Faktor Exacta

Vol. 17, No. 2, June 2024, pp. 131~140
eISSN: 2502-339X, pISSN: 1979-276X, DOI: 10.30998/faktorexacta.v17i2.20625

Penerapan Data Mining untuk Prediksi Minat Jurusan Mahasiswa Baru dengan Pendekatan Algoritma Naive Bayes

Niken Harsanti[1],Arief Wibowo[2]
[1] Budiluhur University Master computer science Student
[2]Department of Information System, Budiluhur University

Article Info

Article history:

Received Oct 10, 2023
Revised Jun 08, 2024
Accepted Jun 24, 2024

Keywords:

Data Mining, Prediction, Student Interest, Naïve Bayes algorithm

ABSTRACT

This research makes a positive contribution to the development of data mining applications in the field of higher education, which has the potential to help students and universities improve the efficiency of major selection. The aim of this research is to apply data mining techniques using the Naïve Bayes algorithm to predict new students' majors. Accurate predictions can help new students make better decisions. Data obtained is based on historical data about past students, including information about academic grades, interests, and other factors that influence major selection. The Naïve Bayes algorithm is used to process this data and produce a prediction model that can identify majors that best suit the characteristics of new students. The method using the Naïve Bayes algorithm forms a grouping model using the Confusion Matrix matrix table and classifies positive and negative values. The Naive Bayes algorithm model obtained can be implemented in the form of an application designed to predict new students' majors in determining the study program they will take. The Naïve Bayes algorithm is able to provide quite accurate predictions, so it can be used as a guide for new students in choosing their major. The results of data processing for new students obtained accuracy values with the Naïve Bayes algorithm model of 98.55%, precision of 99.97%, and recall of 98.55%.

Corresponding Author:

Niken Harsanti,
Budiluhur University Master computer science Student
Jl. Ciledug Raya , Petukangan Utara Jakarta Selatan.
Email: 2211600271@student.budiluhur.ac.id

1. PENDAHULUAN

Tingkat perguruan tinggi menjadi langkah pemikiran dan berpola pikir yang tumbuh dikalangan siswa , khususnya selesai jenjang pendidikan tingkat atas menyiapkan bekal sedini mungkin untuk melanjutkan perguruan tinggi. Mahasiswa baru bersiap menentukan pilihan dengan melihat kesesuaian yang diinginkan dengan melihat rambu rambu aturan yang berlaku untuk mengambil jurusan yang mampu ditempuhnya. Fakultas diperguruan tinggi yang peminatnya digandrungi siswa biasanya yang praktis dan tren dikalangan siwa tingkat atas. [1]. Sistem penerimaan mahasiswa baru dilingkup perguruan tinggi beragam model penerapan serta ketentuan yang berlaku di kampusnya masing masing. Penelitian serupa di STMIK Royal dengan memberikan reward pada mahasiswa baru yang diprediksi masuk dalam program studi dengan peminatan tertentu dengan menggunakan naïve bayes, dimana di perguruan tinggi tersebut melihat semakin menurunnya peminat pada program studi tertentu.[2] . Profil mahasiwa dalam pemetaan peminatan studi tugas akhir juga diteliti dengan menggunakan naïve bayes dan Term frequency invers document frequency, hasilnya sebagai rekomendasi berdasarkan profil siswa tersebut.[3]. Penelitian serupa juga dilakukan latifa dengan menggunakan metode naïve bayes sebagai penunjang promosi program studi diperguruan tinggi UPGRI S dengan klasifikasi data mining [4] . Penelitian dengan membandingkan antara algoritma naïve bayes dan SVM juga dilakukan untuk prediksi mahasiswa yang belum registrasi dimana jumlah tersebut dikelompokkan

menjadi beberapa kelompok, terbentuk kluster yang akan dijadikan solusi akhir perbandingan tersebut pada fakultas ilmu komputer diteliti oleh megira.[5]. Penelitian yang dilakukan mirza penerapan algoritma naïve bayes dalam upaya menentuka trategi promosi mahasiswa baru , dimana hasil penelitian ini didapat peminat yang paling banyak jumlahnya sebagai program studi tren dan menjadi ketertarikan di bidang promosi.[6]
Penelitian ini dilakukan dengan melihat dalam setiap tahunnya perubahan dan minimnya peminat prodi tertentu yang dirasa kurang diminati, maka diperlukan pemetaan konsep kebutuhan dari berbagai atribut program studi yang banyak peminatnya dengam prediksi peminatan mahasiswa baru .
Ragam model peminatan mahasiwa baru yang digunakan ditingkat perguruan tinggi dengan melihat tren serta kekinian yang dapat diprediksi dengan beberapa pendekatan. Data mining bagian dari konsep data dengan tujuan menggali data menjadi informasi yang digunakan dalam basis data dan merupakan bagian dari Knowledge Discovery in Databased (KDD) agar memperoleh model informasi serta pola keterkaitan yang dapat digunakan dalam penambangan data. [7]. Data mining menggali data untuk memperoleh informasi yang diinginkan, bernilai serta manfaat untuk pengelompokan data dimana unsur lain terlibat didalamnya seperti pengguna dan perangkat yang saling terkoneksi secara online.[8]
Penambangan data bagian dari proses yang diperoleh dari dara mining cluster atau pengelompokan yang dibutuhkan para peneliti diambil dari pola relation di database dalam jumlah yang sangat besar. Lingkup informatika sangat tertantang tentang data mining untuk mengkaji secara mendalam apa yang ingin di teliti di berbagai informasi data yang ingin di ambil. Pengambilan dalam penambangan data mendapatkan informasi knowledge yang baru di kelompokan ke dalam kumpulan data.[9] Pengelompokan data mining terdiri dari beberapa metode yang dimiliki dengan tujuan dapat memanfaatkan ragam himpunan data diantarannya durasi, perkiraan, pengklasifikasian, klasterisasi serta asosiasi.[10]
Bagian yang paling khusus dalam teknik data mining dan wajib diperhatikan ialah penerapan aturan agar mendapatkan model terpola frekuensi tertinggi diantara himpunan dari atribut ,item set biasa disebut sebagai Association Rule (Aturan Asosiasi).[11]. Diantara ragam penggunaan dalam algoritma masuk kedalam aturan asosiasi ialah AIS Algorithm, DHP Algorithm, Partition Algorithm, dan Naivebayes Algorithm. Jenis jenis algoritma itu terdapat satu algoritma yang sangat tepat dan valid saat digunakan dan cocok dengan data mining dalam menganalisa prediksi peminatan mahasiswa baru dalam menggambil program studi yang sesuai dengan talenta yang dimiliki sejak awal yaitu algoritma Naivebayes . [12]

Gambar 1. Peminatan Mahasiswa Baru 2022

Peminatan mahasiswa baru dalam setiap semesternya mengalami perubahan yang agak sulit diprediksi dengan kasat mata , disebabkan dalam mengambil jurusan mahasiswa dihadapkan pada keadaan yang sulit dipahami peminataanya dengan kealhilan yang dimiliki. Calon mahasiswa baru dalam memilih jurusan terkadang melihat opini dari kakak kelas dan iklan iklan yang bertaburan di media online yang memikat calon mahasiwa baru dalam mengambil jurusan.[13]
Algoritma Naive Bayes bagian dari klasifikasi algoritma biasa digunakan dengan teorema Bayes diasumsikan naif (naive) dimana semua fitur yang ada didalam data disebut independen antara satu dengan yang lainnya. Walaupun terkait asumsi tersebut adakalanya tidak terpenuhi berdasarkan hasilnya, Naive Bayes bagian dari algoritma paling miniati pada data sains untuk pembelajaran mesin inferensi karena kecepatan serta keakuratannya terjamin. [14]
Cara kerja algoritma Naive Bayes adalah sebagai berikut:

1. Estimasi Probabilitas Kelas: Pertama, algoritma Naive Bayes menghitung probabilitas masing-masing kelas dalam dataset. Ini dilakukan dengan menghitung frekuensi kemunculan setiap kelas dalam data pelatihan dan membaginya dengan jumlah total sampel.
2. Estimasi Probabilitas Fitur: Selanjutnya, algoritma Naive Bayes mengestimasi probabilitas setiap fitur (variabel independen) untuk setiap kelas. Ini melibatkan menghitung frekuensi kemunculan setiap nilai fitur dalam setiap kelas dan membaginya dengan jumlah total sampel dalam kelas tersebut.
3. Asumsi Independensi Fitur: Asumsi independensi adalah asumsi kunci dalam Naive Bayes. Algoritma menganggap bahwa semua fitur dalam data adalah independen satu sama lain, meskipun asumsi ini sering kali tidak realistis di dunia nyata.
4. Menghitung Probabilitas Posterior: Setelah probabilitas kelas dan probabilitas fitur telah diestimasi, Naive Bayes menggunakan teorema Bayes untuk menghitung probabilitas posterior. Probabilitas posterior adalah probabilitas bahwa sampel tertentu termasuk dalam suatu kelas berdasarkan fitur-fiturnya.
5. Klasifikasi: Setelah probabilitas posterior dihitung, algoritma Naive Bayes dapat digunakan untuk memprediksi kelas (variabel dependen) dari sampel yang tidak diketahui. Algoritma akan memilih kelas dengan probabilitas posterior tertinggi sebagai prediksi kelasnya.
6. Naive Bayes memiliki keuntungan dalam kecepatan komputasi dan kebutuhan data pelatihan yang relatif sedikit. Namun, asumsi independensi fitur dapat menjadi keterbatasan jika ada dependensi yang signifikan antar fitur dalam data. Meskipun demikian, Naive Bayes tetap efektif dalam banyak kasus dan sering digunakan dalam klasifikasi teks, pengenalan pola, dan banyak aplikasi lainnya.

Naive Bayes bagian dari salah satu algoritma yang memiliki akurasi baik dalam teknik pengklasifikasian serta perhitungan untuk menilai cluster tertinggi yang menjadi skala pengukur dalam prediksi berbagai penentu penilaian.[15] Teorema Bayes dengan rumus sebagai berikut :
Rumus Teorema Bayes

$$\mathbf{P(C \mid X)} = \frac{\mathbf{P(X|C).P(C)}}{\mathbf{P(X)}}$$

Keterangan.
X = Sampel data yang memiliki class (label) yang tidak diketahui.
C = Hipotesis bahwa X adalah data class (label).
P(C) = Probabilitas hipotesis C.
P(X) = Peluang dari data sampel yang diamati (probabilitas C).
P(X|C) = Probabilitas berdasarkan kondisi pada hipotesis.
Konsep kerja dari alur metode naive bayes antara lain.

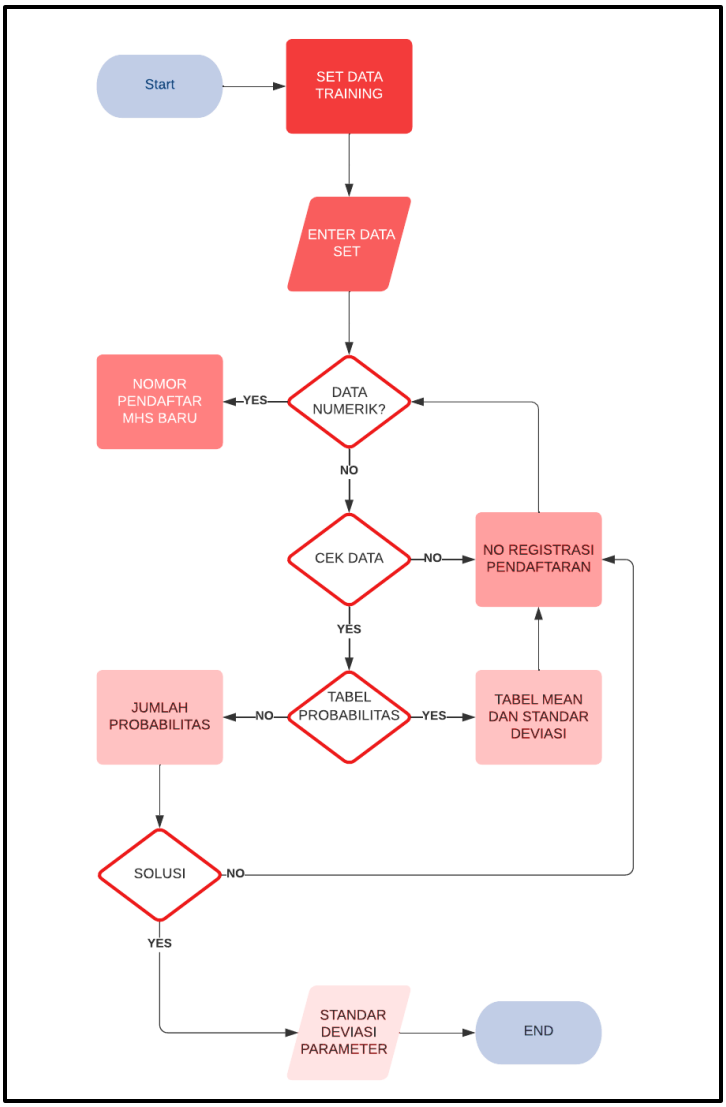
1. Menghitung nilai peluang kasus baru dari setiap hipotesa dengan class (label) yang ada di P(Ci).
2. Menghitung nilai akumulasi peluang dari setiap kelas P(X|Ci).
3. Menghitung nilai P(X|Ci) x P(Ci).
4. Menentukan class dari kasus baru tersebut.

Dalam menentukan mahasiswa baru dipilih peminatan program studi dan dihimpun dalam 10 aturan data yang berlaku. Terdiri dari 4 atribut yang akan digunakan
Jumlah pendaftar mahasiswa pada program studi di Fakultas Teknik (C1).
Tingkat kelulusan tes masuk program studi peminatan (C2).
Daftar ulang dan registrasi mahasiswa baru (C3).
Keputusan memilih program studi yang diminati (C4).

Alur konsep metode Naïve Bayes sebagai berikut:

a. Baca data training
b. Hitung jumlah data probabilitas, namun apabila data numerik maka:
c. Cari nilai mean dan standar deviasi dari masing-masing parameter yang merupakan data numerik

d. Cari nilai probabilitas dengan cara menghitung jumlah data yang sesuai dari dari kategori yang sama dibagi dengan jumlah data pada kategori tersebut.
e. Mendapatkan nilai dalam table mean, standar deviasi dan probabilitas



Gambar 2. Alur Naïve Bayes Prediksi Jurusan Peminatan Mahasiswa Baru

2. METODE

Naïve bayes menggunakan metode Confusion Matrix yang mempresentasikan hasil setelah evaluasi model dengan menggunakan tabel matrik. Dataset terdiri dari 2 kelas, kelas pertama positif dan kelas kedua negatif. Evaluasi menggunakan confusion matrix menghasilkan nilai Akurasi, Precision, Recall, serta F-Measure. Akurasi mendapatkan klasifikasi presentasi ketepatan record data diklasifikasikan secara benar setelah dilakukan pengujian pada hasil klasifikasi. Precision merupakan proposikasi yang diprediksi positif yang juga positif benar pada data sebenarnya. Recall merupakan proporsi kasus positif yang sebenarnya diprediksi positif secara benar.

True Positive (TP) jumlah record postif dalam dataset yang diklasifikasikan positif. True Negative (TN) merupakan jumlah record negative dalam dataset yang diklasifikasikan positif. False Positive (FP) merupakan jumlah record negative dalam dataset yang diklasifikasikan positif. False Negative (FN) merupakan jumlah record positif dalam dataset yang diklasifikasikan negatif.

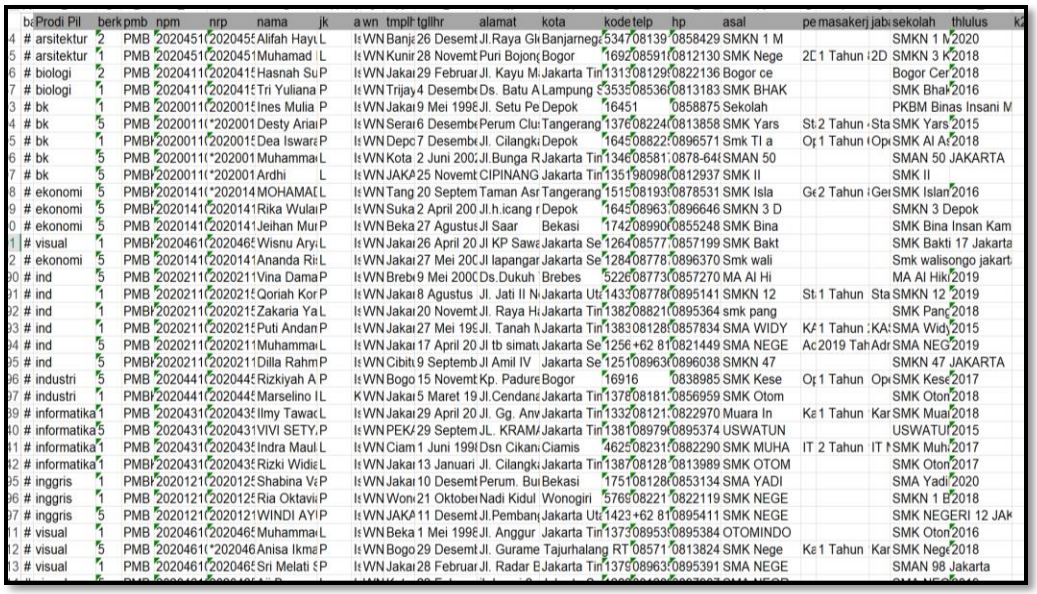
Rumus persamaan Confusion Matrix:

$$\textbf{Accuracy} = \frac{\textbf{TP+TN}}{\textbf{TP+TN+FP+FN}}$$

3. HASIL DAN PEMBAHASAN

Data yang didapat sebanyak 2400 , namun dari data tersebut yang dapat digunakan sebanyak 680 data yang lengkap dan dapat diproses untuk berlanjut . Sebanayk 544 digunakan sebagai data training dan 236 digunakan sebagai data testing. Atribut yang didapatkan ada 8 terdiri dari , asal sekolah, jurusan, tahun lulus, pekerjaan, masa kerja, jenis kelamin, prodi pilihan, prodi alternatif.

3.1. Data Set



	Prodi Pil	berk	pmb	npm	nrp	nama	jk	a	wn	tmplh	tgllhr	alamat	kota	kode	telp	hp	asal	pe	masakerj	jab	sekolah	thlulus
4	# arsitektur	2	PMB	2020451	202045	Alifah Hayu	L	I	WN	Banja	26 Desem	Jl.Raya Gl	Banjarnega	53	47081	39 0858429	SMKN 1 M				SMKN 1 M	2020
5	# arsitektur	1	PMB	2020451	202045	Muhamad	L	I	WN	Kunir	28 Novemt	Puri Bojon	Bogor	16	92	08591	0812130	SMK Nege	2	1 Tahun	2D	SMKN 3 K	2018
6	# biologi	2	PMB	2020411	202041	Hasnah Su	P	I	WN	Jakar	29 Februar	Jl. Kayu M	Jakarta Tin	13	13	08129	0822136	Bogor ce				Bogor Cer	2018
7	# biologi	1	PMB	2020411	202041	Tri Yuliana	P	I	WN	Trijay	4 Desemb	Ds. Batu A	Lampung S	35	35	08536	0813183	SMK BHAK				SMK Bhak	2016
3	# bk	1	PMB	2020011	202001	Ines Mulia	P	I	WN	Jakar	9 Mei 199	Jl. Setu Pe	Depok	16	45	1	0858875	Sekolah				PKBM Binas Insani M	
4	# bk	5	PMB	2020011	*202001	Desty Aria	P	I	WN	Serar	6 Desemb	Perum Clu	Tangerang	13	76	08224	0813858	SMK Yars		St	2 Tahun	Sta	SMK Yars	2015
5	# bk	1	PMB	2020011	202001	Dea Iswara	P	I	WN	Depc	7 Desemb	Jl. Cilangk	Depok	16	45	08822	0896571	Smk TI a		Or	1 Tahun	Or	SMK Al Ac	2018
6	# bk	5	PMB	2020011	*202001	Muhamma	L	I	WN	Kota	2 Juni 200	Jl.Bunga R	Jakarta Tin	13	46	08581	0878-64	SMAN 50				SMAN 50 JAKARTA	
7	# bk	5	PMB	2020011	*202001	Ardhi	L	I	WN	JAKA	25 Novemt	CIPINANG	Jakarta Tin	13	51	98098	0812937	SMK II				SMK II	
8	# ekonomi	5	PMB	2020141	*202014	MOHAMAI	L	I	WN	Tang	20 Septem	Taman Asr	Tangerang	15	15	08193	0878531	SMK Isla		Ge	2 Tahun	Ge	SMK Islan	2016
9	# ekonomi	5	PMB	2020141	202014	Rika Wula	P	I	WN	Suka	2 April 200	Jl.h.icang r	Depok	16	45	08963	0896646	SMKN 3 D				SMKN 3 Depok	
0	# ekonomi	5	PMB	2020141	202014	Jeihan Mu	P	I	WN	Beka	27 Agustu	Jl Saar	Bekasi	17	42	08990	0855248	SMK Bina				SMK Bina Insan Kam	
1	# visual	1	PMB	2020461	202046	Wisnu Ary	L	I	WN	Jakar	26 April 20	Jl KP Sawa	Jakarta Se	12	64	08577	0857199	SMK Bakt				SMK Bakti 17 Jakarta	
2	# ekonomi	5	PMB	2020141	202014	Ananda Ri	L	I	WN	Jakar	27 Mei 20	Jl lapangar	Jakarta Se	12	84	08778	0896370	Smk wali				Smk walisongo jakart	
0	# ind	5	PMB	2020211	202021	Vina Dama	P	I	WN	Breb	9 Mei 200	Ds.Dukuh	Brebes	52	26	08773	0857270	MA Al Hi				MA Al Hik	2019
1	# ind	1	PMB	2020211	202021	Qoriah Kor	P	I	WN	Jakar	8 Agustus	Jl. Jati II N	Jakarta Ut	14	33	08778	0895141	SMKN 12		St	1 Tahun	Sta	SMKN 12	2019
2	# ind	1	PMB	2020211	202021	Zakaria Ya	L	I	WN	Jakar	20 Novemt	Jl. Raya Hi	Jakarta Tin	13	82	08821	0895364	smk pang				SMK Pang	2018
3	# ind	1	PMB	2020211	202021	Puti Andar	P	I	WN	Jakar	27 Mei 19	Jl. Tanah M	Jakarta Tin	13	83	08128	0857834	SMA WIDY		KA	1 Tahun	KA	SMA Widy	2015
4	# ind	5	PMB	2020211	202021	Muhamma	L	I	WN	Jakar	17 April 20	Jl tb simatu	Jakarta Se	12	56	+62 81	0821449	SMA NEGE		Ac	2019 Tah	Adr	SMA NEG	2019
5	# ind	5	PMB	2020211	202021	Dilla Rahm	P	I	WN	Cibit	9 Septemb	Jl Amil IV	Jakarta Se	12	51	08963	0896038	SMKN 47				SMKN 47 JAKARTA	
6	# industri	5	PMB	2020441	202044	Rizkiyah A	P	I	WN	Bogo	15 Novemt	Kp. Padure	Bogor	16	916		0838985	SMK Kese		Or	1 Tahun	Or	SMK Kese	2017
7	# industri	1	PMB	2020441	202044	Marselino	L	K	WN	Jakar	5 Maret 19	Jl.Cendana	Jakarta Tin	13	78	08181	0856959	SMK Otom				SMK Oton	2018
9	# informatika	1	PMB	2020431	202043	Ilmy Tawa	L	I	WN	Jakar	29 April 20	Jl. Gg. Anv	Jakarta Tin	13	32	08121	0822970	Muara In		Ka	1 Tahun	Kar	SMK Muar	2018
0	# informatika	5	PMB	2020431	202043	VIVI SETY	P	I	WN	PEK/	29 Septem	JL. KRAM/	Jakarta Tin	13	81	08979	0895374	USWATUN				USWATUI	2015
1	# informatika	1	PMB	2020431	202043	Indra Maul	L	I	WN	Ciam	1 Juni 199	Dsn Cikan	Ciamis	46	25	08231	0882290	SMK MUHA		IT	2 Tahun	IT M	SMK Muh	2017
2	# informatika	1	PMB	2020431	202043	Rizki Widia	L	I	WN	Jakar	13 Januari	Jl. Cilangki	Jakarta Tin	13	87	08128	0813989	SMK OTOM				SMK Oton	2017
5	# inggris	1	PMB	2020121	202012	Shabina Va	P	I	WN	Jakar	10 Desemt	Perum. Bui	Bekasi	17	51	08128	0853134	SMA YADI				SMA Yadi	2020
6	# inggris	1	PMB	2020121	202012	Ria Oktavi	P	I	WN	Won	21 Oktobe	Nadi Kidul	Wonogiri	57	69	08221	0822119	SMK NEGE				SMKN 1 B	2018
7	# inggris	5	PMB	2020121	202012	WINDI AY	P	I	WN	JAKA	11 Desemt	Jl.Pembar	Jakarta Ut	14	23	+62 81	0895411	SMK NEGE				SMK NEGERI 12 JAK	
1	# visual	1	PMB	2020461	202046	Muhamma	L	I	WN	Beka	1 Mei 199	Jl. Anggur	Jakarta Tin	13	73	08953	0895384	OTOMINDO				SMK Oton	2016
2	# visual	5	PMB	2020461	*202046	Anisa Ikm	P	I	WN	Bogo	29 Desemt	Jl. Gurame	Tajurhalang	RT	08571		0813824	SMK Nege		Ka	1 Tahun	Kar	SMK Nege	2018
3	# visual	1	PMB	2020461	202046	Sri Melati	P	I	WN	Jakar	28 Februar	Jl. Radar E	Jakarta Tin	13	79	08963	0895391	SMA NEGE				SMAN 98 Jakarta	

Gambar 3. Dataset yang digunakan

Terlihat dalam Gambar 3 ada data yang siap digunakan sebanyak data ada 544 baris, dan data yang disajikan tersebut akan dijadikan sebagai data training maupuan data testing. Data tersebut selanjutnya akan melewati proses perhitungan untuk menentukan nilai dari setiap atribut yang sudah didapatkan, antara lain mean dan standar deviasi.

Tabel 1. Perhitungan Mean

Prodi Peminatan	Asal Sekolah	Jurusan	Tahun Lulus	Pekerjaan	Prodi pilihan	Prodi alternatif
Arsitek	77.678	76.793	77.324	77.123	76.346	70.321
Informatika	83.378	82.458	82.645	81.278	82.543	81.123
Visual	81.934	80.234	80.164	80.268	81.268	80.326

Terlihat dalam Pada Tabel 1 didapatkan hasil perhitungan mean dari data training yang telah disediakan.

Tabel 2. Perhitungan Standar Deviasi

Prodi Peminatan	Asal Sekolah	Jurusan	Tahun Lulus	Pekerjaan	Prodi pilihan	Prodi alternatif
Arsitek	4.011	3.893	3.734	3.586	3.044	4.246
Informatika	4.236	4.680	3.562	4.68	4.446	4.551
Visual	4.840	4.840	4.840	4.840	4.840	4.840

Terlihat dalam Pada Tabel 2 didapatkan hasil perhitungan standar deviasi dari data training yang telah disediakan.

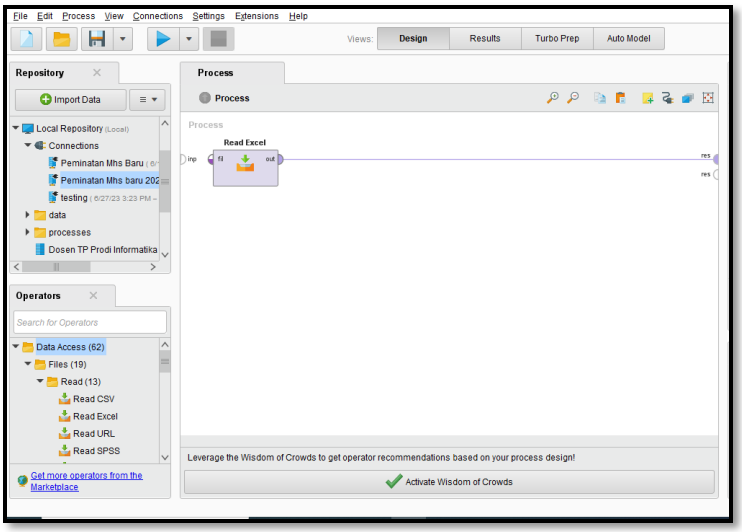
Tabel 3. Probabilitas

Prodi Peminatan	Nilai
Arsitek	0.3873434783443
Informatika	0.4273454795435
Visual	0.2542367881123

Terlihat dalam Pada Tabel 3 didapatkan hasil nilai probabilitas dari data training yang telah disediakan.

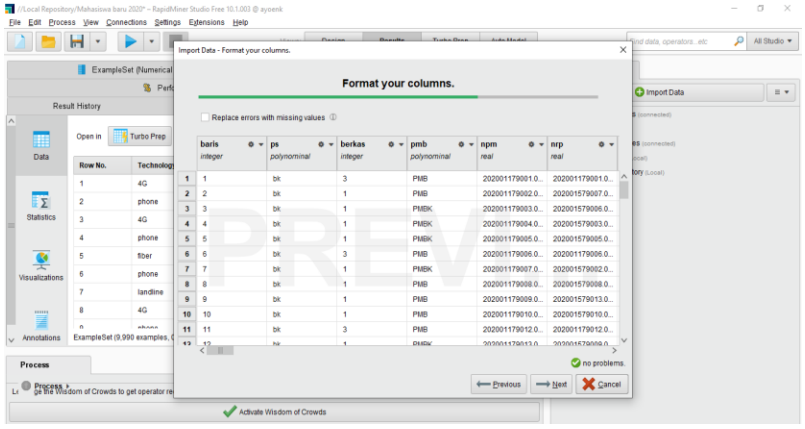
IMPLEMENTASI SISTEM

Setelah serangkaian perhitungan data mining dilakukan dan didapatkan sebuah hasil, maka dilanjutkan perhitungan tersebut dapat di uji dengan menggunakan aplikasi RapidMiner yang dibuat. Berikut bagian awal tampilan dari prediksi jurusan peminatan mahasiswa baru dengan menggunakan algoritma Naïve Bayes. Mulai dari testing koneksi RapidMiner, impor data, data training, data testing, data baru, perhitungan hasil prediksi, klasifikasi dan akurasi.



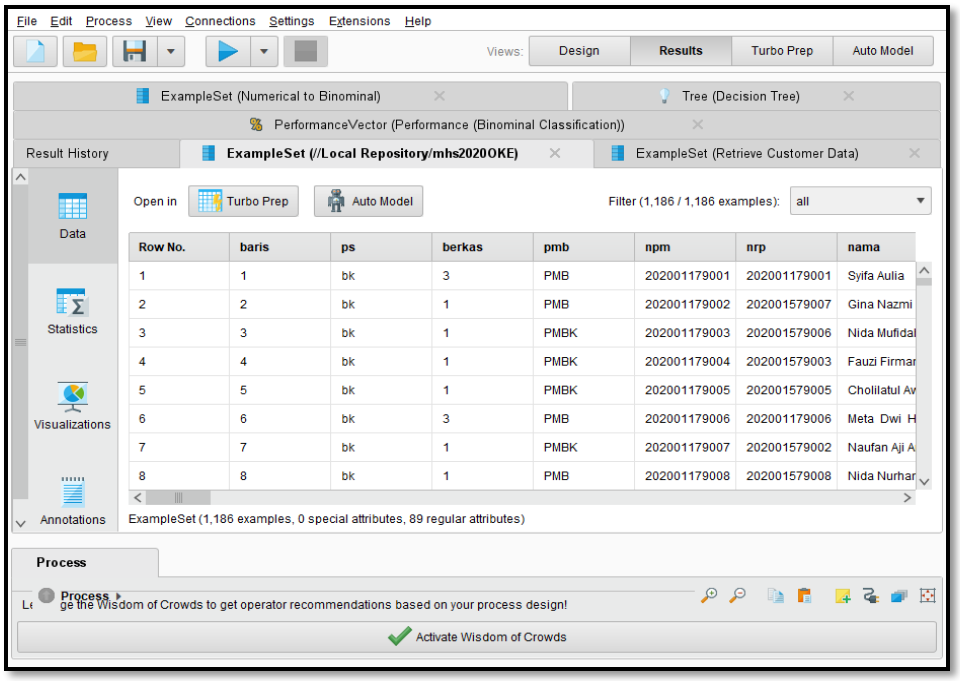
Gambar 4. Testing koneksi rapidminer

Pada gambar 4, tersebut diatas bagian tampilan dari Testing koneksi RapidMiner dimana terlihat fungsi, koneksi dapat dilanjutkan .Dilanjutkan untuk proses data tersebut nanti nya akan diproses untuk perhitungan Naïve Bayes.



Gambar 5. Import Data

Pada gambar 5, merupakan tampilan dari import data dimana terdapat fungsi untuk menambah data atau pun mengimport data training ulang, yang dimana data training tersebut nantinya akan diproses untuk perhitungan Naïve Bayes.



Gambar 6. Tampilan Data Testing

Pada gambar 6 merupakan tampilan dari data training dimana fungsi yang terdapat didalam tampilan data training ini hampir sama dengan gambar 2. Data testing ini nantinya akan digunakan untuk menentukan hasil prediksi dan klasifikasi yang tepat dengan menggunakan algoritma Naïve Bayes. Pada gambar 5 diatas terlihat proses hasil perhitungan prediksi menggunakan algoritma Naïve Bayes berdasarkan data training set yang sudah di import. Setelah proses perhitungan selesai maka hasil akan terlihat prediksi tersebut. Tingkat accuracy : 98,52% +/- 0,58% (micro average : 98,52%)

	true false	true true	class precision
pred.false	9824	3	99,97%
pred.true	145	18	11,04%
class recall	98,55%	85,71%	

Gambar 7. Tampilan Uji Akurasi

Pada gambar 7 tampilan terlihat hasil dari pengolahan data testing maka akan diperoleh hasil pengujian menggunakan metode *Confusion Matrix* untuk melihat nilai dari akurasi, class *precision*, dan class *recall*.

UJI AKURASI

Setelah dilakukan pengolahan terhadap data training set maka akan diperoleh akurasi pada pelatihan data. Untuk perhitungan akurasi menggunakan Confusion Matrix dilakukan dengan cara menjumlah nilai True Positive dengan True Negative kemudian dibagi dengan total keseluruhan data yang ada. Perhitungan tersebut berdasarkan data didapat nilai sebagai berikut :

$$\frac{145 + 3}{145 + 3 + 2 + 0} * 100 = 98.52\%$$

Nilai presisi didapatkan dengan melihat data dan membagi *True false* dengan total penjumlah dari *False Positive* dan *True true* dari data yang ada menjadi data yang ada menjadi class precision

$$\frac{148}{145+1} * 100 = 99.97\%$$

Nilai *recall* didapat dengan cara menghitung nilai *True Positive* yang dibagi dengan total data yang dijumlahkan antara *False Negative* dengan *True Positive* dari data, yaitu sebesar

$$\frac{148}{148+1} * 100 = 98.55\%$$

PENUTUP

Berdasarkan perhitungan , prediksi hasil serta pembahasan tersebut diatas kesimpulan didapat prediksi jurusan peminatan mahasiswa baru dilakukan dengan menggunakan algoritma Naïve Bayes sangat akurat. Hasil pemrosesan data mahasiswa baru diperoleh nilai akurasi dengan model algoritma Naïve Bayes sebesar 98.55%, *precision* sebesar 99.97%, dan *recall* sebesar 98.55%. Model algoritma naive bayes yang diperoleh dapat diimplementasikan ke dalam bentuk aplikasi yang dirancang untuk prediksi peminatan jurusan mahasiswa baru dalam menentukan program studi yang akan diambil.Selain itu dapat disimpulkan algoritma naive bayes tingkat akurasinya sangat cocok digunakan untuk prediksi peminatan mahasiswa baru , dapat digunakan sebagai konsep peminatan jurusan di perguruan tinggi lain.

UCAPAN TERIMAKASIH

Terimakasih kepada Bapak Dr. Arief Wibowo Pengampu Mata kuliah Data Mining di program pasca sarjana ilmu komputer Universitas Budiluhur yang telah memberikan ilmunya , mensupport serta memberikan data, informasi yang valid dan terbaik dalam tulisan ini , serta rekan rekan seperjuangan .

DAFTAR PUSTAKA

[1] A. Yobioktabera and A. W. Wibowo, "Penerapan Data Mining Untuk Memprediksi Penerimaan Calon Mahasiswa Baru Fakultas Kedokteran Menggunakan Algoritma K-NN," *JTET (Jurnal Tek. Elektro Ter.*, vol. 10, no. 1, 2021.

[2] W. Handoko and M. Iqbal, "PREDIKSI PEMINATAN PROGRAM STUDI PADA PENERIMAAN MAHASISWA BARU STMIK ROYAL MENGGUNAKAN NAÏVE BAYES," *J. Sci. Soc. Res.*, vol. 4, no. 2, 2021, doi: 10.54314/jssr.v4i2.661.

[3] H. Ar - Rasyid, S. F. Pane, and M. Y. H. Setyawan, "Pemetaan Profil Mahasiswa Untuk Memprediksi Peminatan Mahasiswa," *PETIR*, vol. 16, no. 1, 2023, doi: 10.33322/petir.v16i1.1337.

[4] K. Latifah, "ANALISIS DAN PENERAPAN ALGORITHMA C45 DALAM DATA MINING UNTUK MENUNJANG STRATEGI PROMOSI PRODI INFORMATIKA UPGRIS," *J. Tek. Inform.*, vol. 11, no. 2, 2018, doi: 10.15408/jti.v11i2.6706.

[5] S. Megira, Kusrini, and E. T. Luthfi, "PERBANDINGAN KINERJA NAIVE BAYES DAN SUPPORT VECTOR MACHINE UNTUK PREDIKSI HERREGISTRASI | Jurnal Sistem Informasi Komputer dan Teknologi Informasi (SISKOMTI)," *J. Siskomti*, vol. 3, no. 2, 2020.

[6] A. H. Mirza, "Application of Naive Bayes Classifier Algorithm in Determining New Student Admission Promotion Strategies Penerapan Algoritma Naive Bayes Classifier Dalam Menentukan Strategi Promosi Penerimaan Mahasiswa Baru," *J. Inf. Syst. Informatics*, vol. 1, no. 1, 2019.

[7] Y. E. Fadrial, "Algoritma Naive Bayes Untuk Mencari Perkiraan Waktu Studi Mahasiswa," *INTECOMS J. Inf. Technol. Comput. Sci.*, vol. 4, no. 1, 2021, doi: 10.31539/intecoms.v4i1.2219.

[8] J. Dongga, A. ` Sarungallo, N. Koru, and G. Lante, "Implementasi Data Mining Menggunakan Algoritma Apriori Dalam Menentukan Persediaan Barang (Studi Kasus: Toko Swapen Jaya Manokwari)," *G-Tech J. Teknol. Terap.*, vol. 7, no. 1, 2023, doi: 10.33379/gtech.v7i1.1938.

[9] A. Andika, S. Syarli, and C. R. Sari, "DATA MINING KLASIFIKASI KELULUSAN MAHASISWA MENGGUNAKAN METODE NAÏVE BAYES," *J. Peqguruang Conf. Ser.*, vol. 4, no. 1, 2022, doi: 10.35329/jp.v4i1.2358.

[10] M. Arifin, "Implementasi Data Mining Pada Prediksi Pemesanan Menggunakan Algoritma Apriori (Studi Kasus : Kimia Farma)," *J. Pelita Inform.*, vol. 8, no. 3, 2020.

[11] I. P. Astuti, "Algoritma Apriori Untuk Menemukan Hubungan Antara Jurusan Sekolah Dengan Tingkat Kelulusan Mahasiswa," *J. Tek. Inform.*, vol. 12, no. 1, 2019, doi: 10.15408/jti.v12i1.10525.

[12] A. Syahrul and A. Solichin, "Rekomendasi Pemilihan Mata Kuliah dalam Pengisian Rencana Studi Mahasiswa dengan Penerapan Algoritma Apriori," *J. Eltikom*, vol. 6, no. 1, 2022, doi: 10.31961/eltikom.v6i1.522.

[13] "Implementasi Data Mining Menggunakan Algoritme Naive Bayes Classifier dan C4.5 untuk Memprediksi Kelulusan Mahasiswa," *Telematika*, vol. 13, no. 1, 2020, doi: 10.35671/telematika.v13i1.881.

[14] S. Raschka and Vahid Mirjalili, "Python Machine Learning Second Edition Machine Learning and Deep Learning with Python, Scikit-learn and TensorFlow," *Taiwan Rev.*, vol. 69, no. 4, 2019.

[15] Z. Zulfauzi and M. N. Alamsyah, "Penerapan Algoritma Naive Bayes Untuk Prediksi Penerimaan Mahasiswa Baru Studi Kasus Universitas Bina Insan Fakultas Komputer," *J. Teknol. Inf. Mura*, vol. 12, no. 02, 2020, doi: 10.32767/jti.v12i02.1096.