



Klasifikasi Mahasiswa Berpotensi Putus studi menggunakan Algoritma Naive Bayes pada Fakultas Teknik Unismuh Makassar

<u>INFO PENULIS</u>	<u>INFO ARTIKEL</u>
Rianita Kamsurya Universitas Muhammadiyah Makassar 105841102620@student.unismuh.ac.id	ISSN: 3026-3603 Vol. 2, No. 2 Oktober 2024 http://jurnal.ardenjaya.com/index.php/ajst
Muhyddin A.M Hayat Universitas Muhammadiyah Makassar	
Rizki Yusliana Bakti Universitas Muhammadiyah Makassar	

© 2024 Arden Jaya Publisher All rights reserved

Saran Penulisan Referensi:

Kamsurya, R., Hayat, M. A. M., & Bakti, R. Y. (2024). Klasifikasi Mahasiswa Berpotensi Putus Studi menggunakan Algoritma Naive Bayes pada Fakultas Teknik Unismuh Makassar. *Arus Jurnal Sains dan Teknologi*, 2 (2), 511-517.

Abstrak

Penelitian ini bertujuan untuk mengidentifikasi dan menganalisis faktor-faktor yang mempengaruhi potensi putus studi mahasiswa di Fakultas Teknik Universitas Muhammadiyah Makassar. Penelitian ini dilakukan di Universitas Muhammadiyah Makassar, khususnya di Fakultas Teknik, pada periode 2013 sampai 2015 Metode penelitian yang digunakan meliputi pengumpulan data, praproses data, pembagian data, klasifikasi menggunakan algoritma Naive Bayes, dan pengujian sistem. Dari hasil yang didapat, algoritma naive bayes memiliki hasil peforma yang cukup tinggi dengan akurasi sebesar 97%. Sehingga algoritma ini menjadi salah satu algoritma yang baik dalam mengklasifikasi mahasiswa potensi putus studi dan tidak potensi pada jurusan Teknik Informatika Universitas Muhammadiyah Makassar. Pengujian menggunakan 10-fold cross-validation menunjukkan rata-rata akurasi 95,38%. Hasil ini mengindikasikan bahwa algoritma Naive Bayes dapat memberikan hasil yang konsisten dan andal dalam mengklasifikasikan mahasiswa berpotensi putus studi.

Kata kunci: Naive Bayes, Klasifikasi, Preprocessing Data, Cross-Validation, Status Putus Studi.

Abstract

This study aims to identify and analyze the factors that influence the potential for students to drop out of study at the Faculty of Engineering, Muhammadiyah University of Makassar. This research was conducted at Muhammadiyah University of Makassar, specifically at the Faculty of Engineering, in the period of 2013 to 2016. The research methods used include data collection, data preprocessing, data division, classification using the Naive Bayes algorithm, and system testing. From the results obtained, the naive bayes algorithm has a fairly high performance with an accuracy of 97%. So this algorithm is one of the good algorithms in classifying students with the potential to drop out of study and those without potential in the Informatics Engineering Department, Muhammadiyah University of Makassar. Testing using 10-fold cross-validation showed average accuracy of 95,38%. These results indicate that the Naive Bayes algorithm can provide consistent and reliable results in classifying students with the potential to drop out of study.

Keywords: Naive Bayes, Classification, Data Preprocessing, Cross-Validation, Study Dropout Status.

A. Pendahuluan

Banyaknya mahasiswa yang putus studi adalah masalah yang saat ini dihadapi oleh banyak institusi pendidikan tinggi di berbagai negara. Mahasiswa yang putus studi dapat menghambat pertumbuhan ekonomi, daya saing, dan produktivitas, yang berdampak pada siswa dan masyarakat secara keseluruhan (Realinho, Machado, Baptista, & Martins, 2022). (Wijayaningrum et al., 2019). Permasalahan mahasiswa putus studi di Universitas Muhammadiyah Makassar merupakan isu penting dan kompleks yang dipengaruhi oleh berbagai faktor, baik internal maupun eksternal. Aspek internal meliputi rendahnya motivasi belajar, kesulitan akademik dan masalah kesehatan mental, sedangkan faktor eksternal mencakup masalah keuangan, kondisi keluarga yang tidak mendukung, lingkungan sosial yang tidak kondusif, dan tekanan untuk bekerja daripada melanjutkan studi. Penentuan mahasiswa yang putus studi saat ini masih dilakukan secara manual yang tidak hanya memakan waktu dan tenaga tetapi juga rentan terhadap kesalahan dan kurang efisien dalam mengidentifikasi mahasiswa yang berisiko. Proses manual ini sering kali tidak mampu memberikan gambaran akurat tentang berbagai faktor penentu. Untuk mengatasi masalah ini diperlukan pendekatan yang lebih sistematis dan berbasis data. Salah satu solusi yang diusulkan adalah melakukan pengembangan sistem klasifikasi tingkat putus studi mahasiswa.

Klasifikasi adalah proses dalam penggalan data yang bertujuan untuk menemukan model atau fungsi yang mendeskripsikan data dan membedakan data ke dalam kelas-kelas. Proses ini melibatkan memeriksa karakteristik objek untuk memastikan bahwa objek tersebut dimasukkan ke dalam salah satu kelas yang sudah didefinisikan sebelumnya. Praproses data adalah proses persiapan data mentah sebelum dilakukan proses klasifikasi data (Salmawati, Yuyun, 2021).

Salah satu algoritma yang digunakan untuk klasifikasi data adalah Algoritma Naive Bayes. Naive Bayes merupakan salah satu metode pembelajaran mesin yang memanfaatkan perhitungan probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi probabilitas dimasa depan berdasarkan pengalaman di masa sebelumnya (Samasil et al., 2022). Terdapat penelitian terdahulu yang menggunakan perbandingan metode Deep Learning, Naive Bayes, dan Random Forest untuk menerapkan data mining pada penerimaan mahasiswa baru di Universitas Islam Negeri Raden Fatah. Hasil menunjukkan bahwa metode Naive Bayes menghasilkan akurasi sebesar 99.79%, jauh lebih tinggi daripada metode lainnya, seperti Deep Learning (52,65%) dan Random Forest (44,65%). Algoritma Naive Bayes juga dianggap lebih akurat dan sering menjadi pilihan terbaik. (Ulinuha & Fanani, 2022).

Berdasarkan latar belakang diatas, penelitian ini akan fokus pada pengembangan sistem klasifikasi tingkat putus studi mahasiswa di UNISMUH Makassar dengan menggunakan algoritma Naive Bayes. Penelitian ini diharapkan dapat memberikan kontribusi signifikan dalam upaya peningkatan kualitas pendidikan dan memberikan kontribusi signifikan dalam upaya peningkatan kualitas pendidikan dan pengurangan angka putus studi.

B. Metodologi

1. Klasifikasi

Dalam proses pembelajaran mesin, klasifikasi adalah tugas memprediksi label kelas dari sampel yang diberikan berdasarkan sekumpulan fitur atau karakteristik (Syahril Dwi Prasetyo et al., 2022)

2. Mahasiswa berpotensi Putus studi

Secara umum ukuran keberhasilan studi adalah ketika mahasiswa mampu menyelesaikan program pendidikan sesuai kecepatan belajar masing-masing dengan tidak melebihi ketentuan batas waktu yang ditetapkan perguruan tinggi. Dalam rangka memaksimalkan tugas perguruan tinggi dalam mengembangkan potensi untuk keberhasilan studi, identifikasi dini atas potensi mahasiswa perlu dilakukan guna mengetahui kemungkinan keberhasilan mahasiswa dalam penyelesaian studi dan mengantisipasi kemungkinan ketidakberhasilan. Penting untuk memprediksi, apakah seorang mahasiswa yang masuk perguruan tinggi akan dapat keluar atau menyelesaikan studi (Yuniarti et al., 2020)

3. Algoritma Naive Bayes

Naive Bayes adalah salah satu metode yang digunakan untuk mengklasifikasikan data.

Bayesian classification merupakan pengklasifikasian statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu class. *Naive bayes* dianggap sebagai penyederhanaan nilai atribut secara kondisional saling bebas jika diberi nilai output. Keuntungan menggunakan *Naive bayes* adalah metode ini hanya membutuhkan dua data yaitu data latih dan data uji untuk menguji data yang diinginkan. *Naive bayes* sering bekerja jauh lebih baik dari yang diharapkan dalam situasi dunia nyata yang paling kompleks (Zarti et al., 2022).

4. Cross Validation

Metode atau algoritma cross-validation adalah tindakan pembuktian. Ini dilakukan dengan membagi data menjadi data pelatihan dan data tes dengan komposisi yang berbeda. Pembagian data acak menjadi sepuluh bagian, dengan satu bagian untuk pemeriksaan data dan sembilan bagian lainnya untuk pelatihan data, adalah pembagian yang paling sering digunakan dalam penelitian klasifikasi data mining. Salah satu nama validasi ini adalah k-fold Cross Validation, dengan k=10. Menurut penelitian Mulajati (2017),

5. Confusion matrix

Konfussion matrix adalah hasil evaluasi klasifikasi data mining yang ditampilkan dalam sebuah tabel dan berisi perhitungan jumlah objek pengujian data yang diproyeksikan masuk ke dalam sebuah kelas dengan klasifikasi yang sebenarnya. Konfussion matrix menghasilkan pengukuran kinerja klasifikasi dalam bentuk nilai ketepatan dan nilai precision dan recall. Nilai ketepatan dan recall ini berkisar antara 0-1. Semakin besar nilai yang diterima, semakin baik (Salmawati, Yuyun, 2021).

C. Hasil dan Pembahasan

1. Analisis Data Mentah

Data yang diolah dalam penelitian ini berasal dari mahasiswa Program Studi Teknik Informatika di Fakultas Teknik Universitas Muhammadiyah Makassar, yang terdiri dari angkatan 2013-2015. Data set ini terdiri dari berbagai atribut seperti mahasiswa (nim, kodeProdi, angkatan, nama, jenisKelamin, semesterAwal, tahunAkademikLulus, lulus, masaStudi) ayah (pekerjaan, penghasilan, status) ibu (pekerjaan, penghasilan, status) wali (pekerjaan, penghasilan) khs (tahunAkademik, ips, sksSmt, ipk, sksTotal, statusMahasiswa).

kodeProdi	angkatan	semesterAwal	nim	nama	jenisKelamin	tahunAkademikLulus	tanggalLulus	lulus	masaStudi	ayah_pekerjaan	ayah_penghasilan	ayah_status	ibu_pekerjaan	ibu_penghasilan
0	22202	2013	20131_10581182313	MUHAMMAD FARID HIKSYAL	L	None	None	False	None	None	None	None	None	None
1	22202	2013	20131_10581182413	MUHAMMAD SAKIR	L	20182	None	True	5 Tahun, 11 Bulan	TANI	500000	HIDUP	-	500000
2	22202	2013	20131_10581182513	FAISAL AKBAR	L	None	None	False	None	None	None	None	None	None
3	22202	2013	20131_10581182613	MUNIR MUHAIDIR	L	None	None	False	None	None	None	None	None	None
4	22202	2013	20131_10581182713	ST FATIMA	P	20172	None	True	4 Tahun, 8 Bulan	WIRASWASTA	500001-1000000	HIDUP	WIRASWASTA	500001-1000000
5	22202	2013	20131_10581182813	SAIFUL ROZQAQ	L	20191	None	True	6 Tahun, 2 Bulan	WIRASWASTA	500001-1000000	HIDUP	WIRASWASTA	500000
6	22202	2013	20131_10581182913	JUPRI	L	None	None	False	None	None	None	None	None	None
7	22202	2013	20131_10581183013	HENRA ADI SAPUTRA	L	20182	None	True	5 Tahun, 9 Bulan	WIRASWASTA	2000001-3000000	HIDUP	IRT	500000
8	22202	2013	20131_10581183113	MOHAMMAD HABIR	L	None	None	False	None	TANI	500000	MENINGGAL	-	500001-1000000
9	22202	2013	20131_10581183213	HILMAN ARSANDI TASLIM	L	None	None	False	None	TANI	500001-1000000	HIDUP	WIRASWASTA	500001-1000000

Gambar 1. Data Mentah

2. Data Cleaning

Data cleaning atau pembersihan data pada penelitian ini dilakukan dengan menggunakan nilai mean atau nilai yang valid. *Data missing value* merupakan data yang bernilai kosong (*null*) yang tidak dapat digunakan dalam penelitian. Jumlah data *missing value* pada tiap atribut dapat dilihat pada tabel berikut.

Tabel 1. Jumlah Data Missing Value

Atribut	Jumlah Data Missing Value
angkatan	0
semesterAwal	0
jenisKelamin	0
ayah_pekerjaan	410

ayah_penghasilan	421
ayah_status	416
ibu_pekerjaan	411
ibu_penghasilan	428
ibu_status	416
wali_pekerjaan	1266
wali_penghasilan	1298
lulus	0
semesterTerakhir	0
kategori_IPK	4
kategori_SKS	0
status_putus_studi	0

Sumber: Pengolahan data, 2024

Jumlah *missing value* diatas merupakan jumlah data kosong pada tiap kolom atribut. Dalam satu baris data terdapat lebih dari satu atribut yang memiliki data *missing value*. Dari 16 atribut yang digunakan, 8 diantaranya memiliki data *missing value* yaitu atribut atribut ayah_pekerjaan sebesar 410 data, atribut ayah_penghasilan sebesar 421 data, atribut ayah_status sebesar 416 data, atribut ibu_pekerjaan sebesar 411 data, atribut ibu_penghasilan sebesar 428 data, atribut ibu_status sebesar 416 data, atribut wali_pekerjaan sebesar 1266 data, atribut wali_penghasilan

ibu_pekerjaan	ibu_penghasilan	ibu_status	wali_pekerjaan	wali_penghasilan	lulus	semesterTerakhir	k
NaN	NaN	NaN	NaN	NaN	False		8
500000	500000	HIDUP	NaN	NaN	True		12
NaN	NaN	NaN	NaN	NaN	False		3
NaN	NaN	NaN	NaN	NaN	False		3
WIRASWASTA	500001-1000000	HIDUP	NaN	NaN	True		10

sebesar 1298 data. Berikut contoh data *missing value*:"

Gambar 2. Data Missing Value

Tabel 2. Jumlah Data Missing Value Setelah menggunakan Mean/Data Valid

Atribut	Jumlah Data <i>Missing Value</i>
angkatan	0
semesterAwal	0
jenisKelamin	0
ayah_pekerjaan	0
ayah_penghasilan	0
ayah_status	0
ibu_pekerjaan	0
ibu_penghasilan	0
ibu_status	0
wali_pekerjaan	0
wali_penghasilan	0
lulus	0
semesterTerakhir	0
kategori_IPK	0
kategori_SKS	0
status_putus_studi	0

Sumber: Pengolahan data, 2024

Dari gambar diatas dapat dilihat bahwa tidak ada lagi data *missing value* atau data yang kosong. Jumlah data pada proses pembersihan data dapat dilihat pada tabel berikut:

Tabel 3. Jumlah Data Hasil Proses Pembersihan

	Jumlah Data
Data awal	1493
Data <i>missing</i>	5.070
Data bersih	1493

Sumber: Pengolahan data, 2024

3. Pembagian Data

Dalam penelitian ini data akan di bagi dengan proporsi 80% data training dan 20% data testing. Berikut jumlah data setelah dilakukan pembagian:

Tabel 4. Pembagian Data

Klasifikasi	Jumlah Data	Data Training (80%)	Data Testing (20%)
Potensial Putus Studi	722	582	140
Tidak Potensial	771	612	159
Total	1.493	1.194	299

Sumber: Pengolahan data, 2024

Dari tabel diatas diketahui bahwa jumlah data training yang dihasilkan adalah 1.194 data dan data testing berjumlah 299 data. Berikut kode program dan hasilnya:

4. Klasifikasi Menggunakan Algoritma *Naïve Bayes*

Setelah data dibagi menjadi data training dan data testing, tahap selanjutnya ialah proses klasifikasi data. Pada tahap klasifikasi data, akan dilakukan pembangunan model menggunakan algoritma *naïve bayes*. Pada proses ini data training akan mempelajari pola dari algoritma sehingga didapat model yang dihasilkan. Setelah itu dapat dilihat performa dari algoritma yang telah dilatih berdasarkan akurasi yang dihasilkan dari data testing atau data uji. Berikut hasil akurasi dari algoritma *naïve bayes*:

Tabel 5. Hasil Akurasi

Algoritma	Akurasi (%)
<i>Naïve Bayes</i>	97%

Sumber: Pengolahan data, 2024

5. Pengujian Menggunakan *Cross Validation*

Berikut Hasilnya :

Tabel 6. Hasil pengujian 10-fold cross validation

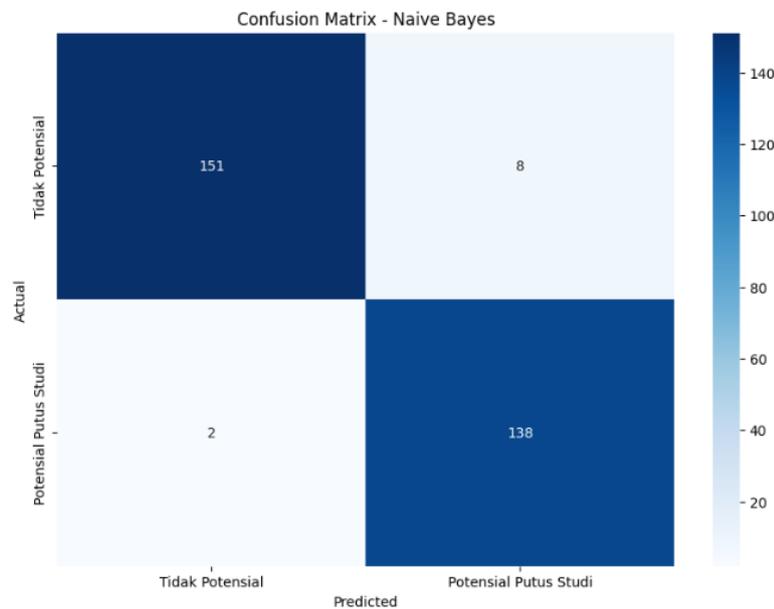
<i>K-Fold</i>	Akurasi (%)
1	95,33%
2	95,33%
3	95,33%
4	97,99%
5	96,64%
6	95,30%
7	93,29%
8	96,64%
9	95,30%
10	92,62%
Rata-Rata	95,38%

Sumber: Pengolahan data, 2024

Berdasarkan hasil dari pengujian diatas dapat dilihat bahwa akurasi algoritma *naïve bayes* sama dengan hasil akurasi sebelum dilakukan pengujian. Rata-rata akurasi dari setiap *fold* yang didapat ialah sebesar **95,38%**

6. Evaluasi Hasil Klasifikasi

Evaluasi hasil klasifikasi pada algoritma *naïve bayes* juga menggunakan metode *confusion matrix*, hasil evaluasi dapat dilihat pada gambar berikut:



Gambar 3. Confussion Matrix Pada Algoritma Naive Bayes

Dari gambar diatas, hasil evaluasi menggunakan *confussion matrix* menghasilkan jumlah data TP sebesar 151 data, jumlah FN sebesar 8 data, data FP berjumlah 2 data dan TN berjumlah 138 data. Lebih jelasnya dapat dilihat pada tabel berikut:

Tabel 7. Confussion Matrix Pada Algoritma Naive Bayes

Data Sebenarnya	Klasifikasi	
	Potensial Putus Studi	Tidak Potensial
Potensial Putus Studi	151 (TP)	8 (FN)
Tidak Potensial	2 (FP)	138 (TN)

Sumber: Pengolahan data, 2024

Pengukuran kinerja algoritma *naive bayes* dapat dilihat pada nilai *akurasi*, *precision*, *recall*, dan *f1-score* yang dihasilkan berdasarkan data dari *confussion matrix*. Berikut nilai, *precision*, *recall* dan *f1-score* yang dihasilkan

```
Confusion Matrix:
[[151  8]
 [  2 138]]
Classification Report:
              precision    recall  f1-score   support

     0       0.99      0.95      0.97       159
     1       0.95      0.99      0.97       140

 accuracy          0.97          299
 macro avg          0.97          299
weighted avg          0.97          299
```

Gambar 4. Hasil Kinerja Algoritma

Pada gambar di

dilihat bahwa hasil *precision* untuk data Potensial Putus Studi dan Tidak Potensial masing-masing sebesar 99% dan 95%. Nilai *recall* dari data Potensial Putus Studi dan Tidak Potensial yang dihasilkan yaitu 97% dan 95%. Selanjutnya untuk nilai *f1-score* dari data Potensial Putus Studi dan Tidak Potensial yang didapat yaitu 97% dan 97%. Nilai rata-rata (*macro average*) dari *precision*, *recall* dan *f1-score* masing-masing sebesar 97%, 97% dan 97%.

```
Rata-rata Precision: 0.9660667920136091
Rata-rata Recall: 0.9676999101527404
Rata-rata F1-Score: 0.9664918414918415
```

Pengukuran *Naive Bayes* atas dapat

D. Kesimpulan

Algoritma Naive Bayes telah berhasil diterapkan untuk mengklasifikasikan mahasiswa yang berpotensi putus studi. Proses klasifikasi dilakukan dengan mengolah dataset yang terdiri dari atribut-atribut seperti pekerjaan dan penghasilan orang tua/wali, IPK, SKS, serta variabel lainnya yang relevan dengan status akademik mahasiswa. Penggunaan Naive Bayes memungkinkan pemodelan probabilistik yang sederhana namun efektif dalam mengidentifikasi mahasiswa dengan risiko putus studi berdasarkan pola yang ada dalam data.

Dari hasil yang didapat, algoritma *naive bayes* memiliki hasil peforma yang cukup tinggi dengan akurasi sebesar 99%. Sehingga algoritma ini menjadi salah satu algoritma yang baik

dalam mengklasifikasi mahasiswa potensi putus studi dan tidak potensi pada jurusan Teknik Informatika Universitas Muhammadiyah Makassar.

Saran

Saran untuk penelitian selanjutnya yaitu:

1. Untuk proses klasifikasi dapat menggunakan algoritma yang lain seperti *Decision Tree C4.5*, *KNearest Neighbour* dan *Support Vector Machines (SVM)*. Bisa juga dengan melakukan 2 perbandingan algoritma klasifikasi atau bahkan lebih.
2. Untuk pengukuran kinerja klasifikasi dapat ditambah dengan metode lain agar hasil lebih optimal.

E. Referensi

- Andreansyach** , C. R., Fardiansyah, T. S., Apriani, D., Sani, A., & Kunci, K. (2022). *Prediksi Persentase Kelulusan Mahasiswa STMIK Widuri Menggunakan Algoritma Naïve Bayes*. 75-84.
- Arum, L., Putri, B., Matematika, P. S., Sains, F., Teknologi, D. A. N., Islam, U., Maulana, N., & Ibrahim, M. (2022). *KLASIFIKASI FAKTOR PENYEBAB SISWA PUTUS SEKOLAH MENGGUNAKAN METODE DECISION TREE ID3* *KLASIFIKASI FAKTOR PENYEBAB SISWA PUTUS SEKOLAH MENGGUNAKAN METODE DECISION TREE ID3*.
- Kamila, V. Z., & Subastian, E. (2019). KNN vs Naive Bayes Untuk Deteksi Dini Putus Kuliah Pada Profil Akademik Mahasiswa. *Jurnal Rekayasa Teknologi Informasi (JURTI)*, 3(2), 116. <https://doi.org/10.30872/jurti.v3i2.3097>
- Mahanggara, A., & Laksito, A. D. (2019). Prediksi Pengunduran Diri Mahasiswa Universitas Amikom Yogyakarta Menggunakan Metode Naive Bayes. *Simetris: Jurnal Teknik Mesin, Elektro dan Ilmu Komputer*, 10(1), 273-280. <https://doi.org/10.24176/simet.v10i1.2967>
- Mulyadi, C., & Juniadi, M. N. (2019). *Prediksi Keaktifan Studi Mahasiswa Baru dengan Algoritma Naive Bayes*. 299-303.
- Nuralia, S., Harliana, H., & Prabowo, T. (2022). Implementasi Naïve Bayes Classifier Dalam Memprediksi Kelulusan Mahasiswa. *Journal Automation Computer Information System*, 3(1), 63-72. <https://doi.org/10.47134/jacis.v3i1.57>
- Rayuwati, Husna Gemasih, & Irma Nizar. (2022). IMPLEMENTASI ALGORITMA NAIVE BAYES UNTUK MEMPREDIKSI TINGKAT PENYEBARAN COVID. *Jural Riset Rumpun Ilmu Teknik*, 1(1), 38-46. <https://doi.org/10.55606/jurritek.v1i1.127>
- Rosaly, R., & Prasetyo, A. (2020). Flowchart Beserta Fungsi dan Simbol-Simbol. *Journal of Chemical Information and Modeling*, 2(3), 5-7.
- Salmawati, Yuyun, H. (2021). Klasifikasi Mahasiswa Berpotensi Drop Out Menggunakan Algoritma Decision Tree C4 . 5 Dan Naive Bayes Di Universitas Jambi. *Jurnal Ilmiah Ilmu Komputer*, 8(2), 115032. <https://repository.unja.ac.id/25341/>
- Samasil, S., Yuyun, Y., & Hazriani, H. (2022). KLASIFIKASI MAHASISWA BERPOTENSI DROP OUT MENGGUNAKAN ALGORITMA NAIVE BAYES DAN DECISION TREE. *JURNAL ILMIAH ILMU KOMPUTER*, 8(2), 108-114. <https://doi.org/10.35329/jiik.v8i2.242>
- Syahril Dwi Prasetyo, Shofa Shofiah Hilabi, & Fitri Nurapriani. (2022). Analisis Sentimen Relokasi Ibukota Nusantara Menggunakan Algoritma Naïve Bayes dan KNN. *Jurnal KomtekInfo*, 10, 1-7. <https://doi.org/10.35134/komtekinfo.v10i1.330>
- Ulinnuha, N., & Fanani, A. (2022). Klasifikasi Status Drop Out Mahasiswa Menggunakan Naïve Bayes dengan Seleksi Fitur Information Gain. *Techno.Com*, 22(4), 1014-1025. <https://doi.org/10.33633/tc.v22i4.9004>
- Wijayaningrum, V. N., Putri, I. K., Kirana, A. P., Mubarak, M. R., Harahap, D. M., Hamesha, B. R., Informasi, J. T., Malang, P. N., Correlation, P., & Syaraf, J. (2019). *Analisis performa seleksi atribut untuk menentukan potensi mahasiswa putus studi*. 237-244.
- Yuniarti, W. D., Faiz, A. N., & Setiawan, B. (2020). Identifikasi Potensi Keberhasilan Studi Menggunakan Naïve Bayes Classifier. *Walisongo Journal of Information Technology*, 2(1), 1. <https://doi.org/10.21580/wjit.2020.2.1.5204>
- Zarti, M. N., Sahputra, E., Sonita, A., & Apridiansyah, Y. (2022). Application Of Data Mining Using The Naïve Bayes Classification Method To Predict Public Interest Participation In The 2024 Elections. *Jurnal Komputer, Informasi dan Teknologi*, 3(1), 105-114. <https://doi.org/10.53697/jkomitek.v3i1.1192>