

PENGEMBANGAN SISTEM CERDAS UNTUK PREDIKSI DAFTAR KEMBALI MAHASISWA BARU DENGAN METODE NAIVE BAYES (STUDI KASUS: UNIVERSITAS PENDIDIKAN GANESHA)

Komang Aditya Pratama¹, Gede Aditra Pradnyana², I Ketut Resika Arthana³

¹Prodi Pendidikan Teknik Informatika, Jurusan Teknik Informatika, Fakultas Teknik dan Kejuruan, Universitas Pendidikan Ganesha
Jl. Udayana No.11, Banyuasri, Kec. Buleleng, Kabupaten Buleleng, Bali 81116 Bali, Indonesia

e-mail: komang.aditya.pratama@undiksha.ac.id¹, gede.aditra@undiksha.ac.id², resika@undiksha.ac.id³

Received : February, 2020

Accepted : March, 2020

Published : March, 2020

Abstract

Ganesha University of Education or Undiksha is one of the state universities in Bali, precisely in the city of Singaraja. In the admission of new students, Undiksha applies 3 admissions paths, as follows the State University National Admission Selection (SNMPTN), State University Joint Entrance Test (SBMPTN), and Independent Entrance Test (SMBJM) consisting of 2 parts namely Computer Based Test (CBT) and Interests and Talents. Each year the committees are busy with the re-registration of prospective students. In determining the number of students quota for re-registration, they are still using the manual method in form of an excel file, so they want to use a system to do the process. These problems can be overcome by using "Intelligent System for Re-Registration of New Students Prediction using the Naive Bayes Method (Case Study: Ganesha University of Education)". The Naive Bayes method is used to determine the re-register probability of the new students so that the number of students who re-register can be determining the new students quota. In developing the system, the researcher use the CRISP-DM methodology as a standard of data mining process as well as a research method. The results of this prediction system research show that the system can predict well with the average predictive system accuracy value of 75.56%.

Keywords: intelligent system, data mining, prediction, naive bayes, CRISP-DM

Abstrak

Universitas Pendidikan Ganesha atau Undiksha merupakan salah satu universitas negeri yang berada di Bali, tepatnya di kota Singaraja. Dalam penerimaan mahasiswa baru, Undiksha menerapkan 3 jalur penerimaan, antara lain jalur Seleksi Nasional Masuk Perguruan Tinggi Negeri (SNMPTN), Seleksi Bersama Masuk Perguruan Tinggi Negeri (SBMPTN), dan Seleksi Mahasiswa Baru Jalur Mandiri (SMBJM) yang terdiri 2 bagian yaitu Computer Based Test (CBT) dan Minat dan Bakat. Setiap tahunnya para panitia disibukkan dengan pendaftaran kembali calon mahasiswa yang berjumlah puluhan ribu. Dalam menentukan jumlah mahasiswa pendaftar kembali yang diterima, mereka masih menggunakan cara manual berupa file excel, sehingga mereka menginginkan untuk menggunakan sistem dalam pengerjaannya. Permasalahan tersebut dapat diatasi dengan menggunakan "Sistem Cerdas untuk Prediksi Daftar Kembali Penerimaan Mahasiswa Baru dengan Metode Naive Bayes (Studi Kasus: Universitas Pendidikan Ganesha)". Metode Naive Bayes digunakan untuk menentukan probabilitas daftar kembali mahasiswa baru sehingga dari jumlah mahasiswa yang mendaftar kembali dapat ditentukan berapa jumlah mahasiswa baru yang diluluskan. Dalam mengembangkan sistem, peneliti menggunakan Metodologi CRISP-DM sebagai standar proses data mining sekaligus metode penelitian.

Hasil penelitian sistem prediksi ini menunjukkan bahwa sistem dapat memprediksi dengan baik dengan rata-rata nilai akurasi sistem prediksi sebesar 75.56%.

Kata Kunci: sistem cerdas, data mining, prediksi, naive bayes, CRISP-DM

1. PENDAHULUAN

Universitas Pendidikan Ganesha atau Undiksha merupakan salah satu universitas negeri yang berada di Bali, tepatnya di kota Singaraja. Dalam penerimaan mahasiswa baru, Undiksha menerapkan 3 jalur penerimaan, antara lain jalur Seleksi Nasional Masuk Perguruan Tinggi Negeri (SNMPTN), Seleksi Bersama Masuk Perguruan Tinggi Negeri (SBMPTN), dan Seleksi Mahasiswa Baru Jalur Mandiri (SMBJM) yang terdiri 2 bagian yaitu *Computer Based Test* (CBT) dan Minat dan Bakat.

Setiap tahunnya para panitia disibukkan dengan pendaftaran kembali calon mahasiswa yang berjumlah puluhan ribu. Dalam pelaksanaannya, tim penerimaan tidak mengalami permasalahan dalam menentukan jumlah mahasiswa baru yang diluluskan, namun mereka merasa perlu adanya prediksi yang lebih akurat terhadap jumlah mahasiswa baru yang mendaftar kembali. Mekanisme yang telah mereka lakukan adalah dengan melihat dari persentase daftar kembali tahun sebelumnya untuk mempertimbangkan kelulusan berikutnya. Dalam menentukan jumlah mahasiswa baru yang diluluskan, mereka masih menggunakan cara manual berupa file excel, sehingga mereka juga menginginkan untuk menggunakan sistem dalam pengerjaannya.

Dari permintaan tersebut peneliti membuat sistem dengan memilih metode Naive Bayes sebagai metode perhitungan data, karena metode ini dapat menemukan faktor-faktor penting untuk memprediksi dalam sistem cerdas dengan tingkat akurasi yang tinggi dan dapat berjalan dalam database yang besar dengan cepat [1].

Peneliti juga lebih mempertimbangkan untuk memilih metode Naive Bayes karena metode naive bayes memiliki kinerja yang lebih baik jika dibandingkan dengan metode lain, kinerja naive bayes masih tetap unggul ketika pengujian dilakukan pada tipe data kategori [2], sebagaimana yang didukung oleh penelitian yang dilakukan oleh Dian Oktafia dan D. L.

Crispina Pardede dengan judul “Perbandingan Kinerja Algoritma Decision Tree dan Naive Bayes dalam Prediksi Kebangkrutan” [2]. Dalam penelitian tersebut, nilai akurasi algoritma decision tree untuk data numeric adalah sebesar 96.97% dan data kategori sebesar 84.85%. Sedangkan algoritma naive bayes menghasilkan nilai yang lebih besar dibandingkan algoritma decision tree yaitu 100% untuk data numeric dan 87.88% untuk data kategori. Pada penelitian lain yang dilakukan oleh Supardi Salmu dan Achmad Solichin dengan judul “Prediksi Tingkat Kelulusan Mahasiswa Tepat Waktu Menggunakan Naive Bayes: Studi Kasus UIN Syarif Hidayatullah Jakarta”. Penelitian ini bertujuan untuk menemukan faktor-faktor yang mempengaruhi prediksi kelulusan mahasiswa melalui data kinerja akademik mahasiswa pada semester satu sampai semester empat. Hasil akhir penelitian tersebut menunjukkan bahwa metode *Naive bayes* dapat dilakukan untuk memprediksi kelulusan mahasiswa tepat waktu pada UIN Syarif Hidayatullah Jakarta, dengan akurasi pengujian data yang diperoleh sebesar 80,72% dari 1162 data yang digunakan untuk *data train* dan 587 data untuk *data test* [3]. Penelitian lainnya dilakukan oleh Mujib Ridwan, Hadi Suyono, dan M. Sarosa dengan judul “Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier”. Penelitian ini bertujuan untuk mengembangkan suatu sistem yang dapat mengklasifikasi kelulusan mahasiswa Universitas Brawijaya pada tahun ke-2 dengan melihat pola kelulusan mahasiswa beberapa periode sebelumnya. Hasil akhir penelitian tersebut menunjukkan bahwa NBC dapat menentukan faktor yang paling berpengaruh dalam penentuan klasifikasi kinerja akademik mahasiswa yaitu Indeks Prestasi Kumulatif (IPK), Indeks Prestasi (IP) semester 1, IP semester 4, dan jenis kelamin. Sehingga faktor-faktor tersebut dapat digunakan sebagai bahan evaluasi bagi pihak pengelola perguruan tinggi [4]. Penelitian lainnya dilakukan oleh Gede Aditra Pradnyana dan Agus Aan Jiwa Permana. dengan judul “Sistem Pembagian Kelas Kuliah Mahasiswa dengan Metode K-Means dan K-

Nearest Neighbors untuk Meningkatkan Kualitas Pembelajaran". Penelitian ini bertujuan untuk membandingkan kinerja dari metode *hierarchical clustering* dan *K-means clustering* dalam pembentukan kelas siswa serta melihat karakteristik siswa dari kelas yang berhasil dibentuk. Dari hasil penelitian yang dilakukan diperoleh kesimpulan bahwa metode *K-means* lebih tepat digunakan untuk mengelompokkan siswa dan menentukan karakter kelompok siswa tersebut [5]. Penelitian lainnya dilakukan oleh Alvino Dwi Rachman Prabowo dan Muljono dengan judul "Prediksi Nasabah Yang Berpotensi Membuka Simpanan Deposito Menggunakan Naive Bayes Berbasis Particle Swarm Optimization". Tujuan dari penelitian ini adalah untuk menentukan strategi pemasaran dan promosi bank-bank yang lebih efisien dengan tidak terlalu banyak menggunakan biaya sehingga masyarakat tertarik untuk berinvestasi pada produk deposito dari bank tersebut. Hasil dari prediksi nasabah yang berpotensi membuka deposito dengan menggunakan Naive Bayes memiliki akurasi sebesar 82,19%. Sedangkan prediksi yang menggunakan Naive Bayes berbasis PSO memiliki akurasi sebesar 89,70%. Penggunaan algoritma PSO ternyata meningkatkan akurasi sebesar 7,51% dan algoritma Naive Bayes berbasis PSO tersebut dapat digunakan untuk decision support system nasabah yang berpotensi membuka deposito karena menjadi model algoritma yang terbaik [6]. Penelitian lainnya dilakukan oleh Alfa Saleh dengan judul "Implementasi metode klasifikasi naive bayes dalam memprediksi besarnya penggunaan listrik rumah tangga". Tujuan dari penelitian ini adalah untuk mampu memprediksi besarnya penggunaan listrik tiap rumah tangga agar lebih mudah mengatur penggunaan listrik. Hasil dari prediksi menunjukkan bahwa dari 60 data penggunaan listrik rumah tangga yang diuji dengan metode naive bayes, maka diperoleh hasil persentase 78,3333% untuk keakuratan prediksi, di mana dari 60 data penggunaan listrik rumah tangga yang diuji terdapat 47 data penggunaan listrik rumah tangga yang berhasil diklasifikasikan dengan benar. [7]. Penelitian lainnya dilakukan oleh Hera Wasiati dan Dwi Wijayanti dengan judul "Sistem Pendukung Keputusan Penentuan Kelayakan Calon Tenaga Kerja Indonesia Menggunakan Metode Naive Bayes (Studi Kasus: Di PT Karyatama Mitra Sejati Yogyakarta)". Tujuan dari penelitian ini adalah untuk membantu Staf dalam

menentukan siapa calon tenaga kerja Indonesia yang layak diterima atau tidak. Dari hasil pengujian yang dilakukan dengan menggunakan data sebanyak 542 dengan 362 sebagai data training dan 180 sebagai data tes, akurasi polanya sebesar 73.89 % dan errornya 26.11% jadi jumlah data yang tepat sebanyak 133 dan yang tidak tepat 47. [8].

Berdasarkan uraian di atas, maka peneliti tertarik untuk mengembangkan sistem prediksi yang berjudul "Pengembangan Sistem Cerdas untuk Prediksi Daftar Kembali Penerimaan Mahasiswa Baru dengan Metode Naive Bayes (Studi Kasus: Universitas Pendidikan Ganesha)". Rumuskan permasalahan yang perlu dicarikan solusi pada sistem tersebut adalah sebagai berikut:

1. Bagaimana merancang sistem cerdas untuk prediksi daftar kembali penerimaan mahasiswa baru dengan metode naive bayes (Studi Kasus: Universitas Pendidikan Ganesha)?
2. Bagaimana implementasi rancangan sistem cerdas untuk prediksi daftar kembali penerimaan mahasiswa baru dengan metode naive bayes (Studi Kasus: Universitas Pendidikan Ganesha)?
3. Bagaimana kinerja dari sistem cerdas untuk prediksi daftar kembali penerimaan mahasiswa baru dengan metode naive bayes (Studi Kasus: Universitas Pendidikan Ganesha)?

Berdasarkan permasalahan yang sudah dirumuskan, adapun tujuan yang ingin dicapai adalah sebagai berikut:

1. Untuk merancang sistem cerdas prediksi daftar kembali penerimaan mahasiswa baru dengan metode naive bayes (Studi Kasus: Universitas Pendidikan Ganesha).
2. Untuk mengimplementasikan rancangan sistem cerdas prediksi daftar kembali penerimaan mahasiswa baru dengan metode naive bayes (Studi Kasus: Universitas Pendidikan Ganesha).
3. Untuk mengetahui kinerja dari sistem cerdas prediksi daftar kembali penerimaan mahasiswa baru dengan metode naive bayes (Studi Kasus: Universitas Pendidikan Ganesha).

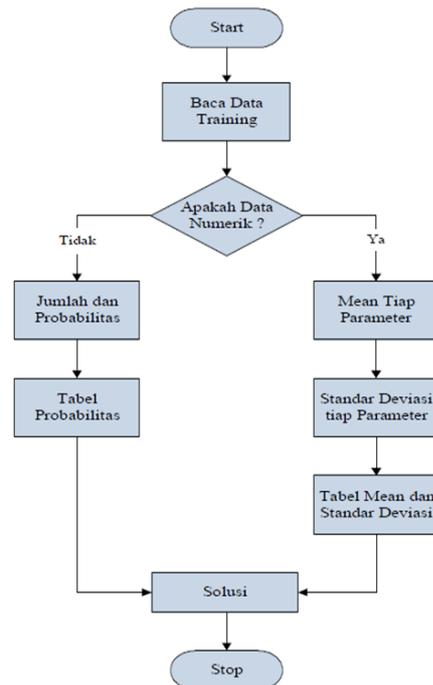
Adapun batasan masalah dari penelitian ini menggunakan dataset penerimaan mahasiswa Undiksha tahun 2017-2019.

pada tahap Modelling. Sistem tahap ini dikembangkan berbasis web dengan menggunakan framework Laravel, template AdminLTE, bahasa pemrograman python pada sistem prediksi, dan bisa diintegrasikan dengan sistem DSS Undiksha.

3. HASIL DAN PEMBAHASAN

3.1 Penerapan Metode Naive Bayes

Naive Bayes merupakan sebuah pengklasifikasian probabilistik sederhana yang menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset yang diberikan. Algoritma menggunakan teorema Bayes dan mengasumsikan semua atribut independen atau tidak saling ketergantungan yang diberikan oleh nilai pada variabel kelas (Patil & Sherekar, 2013). Dalam metode Naive Bayes data string yang bersifat konstan dibedakan dengan data numerik yang bersifat kontinu, perbedaan ini akan terlihat pada saat menentukan nilai probabilitas setiap kriteria baik itu kriteria dengan nilai data string maupun kriteria dengan nilai data numerik. Adapun penerapan metode Naive Bayes sebagai berikut.



Gambar 3.1. Alur Metode Naive Bayes [10]

3.1.1. Baca Data Training

Untuk menentukan data yang nantinya akan dianalisis dengan metode Naive Bayes maka langkah pertama yang dilakukan adalah membaca data train. Misalnya data train pada jalur SBMPTN.

Tabel 3.1: Data Train SBMPTN

nama	nama sekolah	ukt	pilihan 1	pilihan 2	pilihan 3	prodi pilihan	daftar kembali
Mhs 1	SMAS KATOLIK SANTO YOSEPH	UKT Kelompok 4	-	Pendidikan Matematika	Pendidikan Biologi	3	TIDAK
Mhs 2	SMAN 3 TEBING TINGGI	UKT Kelompok 4	-	Pendidikan Matematika	Pendidikan Kimia	3	YA
Mhs 3	SMAN 18 MEDAN	UKT Kelompok 1	-	-	Pendidikan Biologi	3	YA
Mhs 4	SMAN 1 SIDIKALANG	UKT Kelompok 4	-	-	Pendidikan Biologi	3	YA
Mhs 5	SMAS PELITA	UKT Kelompok 4	Pendidikan Teknik Informatika	Pendidikan Kimia	-	1	YA
Mhs 6	SMA SWASTA SANTA MARIA KABANJAHE	UKT Kelompok 4	-	Pendidikan Biologi	-	2	YA
Mhs 7	SMAN 1 SILAEN	UKT Kelompok 4	-	-	Pendidikan Fisika	3	YA
Mhs 8	-	UKT Kelompok 4	-	-	Pendidikan Biologi	3	TIDAK
Mhs 9	SMA SWASTA SANTA MARIA KABANJAHE	UKT Kelompok 4	-	-	Pendidikan Kimia	3	YA
...
Mhs 4007	SMKN 3 SINGARAJA	UKT Kelompok 4	Pendidikan Teknik Informatika	Ilmu Hukum	-	1	YA

3.1.2. Kriteria dan Probabilitas

Pada data train sebelumnya kemudian dicari nilai probabilitas daftar kembali masing-masing

kriteria. Adapun nilai probabilitas masing-masing kriteria adalah sebagai berikut.

Tabel 3.2: Nilai kriteria nama sekolah

nama_sekolah	YA	TIDAK	prob YA	prob TIDAK
-	0	62	0.000285063	0.029871977
SMA ISLAM TERPADU MENTARI ILMU KARAWANG	1	0	0.000570125	0.000474158
SMA PGRI 1 PATI	1	0	0.000570125	0.000474158
SMAN 1 BABAKAN	1	0	0.000570125	0.000474158
SMAN 1 CIKARANG BARAT	0	1	0.000285063	0.000948317
SMAN 1 JEMBER	1	0	0.000570125	0.000474158
SMAN 1 KALITIDU	1	0	0.000570125	0.000474158
SMAN 1 KENDAL	1	0	0.000570125	0.000474158
SMAN 1 LAHAT	1	0	0.000570125	0.000474158
...
SMKS WIRA HARAPAN	3	3	0.001140251	0.001896633
Jumlah	2703	1304	0.674569503	0.325430496

Tabel 3.3: Nilai kriteria ukt

ukt	YA	TIDAK	prob YA	prob TIDAK
Bidikmisi	552	107	0.20398377	0.082317073
UKT Kelompok 1	116	51	0.043157506	0.039634146
UKT Kelompok 2	152	60	0.056436739	0.046493902
UKT Kelompok 3	282	156	0.104389524	0.119664634
UKT Kelompok 4	587	382	0.216894135	0.291920732
UKT Kelompok 5	37	5	0.014016968	0.004573171
UKT Kelompok 6	289	23	0.106971597	0.018292683
UKT Kelompok 7	688	520	0.25414976	0.397103659
Jumlah	2703	1304	0.674569503	0.325430496

Tabel 3.4: Nilai kriteria pilihan 1

pilihan_1	YA	TIDAK	prob YA	prob TIDAK
-	1010	796	0.368843488	0.593889717
Akuakultur	9	1	0.003648304	0.001490313
Akuntansi (S1)	273	69	0.099963517	0.052160954
Bimbingan dan Konseling	19	12	0.007296607	0.009687034
Biologi (S1)	2	3	0.001094491	0.002980626
Ilmu Hukum	66	18	0.024443634	0.014157973
Ilmu Keolahragaan	9	2	0.003648304	0.002235469
Ilmu Komputer	9	4	0.003648304	0.003725782
Kedokteran	27	9	0.01021525	0.007451565
...
Teknologi Pendidikan	30	8	0.011309741	0.006706408
Jumlah	2703	1304	0.674569503	0.325430496

Tabel 3.5: Nilai kriteria pilihan 2

pilihan_2	YA	TIDAK	prob YA	prob TIDAK
-	778	533	0.284099198	0.397617275
Akuakultur	6	8	0.002552881	0.006701415
Akuntansi (S1)	167	58	0.061269147	0.043931497
Analisis Kimia	99	16	0.03646973	0.012658228
Bimbingan dan Konseling	56	32	0.020787746	0.024571854
Biologi (S1)	8	4	0.003282276	0.003723008
Ilmu Hukum	115	45	0.042304887	0.034251675
Ilmu Keolahragaan	3	7	0.001458789	0.005956813
Ilmu Komputer	25	5	0.00948213	0.00446761
...
Teknologi Pendidikan	49	32	0.018234865	0.024571854
Jumlah	2703	1304	0.674569503	0.325430496

Tabel 3.6: Nilai kriteria pilihan 3

pilihan_3	YA	TIDAK	prob YA	prob TIDAK
-	1597	604	0.583424608	0.451492537
Akuakultur	6	6	0.002555677	0.005223881
Akuntansi (S1)	89	45	0.032858708	0.034328358
Analisis Kimia	209	62	0.076670318	0.047014925
Bimbingan dan Konseling	39	20	0.01460387	0.015671642
Biologi (S1)	6	7	0.002555677	0.005970149
Ilmu Hukum	74	43	0.027382256	0.032835821
Ilmu Keolahragaan	3	0	0.001460387	0.000746269
Kimia	7	11	0.002920774	0.008955224
...
Teknologi Pendidikan	33	18	0.01241329	0.014179104
Jumlah	2703	1304	0.674569503	0.325430496

Tabel 3.7: Nilai kriteria prodi pilihan

prodi_pilihan	YA	TIDAK	prob YA	prob TIDAK
1	1259	266	0.465631929	0.204284621
2	952	532	0.35218034	0.407804132
3	492	506	0.182187731	0.387911247
Jumlah	2703	1304	0.674569503	0.325430496

3.1.3. Prediksi Data Test

Pada probabilitas kriteria sebelumnya kemudian digunakan untuk memprediksi daftar kembali pada data test dengan kriteria yang

sama. Adapun data test dengan jalur yang sama dengan data train (SBMPTN) adalah sebagai berikut.

Tabel 3.8: Data Test SBMPTN

nama	nama sekolah	ukt	pilihan 1	pilihan 2	pilihan 3	prodi pilihan	daftar kembali
Test 1	SMAN 1 SAPEKEN	UKT Kelompok 2	Akuntansi (S1)	Pendidikan IPA	-	2	???
Test 2	MAN PATAS	Bidikmisi	Pendidikan Seni Rupa	Pendidikan Guru Sekolah Dasar	-	1	???
Test 3	SMAN 1 BANJAR	Bidikmisi	Pendidikan Teknik Informatika	Pendidikan Bahasa dan Sastra Indonesia	Pendidikan Bahasa Bali	2	???
Test 4	SMAN BALI MANDARA	UKT Kelompok 1	-	-	Pendidikan Guru Sekolah Dasar	3	???
Test 5	SMKN 2 KINTAMANI	Bidikmisi	Pendidikan IPA	Pendidikan Guru Sekolah Dasar	-	1	???
Test 6	SMAN 1 PAYANGAN	UKT Kelompok 4	Pendidikan Guru Sekolah Dasar	Pendidikan Teknik Informatika	Bimbingan dan Konseling	3	???
Test 7	SMAN 4 SINGARAJA	UKT Kelompok 3	-	-	Akuntansi (S1)	3	???
Test 8	SMKN 2 NEGARA	UKT Kelompok 4	Pendidikan Kepelatihan Olahraga	Pendidikan Guru Sekolah Dasar	Ilmu Keolahragaan	1	???
Test 9	SMKN 3 SINGARAJA	UKT Kelompok 4	Pendidikan Teknik Elektro	Pendidikan Teknik Mesin	Akuntansi (S1)	1	???
...
Test 400	SMAN 7 DENPASAR	UKT Kelompok 7	-	Pendidikan Guru Pendidikan Anak Usia Dini	-	2	???

Kemudian nilai kriteria pada data test dilihat nilai probabilitas pada tabel probabilitas sebelumnya. Sebagai contoh perhitungan akan dilakukan untuk mahasiswa baru data test yang bernama Test 1. Untuk perhitungan mahasiswa baru yang lain dapat dilakukan dengan cara yang sama.

$$P(YA) = 2703/4007 = 0.6745$$

$$P(\text{nama sekolah} = \text{SMKN 1 Sapeken} \mid YA) = 0.0051$$

$$P(\text{ukt} = \text{UKT Kelompok 2} \mid YA) = 0.0564$$

$$P(\text{pilihan 1} = \text{Akuntansi (S1)} \mid YA) = 0.0999$$

$$P(\text{pilihan 2} = \text{Pendidikan IPA} \mid YA) = 0.0149$$

$$P(\text{pilihan 3} = - \mid YA) = 0.0003$$

$$P(\text{prodi pilihan} = 2 \mid YA) = 0.3521$$

$$P(\text{TIDAK}) = 1304/4007 = 0.3254$$

$$P(\text{nama sekolah} = \text{SMKN 1 Sapeken} \mid \text{TIDAK}) = 0.0028$$

$$P(\text{ukt} = \text{UKT Kelompok 2} \mid \text{TIDAK}) = 0.0464$$

$$P(\text{pilihan 1} = \text{Akuntansi (S1)} \mid \text{TIDAK}) = 0.0521$$

$$P(\text{pilihan 2} = \text{Pendidikan IPA} \mid \text{TIDAK}) = 0.0089$$

$$P(\text{pilihan 3} = - \mid \text{TIDAK}) = 0.0007$$

$$P(\text{prodi pilihan} = 2 \mid \text{TIDAK}) = 0.4078$$

Terakhir, menghitung probabilitas akhir dengan menggunakan persamaan teorema Bayes.

$$P(H|X) = \frac{P(X|H) \cdot P(H)}{P(X)}$$

Dimana:

X: Data dengan *class* yang belum diketahui

H: Hipotesis data merupakan suatu *class* spesifik

P(H|X): Probabilitas hipotesis H berdasar kondisi X (posteriori probabilitas)

P(H): Probabilitas hipotesis H (prior probabilitas)

P(X|H): Probabilitas X berdasarkan kondisi pada hipotesis H

P(X): Probabilitas X

P(X) disini dapat dihilangkan karena nilai P(X) pada kedua probabilitas akhir memiliki nilai yang sama sehingga tidak mempengaruhi perbandingan nilai probabilitas akhir keduanya.

$$P(YA|daftar kembali) = 0.6745 \times 0.0051 \times 0.0564 \times 0.0999 \times 0.0149 \times 0.0003 \times 0.3521 = 3.05 \times 10^{-11}$$

$$P(TIDAK|daftar kembali) = 0.3254 \times 0.0028 \times 0.0464 \times 0.0521 \times 0.0089 \times 0.0007 \times 0.4078 = 5.59 \times 10^{-12}$$

Maka didapat kesimpulan bahwa nilai daftar kembali pada data mahasiswa Test 1 adalah YA. Berikut merupakan nilai daftar kembali untuk perhitungan keseluruhan data test dengan melakukan cara yang sama.

Tabel 3.9: Data Test SBMPTN Hasil Prediksi

nama	nama sekolah	ukt	pilihan 1	pilihan 2	pilihan 3	prodi pilihan	daftar kembali
Test 1	SMAN 1 SAPEKEN	UKT Kelompok 2	Akuntansi (S1)	Pendidikan IPA	-	2	YA
Test 2	MAN PATAS	Bidikmisi	Pendidikan Seni Rupa	Pendidikan Guru Sekolah Dasar	-	1	YA
Test 3	SMAN 1 BANJAR	Bidikmisi	Pendidikan Teknik Informatika	Pendidikan Bahasa dan Sastra Indonesia	Pendidikan Bahasa Bali	2	YA
Test 4	SMAN BALI MANDARA	UKT Kelompok 1	-	-	Pendidikan Guru Sekolah Dasar	3	TIDAK
Test 5	SMKN 2 KINTAMANI	Bidikmisi	Pendidikan IPA	Pendidikan Guru Sekolah Dasar	-	1	YA
Test 6	SMAN 1 PAYANGAN	UKT Kelompok 4	Pendidikan Guru Sekolah Dasar	Pendidikan Teknik Informatika	Bimbingan dan Konseling	3	TIDAK
Test 7	SMAN 4 SINGARAJA	UKT Kelompok 3	-	-	Akuntansi (S1)	3	TIDAK
Test 8	SMKN 2 NEGARA	UKT Kelompok 4	Pendidikan Keperawatan Olahraga	Pendidikan Guru Sekolah Dasar	Ilmu Keolahragaan	1	YA
Test 9	SMKN 3 SINGARAJA	UKT Kelompok 4	Pendidikan Teknik Elektro	Pendidikan Teknik Mesin	Akuntansi (S1)	1	YA
...
Test 400	SMAN 7 DENPASAR	UKT Kelompok 7	-	Pendidikan Guru Pendidikan Anak Usia Dini	-	2	TIDAK

3.2 Pengujian Metode Naive Bayes

Dari nilai-nilai probabilitas yang telah dicari kemudian dilakukan proses testing untuk mengetahui bug/masalah yang terdapat pada sistem yang dikembangkan, proses testing yang dilakukan antara lain blackbox testing, whitebox testing, uji kesesuaian, uji akurasi, uji kompatibilitas, dan uji respon pengguna.

1. Blackbox Testing

Pelaksanaan pengujian *blackbox* perangkat lunak ini dilakukan untuk menguji seluruh fungsional pada setiap fungsional sistem. Pengujian ini dilakukan oleh Bapak Anom selaku pengguna pada hari Jumat, 31 Januari 2020. Pengujian yang telah

dilakukan menunjukkan hasil yang baik ditunjukkan dengan semua fungsional berjalan sesuai dengan kebutuhan fungsional yang telah dirancang sebelumnya.

2. Whitebox Testing

Pelaksanaan pengujian *whitebox* berupa pengujian algoritma dilakukan oleh peneliti pada hari Selasa-Rabu, 28-29 Januari 2020. Dari hasil pengujian *whitebox* untuk uji algoritma tersebut dapat dinyatakan bahwa semua algoritma yang digunakan dalam pengembangan sistem cerdas untuk prediksi daftar kembali penerimaan mahasiswa baru di undiksha dengan

metode naive bayes sudah benar dan *output* sudah sesuai dengan tujuan dari masing-masing algoritma tersebut.

3. Uji Kesesuaian

Pelaksanaan pengujian kesesuaian dari perhitungan manual dengan hasil perhitungan yang terdapat pada sistem telah dilakukan oleh peneliti. Dari hasil pengujian tersebut dapat dinyatakan bahwa hasil perhitungan sistem cerdas untuk prediksi daftar kembali penerimaan mahasiswa baru di undiksha dengan metode naive bayes dengan perhitungan manual sudah benar dan *output* sudah sesuai.

4. Uji Akurasi

Pelaksanaan uji akurasi atau performance sistem dari perhitungan manual dengan hasil perhitungan yang terdapat pada sistem telah dilakukan oleh peneliti. Pengujian akurasi dilakukan dengan menggunakan

metode naive bayes, k-fold cross validation dan confusion matrix. K-fold cross validation merupakan salah satu metode yang digunakan untuk mengetahui rata-rata keberhasilan dari suatu sistem dengan cara melakukan perulangan dengan mengacak atribut masukan sehingga sistem tersebut teruji untuk beberapa atribut input yang acak, sedangkan confusion matrix merupakan suatu metode perhitungan yang mengelompokan jumlah data train yang benar diklasifikasikan dan jumlah data train yang salah diklasifikasikan. Berikut merupakan hasil pengujian performance berupa akurasi, presisi, recall, specificity, dan F1 yang dilakukan terhadap data train setiap prodi pada masing-masing jalur penerimaan. (Hasil pengujian data train setiap prodi merupakan rata-rata dari hasil penerapan K-fold cross validation dengan k=10 pada setiap prodinya

Tabel 3.10: Pengujian Performance Jalur SNMPTN

No	Prodi	Total	Salah	Akurasi	Presisi	Recall	Specificity	F1 Score	Pilihan 1	Pilihan 2	Pilihan 3
1	Kedokteran	11	0	100.00 %	100.00 %	100.00 %	NaN	100.00 %	100.00 %	NaN	NaN
2	Matematika (S1)	36	10	73.33 %	90.74 %	71.67 %	87.50%	79.89 %	29.17 %	100.00 %	NaN
3	Biologi (S1)	18	5	70.00 %	83.33 %	57.14 %	80.00%	86.67 %	40.00 %	NaN	NaN
4	Kimia	46	27	40.00 %	42.59 %	35.83 %	43.33%	53.33 %	41.33 %	66.67 %	NaN
5	Akuakultur	17	9	50.00 %	50.00 %	66.67 %	30.00%	80.00 %	61.11 %	50.00 %	NaN
6	Ilmu Komputer	10	1	90.00 %	90.00 %	100.00 %	0.00%	100.00 %	100.00 %	NaN	NaN
7	Sistem Informasi	50	21	58.00 %	70.83 %	64.81 %	65.42%	65.32 %	44.83 %	100.00 %	NaN
8	Manajemen	272	64	76.42 %	78.59 %	92.08 %	33.76%	84.38 %	85.29 %	NaN	NaN
9	Akuntansi (S1)	286	50	82.56 %	83.76 %	98.01 %	16.44%	89.79 %	95.84 %	NaN	NaN
...
34	Pendidikan Seni Rupa	23	5	80.00 %	85.00 %	94.44 %	0.00%	92.59 %	95.00 %	NaN	NaN
	Rata-Rata	89.12	23.53	70.91 %	74.52 %	78.31 %	44.88%	77.34 %	67.74 %	49.86 %	NaN

Tabel 3.11: Pengujian Performance Jalur SBMPTN

No	Prodi	Total	Salah	Akurasi	Presisi	Recall	Specificity	F1 Score	Pilihan 1	Pilihan 2	Pilihan 3
1	Kedokteran	20	0	100.00 %	100.00 %	100.00 %	NaN	100.00 %	100.00 %	100.00 %	NaN
2	Matematika (S1)	16	10	45.00 %	50.00 %	57.14 %	33.33%	100.00 %	50.00 %	100.00 %	25.00 %

3	Biologi (S1)	27	17	38.33 %	40.48 %	37.50 %	48.15%	66.67 %	40.00 %	80.00 %	33.33 %
4	Kimia	30	12	60.00 %	61.90 %	54.76 %	72.22%	73.33 %	100.00 %	50.00 %	0.00 %
5	Akuakultur	27	9	68.33 %	80.95 %	55.56 %	85.00%	85.56 %	100.00 %	0.00%	20.00 %
6	Ilmu Komputer	31	10	66.67 %	68.33 %	85.00 %	0.00%	82.22 %	92.86 %	94.44 %	NaN
7	Sistem Informasi	70	19	72.86 %	81.08 %	73.50 %	65.00%	77.30 %	100.00 %	54.76 %	21.11 %
8	Manajemen	248	59	76.20 %	83.36 %	83.41 %	36.21%	83.03 %	100.00 %	81.52 %	0.00 %
9	Akuntansi (S1)	299	92	69.20 %	84.46 %	66.50 %	53.69%	70.34 %	100.00 %	39.49 %	0.00 %
...
37	PENDIDIKAN VOKASIONAL SENI KULINER	21	8	61.67 %	83.33 %	55.00 %	78.57%	83.33 %	100.00 %	11.11 %	NaN
	Rata-Rata	108.3	35.24	66.95 %	72.75 %	71.70 %	49.40%	75.55 %	94.69 %	65.95 %	16.13 %

Tabel 3.10: Pengujian Performance Jalur CBT

No	Prodi	Total	Salah	Akurasi	Presisi	Recall	Specificity	F1 Score	Pilihan 1	Pilihan 2	Pilihan 3
1	Kedokteran	72	7	90.00%	90.00%	100.00 %	0.00%	94.49%	100.00 %	NaN	NaN
2	Kebidanan (D3)	82	17	79.17%	82.56%	95.83%	0.00%	87.50%	98.75%	50.00%	NaN
3	Teknik Elektronika	12	0	100.00 %	100.00 %	100.00 %	NaN	100.00 %	100.00 %	100.00 %	NaN
4	Analisis Kimia	16	4	70.00%	66.67%	100.00 %	20.00%	90.48%	100.00 %	0.00%	NaN
5	Survey & Pemetaan (D3)	23	6	76.67%	81.67%	95.00%	0.00%	83.33%	95.00%	NaN	NaN
6	Matematika (S1)	10	2	80.00%	75.00%	100.00 %	50.00%	100.00 %	100.00 %	0.00%	NaN
7	Biologi (S1)	21	5	76.67%	69.44%	75.00%	72.22%	82.67%	100.00 %	0.00%	NaN
8	Kimia	11	6	50.00%	0.00%	0.00%	71.43%	NaN	25.00%	0.00%	NaN
9	Akuakultur	19	7	60.00%	72.22%	83.33%	0.00%	83.33%	100.00 %	50.00%	0.00%
...
40	Perhotelan	56	14	75.00%	76.33%	98.00%	4.17%	84.52%	98.00%	100.00 %	0.00%
	Rata-Rata	69.4	10.32	76.78%	77.82%	89.16%	13.32%	92.40%	95.25%	39.12%	43.75 %

Tabel 3.11: Pengujian Performance Jalur Minat dan Bakat

No	Prodi	Total	Salah	Akurasi	Presisi	Recall	Specificity	F1 Score	Pilihan 1	Pilihan 2	Pilihan 3
1	Kebidanan (D3)	152	36	76.46 %	82.88 %	91.01 %	20.37%	85.76 %	90.42%	NaN	NaN
2	Teknik Elektronika	39	10	74.17 %	74.17 %	86.67 %	39.58%	78.90 %	80.00%	NaN	NaN
3	Analisis Kimia	38	17	55.00 %	64.81 %	76.67 %	33.33%	64.44 %	72.50%	NaN	NaN

4	Survey & Pemetaan (D3)	39	9	77.50 %	81.48 %	85.19 %	53.33%	80.00 %	72.50%	NaN	NaN
5	Manajemen Informatika	111	32	71.14 %	77.22 %	80.25 %	52.00%	74.93 %	67.05%	NaN	NaN
6	Akuntansi (D3)	58	27	53.67 %	60.00 %	69.17 %	35.19%	67.54 %	68.00%	100.00 %	NaN
7	Perpustakaan (D3)	38	15	61.67 %	62.50 %	86.67 %	20.37%	70.57 %	87.50%	0.00%	NaN
8	Bahasa Inggris	98	22	77.00 %	83.63 %	83.33 %	65.83%	80.60 %	60.67%	NaN	NaN
9	Desain Komunikasi Visual (D3)	36	5	86.67 %	89.17 %	97.50 %	0.00%	92.29 %	100.00 %	0.00%	NaN
10	Perhotelan	92	12	87.22 %	91.67 %	88.21 %	82.17%	89.42 %	64.78%	0.00%	NaN
	Rata-Rata	70.1	18.5	72.05 %	76.75 %	84.47 %	40.22%	78.45 %	76.34%	25.00%	NaN

Berdasarkan pengujian yang telah dilakukan terhadap 10582 data train yang terdiri dari 4 jalur pendaftaran, didapat hasil rata-rata

performance keseluruhan dari rata-rata masing-masing jalur pendaftaran

Tabel 3.12: Pengujian Performance Rata-Rata

No	Jalur	Total	Salah	Akurasi	Presisi	Recall	Specificity	F1 Score	Pilihan 1	Pilihan 2	Pilihan 3
1	SNMPTN	89.12	23.53	70.91 %	74.52 %	78.31 %	44.88%	77.34 %	67.74 %	49.86 %	NaN
2	SBMPTN	108.3	35.24	66.95 %	72.75 %	71.70 %	49.40%	75.55 %	94.69 %	65.95 %	16.13 %
3	CBT	69.4	10.32	76.78 %	77.82 %	89.16 %	13.32%	92.40 %	95.25 %	39.12 %	43.75 %
4	Minat dan Bakat	70.1	18.5	72.05 %	76.75 %	84.47 %	40.22%	78.45 %	76.34 %	25.00 %	NaN
	Rata-Rata	84.23	21.90	71.67 %	75.46 %	80.91 %	36.96%	80.94 %	83.51 %	44.98 %	29.94 %

Sehingga sistem prediksi memiliki rata-rata akurasi sebesar 71,67%, rata-rata presisi sebesar 75,46%, rata-rata recall sebesar 80,91%, rata-rata specificity sebesar 36,96%, dan rata-rata f1 score sebesar 80,94%. Disini bisa dilihat juga nilai probabilitas secara keseluruhan mahasiswa untuk mendaftar kembali di pilihan 1 sebesar 83,51%, probabilitas di pilihan 2 sebesar 44,98%, dan probabilitas di pilihan 3 sebesar 29,94%. Dari hasil pengujian tersebut dapat dinyatakan bahwa hasil akurasi sistem prediksi ini tergolong *cukup baik* dalam memprediksi daftar kembali pada mahasiswa baru

5. Uji Respon Pengguna

Pelaksanaan pengujian respon pengguna dilakukan dengan menggunakan *System Usability Scale (SUS)* kepada pengguna yang dapat mengakses perangkat lunak yaitu

administrator. Berdasarkan hasil pengujian yang telah dilakukan terhadap responden diperoleh rata – rata nilai SUS sebesar 85%. Hal ini menunjukkan bahwa tingkat usability sistem tergolong dalam kategori *excellent* atau sangat baik dan dapat diterima serta digunakan dengan mudah oleh pengguna.

4. KESIMPULAN

4.1 Kesimpulan

Berdasarkan penelitian Pengembangan Sistem Cerdas untuk Prediksi Daftar Kembali Penerimaan Mahasiswa Baru dengan Metode Naive Bayes (Studi Kasus: Universitas Pendidikan Ganesha) dapat ditarik beberapa kesimpulan sebagai berikut.

1. Sistem ini dikembangkan dengan menggunakan bahasa pemrograman PHP dengan *framework laravel* dan bahasa

pemrograman python, dan telah mampu memprediksi daftar kembali atau tidaknya mahasiswa baru yang telah lulus jalur penerimaan.

2. Berdasarkan pengujian yang telah dilakukan terhadap 10582 data uji yang terdiri dari 4 jalur pendaftaran. Maka dari rata-rata keempat jalur pendaftaran tersebut, sistem prediksi memiliki rata-rata akurasi sebesar 71,67%, rata-rata presisi sebesar 75,46%, rata-rata recall sebesar 80,91%, rata-rata specificity sebesar 36,96%, dan rata-rata f1 score sebesar 80,94%. Disini bisa dilihat juga nilai probabilitas secara keseluruhan mahasiswa untuk mendaftar kembali di pilihan 1 sebesar 83,51%, probabilitas di pilihan 2 sebesar 44,98%, dan probabilitas di pilihan 3 sebesar 29,94%. Dari hasil pengujian tersebut dapat dinyatakan bahwa hasil akurasi sistem prediksi ini tergolong *cukup baik* dalam memprediksi daftar kembali pada mahasiswa baru

4.2 Saran

Berdasarkan pengamatan peneliti di lapangan, terdapat beberapa hal yang dapat dijadikan bahan pertimbangan untuk pengembangan berikutnya. Diantaranya adalah sebagai berikut.

1. Diharapkan pengembangan selanjutnya dapat memilih metode pengembangan yang lebih efektif dan memilih atribut data yang lebih relevan agar sistem dapat mendapatkan tingkat akurasi yang lebih tinggi.
2. Diharapkan pengembangan selanjutnya dapat menambahkan fitur tambah pengguna agar sistem ini bisa dikelola oleh pengguna selain admin.

Penelitian selanjutnya dapat dilakukan dengan pengujian penelitian eksperimen atau *user experience testing* (UX) terhadap Sistem Cerdas untuk Prediksi Daftar Kembali Penerimaan Mahasiswa Baru dengan Metode Naive Bayes (Studi Kasus: Universitas Pendidikan Ganesha).

DAFTAR PUSTAKA

- [1] Prasetyo, E. Data Mining konsep dan Aplikasi menggunakan MATLAB. Yogyakarta: Andi, 2012.
- [2] Oktafia, D. Perbandingan Kinerja Algoritma Decision Tree dan Naive Bayes dalam Prediksi Kebangkrutan. 3, 1–5, 2014.
- [3] Salmu, S., & Solichin, A. Prediksi Tingkat Kelulusan Mahasiswa Tepat Waktu

Menggunakan Naïve Bayes : Studi Kasus UIN Syarif Hidayatullah Jakarta. *Prosiding Seminar Nasional Multidisiplin Ilmu Universitas Budi Luhur*, (April), 701–709, 2017.

- [4] Ridwan, M., Suyono, H., & Sarosa, M. Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier. *Eecis*, 7(1), 59–64. <https://doi.org/10.1038/hdy.2009.180>. 2013.
- [5] Pradnyana G. A. & Permana A. A. J. Sistem Pembagian Kelas Kuliah Mahasiswa dengan Metode K-Means dan K-Nearest Neighbors untuk Meningkatkan Kualitas Pembelajaran. *JUTI: Jurnal Ilmiah Teknologi Informasi*. [Online]. 16(1), hal. 59–68. Tersedia: <http://juti.if.its.ac.id/index.php/juti/article/view/696>. 2018.
- [6] Prabowo, A.D.R., and Muljono, Prediksi Nasabah Yang Berpotensi Membuka Simpanan Deposito Menggunakan Naive Bayes Berbasis Particle Swarm Optimization, Semarang: Universitas Dian Nuswantoro. *Techno.COM*, Vol. 17, No. 2, 208-219, 2018.
- [7] Saleh, Alfa. "Implementasi metode klasifikasi naive bayes dalam memprediksi besarnya penggunaan listrik rumah tangga." *Creative Information Technology Journal* 2.3 (2015): 207-217.
- [8] Wasiati, Hera, and Dwi Wijayanti. "Sistem Pendukung Keputusan Penentuan Kelayakan Calon Tenaga Kerja Indonesia Menggunakan Metode Naive Bayes (Studi Kasus: Di PT Karyatama Mitra Sejati Yogyakarta)." *IJNS-Indonesian Journal on Networking and Security* 3.2 (2014).
- [9] Taylor. James. "Four Problems in Using CRISP-DM and How to Fix Them". Internet: <https://www.kdnuggets.com/2017/01/four-problems-crisp-dm-fix.html>, 29 Januari, 2017 [Juni. 8, 2019].
- [10] Bustami, Bustami. "Penerapan Algoritma Naive Bayes Untuk Mengklasifikasi Data Nasabah Asuransi." *TECHSI-Jurnal Teknik Informatika* 5.2 (2013).