

Klasifikasi Bidang Minat Mahasiswa Elektronika Dalam Menentukan Topik Tugas Akhir Menggunakan Algoritma *Naïve Bayes Classifier* (Studi Kasus: Prodi Pendidikan Teknik Informatika FT-UNP)

Fakhri Malja¹, Denny Kurniadi², Khairi Budayawan³, Resmi Darni⁴

¹Program Studi Pendidikan Teknik Informatika, Universitas Negeri Padang

²Departemen Teknik Elektronika, Universitas Negeri Padang
e-mail: maljafakhri@gmail.com¹, dennykurniadi@ft.unp.ac.id²,
khairi@ft.unp.ac.id³, resmidarni@ft.unp.ac.id⁴

Abstrak

Konsentrasi bidang studi merupakan pengkhususan studi yang harus diambil seorang mahasiswa dalam sebuah jurusan atau bidang studi yang nantinya akan mengarahkan mahasiswa ke bidang yang lebih spesifik. Mata kuliah konsentrasi merupakan sarana untuk menguasai pengetahuan dan keterampilan yang terarah pada penyusunan tugas akhir. Tidak sedikit mahasiswa yang tidak menyelesaikan tugas akhir karena topik yang diambil tidak sesuai dengan peminatannya. Penelitian ini bertujuan untuk memberikan rekomendasi konsentrasi bidang studi mahasiswa program studi Pendidikan Teknik Informatika dan Komputer Universitas Negeri Padang menggunakan hasil model klasifikasi *naïve bayes classifier* dan dibantu dengan aplikasi RapidMiner. Klasifikasi ini menggunakan 24 atribut yang merupakan transkrip nilai semester 1 sampai 4. Hasil klasifikasi dari aplikasi ini menyatakan bahwa dari 76 set data testing memperoleh hasil accuracy 88.16%, precision 100%, recall 41,67%.

Kata kunci: *Klasifikasi, Konsentrasi Bidang Studi, Data Mining, Naïve Bayes Classifier*

Abstract

The concentration of study field is a specialization that a student must take in a department or field of study, which will later guide the student towards a more specific field. Concentration courses serve as a means to acquire knowledge and skills directed towards the completion of a final project. Many students fail to complete their final projects because the topics they choose are not aligned with their interests. This research aims to provide recommendations for the concentration of study fields for students in the Educational Technology and Computer Science program at Padang State University, using the results of a naive Bayes classifier model and assisted by the

RapidMiner application. This classification utilizes 24 attributes, which are the transcripts of the first to fourth semester grades. The classification results from this application indicate that out of 76 sets of testing data, an accuracy of 88.16%, precision of 100%, and recall of 41.67% were achieved.

Keywords : *Classification, Concentration Field of Study, Data Mining, Naïve Bayes Classifier*

PENDAHULUAN

Konsentrasi bidang studi merupakan pengkhususan studi yang harus diambil seorang mahasiswa dalam sebuah jurusan atau bidang studi (Riska Fitri Nur Alifah, 2023). Konsentrasi nantinya akan mengarahkan mahasiswa ke bidang yang lebih spesifik. Mata kuliah konsentrasi merupakan sarana untuk menguasai pengetahuan dan keterampilan yang terarah pada penyusunan tugas akhir.

Problema studi yang sering dihadapi mahasiswa dalam penyelesaian tugas akhir yaitu kesulitan dalam memilih program studi, konsentrasi pilihan mata kuliah yang sesuai dengan kemampuan dan waktu yang tersedia. Tidak sedikit mahasiswa yang tidak menyelesaikan tugas akhir karena topik yang diambil tidak sesuai dengan peminatannya (Syaifuddin Azwar, 2006).

Konsentrasi bidang studi mahasiswa dapat diketahui dari nilai pada mata kuliah tertentu yang diperoleh selama perkuliahan, karena nilai yang diperoleh dapat memberikan gambaran terkait konsentrasi bidang atau kompetensi yang dimiliki oleh mahasiswa tersebut. Hal yang sama juga berlaku dalam penentuan topik tugas akhir, dimana konsentrasi bidang studi akan mengarahkan mahasiswa dalam menentukan topik tugas akhir sesuai dengan konsentrasi bidang yang dimiliki.

Salah satu teknik yang bisa digunakan untuk mengoptimalkan data agar memiliki nilai manfaat serta menjadi informasi dan pengetahuan adalah dengan *data mining*. Penggunaan *data mining* dapat digunakan sebagai pertimbangan dalam pengambilan keputusan lebih lanjut tentang penentuan topik skripsi atau tugas akhir mahasiswa.

Data Mining diperlukan karena adanya kumpulan data yang sangat besar dimana data tersebut dapat digunakan untuk menghasilkan informasi dan *knowledge* yang berguna. Informasi dan *knowledge* yang didapat tersebut dapat digunakan pada banyak bidang, mulai manajemen bisnis, control produksi, kesehatan, dan lain-lain. Data mining sering dianggap sebagai bagian dari *Knowledge Discovery in Database* (KDD) yaitu sebuah proses mencari pengetahuan yang bermanfaat dari data (Yuli Mardi, 2016).

Metode dalam *data mining* diantaranya yaitu prediksi, klasifikasi, pengklusteran, dan asosiasi. Teknik *data mining* yang akan digunakan pada penelitian kali ini adalah Algoritma *Naïve Bayes Classifier* (NBC) yang merupakan sebuah pengklasifikasi probabilitas sederhana yang mengaplikasikan Teorema Bayes. Ide dari Teorema

Bayes adalah menangani masalah yang bersifat hipotesis yakni mendesain suatu klasifikasi untuk memisahkan objek (Santoso B, 2007).

Algoritma Naïve Bayes merupakan algoritma data mining yang dapat digunakan untuk melakukan klasifikasi, dan merupakan algoritma yang akan memberikan hasil kemungkinan yang akan terjadi di masa akan datang dari hasil klasifikasi (Gunawan et al., 2021).

Dalam menerapkan perhitungan Algoritma *Naïve Bayes Classifier* untuk probabilitas tiap kelas dalam pengklasifikasian kompetensi atau konsentrasi bidang studi menggunakan beberapa atribut pendukung. Dengan menggunakan Algoritma *Naïve Bayes Classifier* untuk klasifikasi kompetensi atau konsentrasi bidang studi, dapat memprediksi probabilitas sebuah class, sehingga dapat menghasilkan rekomendasi berdasarkan data pembelajaran atau *data training*.

METODE

Dalam menerapkan perhitungan Algoritma Naïve Bayes Classifier untuk probabilitas tiap kelas dalam pengklasifikasian kompetensi atau konsentrasi bidang studi menggunakan beberapa atribut pendukung. Dengan menggunakan Algoritma Naïve Bayes Classifier untuk klasifikasi kompetensi atau konsentrasi bidang studi, dapat memprediksi probabilitas sebuah class, sehingga dapat menghasilkan rekomendasi berdasarkan data pembelajaran atau data training.

Proses Analisis data

Proses analisis data yang dilakukan dengan memasukkan data latih berupa data Mahasiswa angkatan 2013, 2014 dan 2015 program studi Pendidikan Teknik Informatika yang diperoleh dari SIA (Sistem Informasi Akademik) Universitas Negeri Padang, data tersebut diklasifikasikan menggunakan algoritma *naïve bayes*, sehingga dari proses klasifikasi *naïve bayes* menghasilkan suatu model yang akan dilakukan penerapannya pada data uji mahasiswa angkatan 2013, 2014 dan 2015 yang nantinya akan menghasilkan suatu keluaran data uji tersebut.

Perhitungan Naïve Bayes Classifier

Tahapan dalam perhitungan menggunakan algoritma naïve bayes yaitu (a) Menyiapkan dataset yang akan digunakan sebagai bahan pengujian (data training), (b) menghitung jumlah dan probabilitas setiap kelas dan fitur, (c) Membandingkan probabilitas akhir.

HASIL DAN PEMBAHASAN

Analisis Data Mahasiswa Menggunakan Algoritma Naïve Bayes Classifier

Adapun atribut yang digunakan dalam penelitian ini terangkum dalam tabel berikut:

Tabel 1. Atribut yang digunakan dalam penelitian

No	Atribut	Keterangan
1.	Nama	Nama Mahasiswa
2.	X1	Mata kuliah Algoritma Pemrograman
3.	X2	Mata kuliah Disain Grafis
4.	X3	Mata kuliah Konsep Teknologi Informasi
5.	X4	Mata kuliah Sistem Informasi
6.	X5	Mata kuliah Matematika Diskrit
7.	X6	Mata kuliah Pratikum Algoritma dan Struktur Data
8.	X7	Mata kuliah Sistem Basis Data
9.	X8	Mata kuliah Struktur Data
10.	X9	Mata kuliah Jaringan Komputer dan Komunikasi Data
11.	X10	Mata kuliah Pemograman Berorientasi Objek
12.	X11	Mata kuliah Pemrograman Berbasis Web
13.	X12	Mata kuliah Praktikum Basis Data
14.	X13	Mata kuliah Praktikum Pemograman Berorientasi Objek
15.	X14	Mata kuliah Sistem Operasi
16.	X15	Mata kuliah Teknik Komputasi
17.	X16	Mata kuliah Analisis Perancangan Sistem
18.	X17	Mata kuliah Interaksi Manusia dan Komputer
19.	X18	Mata kuliah Kecerdasan Buatan
20.	X19	Mata kuliah Media Pendidikan
21.	X20	Mata kuliah Pemograman Visual
22.	X21	Mata kuliah Perancangan sistem Basis Data
23.	X22	Mata kuliah Praktikum Instalasi dan Jaringan Komputer
24.	X23	Mata kuliah Teknik Multimedia dan Animasi
25.	X24	IPK Mahasiswa
26.	Konsentrasi Bidang	Multimedia, Rekayasa Perangkat Lunak, Teknologi Informasi-Sistem Informasi, Teknik Komputer dan Jaringan

Klasifikasi dengan *Naïve Bayes Classifier*

Berdasarkan *data training*, terdapat 4 kelas klasifikasi yang dibentuk dari probabilitas kelas bidang minat atau topik tugas akhir, yaitu :

Bidang Minat "MM" = 20

P (Bidang Minat "MULTIMEDIA") = $20/126 = 0.158730$

Bidang Minat "RPL" = 60

P (Bidang Minat "RPL") = $60/126 = 0.476190$

Bidang Minat "TI-SI" = 43

P (Bidang Minat "TI-SI") = $43/126 = 0.341269$

Bidang Minat "TKJ" = 3

P (Bidang Minat "TKJ") = $3/126 = 0.023809$

Menghitung Nilai Mean Dan Standar Deviasi

Menghitung probabilitas untuk setiap fitur berdasarkan atribut dari kasus diatas akan dilakukan perhitungan dengan probabilitas dari variabel yang bersifat kontinu. Langkah pertama yang harus dilakukan adalah menghitung nilai mean dan standar deviasi setiap atribut pada kelas "MM", "RPL", "TI-SI", dan "TKJ".

Persamaan yang digunakan untuk menghitung nilai rata-rata hitung (mean) dapat dilihat sebagai berikut :

$$\mu = \frac{\sum_{i=1}^n x_i}{n} \quad (1)$$

Dimana :

μ : Rata-rata hitung (*mean*)

x_i : Nilai sample ke-l

n : Jumlah sample

Adapun persamaan untuk menghitung nilai simpangan baku (standar deviasi) dapat dilihat sebagai berikut :

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \mu)^2}{n - 1}} \quad (2)$$

Dimana :

σ : Standar deviasi

x_i : Nilai x ke-i

μ : Rata-rata hitung

n : Jumlah sampel

Nilai mean dan standar deviasi pada masing-masing kelas dari semua atribut terangkum dalam tabel di bawah:

Tabel 2. Nilai Mean Setiap Kelas

Nama Atribut	Mean			
	MM	RPL	TI-SI	TKJ
X1	2.16	2.541666	2.941860	3.3
X2	3.07	3.16	3.190697	3.5
X3	3.315	3.036666	3.011627	2.666666
X4	3.265	3.265	3.353488	3.733333
X5	3.185	3.271666	3.337209	3.3
X6	2.755	2.988333	2.897674	2.833333
X7	2.65	3.085	3.239534	3.433333
X8	3.165	3.123333	3.193023	3.3
X9	3.28	3.17	3.334883	2.966666
X10	3.37	3.248333	3.253488	3.2
X11	3.365	3.328333	3.288372	3.2
X12	3.15	3.235	3.344186	3.4
X13	3.525	3.408333	3.3	3.166666
X14	3.37	3.328333	3.488372	2.733333
X15	3.36	3.268333	3.297674	3.866666
X16	3.175	3.238333	3.474418	3.633333
X17	3.225	3.41	3.376744	3.2
X18	3.36	3.341666	3.225581	2.833333
X19	3.385	3.336666	3.351162	3.3
X20	3.36	3.381666	3.337209	3.3
X21	3.44	3.413333	3.293023	3.733333
X22	3.305	3.238333	3.451162	2.966666
X23	3.4	3.338333	3.302325	2.866666
X24	3.289	3.333	3.330930	3.34

Tabel 3. Nilai Standar Deviasi Setiap Kelas

Nama Atribut	Standar Deviasi			
	MM	RPL	TI-SI	TKJ
X1	0.832105	0.926695	0.521729	0.244948
X2	0.360693	0.368872	0.662558	0.141421
X3	0.455274	0.658018	0.720532	0.942809
X4	0.444156	0.436205	0.459701	0.188561
X5	0.658995	0.486857	0.502099	0.244948

X6	0.769074	0.700497	0.810824	0.880656
X7	0.575760	0.651108	0.720187	0.418993
X8	0.648286	0.691222	0.831472	0.571547
X9	0.380263	0.520992	0.471456	0.531245
X10	0.298496	0.516233	0.555039	0.141421
X11	0.310282	0.452802	0.597751	0.141421
X12	0.663701	0.676221	0.497576	0.141421
X13	0.578683	0.571341	0.531343	0.418993
X14	0.544150	0.568357	0.458110	0.418993
X15	0.285306	0.483904	0.674619	0.188561
X16	0.617150	0.530123	0.422117	0.286744
X17	0.854912	0.404021	0.527317	0.588784
X18	0.385227	0.505731	0.533131	0.555777
X19	0.469334	0.456057	0.401955	0.571547
X20	0.654522	0.588923	0.485617	0.244948
X21	0.592790	0.475908	0.609382	0.188561
X22	0.580925	0.658809	0.515062	0.740870
X23	0.461519	0.508622	0.677714	0.418993
X24	0.196211	0.171048	0.176144	0.156418

Menghitung Probabilitas Akhir Setiap Kelas

Nilai probabilitas setiap fitur mengacu kepada nilai yang telah diperoleh pada tabel 2 dan 3 . Notasi yang dipakai adalah

$$P(X_i = x_1 | Y = y_j) = \frac{1}{\sqrt{2\pi} \sigma_{ij}} \exp \frac{(x_1 - \mu_{ij})^2}{2 \sigma_{ij}^2} \quad (3)$$

Tabel 4. Perhitungan Probabilitas Akhir Setiap Atribut

Nama Atribut	Probabilitas Akhir			
	MM	RPL	Ti-Si	TKJ
X1 = 3.6	0.097862	0.215938	0.249319	0.380856
X2 = 3.6	0.225738	0.322565	0.405078	0.826397
X3 = 2	0.009126	0.142213	0.175448	0.320062
X4 = 4	0.152267	0.146101	0.218922	0.338063
X5 = 3.6	0.403146	0.455574	0.491068	0.380856
X6 = 1.6	0.147325	0.066889	0.123127	0.159485

X7 = 3.3	0.278062	0.468291	0.468561	0.586039
X8 = 3.3	0.484976	0.464543	0.434011	0.527829
X9 = 3	0.493451	0.524184	0.451582	0.546408
X10 = 3	0.338773	0.494706	0.482576	0.390362
X11 = 3.3	0.700829	0.591854	0.516033	0.826397
X12 = 3.3	0.477465	0.483024	0.563479	0.826397
X13 = 3.3	0.486375	0.518519	0.547435	0.586039
X14 = 2.3	0.078259	0.103005	0.020384	0.361119
X15 = 4	0.060352	0.182897	0.282579	0.715681
X16 = 4	0.207865	0.195241	0.282918	0.329009
X17 = 2.6	0.330371	0.084142	0.185706	0.309416
X18 = 3.6	0.529514	0.492493	0.427071	0.206711
X19 = 2.6	0.143814	0.160301	0.109796	0.249329
X20 = 3.6	0.461170	0.485451	0.494633	0.380856
X21 = 3.6	0.499747	0.535613	0.450267	0.715681
X22 = 2.3	0.117236	0.178294	0.045750	0.309258
X23 = 2.3	0.034305	0.069638	0.162372	0.247018
X24 = 3.21	0.830723	0.745031	0.751166	0.714307

Langkah selanjutnya yaitu menghitung probabilitas akhir setiap kelas dengan cara mengalikan probabilitas akhir setiap fitur berdasarkan kelas masing-masing. Hasil $P(X | \text{Bidang Minat "MM"}) = 1.27235 \cdot 10^{-16}$, $P(X | \text{Bidang Minat "RPL"}) = 2.16738 \cdot 10^{-14}$, $P(X | \text{Bidang Minat "TI-SI"}) = 2.91815 \cdot 10^{-14}$, dan $P(X | \text{Bidang Minat "TKJ"}) = 1.10168 \cdot 10^{-9}$.

Selanjutnya nilai-nilai tersebut dimasukkan untuk mendapatkan probabilitas akhir :

$$P(X | \text{Bidang Minat "MM"}) = \alpha \times 0.158730 \times 1.27235 \cdot 10^{-16} = 2.01961 \cdot 10^{-17}$$

$$P(X | \text{Bidang Minat "RPL"}) = \alpha \times 0.476190 \times 2.16738 \cdot 10^{-14} = 1.03208 \cdot 10^{-14}$$

$$P(X | \text{Bidang Minat "TI-SI"}) = \alpha \times 0.341269 \times 2.91815 \cdot 10^{-14} = 9.95878 \cdot 10^{-15}$$

$$P(X | \text{Bidang Minat "TKJ"}) = \alpha \times 0.023809 \times 1.10168 \cdot 10^{-9} = 2.62305 \cdot 10^{-11}$$

Berdasarkan Perhitungan diatas, nilai terbesar ada di kelas Bidang Minat "TKJ" = $2.62305 \cdot 10^{-11}$. Sehingga data uji mahasiswa dengan ID = 126 diklasifikasikan ke dalam Bidang Minat "TKJ". Hasil pengujian atas nama SANDI SAPUTRA ID 126 dapat dilihat pada tabel berikut :

Tabel 5. Hasil Akhir Klasifikasi

ID Mahasiswa	Nama Mahasiswa	Kelas Bidang Minat		Hasil
		<i>Actual</i>	<i>Prediksi Naïve Bayes</i>	
126	SANDI SAPUTRA	TKJ	TKJ	AKURAT

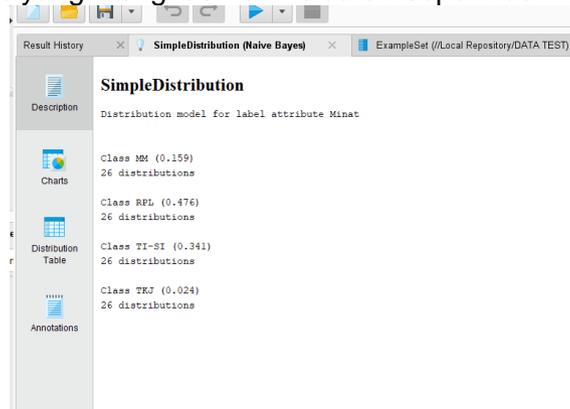
Perhitungan Menggunakan Tools RapidMiner

Pengujian ini bertujuan untuk mengukur tingkat akurasi dari klasifikasi bidang minat mahasiswa menggunakan algoritma *naïve bayes classifier*. Data yang dianalisis adalah data mahasiswa prodi Pendidikan Teknik Informatika dan Komputer Universitas Negeri Padang angkatan 2013 sampai dengan 2015 yang didapat dari SIA UNP.



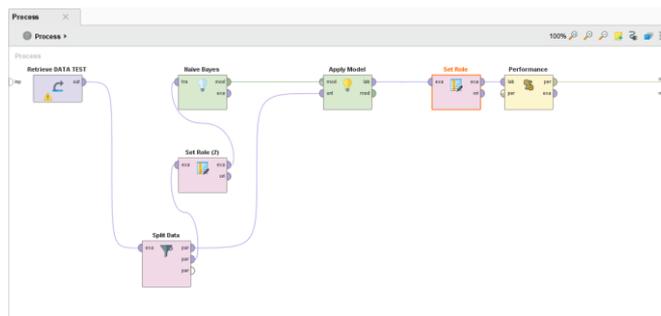
Gambar 1. Proses Training data

Gambar di atas menunjukkan proses pelatihan data training menggunakan algoritma *naïve bayes* yang menghasilkan keluaran seperti berikut:



Gambar 2. Hasil Proses Training Data

Gambar di atas merupakan model distribusi pengujian dengan algoritma *naïve bayes*. Hasil perincian kelas MM = 0.159 dengan 26 distribusi, kelas RPL = 0.476 dengan 26 distribusi, kelas TI-SI = 0.341 dengan 26 distribusi, dan kelas TKJ = 0.024 dengan 26 distribusi.



Gambar 3. Implementasi *naïve bayes* untuk uji *performance data testing*.

Gambar di atas merupakan suatu pengaplikasian model *naïve bayes* menggunakan tools RapidMiner pada data testing.

Tabel 6. Hasil pengujian *data testing*

Prediksi	Bidang Minat			
	MM	RPL	TI-SI	TKJ
MM	5	0	0	0
RPL	7	41	1	0
TI-SI	0	0	21	1
TKJ	0	0	0	0

Setelah dilakukan pengujian, diperoleh nilai akurasi sebesar 88.16%, nilai presisi sebesar 100%, dan nilai recall 41.67%.

Tabel 7. Rangkuman hasil eksperimen *naïve bayes*

Sampel Testing	Jumlah Data	Accuracy	Precision	Recall
Data Uji Mahasiswa	76	88.16%	100%	41.67%

SIMPULAN

Berdasarkan analisis dan pengujian data mahasiswa prodi Pendidikan Teknik Informatika dan Komputer Universitas Negeri Padang angkatan 2013 sampai dengan 2015 dapat diambil kesimpulan sebagai berikut:Metode *data mining* menggunakan algoritma *naïve bayes* dapat diterapkan kedalam klasifikasi bidang minat mahasiswa prodi Pendidikan Teknik Informatika dan Komputer Universitas Negeri Padang.Tingkat akurasi yang dihasilkan dari suatu proses klasifikasi dengan menggunakan algoritma *naïve bayes classifier* sangat ditentukan dari pemilihan data training dan pemilihan atribut inputpun juga berpengaruh terhadap hasil output (hasil prediksi). Data keluaran atau prediksi bidang minat mahasiswa yang dihasilkan oleh program adalah suatu keputusan apakah mahasiswa tersebut termasuk dalam klasifikasi bidang minat “MM”,

“RPL”, “TI-SI”, atau “TKJ”, sehingga data tersebut menjadi tolak ukur dalam menunjang strategi konsentrasi bidang minat di program studi Pendidikan Teknik Informatika. Dari hasil pengujian untuk data testing diperoleh hasil dengan tingkat *accuracy* sebesar 88.16%, tingkat *precision* kelas “MM” sebesar 100%, dan tingkat *recall* kelas “MM” sebesar 41.67%.

DAFTAR PUSTAKA

- Alifah, Riska Fitri Nur., Fauzan, Abd. Charis. 2023. “Implementasi Algoritma K-Means Clustering Berbasis Jarak Manhattan untuk Klasterisasi Konsentrasi Bidang Mahasiswa”. 5 (1) (2023), April 2023.
- Azwar, Saifuddin. 2006. *Reabilitas dan Validitas*. Yogyakarta: Pustaka Belajar.
- Fadillah, Annisa Paramitha dan Hardiyana, Bella. 2018. “Penerapan Naive Bayes Classifier Untuk Pemilihan Konsentrasi Mata Kuliah”, *JATI*, 8 (2). Hlm. 100-107, September 2018.
- Farid., Enri, Ultach., dan Umaidah, Yuyun. 2021. “Sistem Pendukung Keputusan Rekomendasi Topik Skripsi Menggunakan *Naïve Bayes Classifier*”, 6 (1)
- Gunawan, M., Zarlis, M., & Roslina, R. 2021. Analisis Komparasi Algoritma Naïve Bayes dan K-Nearest Neighbor Untuk Memprediksi Kelulusan Mahasiswa Tepat Waktu. *Jurnal Media Informatika Budidarma*, 5(2), 513. <https://doi.org/10.30865/mib.v5i2.2925>
- Mardi, Yuli. 2016. “Data Mining : Klasifikasi Menggunakan Algoritma C4.5”. 2 (2), <https://ejournal.upgrisba.ac.id/index.php/eDikInformatika/article/view/1465/789>
- Pratama, Teddy dan Yulmaini. 2018. “Implementasi Algoritma Naïve Bayes Dalam Menentukan Konsentrasi Skripsi dan Rekomendasi Bahasa Pemograman”. *Jurnal Informatika*, 18 (1). Hlm. 1–13. Juni 2018.
- Saleh, Alfa. 2015. “Implementasi Metode Klasifikasi Naïve Bayes Dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga”. *Creative Information Technology Journal*, 2(3), Hlm. 207–217.
- Santoso, Budi. 2007. *Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis*. Yogyakarta: Graha Ilmu.
- Saputra, Hadi Kurnia. 2018. “Analisis *Data Mining* Untuk Pemetaan Mahasiswa Yang Membutuhkan Bimbingan Dan Konseling Menggunakan Algoritma *Naïve Bayes Classifier*”, *Jurnal Teknologi Informasi dan Pendidikan*. 11(2).