Implementasi *Naive Bayes* Classifier Dalam Memprediksi Kelulusan Mahasiswa

Implementation of Naive Bayes Classifier in Predicting Student Graduation

Siti Nuralia*[1], Harliana[2], Tito Prabowo[3]
[1,2,3] Program Studi Ilmu Komputer, Fakultas Ilmu Eksakta, Universitas Nahdlatul Ulama Blitar
*e-mail: *[1]snabelieve@gmail.com, [2]harliana@unublitar.ac.id, [3]titoprabowo@unublitar.ac.id*

Abstrak

Saat ini mutu pendidikan suatu perguruan tinggi dapat dilihat melalui keberhasilan ataupun kegagalan mahasiswa dalam menyelesaikan studinya. Beberapa penelitian mengenai prediksi kelulusan mahasiswa sudah banyak dilakukan, baik yang datasetnya berasal dari tempat penelitian ataupun Kaggle, namun pada penelitian ini penulis melakukan prediksi kelulusan mahasiswa yang datasetnya berasal data Kaggle dengan algoritma Naive Bayes Classifier. Adapun tujuan dari penelitian ini yaitu mengetahui akurasi yang dihasilkan oleh Naive Bayes Classifier dalam melakukan prediksi terhadap lama studi mahasiswa. Untuk mengetahui tingkat akurasi tersebut, penelitian ini akan membagi 500 dataset menjadi 3 skenario pengujian yang berbeda, yaitu scenario I dengan perbandingan antara data training : data testing adalah 80:20, scenario II 50:50 dan scenario II dengan perbandingan 20:80. Analisis terhadap hasil pengujian selanjutnya akan dianalisis menggunakan confusion matrix. Berdasarkan hasil pengujian 3 skenario tersebut didapatkan bahwa scenario I mampu menghasilkan nilai akurasi tertinggi dengan nilai f1-score yang dihasilkanpun diatas 90%.
Kata kunci: *Prediksi, Naive Bayes Classifier, Kelulusan Mahasiswa*

Abstrack

Currently the quality of higher education can be seen based on the success or failure of students in completing their studies. Several studies have been conducted on predicting student graduation, both datasets originating from research sites and Kaggle, but in this study the authors predicted student graduation whose datasets were derived from Kaggle data using the Naive Bayes Classifier algorithm. The purpose of this study was to determine the accuracy produced by the Naive Bayes Classifier in predicting student length of study. To find out the level of accuracy, this study will divide the 500 datasets into 3 different test scenarios, namely scenario I with a comparison between training data: data testing 80:20, scenario II 50:50 and scenario II with a ratio of 20:80. Analysis of the test results will then be analyzed using the confusion matrix. Based on the test results of the 3 scenarios, it is known that scenario I is able to produce the highest accuracy value with an f1-score above 90%.
Keyword: *Prediction, Naive Bayes Classification, Student Graduation*

1. PENDAHULUAN

Perguruan tinggi merupakan lembaga pendidikan tinggi yang menyelenggarakan pendidikan akademik bagi mahasiswa sedangkan mahasiswa merupakan sebagian darikelompok masyarakat yang memiliki tingkat intelektualitas tinggi yang diharapkan dapat menjadi bibit calon pemimpin bangsa kelak seperti sesuai dengan undang- undang nomor 12 tahun 2012. Kualitas suatu perguruan tinggi dapat ditinjau dari tingkatnya kelulusan mahasiswa dan kelulusan mahasiswa menjadi tolok ukur dalam meningkatkan penilaian akreditasi perguruan tinggi juga[1]. Mahasiswa dapat dinyatakan lulus apabila dapat memenuhi syarat atau standar kelulusan yang berlaku disuatu perguruan tinggi[2]. Menurut PermendikbudNo.49 Tahun 2014 tentang Standar Nasional Pendidikan Tinggi (SNP) memberikan paparan bahwa salah satu syarat kelulusan mahasiswa yaitu mahasiswa dapat menempuh kuliah selama 8 semester dengan beban studi minimal 144 SKS. Faktor pendukung dari keberhasilan atau kelulusan mahasiswa dapat ditinjau

History of article:
Received: April, 2023 : Accepted:Mei, 2023

dari kualitas pengajar dan materi pembelajaran yang baik juga proses yang telah tertata dengan baik pula, pengukuran keberhasilan, dan sarana penunjang pendidikan yang dengan mudah dimanfaatkan[3].

Selain kelulusan mahasiswa yang memiliki pengaruh terhadap mutu pendidikan pada perguruan tinggi, nyatanya kegagalan mahasiswa juga dapat memberikan pengaruh terhadap mutu pendidikan pada perguruan tinggi. Berdasarkan SPT 2019 mahasiswa mengalami putus studi terbanyak berdasarkan pulau berada pada pulau Jawa tercatat sebanyak 414.901 disusul pulau Sumatera sebanyak 130.644[4]. Sedangkan Berdasarkan SPT 2020 menyatakan bahwa tercatat mahasiswa yang mengalami DO tercatat sebanyak 602.208 dimana 7% tersebut merupakan total dari mahasiswa terdaftar (8.483.213), hal tersebut menunjukkan bahwa angka putus studi lebih rendah dibandingkan dengan tahun sebelumnya yaitu sebesar 8%[5].

Selain di Indonesia, terdapat juga beberapa masalah terkait dengan adanya putus studi salah satunya adalah putus studi atau *dropout* yang terjadi dibeberapa negara, salah satunya di Eropa dimana pada penelitian tersebut memberikan 3 pertanyaan mengenai pengertian *dropout*, penyebab terjadinya *dropout,* dan tindakan yang harus dilakukan dalam mengurangi atau mencegah adanya *dropout*[6]. Dari pertanyaan tersebut memberikan jawaban bahwa *dropout* dapat terjadi karena beberapa faktor mulai dari latar belakang sosiodemografi yakni meliputi tingkat pendidikan orang tua, status pekerjaan yang dimiliki orang tua. Dari segi karakteristik pribadi mahasiswa dapat ditinjau dari usia, jenis kelamin, dan prestasi akademik sebelumnya dan untuk menjawab pertanyaan ketiga mengenai cara mencegah putus studi (*dropout)* pada penelitian ini memberikan paparan bahwa untuk mencegah adanya *dropout* adalah dengan meningkatkan integrasi akademik dan sosial juga motivasi belajar merupakan salah satu cara dalam mencegah adanya *dropout.* Sehingga untuk memperbaiki mutu pendidikan sebuah perguruan tinggi adalah dengan mengetahui kelulusan mahasiswa secara dini dan meminimalisir adanya *dropout* maka penelitian ini memanfaatkan algoritma *Naive Bayes* dalam melakukan klasifikasi.

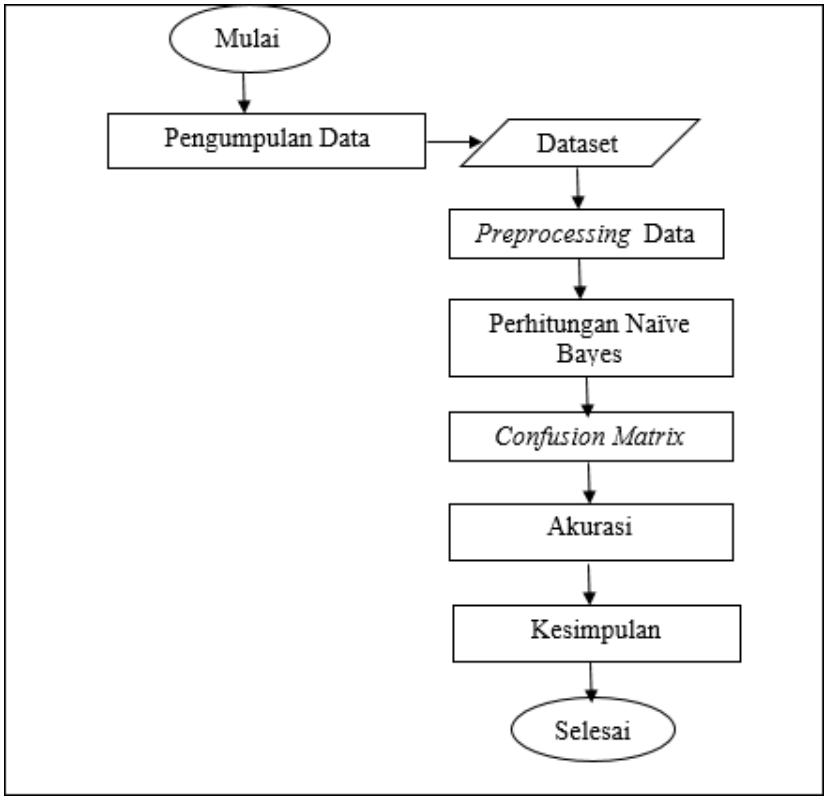
Beberapa penelitian sebelumnya yang telah dilakukan mengenai kelulusan mahasiswa dengan menggunakan algoritma *Naive Bayes* dalam mengklasifikasi kelulusan mahasiswa menggunakan 379 data dengan membagi menjadi data training sebanyak 303 data dan data testing 76 data, atribut atau variabel yang digunakan meliputi nama, status mahasiswa, status perkawinan, IPS, IPK, dan status kelulusan hingga dari penelitian tersebut didapatkan akurasi 88,16%[7]. Penelitian selanjutnya dilakukan perbandingan 4 metode data mining yaitu algoritma C4.5, *Naive Bayes*, KKN, dan SVM dimana topik tersebut mengenai klasifikasi status kelulusan mahasiswa dari sebuah program studi dengan membutuhkan 35 dataset yang akan akan dilakukan pengujian dimana terdapat 12 atribut meliputi nama, jenis kelamin, nim, usia, IPS dari semester 1-8 dan 1 atribut tujuan atau target yaitu status kelulusan dan dari atribut diatas penulis menggunakan atribut usia dan nilai IPS ke 1 sampai 8, sehingga dari 4 algoritma tersebut yang memiliki nilai akurasi tinggi yaitu *Naive Bayes*dengan nilai akurasi 76,79%, untuk SVM memiliki akurasi 74,04%, KNN dengan nilai akurasi 68,05%, dan algoritma C4.5 dengan nilai akurasi 75,96%[8]. Selanjutnya penelitian terdahulu mengenai kelulusan mahasiswa dengan menerapkan algoritma *Naive Bayes* berbasis web dilakukan dengan 20 dataset mahasiswa yang luluspada tahun 2019, dimana pada penelitian tersebut dataset dijadikan sebagai data testingdengan atribut yang digunakan meliputi nama mahasiswa, NPM, jalur masuk, IPK dan jenis kelamin dengan output berupa prediksi terlambat dan tepat waktu dan dari penelitian tersebut didapatkan akurasi 90%[9]. Dalam penelitian ini penulis menggunakan algoritma *Naive Bayes* dikarenakan memiliki nilai akurasi dan kecepatan tinggi saat diaplikasikan ke dalam database dengan data yang besar[10].

Adapun tujuan dari penelitian ini yaitu mengetahui akurasi yang dihasilkan oleh *Naive Bayes* Classifier dalam melakukan prediksi terhadap lama studi mahasiswa berdasarkan dataset kaggle dengan Variabel yang digunakan hanyalah *course, tuition fees up to date, gender, scholarship holder, curicular unit 1st sem(approved). Curicular unit 2nd sem(aprroved). Curicular unit 1st sem(grade)*, dan *curicular unit 2nd sem(grade)*

2. METODE PENELITIAN

Penelitian ini menggunakan data sekunder yang dimana penulis mendapatkandata secara tidak langsung namun dari pihak lain. Data yang digunakan terkaitdengan keberhasilan dan kegagalan mahasiswa yang didapatkan dari situs kaggledengan judul "*Predict Student's Dropout and Academic Success"* dengan link URL https://bit.ly/45fwDP4

Setelah dilakukan pengumpulan data tahap selanjutnya adalah melakukan analisis data. Dimana pada teknik analisis data dilakukan *preprocessing*. Pada penelitian ini dilakukan transformasi data yang mana merupakan salah satu bagian dari *preprocessing* data. Transformasi data adalah upaya yang dilakukan merubah data aslikedalam data mining[11]. Karena dataset yang didapatkan tidak terjadi *missing value* dan *out*lier. Namun pada atribut kelas/label masih berupa data kategorikal, maka untuk itu dilakukan inisialisasi guna memudahkan dalam melakukan perhitungan yaitu memberikan kelas graduate dengan nilai 0 dan *dropout* dengan nilai 1. Gambaran mengenai alur penelitian tergambar pada Gambar 1 yaitu flowchart dari penelitian.



Gambar 1. Flowchart alur penelitian

Berdasarkan Gambar 1, dapat diketahui bahwa penelitian yang akan dilakukan adalah mengumpulkan dataset terlebih dahulu, setelah itu data dilakukan *preprocessing* data guna memeriksa terjadi *missing* value. Setelah *preprocessing* data telah dilakukan, tahap selanjutnya adalah menghitung algoritma *Naive Bayes* dan melakukan evaluasi dengan mengunakan *confusion matrix* dari evaluasi ini akandidapatkan akurasi sehingga tahap terakhir adalah menarik kesimpulan.

3. HASIL DAN PEMBAHASAN

Dataset:

Data yang digunakan pada penelitian ini merupakan data publik yang diperoleh dari situs kaggle yang dikembangkan di *Polytechnic Institute of Portalegre* memberikan informasi terkait

dengan keberhasilan dan kegagalan mahasiswa dimana sumber data yang didapat mencakup data demografis, social ekonomi, kinerja akademik semester pertama dan kedua dengan 34 variabel meliputi *Marital status, Application mode, Application order, Course, Daytime/evening attendance, Previous qualification in higher education, Nacionality, Mother's qualification, Father's qualification, Mother's occupation, Father's occupation, Displaced, Educational special needs, Debtor, Tuition fees up to date, Gender, Scholarship holder, Age at enrollment, International, Curricularunits 1st sem dan 2nd sem (credited), Curricular units 1st sem dan 2nd sem (enrolled), Curricular units 1st dan 2nd sem (evaluations), Curricular units 1st dan 2nd sem (approved), Curricular units 1st dan 2nd sem (without evaluations), inflation rate, GDP*.Dari 34 variabel tersebut hanya diambil beberapa atribut. Penjelasan mengenai atribut yang digunakan terangkum pada Tabel 1.

Tabel 1. Atribut / variable yang digunakan dan penjelasannya.

Variabel/Atribut	Nilai
*Course(*Jurusan)	*1.Biofuel Production Technologies 2.Animation and Multimedia Design 3. Social Service (evening attendance) 4. Agronomy 5.Communication Design 6. Veterinary Nursing 7. Informatics Engineering 8. Equiniculture 9.Management 10. Social Service 11. Tourism 12.Nursing 13. Oral Hygiene 14. Advertising and Marketing Management 15. Journalism and Communication 16. Basic Education 17.Management (evening attendance)*
*Tuitinions Up to date (*Biaya terkini)	0. No 1. Yes
Gender	0. Perempuan 1.Laki-Laki
Sholarship holder	0. No 1.Yes
Curicular 1st sem(approved)	Mata kuliah yang disetujui semester pertama
Curicular 1st sem(grade)	Nilai rata-rata semester pertama
Curicular 2nd sem(approved))	Mata kuliah yang disetujui semester 2
Curicular 2nd sem(grade)	Nilai rata-rata semester 2

Dan untuk nilai atau bobot setiap kelasnya seperti pada tabel

Tabel 2. Kelas (target)

Status	Kelas
Graduate	0
Dropout	1

Seperti pada gambar 1 dan tabel 2 terkait dengan pemilhan atribut yang mana penentuan atribut berdasarkan beberapa penelitian terdahulu. Penentuan berasiswa didasarkan pada penelitian [12] dimana variabel yang digunakan meliputi gender, beasiswa, status mahasiswa dan nilai dan pada penelitian ini dipilih variabel beasiswa sebagai atribut yang akan digunakan. Selanjutnya pemilihan variabel *tuitions up to date* dipilih berdasarkan penelitian [13] dimana penelitian tersebut menjelaskan bahwa penyebab dari kegagalan mahasiswa adalah karena biaya kuliah yang tidak lancer. Dan penelitian [13] dan [14] dimana IPK dan jenis kelamin memiliki pengaruh terhadap keberhasilan dan kegagalan mahasiswa, namun pada penelitian ini atribut yang digunakan adalah *curricular 1st sem (approved).Curricular 2nd sem (approved), curricular 1st sem(grade), curricular 2nd sem(grade),* jenis kelamin, dan *course.s*

Setelah pemilihan atribut yang relevan selanjutnya adalah melakukan *preprocessing data. Preprocessing data* merupakan tahapan sebelum dilakukan proses pengklasifikasian dengan cara membersihkan data, menghilangkan data dengan tujuan agar data yang akan digunakan dapat memberikan nilai yang optimal pada saat digunakan dalam proses pengklasifikasian[15]. *Preprocessing*data pada penelitian ini adalah melakukan pengecekan terkait dengan *missing*

value. Setelah pemilihan atribut dan *preprocessing* data selanjutnya adalah memilih dataset yang akan dijadikan sampel penelitian. *Sampling* yang digunakan pada penelitianini adalah *random under sampling* yaitu metode yang digunakan dalam menangani *imbalance* data. Caranya adalah dengan mengetahui kelas mayoritas dan minoritas setelah itu menyamakan jumlah setiap kelasnya dengan cara mengambil jumlah kelas mayoritas yang sama dengan kelas minoritas secara acak. Setelah melakukan *preprocessing* dan menentukan atribut yang digunakan selanjutnya adalah membagi data menjadi data training dan juga data testing. Proporsi dalam pembagian data bersifat subjektif bergantung kepada penulis [16]. Pada penelitian ini pembagian data dilakukan dengan melakukan 3 scenario pengujian guna mengetahui nilai akurasi dari hasil pengujian 3 skenario yang telah dilakukan

Perhitungan *Naive Bayes* Classification

Dalam perhitungan algoritma *Naive Bayes* ada beberapa langkah yang harus dilakukan sebagai berikut :

a. Menentukan Data Training
   Pada tahapan pertama dalam perhitungan algoritma *Naive Bayes* yaitu menentukan berapa data yang akan dijadikan data training. Pada penelitian ini merupakan pengujian scenario 1 dimana dari 500 data 80% dijadikan datatraining maka data training yang akan disiapkan adalah sebanyak 400 data sebagaimana ditunjukkan pada tabel 3.

Tabel 3 Data Training

No	Course	Tuition up to date	Gender	Scholarship Holder	Curicular 1st sem (approved)	Curicular 1st sem (Grade)	Curicular 2nd sem (approved)	Curicular 2nd sem (Grade)	Target
1	12	1	0	0	7	13,4	8	13,9	0
2	14	1	0	0	4	11,75	3	13,3	0
3	5	1	0	1	6	13,3	6	14,6	0
4	9	1	0	1	5	13,8	5	12	0
5	11	1	0	0	6	13,8	3	12,6	0
…	…	…	…	…	…	…	…	…	…
400	9	1	1	0	1	10	0	-	1

b. Maka tahap selanjutnya yaitu mencari mean (rata-rata) dari setiap variabel dalamtiap kelas dan standar deviasi lalu dibuatkan tabel dari mean(rata-rata) dan jugastandar deviasi tiap kelas.

Tabel 4 Hasil Mean tiap Atribut

	Mean							
Target	**Course**	**Tuition up to date**	**Gender**	**Scholarship Holder**	**Curicular 1st sem (approved)**	**Curicular 1st sem (Grade)**	**Curicular 2nd sem (approved)**	**Curicular 2nd sem (Grade)**
Graduate	9,856436	0,995049505	0,57425743	0,297029703	6,267327	459,3313	6,029703	505,9163
Dropo ut	10,12121	0,717171717	0,27273	0,090909	3,10101	207,3769	2,414141	81,22672

c. Menghitung standar deviasi tiap variabel dari setiap kelas

Tabel 5 Hasil Standar Deviasi Tiap Kelas

	Standar Deviasi							
Target	**Course**	**Tuition up to date**	**Gender**	**Scholarship Holder**	**Curicular 1st sem(approved)**	**Curicular 1st sem(Grade)**	**Curicular 2nd sem(approved)**	**Curicular 2nd sem (Grade)**
Graduate	4.233740706	0.07036	0.438302137	0.458085014	2.928301	2378.296	2.393837	2623.005
Dropout	4.357787189	0.451515	0.4464907	0.288208508	2.675305	1610.053	2.502831	1038.845

d. Tahap keempat adalah menghitung probabilitas kelas sebagaimana pada tabel 6 yang mana merupakan hasil dari perhitungan probabilitas masing-masing label atau target.

Tabel 6. Probabilitas Kelas

kelas	Data Training	Probabilitas
0	202	0,505
1	198	0,495
total	400	1

e. Setelah didapatkan perhitungan mean, standar deviasi, dan juga probabilitas maka tahap selanjutnya yaitu memasukkan data uji atau data testing. Pada penelitian ini data testing yang dijadikan sampel yaitu 20% dari dataset artinya 100 data yang dijadikan data testing.

Tabel 7 Data Testing

No	Course	Tuition up to date	Gender	Scholarship Holder	Curicular 1st sem (approved)	Curicular 1st sem (Grade)	Curicular 2nd sem (approved)	Curicular 2nd sem (Grade)	Target
1	12	1	1	1	7	14,6	8	15,05	0
2	11	1	1	0	6	13,8	5	12,8	0
3	12	12	0	1	7	12,6	7	12,6	0
4	15	1	0	0	10	12,2	9	12,33	0
5	15	1	0	0	6	13,16	6	13	0
6	10	1	0	1	6	11	7	11	0
7	12	1	0	1	7	14,01	8	15,35	0
8	15	1	0	1	4	13	4	11,25	0
9	15	1	0	0	6	14	6	14,66	0
10	12	1	0	1	6	13,8	6	14,16	0
-	-	-	-	-	-	-	-	-	-
100	12	1	1	1	7	13,8	7	13,84	0

Setelah menyiapkan data testing didapatkan bahwa data tersebut berupa data numerik maka tahap selanjutnya yaitu menghitung nilai distribusi normal distribusi Gaussian Hasil distribusigaussian seperti pada tabel 8.

Tabel 8 Hasil Distribusi Gaussian

No	Course	Tuition up to date	Gender	Scholarship Holder	Curicular 1st sem(approved)	Curicular 1st sem(Grade)	Curicular 2nd sem(approved)	Curicular 2nd sem (Grade)
1	12	1	1	1	7	14,6	8	15,05
	0,170605678	1,502520919	0,2941	0,32724	0,22957	0,00811	0,21773	0,00772
	0,182476076	0,538371951	0,30764	0,06179	0,14346	0,00991	0,07261	0,01237
2	11	1	1	0	6	13,8	5	12,8
	0,186988729	1,500662514	0,1435	0,4778	0,23222	0,00804	0,23512	0,00766
	0,187308103	0,488068648	0,15848	0,70723	0,13563	0,00987	0,14791	0,01235
3	12	1	0	1	7	12,6	7	12,6
	0,181897096	1,502520919	0,55294	0,32724	0,22957	0,00811	0,24753	0,00772
	0,182476076	0,538371951	0,54401	0,06179	0,14346	0,00991	0,10897	0,01237
4	15	1	0	0	10	12,2	9	12,33
	0,134092902	1,502520919	0,55294	0,53076	0,15535	0,00811	0,17551	0,00772
	0,13973343	0,538371951	0,54401	0,72504	0,04627	0,00991	0,04467	0,01237
5	15	1	0	0	6	13,16	6	13
	0,092715753	1,500662514	0,50726	0,4778	0,23222	0,00804	0,25789	0,00766
	0,102144144	0,488068648	0,49556	0,70723	0,13563	0,00987	0,09038	0,01235
6	10	1	0	1	6	11	7	11
	0,193880087	1,502520919	0,55294	0,32724	0,23271	0,00811	0,24753	0,00772
	0,191118694	0,538371951	0,54401	0,06179	0,18191	0,00991	0,10897	0,01237
7	12	1	0	1	7	14,01	8	15,35
	0,220456354	1,508109955	0,71621	1,9139	0,24061	0,00833	0,36189	0,00793
	0,209773026	0,722581716	0,71967	107,565	0,7056	0,01002	3,04379	0,01241
8	15	1	0	1	4	13	4	11,25
	0,280485432	1,506244637	0,65703	1,06226	0,2709	0,00825	0,30869	0,00786
	0,2615014	0,655066596	0,65558	8,94169	0,25095	0,00998	0,27887	0,01239
9	15	1	0	0	6	14	6	14,66
	0,405660359	1,508109955	0,71621	0,72752	0,23417	0,00833	0,25793	0,00793
	0,357734533	0,722581716	0,71967	0,78122	0,43884	0,01002	0,70395	0,01241
10	12	1	0	1	7	14,4	7	14,4
	0,220456354	1,508109955	0,71621	1,9139	0,24061	0,00833	0,27999	0,00793
	0,209773026	0,722581716	0,71967	107,565	0,7056	0,01002	1,35149	0,01241
-	-	-	-	-	-	-	-	-
400	12	1	1	1	7	13,8	7	13,84
	0,170605678	1,502520919	0,2941	0,32724	0,22957	0,00811	0,24753	0,00772
	0,174190592	0,538371951	0,30764	0,06179	0,14346	0,00991	0,10897	0,01237

Setelah didapatkan nilai distribusi gaussian tahap selanjutnya yaitu mengalikan probabilitas kelas dengan hasil nilai distribusi gaussian sehingga didapatkan probabilitas akhir seperti pada tabel 9

Tabel 9 Hasil Probabilitas Akhir

No	Graduate (0)	Dropout (1)	Prediksi
1	3,90124E-08	1,18011E-09	0
2	3,2651E-08	1,24112E-08	0
3	8,88929E-08	3,13124E-09	0

4	5,09956E-08	3,72028E-09	0
5	6,27689E-08	1,29309E-08	0
6	9,60316E-08	4,15804E-09	0
-	-	-	-
400	4,43491E-08	1,69049E-09	0

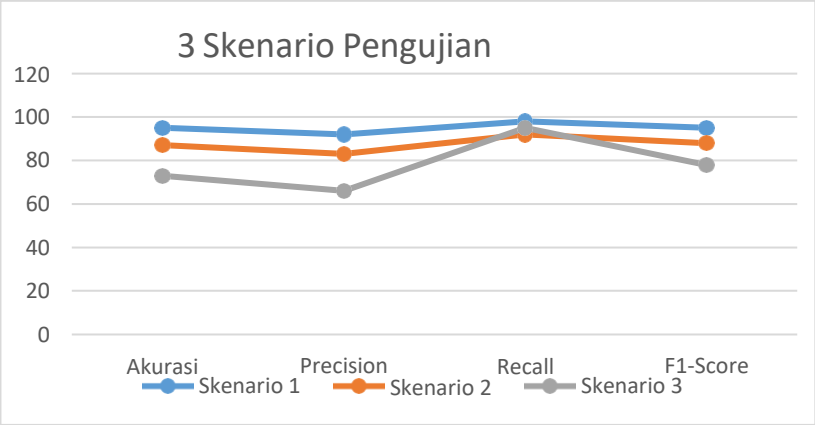
Analisis Hasil Melalui Pengujian Sistem

Untuk mengetahui kinerja *Naive Bayes* dalam memprediksi kelulusan mahasiswa,maka penulis melakukan 3 pengujian terhadap dataset yang digunakan, Adapun rencana pengujian terangkum pada tabel 10.

Tabel 10 Rencana Pengujian

No	Rencana Pengujian	Prosentase Data Training	Prosentase Data Testing
1	Skenario Pengujian 1	80% (400 dataset)	20% (100 dataset)
2	Skenario Pengujian 2	50% (250 dataset)	50% (250 dataset)
3	Scenario Pengujian 3	20% (100 dataset)	80% (400 dataset)

Rangkuman mengenai hasil pengujian 3 skenario terangkum pada gambar 4,18



Gambar 2. Grafik Confusion Matrix 3 Scenario

Berdasarkan gambar 2 terlihat bahwa *Naive Bayes* mampu melakukan prediksidengan tepat melalui berbagai scenario yang telah dilakukan. Hal ini ditunjukkan dengan rata-rata nilai akurasi diatas. Namun dari 3 skenario yang telah dilakukan kinerja terbaik terlihat pada scenario ke 1 dengan hasil akurasi tertinggi yaitu 95%. Selain itu performa dari algoritma *Naive Bayes* diukur juga dengan nilai *precision, recall,* dan *fi-score*. Untuk mempermudah dalam mengetahui masing-masing *precision,recall,* dan *f1-score* dari tiap kelas dari 3 scenario yang telah dilakukan pengujian sebagaimana pada tabel 11.

Tabel 11. Pengujian 3 Skenario

	Presisi		Recall		F1-Score		Akurasi
Skenario	0	1	0	1	0	1	
1	92%	97,87%	98%	92%	95%	94,85%	95%
2	83%	90,99%	92,06%	81,45%	87,55%	85,96%	87%
3	66%	90,83%	95%	50%	78%	65%	73%

Berdasarkan gambar 2 dan tabel 11 terkait hasil dari nilai *precision, recall* dan *f1- score* dengan melakukan pengujian 3 scenario maka didapatkan bahwa nilai akurasi terbaik berada

pada scenario 1 dengan nilai *presisi, recall,*dan *fi-score* dari masing- masing kelas adalah 92%, 98%, 95% untuk kelas graduate(0) sedangkan untuk kelas dropout sebesar 98%, 92%, 95%. Pada scenario 2 mengalami penurunan terhadap akurasi dengan didapatkan sebesar 87% meskipun untuk nilai presisi, recall, dan f1-scoredari masing-masing kelas memiliki persentase yang bagus. Pada scenario 3 akurasi mengalami penurunan dibanding dengan scenario 2, dimana akurasi pada scenario 3 didapatkan sebesar 72% meskipun hasil *recall* dari skenario 3 lebih tinggi daripada skenario 2 sebagaimana pada gambar grafik gambar 2.

Model terbaik yaitu model yang memiliki *f1-score* tertinggi dimana dapat dilihat nilai dari *precision* dan *recall*dari model yang bersangkutan[17]. Dikatakan bahwa semakin kecil nilai TP maka presisi akan semakin besar dan semakin kecil nilai dari FN maka recall akan semakin besar[18]. Dari 3 skenario yang telah dilakukan maka pada scenario 1 merupakan hasil yang baik baik dari nilai presisi, recall, dan f1-score dari masing-msingkelas dengan akurasi yang didapatkan sebesar 95% maka hasil klasifikasi pada scenario 1 termasuk dalam *Excellent Classification* sebagaimana pada pedoman terkait dengan parameter hasil klasifikasi[19].

4. KESIMPULAN

Hasil penelitian implementasi algoritma *Naive Bayes* dalam memprediksi kelulusan mahasiswa dapat disimpulkan bahwa algoritma *Naive Bayes* dapat diimplementasikan dengan data prediksi kelulusan mahasiswa dengan melakukan 3 scenario pengujian dimana scenario 1 mendapatkan hasil akurasi tertinggi yaitu sebesar 95% dengan proporsi perbandingan 80:20, selain itu jugaterdapat dengan nilai rata-rata *precision, recall* dan *fi-score* masing masing sebesar 95,16%, 95%, dan 95%.

5. SARAN

Berikut saran untuk mengembangkan penelitian yang akan datang adalah menambah jumlah data dan beberapa atribut yang digunakan. Program dapat menerima masukan bukan hanya dengan file csv namun jugadalam bentuk file lainnya seperti xlxs. Serta menambahkan beberapa fitur pada streamlit seperti korelasi antar atribut

DAFTAR PUSTAKA

[1] A. Primadewi and M. Hanafi, "Pengelolaan Data Terintegrasi Berdasarkan Instrumen Akreditasi Perguruan Tinggi 3.0 Menggunakan Zachman Framework," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 4, no. 6, pp. 5–10, 2020, doi: 10.29207/resti.v4i6.2540.

[2] Herni, "Manajemen Sistem Penjaminan Mutu Internal (SPMI) Dalam Meningkatkan Mutu Lulusan Perguruan Tinggi," *al-Afkar, J. Islam. Stud.*, no. 7, pp. 281–289, 2022, doi: 10.31943/afkarjournal.v5i4.380.

[3] L. Setiyani, M. Wahidin, D. Awaludin, and S. Purwani, "Analisis Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode Data Mining Naïve Bayes : Systematic Review," *Fakt. Exacta*, vol. 13, no. 1, p. 35, 2020, doi: 10.30998/faktorexacta.v13i1.5548.

[4] W. J. Hussar and T. M. Bailey, "Projections of Education Statistics to 2027 (NCES 2019-001)," 2019.

[5] F. A. Yusuf, "The independent campus program for higher education in indonesia: The role of government support and the readiness of institutions, lecturers and students," *J. Soc. Stud. Educ. Res.*, vol. 12, no. 2, pp. 280–304, 2021.

[6] B. M. Kehm, M. R. Larsen, and H. B. Sommersel, "Student dropout from universities in Europe: A review of empirical literature," *Hungarian Educ. Res. J.*, vol. 9, no. 2, pp. 147–

164, 2020, doi: 10.1556/063.9.2019.1.18.
[7] R. P. S. Putri and I. Waspada, "Penerapan Algoritma C4.5 pada Aplikasi Prediksi Kelulusan Mahasiswa Prodi Informatika," *Khazanah Inform. J. Ilmu Komput. dan Inform.*, vol. 4, no. 1, pp. 1–7, 2018, doi: 10.23917/khif.v4i1.5975.
[8] S. Widaningsih, "Perbandingan Metode Data Mining Untuk Prediksi Nilai Dan Waktu Kelulusan Mahasiswa Prodi Teknik Informatika Dengan Algoritma C4.5, Naïve Bayes, KNN Dan SVM," *J. Tekno Insentif*, vol. 13, no. 1, pp. 16–25, 2019, doi: 10.36787/jti.v13i1.78.
[9] Y. Apridiansyah, N. D. M. Veronika, and E. D. Putra, "Prediksi Kelulusan Mahasiswa Fakultas Teknik Informatika Universitas Muhammadiyah Bengkulu Menggunakan Metode Naive Bayes," *JSAI (Journal Sci. Appl. Informatics)*, vol. 4, no. 2, pp. 236–247, 2021, doi: 10.36085/jsai.v4i2.1701.
[10] R. Hasudungan and W. J. Pranoto, "Implementasi Teorema Naïve Bayes Pada Prediksi Prestasi Mahasiswa," *J. Rekayasa Teknol. Inf.*, vol. 5, no. 1, p. 10, 2021, doi: 10.30872/jurti.v5i1.4996.
[11] H. D. Fahma and A. C. Fauzan, "Prediksi Keberlangsungan Studi Mahasiswa Fakultas Ilmu Pendidikan dan Sosial Universitas Nahdlatul Ulama Blitar Menggunaan Algoritma C4.5," *JACIS J. Autom. Comput. Inf. Syst.*, vol. 1, no. 2, pp. 110–119, 2021, doi: 10.47134/jacis.v1i2.21.
[12] N. Yahya and A. Jananto, "Komparasi Kinerja Algoritma C4.5 Dan Naive Bayes Untuk Prediksi Kegiatan Penerimaan Mahasiswa Baru (Studi Kasus : Universitas Stikubank Semarang)," *Pros. SENDI*, no. 2014, pp. 978–979, 2019, [Online]. Available: https://www.unisbank.ac.id/ojs/index.php/sendi_u/article/view/7389
[13] B. Gunawan Sudarsono and A. Ulan Bani, "Prediksi Mahasiswa Berpotensi Berhenti Kuliah Secara Sepihak Menggunakan Data Mining Algoritma C4.5," *J. Sains Komput. Inform. (J-SAKTI*, vol. 4, no. 2, pp. 359–367, 2020, doi: 10.30645/j-sakti.v4i2.227.
[14] M. Nasir, "Penerapan Algoritma Naive Bayes Classifier Untuk Evaluasi Kinerja Akademik Mahasiswa Universitas Bina Darma," *J. Inform. dan Komputer)*, vol. 5, no. 2, pp. 81–88, 2021, doi: 10.26798/jiko.v5i2.227.
[15] S. Hasanah, I. Purwasih, and I. Santoso, "Analisis Sentimen Terhadap Masyarakat Adanya Uang Kertas Baru Menggunakan Algoritma K-Nearest Neighbor(Knn)," vol. 7, no. 2, pp. 105–114, 2023, [Online]. Available: https://journals.upi-yai.ac.id/index.php/ikraith-informatika/issue/archive
[16] S. AISYAH, S. WAHYUNINGSIH, and F. AMIJAYA, "Peramalan Jumlah Titik Panas Provinsi Kalimantan Timur Menggunakan Metode Radial Basis Function Neural Network," *Jambura J. Probab. Stat.*, vol. 2, no. 2, pp. 64–74, 2021, doi: 10.34312/jjps.v2i2.10292.
[17] N. L. P. C. Savitri, R. A. Rahman, R. Venyutzky, and N. A. Rakhmawati, "Analisis Klasifikasi Sentimen Terhadap Sekolah Daring pada Twitter Menggunakan Supervised Machine Learning," *J. Tek. Inform. dan Sist. Inf.*, vol. 7, no. 1, pp. 47–58, 2021, doi: 10.28932/jutisi.v7i1.3216.
[18] M. Meiriyama, S. Devella, and S. M. Adelfi, "Klasifikasi Daun Herbal Berdasarkan Fitur Bentuk dan Tekstur Menggunakan KNN," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 9, no. 3, pp. 2573–2584, 2022, doi: 10.35957/jatisi.v9i3.2974.
[19] D. P. Pertiwi, W. Wiranto, and R. Anggrainingsih, "Evaluation of Campaign Categories on Kitabisa. Com By Naive Bayes Classifier Method," *ITSMART J. Teknol. dan Inf.*, vol. 8, no. 1, 2019, doi: 10.20961/itsmart.v8i1.27426.