahasiswa dalam menjalani proses pendidikan menjadi tanggung jawab semua *stakeholder* yang terlibat didalamnya termasuk pemerintah, orang tua dan tenaga pengajarnya (Santoso and Yulia 2020). 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 (Issah et al. 2023) serta memahami faktor-faktor yang mempengaruhinya tersebut (Irawan et al. 2023) dapat memberikan penilaian tentang ketepatan waktu studi yang merupakan keberhasilan dalam pengelolaan pendidikan (Jananto et al. 2021). Faktor tersebut merupakan data sumber utama yang dapat dimanfaatkan untuk dikelola, sehingga pola data yang ada dapat dimengerti (Rolansa, Yunita, and Suheri 2020). 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 (Abuzinadah et al. 2023).

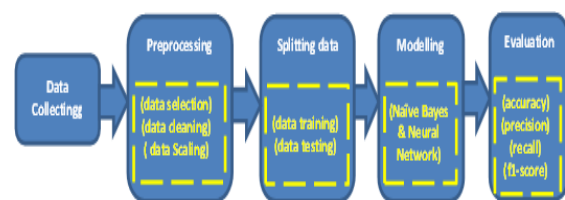
Dalam memprediksi kinerja mahasiswa menggunakan algoritma *machine learning* sudah dilakukan oleh peneliti sebelumnya, seperti pemodelan menggunakan *naïve bayes* yang digunakan untuk klasifikasi data berdasarkan probabilitas dan algoritma *Neural Network* atau jaringan saraf tiruan yang merupakan pemrosesan untuk menggali informasi yang terinspirasi oleh sistem saraf otak manusia (Muhamad Ziaul Haq and Nursalim 2023). Algoritma *naïve bayes* menggunakan atribut umur, kota, pekerjaan ayah, prodi, kelompok kelas, jumlah saudara dalam keluarga dan Iindek Prestasi Kumulatif memiliki tingkat keakuratan pada data test-nya sebesar 57,63% (Azahari et al. 2020), penelitian lainnya menyebutkan hasil akurasi *naïve bayes* sebesar 84,8052% (Amalia et al. 2022). Sementara itu implementasi menggunakan *feature forward selection* ditambah dengan SMOTE dapat mempengaruhi keakurasian sebesar 87.13% (Kurniadi, Nuraeni, and Lestari 2022). Hasil

akurasi algoritma *naïve bayes* yang tertinggi juga diperoleh dengan data perbandingan 80:20 (Nuralia, Harliana, and Prabowo 2023). Sementara itu juga, hasil terbaik terhadap pembagian data train dan data test didapatkan dengan perbandingan 75:25 pada data ulasan film (Awangga and Khonsa' 2022) dan perbandingan 55:45 menggunakan skema *holdout validation* (Oktafiani, Hermawan, and Avianto 2023). Penelitian lain yang digunakan untuk meningkatkan penjualan pada toko bangunan menggunakan *naïve bayes* menghasilkan akurasi sebesar 0.71 (Valentinus et al. 2023).

Untuk penelitian menggunakan *neural network* didapatkan akurasi sebesar 63,63% dengan menggunakan data testing 20% (R Habibah and Solichin 2022). Hasil prediksi kinerja mahasiswa dapat dimanfaatkan dalam melaksanakan monev (monitoring dan evaluasi) terhadap jalanya proses pembelajaran dalam upaya meningkatkan kualitas pembelajaran (Muhammad Haris Diponegoro, Sri Suning Kusumawardani, and Indriana Hidayah 2021). 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*.

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 dapat dilihat pada gambar 1.



Gambar 1: Metode Penelitian

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. Preprocessing

Preprocessing merupakan langkah yang sangat penting karena dapat mempengaruhi kualitas data (Suriyani 2023) 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 (Haryatmi and Pramita Hervianti 2021), pada tahap ini data yang kosong akan dihapus dan tidak diikuti dalam tahap berikutnya. Karena datanya mempunyai perbedaan nilai dan beragam, maka dilakukan proses membuat skala data agar data yang diproses lebih cepat.

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 (Oktafiani et al. 2023). *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*.

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 (Abuzinadah et al. 2023) 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.

5. Evaluation

Tahap terakhir adalah evaluasi menggunakan *confusion matrix* yang merupakan matrix untuk menggambarkan hasil model klasifikasi untuk dataset yang akan diuji (Sutoyo and Almaarif 2020). *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 (Suliztia and Fauzan 2019).

HASIL DAN PEMBAHASAN

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.

2. Preprocessing

a. Data selection

Dari data yang sudah terkumpul, terdapat fitur yang tidak diikuti 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, id 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 merupakan 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
id_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
id_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 x 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,  0.3398012 ],
       ...,
       [  1.35784756, -0.43536477, -0.47518494, ...,  0.21275738,
        -0.71635827,  0.3398012 ],
       [ -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. Splitting Data

Setelah data disiapkan 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 4.

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 90% berbanding 10%, 80% berbanding 20%, 70% berbanding 30%, 60% berbanding 40% dan 50% berbanding 50%.

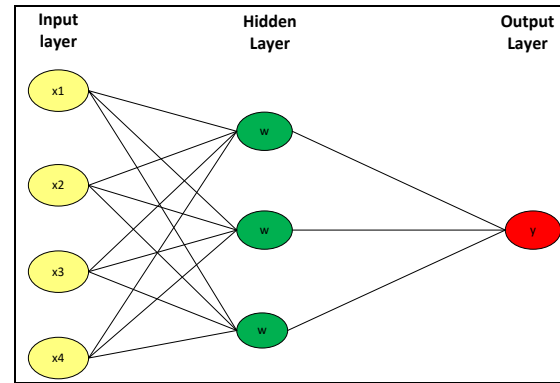
4. Modelling

Tahap *modelling* ini dilaksanakan menggunakan algoritma *machine learning* yaitu *naïve 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 *naïve bayes* telah berhasil. Adapun fungsi algoritma *naïve bayes* tersebut dapat mengasumsikan semua fitur dalam model adalah independen (Abuzinadah et al. 2023). Sehingga dapat digunakan untuk mengklasifikasikan objek menggunakan data yang terdistribusi secara seragam. Penghitungan tersebut menggunakan rumus :

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

$$P(c|x) = P(x_1|x) * \dots, P(x_1|x) * P(c) \quad (2)$$

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 (Wibowo Putra et al. 2022).



Gambar 3 : Model arsitektur neural network

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 naïve 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 terlihat pada tabel 6.

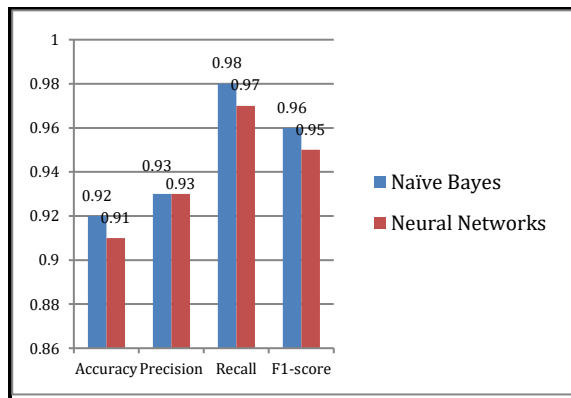
Tabel 6 : Hasil evaluasi algoritma neural network

No	Split	Akurasi	%	Precisi on	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 terlihat pada gambar 4.



Gambar 4 : Grafik hasil evaluasi

Hasil evaluasi perbandingan antara pemodelan *naïve bayes* dan *neural network*, seperti yang terlihat pada grafik gambar 4, menunjukkan bahwa nilai accuracy untuk *naïve bayes* adalah 0,92 sedangkan untuk *neural network* adalah 0,91, perbandingan nilai precision keduanya sama, yaitu 0,93. sementara itu, nilai recall untuk *naïve bayes* adalah 0,98 sedangkan untuk *neural network* adalah 0,97. Perbandingan nilai F1-score untuk *naïve bayes* adalah 0,96, sedangkan untuk *neural network* adalah 0,95.

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% sehingga, Fakultas dapat menggunakan model *naive bayes* untuk melakukan prediksi terkait ketepatan studi mahasiswa. Hal ini bertujuan untuk mengantisipasi dan mengatasi permasalahan yang mungkin timbul terkait kelulusan mahasiswa dengan tepat waktu. 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.

REFERENSI

- Abuzinadah, Nihal, Muhammad Umer, Abid Ishaq, Abdullah Al Hejaili, Shtwai Alsubai, Ala' Abdulmajid Eshmawi, Abdullah Mohamed, and Imran Ashraf. 2023. "Role of Convolutional Features and Machine Learning for Predicting Student Academic Performance from MOODLE Data" edited by M. Hammad. *PLOS ONE* 18(11):e0293061. doi: 10.1371/journal.pone.0293061.
- Amalia, Hilda, Ari Puspita, Ade Fitria Lestari, and Frieyadie Frieyadie. 2022. "Application Of Decision Tree And Naive Bayes On Student Performance Dataset." *Jurnal Pilar Nusa Mandiri* 18(1):53–58. doi: 10.33480/pilar.v18i1.2714.
- Awangga, Rolly Maulana, and Nuha Hanifatul Khonsa'. 2022. "Analisis Performa Algoritma Random Forest Dan Naive Bayes Multinomial Pada Dataset Ulasan Obat Dan Ulasan Film." *InComTech: Jurnal Telekomunikasi Dan Komputer* 12(1):60. doi: 10.22441/incomtech.v12i1.14770.

- Azahari, Azahari, Yulindawati Yulindawati, Dewi Rosita, and Syamsuddin Mallala. 2020. "Komparasi Data Mining Naive Bayes dan Neural Network memprediksi Masa Studi Mahasiswa S1." *Jurnal Teknologi Informasi dan Ilmu Komputer* 7(3):443–52. doi: 10.25126/jtiik.2020732093.
- Haryatmi, Emy, and Sheila Pramita Hervianti. 2021. "Penerapan Algoritma Support Vector Machine Untuk Model Prediksi Kelulusan Mahasiswa Tepat Waktu." *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)* 5(2):386–92. doi: 10.29207/resti.v5i2.3007.
- Hasibuan, Tuhfatul Habibah, and Deni Mahdiana. 2023. "Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Algoritma C4.5 Pada Uin Syarif Hidayatullah Jakarta." *SKANIKA* 6(1):61–74. doi: 10.36080/skanika.v6i1.2976.
- Irawan, Indra, M. Riski Qisthiano, Muhammad Syahril, and Pamuji M. Jakak. 2023. "Optimasi Prediksi Kelulusan Tepat Waktu: Studi Perbandingan Algoritma Random Forest dan Algoritma K-NN Berbasis PSO." *Jurnal Pengembangan Sistem Informasi dan Informatika* 4(4):26–35. doi: 10.47747/jpsii.v4i4.1374.
- Issah, Iddrisu, Obed Appiah, Peter Appiahene, and Fuseini Inusah. 2023. "A Systematic Review of the Literature on Machine Learning Application of Determining the Attributes Influencing Academic Performance." *Decision Analytics Journal* 7:100204. doi: 10.1016/j.dajour.2023.100204.
- Jananto, Arief, Sulastri Sulastri, Eko Nur Wahyudi, and Sunardi Sunardi. 2021. "Data Induk Mahasiswa sebagai Prediktor Ketepatan Waktu Lulus Menggunakan Algoritma CART Klasifikasi Data Mining." *Jurnal Sisfokom (Sistem Informasi dan Komputer)* 10(1):71–78. doi: 10.32736/sisfokom.v10i1.991.
- Kurniadi, Dede, Fitri Nuraeni, and Sri Mulyani Lestari. 2022. "Implementasi Algoritma Naïve Bayes Menggunakan Feature Forward Selection dan SMOTE Untuk Memprediksi Ketepatan Masa Studi Mahasiswa Sarjana." *Jurnal Sistem Cerdas* 5(2):63–82. doi: 10.37396/jsc.v5i2.215.
- Muhamad Ziaul Haq and Nursalim. 2023. "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." *Jurnal Kolaboratif Sains* 6(2):147–53. doi: 10.56338/jks.v6i2.3303.
- Muhammad Haris Diponegoro, Sri Suning Kusumawardani, and Indriana Hidayah. 2021. "Tinjauan Pustaka Sistematis: Implementasi Metode Deep Learning pada Prediksi Kinerja Murid." *Jurnal Nasional Teknik Elektro dan Teknologi Informasi* 10(2):131–38. doi: 10.22146/jnteti.v10i2.1417.
- Nuralia, Siti, Harliana Harliana, and Tito Prabowo. 2023. "Implementasi Naïve Bayes Classifier Dalam Memprediksi Kelulusan Mahasiswa." *Journal Automation Computer Information System* 3(1):63–72. doi: 10.47134/jacis.v3i1.57.
- Oktafiani, Rian, Arief Hermawan, and Donny Avianto. 2023. "Pengaruh Komposisi Split Data Terhadap Performa Klasifikasi Penyakit Kanker Payudara Menggunakan Algoritma Machine Learning." *Jurnal Sains Dan Informatika* 19–28. doi: 10.34128/jsi.v9i1.622.
- R Habibah, Ummu, and Achmad Solichin. 2022. "Prediksi Kelulusan Mahasiswa Dengan Metode Naïve Bayes dan Artificial Neural Network: Studi Kasus Fakultas Teknik UNIS Tangerang." *Faktor Exacta* 15(1):73~83. doi: 10.30998/faktorexacta.v15i1.11816.
- Rolansa, Freska, Yunita Yunita, and Suheri Suheri. 2020. "Sistem prediksi dan evaluasi prestasi akademik mahasiswa di Program Studi Teknik Informatika menggunakan data mining." *Jurnal*

- Pendidikan Informatika dan Sains*
9(1):75. doi:
10.31571/saintek.v9i1.1696.
- Santoso, Leo Willyanto, and Yulia Yulia. 2020. "Predicting Student Performance in Higher Education Using Multi-Regression Models." *TELKOMNIKA (Telecommunication Computing Electronics and Control)* 18(3):1354. doi: 10.12928/telkomnika.v18i3.14802.
- Suliztia, Mega Luna, and Achmad Fauzan. 2019. "Comparing Naive Bayes, K-Nearest Neighbor, And Neural Network Classification Methods Of Seat Load Factor In Lombok Outbound Flights." *Jurnal Matematika, Statistika dan Komputasi* 16(2):187. doi: 10.20956/jmsk.v16i2.7864.
- Suriani, Uci. 2023. "Penerapan Data Mining untuk Memprediksi Tingkat Kelulusan Mahasiswa Menggunakan Algoritma Decision Tree C4.5." *Journal of Computer and Information Systems Ampera* 3(2). doi: 10.51519/journalcisa.v4i2.393.
- Sutoyo, Edi, and Ahmad Almaarif. 2020. "Educational Data Mining untuk Prediksi Kelulusan Mahasiswa Menggunakan Algoritme Naïve Bayes Classifier." *JURNAL RESTI* 4(1):95–101.
- Valentinus, Fernando, Fabian Sujono, Ilham Ariansyah, and Dwi Ade Handayani Capah. 2023. "Implementation Of Data Mining With Classification And Forecasting Method Use Model Gaussian Naïve Bayes For Building Store (Studi Case: Tb Sinar Jaya)." *Jurnal Teknik Informatika (Jutif)* 4(2):413–20. doi: 10.52436/1.jutif.2023.4.2.701.
- Wibowo Putra, Ari, Widiyono Widiyono, Anas Saifudin, Arief Darmawan Soma, and Eko Budihartono. 2022. "Naïve Bayes, Neural Network dan K-Nearest Neighbor untuk Klasifikasi Topik Tugas Akhir." *Smart Comp: Jurnalnya Orang Pintar Komputer* 11(4). doi: 10.30591/smartcomp.v11i4.4251.