

Penerapan *Data Mining* Untuk Memprediksi Kelulusan Mahasiswa Menggunakan Algoritma *Naive Bayes Classifier* (Studi Kasus STMIK Primakara)

Putu Sainanda Cahyani Moonallika^{1*}, Ketut Queena Fredlina², I.B. Kresna Sudiatmika³

^{1,2,3}Jurusan Teknik Informatika, STMIK Primakara, Denpasar

^{1,2,3}Jl. Tukad Badung No. 135 Denpasar, Telp. (0361) 8956085

²naa.queena@gmail.com, ³kresna@primakara.ac.id

*Corresponding Author: cahyanimoonallika97@gmail.com

ABSTRAK

Tingkat ketepatan kelulusan mahasiswa merupakan salah satu indikator penting untuk menilai kualitas mahasiswa maupun perguruan tinggi. Prediksi ketepatan waktu lulus mahasiswa diharapkan dapat membantu pihak manajemen perguruan tinggi untuk menjaga kestabilan tingkat kelulusan mahasiswa, agar kualitas lulusan dapat terus dikendalikan dengan baik. Artikel ini menyajikan penggunaan metode *naive bayes classifier* untuk memprediksi kelulusan mahasiswa. Pengujian akurasi sistem dilakukan dengan mencocokkan hasil prediksi dengan data yang sesungguhnya dengan metode pengujian *confusion matrix*. Hasil uji menunjukkan *recall* sebesar 80%, *accuracy* sebesar 80% dan *precision* sebesar 80%.

Kata Kunci: *Data Mining, Naive Bayes Classifier, Prediksi Kelulusan Mahasiswa*

ABSTRACT

The level of accuracy of student graduation is one important indicator to assess the quality of students and universities. The predicted timeliness of graduating students is expected to help the management of higher education to maintain a stable level of student graduation, so that the quality of graduates can continue to be controlled well. This article presents the use of the Naive Bayes classifier method to predict student graduation. System accuracy testing is done by matching the results of predictions with actual data with the confusion matrix testing method. The test results showed a recall of 80%, an accuracy of 80% and a precision of 80%.

Keywords: *Data Mining, Naive Bayes Classifier, Student Graduation Prediction*

1. Pendahuluan

Tingkat ketepatan kelulusan mahasiswa merupakan salah satu indikator penting bagi mahasiswa maupun perguruan tinggi karena hal tersebut berpengaruh besar terhadap kualitas perguruan tinggi tersebut. Selain perguruan tinggi, program studi juga harus bertindak kritis terhadap tingkat kelulusan mahasiswanya untuk mencegah terjadinya kemunduran waktu studi dari mahasiswa tersebut. Adapun 2 (dua) faktor utama yang mempengaruhi kelulusan mahasiswa yaitu faktor internal dan eksternal. Faktor internal adalah faktor dari dalam diri yang meliputi kecerdasan mahasiswa, penangkapan materi saat dikelas serta kemampuan mahasiswa itu sendiri, sedangkan faktor eksternal adalah faktor dari luar diri yang meliputi kondisi ekonomi, kondisi sosial serta pergaulan. Selain mempengaruhi kelulusan, faktor tersebut juga akan mempengaruhi ekonomi mahasiswa seperti harus membayar uang semester lebih untuk menanggung matakuliah yang tidak lulus atau ketertinggalan mata kuliah.

Berdasarkan buku pedoman akademik STMIK Primakara, besarnya beban studi yang harus diselesaikan oleh mahasiswa adalah setara dengan 144 SKS sampai 160 SKS dengan batas waktu studi 14 semester. Bobot SKS per semester yang mesti diselesaikan oleh mahasiswa berkisar antara 16 sampai 24 SKS [1]. Prediksi Hasil pembelajaran mahasiswa sangat penting dalam sistem pendidikan. Hal ini dianggap penting karena semakin awal mengetahui adanya potensi mahasiswa yang kemungkinan akan mengalami hambatan atau

kesulitan dalam belajarnya, maka akan semakin cepat langkah-langkah antisipatif yang akan dilakukan. Dampak dari kurangnya antisipasi atau pencegahan sejak dini terhadap mahasiswa yang berpotensi terhambat pada saat perkuliahan tampak dari hasil belajar berupa indeks prestasi yang rendah [2].

Naive Bayes adalah sebuah model datamining untuk 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. Keuntungan penggunaan *Naive Bayes* adalah bahwa metode ini hanya membutuhkan jumlah data pelatihan (*training data*) yang kecil untuk menentukan estimasi parameter yang diperlukan dalam proses pengklasifikasian. *Naive Bayes* sering bekerja jauh lebih baik dalam kebanyakan prediksi situasi dunia nyata yang kompleks dari pada yang diharapkan [3]. *Naive Bayes* telah banyak digunakan dalam prediksi berbagai hal. *Naive Bayes* telah digunakan untuk prediksi di bidang perbankan/keuangan [4][5], bidang sosial kemasyarakatan [6][7], serta bidang pendidikan [8][9].

Paper ini menyajikan model algoritma *naive bayes classifier* dalam melakukan prediksi ketepatan kelulusan mahasiswa. Adapun variabel yang penulis gunakan dalam melakukan perancangan ini sudah mewakili dari masing-masing faktor internal dan faktor eksternal, yaitu IPS 1 sampai IPS 4, jenis kelamin, asal sekolah, daerah asal, dan status pekerjaan. Alasan penulis melaksanakan studi kasus di STMIK Primakara karena STMIK Primakara merupakan kampus yang belum banyak mencetak lulusan karena masih tergolong kampus yang baru berdiri.

2. Tinjauan Pustaka

2.1 Data Mining

Data mining dapat diartikan sebagai suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam *database*. *Data mining* adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai *database* besar [10].

Data mining memiliki fungsi dalam beberapa area, yaitu [11]:

1. Deskripsi, yaitu cara untuk menggambarkan pola dan kecenderungan yang terdapat dalam data.
2. Estimasi, hampir sama dengan klasifikasi, kecuali variabel target estimasi lebih ke arah numerik daripada ke arah kategori. Model dibangun menggunakan *record* lengkap yang menyediakan nilai dari variabel prediksi.
3. Prediksi, hampir sama dengan klasifikasi dan estimasi, kecuali bahwa dalam prediksi nilai dari hasil akan ada dimasa mendatang.
4. Klasifikasi, berorientasi pada target variabel kategori. Sebagai contoh, penggolongan pendapatan dapat dipisahkan dalam tiga kategori, yaitu pendapatan tinggi, pendapatan sedang, pendapatan rendah.

Dalam konsep datamining terdapat dua pendekatan yang sering digunakan, yaitu *Supervised Learning* dan *Unsupervised Learning*. Pendekatan *supervised learning* adalah algoritma yang paling sering digunakan dalam dunia *data science* dibandingkan dengan *unsupervised learning*. Perbedaan kedua algoritma tersebut terletak pada bagaimana mereka belajar untuk membuat suatu prediksi maupun klasifikasi. Dalam *supervised learning*, algoritma tersebut seolah-olah dilatih terlebih dahulu agar dapat melakukan prediksi maupun klasifikasi [12].

2.2 Algoritma *Naive Bayes*

Deng, et al. [13] mengemukakan bahwa *Naive Bayes Classifier* dikenal sebagai pengklasifikasi bayesian sederhana dan telah menjadi model probabilistik yang penting dan telah berhasil dalam praktiknya. Walaupun memiliki asumsi independensi yang kuat, *Naive Bayes Classifier* telah terbukti efektif dalam klasifikasi dalam bentuk teks, diagnosa medis dan manajemen kinerja komputer. Teorema Bayes ditunjukkan pada persamaan 1 berikut [14]:

$$P(C|X) = \frac{P(X|C) \cdot P(C)}{P(X)} \quad (1)$$

Keterangan

- X : Data dengan kelas yang belum diketahui
 C : Hipotesis data X merupakan suatu kelas spesifik
 P(C | X) : Probabilitas hipotesis C berdasar kondisi X
 P(C) : Probabilitas hipotesis C (probabilitas prior)
 P(X | C) : Probabilitas X berdasarkan kondisi pada hipotesis C
 P(X) : Probabilitas X

2.3 Penelitian Sejenis

Beberapa penelitian mengenai Prediksi Kelulusan telah dilakukan. Rohman [15] menggunakan *algoritma K-Nearest Neighbor* (K-NN) untuk memprediksi Kelulusan Mahasiswa. Pada penelitian tersebut menguji 1582 siswa dengan jumlah kasus yang “tepat” waktu berjumlah 671 siswa sedangkan kasus yang “terlambat” berjumlah 911 siswa. Dan atribut yang dipakai dalam eksperimen yaitu; fakultas, jenis kelamin, umur, IPK sari semester 1 sampai 4. Kinerja Algoritma K-NN pada penelitian tersebut memiliki akurasi 82,25%.

Kusumawati, Winamo, & Arief [16] telah menguji algoritma *Neural Network* dalam memprediksi kelulusan mahasiswa. Variabel input yang digunakan dalam Prediksi kelulusan mahasiswa ini menggunakan yaitu 30 nilai mata kuliah. Output yang dihasilkan adalah kelulusan Tepat Waktu atau Melebihi Batas Waktu. Adapun tools yang digunakan yaitu MatlabR2010a. Pengujian dilakukan menggunakan perangkat lunak matlab yang diuji dengan beberapa bentuk arsitektur jaringan. Berdasarkan data uji arsitektur dengan konfigurasi terbaik untuk NN terdiri dari 30 lapisan masukan, 25 lapisan tersembunyi dan 1 lapisan keluaran dengan nilai learning rate 0.2, RMSE 0.070 dengan akurasi 78.26%. Nilai RMSE setelah dioptimasi yaitu 0.048 dengan tingkat akurasi 87.31%.

Paper ini menguji algoritma *Naive Bayes* yang merupakan sebuah metoda klasifikasi menggunakan metode probabilitas dan statistik yg dikemukakan oleh ilmuwan Inggris Thomas Bayes. Algoritma Naive Bayes memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya sehingga dikenal sebagai Teorema Bayes. Ciri utama dr Naive Bayes Classifier ini adalah asumsi yg sangat kuat (*naïf*) akan independensi dari masing-masing kondisi / kejadian. Keuntungan penggunaan adalah bahwa metode ini hanya membutuhkan jumlah data pelatihan (*training data*) yang kecil untuk menentukan estimasi parameter yg diperlukan dalam proses pengklasifikasian karena yg diasumsikan sebagai variabel independent, maka hanya varians dari suatu variabel dalam sebuah kelas yang dibutuhkan untuk menentukan klasifikasi, bukan keseluruhan dari matriks kovarians. Penerapan Naive Bayes menggunakan WEKA (*Waikato Environment of Knowledge Analysis*). WEKA adalah suatu perangkat lunak terintegrasi yang berisi implementasi dari metode-metode *data mining*. Dengan mengadopsi konsep *open source software*, menjadikan WEKA dapat digunakan dan dimodifikasi siapapun secara gratis [17].

3. Metodologi

Pada penelitian ini menggunakan 2 sumber data, yaitu data alumni STMIK Primakara dan data mahasiswa aktif STMIK Primakara. Adapun variabel yang diambil dari data tersebut adalah data IPS1-4, jenis kelamin, asal sekolah, daerah asal, dan status pekerjaan. Data tersebut akan dibagi menjadi 3, yaitu data *training*, data validasi, data *testing*. Jumlah data yang digunakan adalah data *training* sebanyak 23 orang dari data alumni STMIK Primakara, data validasi sebanyak 10 orang dari data alumni STMIK Primakara, data *testing* sebanyak 10 orang dari data mahasiswa aktif STMIK Primakara.

4. Hasil dan Pembahasan

4.1. Implementasi Metode

Penelitian ini akan membandingkan hasil prediksi mahasiswa menggunakan algoritma *naïve bayes classifier* dengan kondisi sesungguhnya yang terjadi di lapangan. Perbandingan ini dilakukan dengan cara menginputkan data IPS 1 sampai IPS 4, jenis kelamin, asal sekolah, daerah asal dan status pekerjaan dari objek penelitian ke dalam sistem yang telah dibuat dan hasilnya akan dibandingkan dengan kondisi sesungguhnya yang terjadi di lapangan. Tahap perbandingan akan menggunakan metode *confusion matrix*, yaitu perbandingan dengan menggunakan data hasil prediksi dan data yang sebenarnya.

Penghitungan dilakukan dengan beberapa tahap, dari mulai menentukan variabel sampai penarikan kesimpulan. Adapun tahap-tahap penghitungannya adalah sebagai berikut:

1. Pembobotan IPS

Kasus ini menggunakan pembobotan IPS sebagai berikut:

Jika $IPS\ X \geq 3.50$, maka nilai baik.

Jika $3.50 > X \geq 2.75$, maka nilai cukup.

Jika $IPS\ X < 2.75$, maka nilai kurang.

2. Menentukan data *training*

Data *training* digunakan sebagai data acuan dalam penghitungan ini. Data yang digunakan adalah sebanyak 23 data alumni STMIK Primakara.

Tabel 1. Sampel Data Training

IPS 1	IPS 2	IPS 3	IPS 4	Jenis Kelamin	Asal Sekolah	Daerah Asal	Status Pekerjaan	Keterangan
Baik	Cukup	Baik	Kurang	Laki-laki	Luar Kota Denpasar	Kota	Ya	Tepat Waktu
Cukup	Baik	Baik	Baik	Laki-laki	Luar Kota Denpasar	Luar Kota	Tidak	Tepat Waktu
Baik	Cukup	Baik	Baik	Laki-laki	Luar Kota Denpasar	Kota	Tidak	Tepat Waktu
Baik	Baik	Cukup	Baik	Laki-laki	Luar Kota Denpasar	Luar Kota	Ya	Tepat Waktu
Cukup	Kurang	Cukup	Cukup	Perempuan	Kota Denpasar	Kota	Ya	Terlambat
Cukup	Cukup	Cukup	Cukup	Perempuan	Kota Denpasar	Kota	Ya	Tepat Waktu
Cukup	Cukup	Cukup	Cukup	Perempuan	Kota Denpasar	Kota	Ya	Tepat Waktu
Cukup	Kurang	Cukup	Kurang	Perempuan	Luar Kota Denpasar	Kota	Tidak	Terlambat
Baik	Baik	Baik	Cukup	Laki-laki	Kota Denpasar	Kota	Ya	Tepat Waktu
Cukup	Kurang	Cukup	Kurang	Perempuan	Kota Denpasar	Kota	Ya	Tepat Waktu
Cukup	Cukup	Cukup	Cukup	Perempuan	Kota Denpasar	Kota	Ya	Tepat Waktu
Cukup	Baik	Cukup	Cukup	Perempuan	Kota Denpasar	Kota	Ya	Terlambat
Baik	Baik	Baik	Cukup	Perempuan	Kota Denpasar	Kota	Ya	Tepat Waktu
Kurang	Cukup	Kurang	Cukup	Laki-laki	Kota Denpasar	Kota	Ya	Terlambat
Baik	Cukup	Baik	Cukup	Perempuan	Luar Kota Denpasar	Luar Kota	Ya	Tepat Waktu
Kurang	Cukup	Cukup	Cukup	Laki-laki	Luar Kota Denpasar	Luar Kota	Tidak	Tepat Waktu
Cukup	Baik	Cukup	Cukup	Perempuan	Luar Kota Denpasar	Luar Kota	Ya	Terlambat
Baik	Baik	Baik	Baik	Perempuan	Luar Kota Denpasar	Luar Kota	Tidak	Tepat Waktu
Cukup	Cukup	Cukup	Baik	Laki-laki	Kota Denpasar	Kota	Ya	Tepat Waktu
Baik	Cukup	Kurang	Cukup	Laki-laki	Luar Kota Denpasar	Kota	Ya	Tepat Waktu
Kurang	Cukup	Cukup	Cukup	Laki-laki	Luar Kota Denpasar	Luar Kota	Tidak	Tepat Waktu
Baik	Baik	Cukup	Cukup	Perempuan	Luar Kota Denpasar	Luar Kota	Tidak	Tepat Waktu
Cukup	Cukup	Baik	Cukup	Perempuan	Kota Denpasar	Kota	Ya	Tepat Waktu

3. Menghitung jumlah kelas

Tabel 2. Penghitungan jumlah kelas

Label	Jumlah hipotesis	Jumlah Kasus	Hasil
P C "Tepat Waktu"	18	23	0.782608696
P C "Terlambat"	5	23	0.217391304

4. Menghitung jumlah kasus per kelas pada data *training*

a. Untuk variabel IPS 1

Tabel 3. Penghitungan variabel IPS

IPS 1	Tepat Waktu	Terlambat	P(Tepat Waktu)	P(Terlambat)
Baik	9	0	0.5	0
Cukup	7	4	0.388888889	0.8
Kurang	2	1	0.111111111	0.2
Total	18	5		

b. Untuk variabel IPS 2

Tabel 4. Penghitungan variabel IPS

IPS 2	Tepat Waktu	Terlambat	P(Tepat Waktu)	P(Terlambat)
Baik	6	2	0.33333333	0.4
Cukup	11	1	0.61111111	0.2
Kurang	1	2	0.05555556	0.4
Total	18	5		

c. Untuk variabel IPS 3

Tabel 5. Penghitungan variabel IPS 3

IPS 3	Tepat Waktu	Terlambat	P(Tepat Waktu)	P(Terlambat)
Baik	8	0	0.44444444	0
Cukup	9	4	0.5	0.8
Kurang	1	1	0.05555556	0.2
Total	18	5		

d. Untuk variabel IPS 4

Tabel 6. Penghitungan variabel IPS 4

IPS 4	Tepat Waktu	Terlambat	P(Tepat Waktu)	P(Terlambat)
Baik	5	0	0.27777778	0
Cukup	11	4	0.61111111	0.8
Kurang	2	1	0.11111111	0.2
Total	18	5		

e. Untuk variabel jenis kelamin

Tabel 7. Penghitungan variabel jenis kelamin

Jenis Kelamin	Tepat Waktu	Terlambat	P(Tepat Waktu)	P(Terlambat)
Laki-Laki	9	1	0.5	0.2
Perempuan	9	4	0.5	0.8
Total	18	5		

f. Untuk variabel asal sekolah

Tabel 8. Penghitungan variabel asal sekolah

Asal Sekolah	Tepat Waktu	Terlambat	P(Tepat Waktu)	P(Terlambat)
Kota Denpasar	8	4	0.44444444	0.66666667
Luar Kota Denpasar	10	2	0.55555556	0.33333333
Total	18	6		

g. Untuk variabel daerah asal

Tabel 9. Penghitungan variabel daerah asal

Daerah Asal	Tepat Waktu	Terlambat	P(Tepat Waktu)	P(Terlambat)
Kota	11	4	0.61111111	0.8
Luar Kota	7	1	0.38888889	0.2
Total	18	5		

h. Untuk variabel status pekerjaan

Tabel 10. Penghitungan variabel status pekerjaan

Status Pekerjaan	Tepat Waktu	Terlambat	P(Tepat Waktu)	P(Terlambat)
Ya	12	4	0.66666667	0.8
Tidak	6	1	0.33333333	0.2
Total	18	5		

5. Menentukan data validasi

Data validasi digunakan untuk mengecek validitas dari penghitungan yang berdasarkan data *training*. Data yang digunakan adalah sebanyak 10 data alumni STMIK Primakara.

Tabel 11. Data validasi

IPS 1	IPS 2	IPS 3	IPS 4	Jenis Kelamin	Asal Sekolah	Daerah Asal	Status Pekerjaan	Keterangan
Cukup	Kurang	Baik	Baik	Laki-Laki	Kota Denpasar	Luar Kota	Tidak	Tepat Waktu
Cukup	Cukup	Cukup	Cukup	Perempuan	Kota Denpasar	Kota	Tidak	Tepat Waktu
Cukup	Baik	Cukup	Baik	Perempuan	Luar Kota Denpasar	Luar Kota	Ya	Tepat Waktu
Baik	Cukup	Cukup	Baik	Laki-laki	Kota Denpasar	Kota	Ya	Tepat Waktu
Baik	Baik	Kurang	Baik	Perempuan	Luar Kota Denpasar	Kota	Tidak	Tepat Waktu
Cukup	Kurang	Kurang	Cukup	Laki-Laki	Luar Kota Denpasar	Kota	Ya	Terlambat
Kurang	Kurang	Baik	Baik	Perempuan	Kota Denpasar	Luar Kota	Tidak	Terlambat
Kurang	Baik	Kurang	Cukup	Perempuan	Luar Kota Denpasar	Luar Kota	Ya	Terlambat
Kurang	Kurang	Cukup	Kurang	Perempuan	Luar Kota Denpasar	Luar Kota	Tidak	Terlambat
Cukup	Baik	Kurang	Kurang	Laki-Laki	Luar Kota Denpasar	Kota	Ya	Terlambat

6. Menghitung jumlah kasus perkelas pada data validasi terhadap data *training* dan mengalikan semua variabel kelas

Pada proses ini, seluruh data validasi akan dijabarkan menurut variabel dan kelasnya dengan cara menyesuaikan data validasi yang ada dengan variabel data *training* yang telah dihitung sebelumnya.

Tabel 12. Penghitungan data validasi

6	Cukup	Kurang	Kurang	Cukup	Laki-Laki	Luar Kota Denpasar	Kota	Ya	Hasil
Tepat Waktu	0.38888889	0.05555556	0.05555556	0.61111111	0.5	0.55555556	0.61111111	0.66666667	8.30094E-05
Terlambat	0.8	0.4	0.2	0.8	0.2	0.33333333	0.8	0.8	0.002184533
7	Kurang	Kurang	Baik	Baik	Perempuan	Kota Denpasar	Luar Kota	Tidak	Hasil
Tepat Waktu	0.11111111	0.05555556	0.44444444	0.27777778	0.5	0.44444444	0.38888889	0.33333333	2.19529E-05
Terlambat	0.2	0.4	0	0	0.8	0.66666667	0.2	0.2	0
8	Kurang	Baik	Kurang	Cukup	Perempuan	Luar Kota Denpasar	Luar Kota	Ya	Hasil
Tepat Waktu	0.11111111	0.33333333	0.05555556	0.61111111	0.5	0.55555556	0.38888889	0.66666667	9.0557E-05
Terlambat	0.2	0.4	0.2	0.8	0.8	0.33333333	0.2	0.8	0.000546133
9	Kurang	Kurang	Cukup	Kurang	Perempuan	Luar Kota Denpasar	Luar Kota	Tidak	Hasil
Tepat Waktu	0.11111111	0.05555556	0.5	0.11111111	0.5	0.55555556	0.38888889	0.33333333	1.23485E-05
Terlambat	0.2	0.4	0.8	0.2	0.8	0.33333333	0.2	0.2	0.000136533
10	Cukup	Baik	Kurang	Kurang	Laki-Laki	Luar Kota Denpasar	Kota	Ya	Hasil
Tepat Waktu	0.38888889	0.33333333	0.05555556	0.11111111	0.5	0.55555556	0.61111111	0.66666667	9.0557E-05
Terlambat	0.8	0.4	0.2	0.2	0.2	0.33333333	0.8	0.8	0.000546133
1	Cukup	Kurang	Baik	Baik	Laki-Laki	Kota Denpasar	Luar Kota	Tidak	Hasil
Tepat Waktu	0.38888889	0.05555556	0.44444444	0.27777778	0.5	0.44444444	0.38888889	0.33333333	7.68351E-05
Terlambat	0.8	0.4	0	0	0.2	0.66666667	0.2	0.2	0
2	Cukup	Cukup	Cukup	Cukup	Perempuan	Kota Denpasar	Kota	Tidak	Hasil
Tepat Waktu	0.38888889	0.61111111	0.5	0.61111111	0.5	0.44444444	0.61111111	0.33333333	0.003287171
Terlambat	0.8	0.2	0.8	0.8	0.8	0.66666667	0.8	0.2	0.008738133
3	Cukup	Baik	Cukup	Baik	Perempuan	Luar Kota Denpasar	Luar Kota	Ya	Hasil
Tepat Waktu	0.38888889	0.33333333	0.5	0.27777778	0.5	0.55555556	0.38888889	0.66666667	0.001296593
Terlambat	0.8	0.4	0.8	0	0.8	0.33333333	0.2	0.8	0
4	Baik	Cukup	Cukup	Baik	Laki-laki	Kota Denpasar	Kota	Ya	Hasil
Tepat Waktu	0.5	0.61111111	0.5	0.27777778	0.5	0.44444444	0.61111111	0.66666667	0.003842148
Terlambat	0	0.2	0.8	0	0.2	0.66666667	0.8	0.8	0
5	Baik	Baik	Kurang	Baik	Perempuan	Luar Kota Denpasar	Kota	Tidak	Hasil
Tepat Waktu	0.5	0.33333333	0.05555556	0.27777778	0.5	0.55555556	0.61111111	0.33333333	0.000145536
Terlambat	0	0.4	0.2	0	0.8	0.33333333	0.8	0.2	0

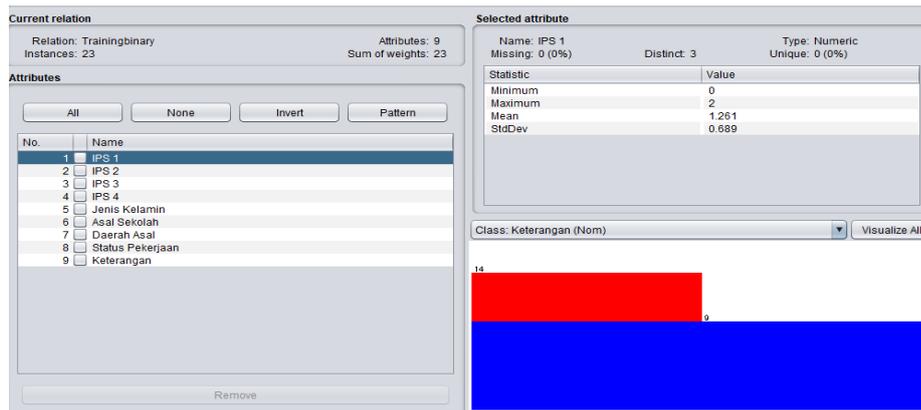
7. Membandingkan hasil perkelas

Berdasarkan data *training* yang ada, didapatkan bahwa 2(dua) data yang tidak sesuai dengan kondisi yang sesungguhnya.

Tabel 13. Perbandingan data

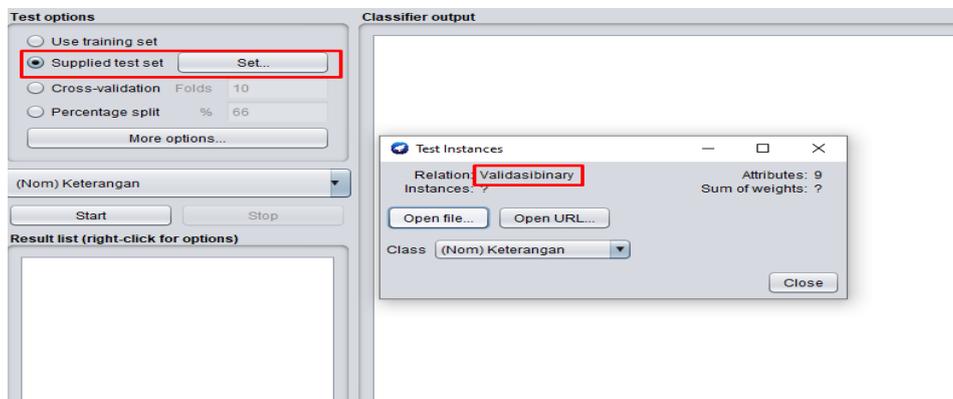
1	VALID	
	Hasil Sesungguhnya	Tepat Waktu
	Hasil Prediksi	Tepat Waktu
2	INVALID	
	Hasil Sesungguhnya	Tepat Waktu
	Hasil Prediksi	Terlambat
3	VALID	
	Hasil Sesungguhnya	Tepat Waktu
	Hasil Prediksi	Tepat Waktu
4	VALID	
	Hasil Sesungguhnya	Tepat Waktu
	Hasil Prediksi	Tepat Waktu
5	VALID	
	Hasil Sesungguhnya	Tepat Waktu
	Hasil Prediksi	Tepat Waktu
6	VALID	
	Hasil Sesungguhnya	Terlambat
	Hasil Prediksi	Terlambat
7	INVALID	
	Hasil Sesungguhnya	Terlambat
	Hasil Prediksi	Tepat Waktu
8	VALID	
	Hasil Sesungguhnya	Terlambat
	Hasil Prediksi	Terlambat
9	VALID	
	Hasil Sesungguhnya	Terlambat
	Hasil Prediksi	Terlambat
10	VALID	
	Hasil Sesungguhnya	Terlambat
	Hasil Prediksi	Terlambat

- Penghitungan menggunakan WEKA
Melakukan penghitungan dengan menggunakan WEKA dimulai dengan mengimport data *training* terlebih dahulu.



Gambar 1. Data *training* pada WEKA

Lalu dilanjutkan dengan mengimport data validasi kedalam aplikasi WEKA.



Gambar 2. Data validasi pada WEKA

```

Classifier output
=== Evaluation on test set ===

Time taken to test model on supplied test set: 0 seconds

=== Summary ===

Correctly Classified Instances      8          80   %
Incorrectly Classified Instances    2          20   %
Kappa statistic                    0.6
Mean absolute error                 0.2827
Root mean squared error             0.4058
Relative absolute error             56.5494 %
Root relative squared error        71.9986 %
Total Number of Instances          10

=== Detailed Accuracy By Class ===

                TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Cls
                0.800   0.200   0.800     0.800   0.800     0.600   0.920   0.927   y
                0.800   0.200   0.800     0.800   0.800     0.600   0.920   0.943   t
Weighted Avg.   0.800   0.200   0.800     0.800   0.800     0.600   0.920   0.935

=== Confusion Matrix ===

a b  <-- classified as
4 1 | a = y
1 4 | b = t

```

Gambar 3. Hasil Penghitungan Algoritma

Setelah data dihitung, selanjutnya akan dilakukan pengujian data. Tahap pengujian dilakukan untuk mengecek kembali tentang keakuratan penghitungan yang dilakukan oleh program yang telah digunakan. Pengujian program dilakukan dengan cara melakukan perbandingan hasil akhir prediksi program dengan data yang sesungguhnya terjadi di lapangan. Adapun metode yang digunakan untuk melakukan pengujian adalah metode *confusion matrix*. Berikut adalah hal yang perlu dilakukan untuk melakukan pengujian dengan metode *confusion matrix*:

Tabel 14. *Confusion matrix*

	Tepat Waktu	Terlambat
Tepat Waktu	4	1
Terlambat	1	4

a. *Recall*

Recall adalah data penghapusan yang berhasil diambil dari data yang relevan dengan *query*

$$Recall = \left(\frac{4}{1 + 4} \right) * 100\% = 80\%$$

b. *Accuracy*

Accuracy(akurasi) adalah presentase dari total data yang diidentifikasi dan dinilai.

$$Accuracy = \left(\frac{4 + 4}{4 + 1 + 1 + 4} \right) * 100\% = 80\%$$

c. *Precision*

Precision (Presisi) adalah data yang diambil berdasarkan informasi yang kurang.

$$Precision = \left(\frac{4}{4 + 1} \right) * 100\% = 80\%$$

Berdasarkan hasil pengujian diatas dapat disimpulkan bahwa hasil keakuratan, kerelevanan dan presisi data terhadap pengujian sebesar 80%

4.2 Hasil Implementasi

Setelah didapatkan hasil dari pengujian menggunakan *confusion matrix*, mengimplementasikan penghitungan, penguji mencoba memakai 10 (sepuluh) sampel untuk mencari tahu hasil dari kesimpulan data yang akan diuji. Adapun data yang akan diuji adalah 10(sepuluh) data yang didapat dari mahasiswa aktif STMIK Primakara yang belum diketahui ketepatan waktu kelulusannya.

Tabel 15. Data *testing*

Data Testing							
IPS 1	IPS 2	IPS 3	IPS 4	Jenis Kelamin	Asal Sekolah	Daerah Asal	Status Pekerjaan
Kurang	Cukup	Baik	Baik	Perempuan	Luar Kota Denpasar	Luar Kota	Tidak
Baik	Cukup	Cukup	Kurang	Laki-laki	Luar Kota Denpasar	Kota	Tidak
Kurang	Baik	Kurang	Cukup	Perempuan	Kota Denpasar	Luar Kota	Ya
Baik	Baik	Baik	Kurang	Perempuan	Kota Denpasar	Luar Kota	Ya
Kurang	Cukup	Cukup	Cukup	Laki-laki	Luar Kota Denpasar	Luar Kota	Tidak
Kurang	Kurang	Kurang	Cukup	Perempuan	Luar Kota Denpasar	Kota	Ya
Kurang	Baik	Kurang	Kurang	Perempuan	Kota Denpasar	Luar Kota	Ya
Cukup	Baik	Baik	Kurang	Laki-laki	Luar Kota Denpasar	Luar Kota	Tidak
Cukup	Cukup	Cukup	Cukup	Laki-laki	Luar Kota Denpasar	Kota	Tidak
Baik	Baik	Kurang	Cukup	Perempuan	Kota Denpasar	Kota	Ya

Tabel 16. Perhitungan data *testing*

1	Kurang	Cukup	Baik	Baik	Perempuan	Luar Kota Denpasar	Luar Kota	Tidak	Hasil
Tepat Waktu	0.11111111	0.61111111	0.44444444	0.27777778	0.5	0.55555556	0.38888889	0.33333333	0.000301852
Terlambat	0.2	0.2	0	0	0.8	0.33333333	0.2	0.2	0
2	Baik	Cukup	Cukup	Kurang	Laki-laki	Luar Kota Denpasar	Kota	Tidak	Hasil
Tepat Waktu	0.5	0.61111111	0.5	0.11111111	0.5	0.55555556	0.61111111	0.33333333	0.000960537
Terlambat	0	0.2	0.8	0.2	0.2	0.33333333	0.8	0.2	0
3	Kurang	Baik	Kurang	Cukup	Perempuan	Kota Denpasar	Luar Kota	Ya	Hasil
Tepat Waktu	0.05555556	0.33333333	0.05555556	0.61111111	0.5	0.44444444	0.38888889	0.66666667	3.62223E-05
Terlambat	0.2	0.4	0.2	0.8	0.8	0.66666667	0.2	0.8	0.001092267
4	Baik	Baik	Baik	Kurang	Perempuan	Kota Denpasar	Luar Kota	Ya	Hasil
Tepat Waktu	0.5	0.33333333	0.44444444	0.11111111	0.5	0.44444444	0.38888889	0.66666667	0.000474182
Terlambat	0	0.4	0	0.2	0.8	0.66666667	0.2	0.8	0
5	Kurang	Cukup	Cukup	Cukup	Laki-laki	Luar Kota Denpasar	Luar Kota	Tidak	Hasil
Tepat Waktu	0.05555556	0.61111111	0.5	0.61111111	0.5	0.55555556	0.38888889	0.33333333	0.000373542
Terlambat	0.2	0.2	0.8	0.8	0.2	0.33333333	0.2	0.2	6.82667E-05
6	Kurang	Kurang	Kurang	Cukup	Perempuan	Luar Kota Denpasar	Kota	Ya	Hasil
Tepat Waktu	0.05555556	0.05555556	0.05555556	0.61111111	0.5	0.55555556	0.61111111	0.66666667	1.18585E-05
Terlambat	0.2	0.4	0.2	0.8	0.8	0.33333333	0.8	0.8	0.002184533
7	Kurang	Baik	Kurang	Kurang	Perempuan	Kota Denpasar	Luar Kota	Ya	Hasil
Tepat Waktu	0.05555556	0.33333333	0.05555556	0.11111111	0.5	0.44444444	0.38888889	0.66666667	6.58587E-06
Terlambat	0.2	0.4	0.2	0.2	0.8	0.66666667	0.2	0.8	0.000273067
8	Cukup	Baik	Baik	Kurang	Laki-laki	Luar Kota Denpasar	Luar Kota	Tidak	Hasil
Tepat Waktu	0.38888889	0.33333333	0.44444444	0.11111111	0.5	0.55555556	0.38888889	0.33333333	0.000230505
Terlambat	0.8	0.4	0	0.2	0.2	0.33333333	0.2	0.2	0
9	Cukup	Cukup	Cukup	Cukup	Laki-laki	Luar Kota Denpasar	Kota	Tidak	Hasil
Tepat Waktu	0.38888889	0.61111111	0.5	0.61111111	0.5	0.55555556	0.61111111	0.33333333	0.004108964
Terlambat	0.8	0.2	0.8	0.8	0.2	0.33333333	0.8	0.2	0.001092267
10	Baik	Baik	Kurang	Cukup	Perempuan	Kota Denpasar	Kota	Ya	Hasil
Tepat Waktu	0.5	0.33333333	0.05555556	0.61111111	0.5	0.44444444	0.61111111	0.66666667	0.000512286
Terlambat	0	0.4	0.2	0.8	0.2	0.66666667	0.8	0.8	0

Berdasarkan pengujian dari data *training* yang telah ditentukan sebelumnya dengan keakuratan 80% yang telah di uji terhadap data validasi dengan metode *confusion matrix*, didapatkan hasil bahwa data testing diprediksi memiliki keterangan sebagai berikut.

1. Kesimpulan kelulusan untuk mahasiswa 1 adalah "Tepat Waktu".
2. Kesimpulan kelulusan untuk mahasiswa 2 adalah "Tepat Waktu".
3. Kesimpulan kelulusan untuk mahasiswa 3 adalah "Terlambat".

4. Kesimpulan kelulusan untuk mahasiswa 4 adalah “Tepat Waktu”.
5. Kesimpulan kelulusan untuk mahasiswa 5 adalah “Tepat Waktu”.
6. Kesimpulan kelulusan untuk mahasiswa 6 adalah “Terlambat”.
7. Kesimpulan kelulusan untuk mahasiswa 7 adalah “Terlambat”.
8. Kesimpulan kelulusan untuk mahasiswa 8 adalah “Tepat Waktu”.
9. Kesimpulan kelulusan untuk mahasiswa 9 adalah “Terlambat”.
10. Kesimpulan kelulusan untuk mahasiswa 10 adalah “Tepat Waktu”.

5. Kesimpulan

Penarikan kesimpulan prediksi terhadap IPS 1 sampai IPS 4, jenis kelamin, asal sekolah, daerah asal dan status pekerjaan dapat dilihat dari kedua akumulasi probabilitas baik yang tepat waktu maupun yang terlambat dengan cara melihat nilai akumulasi probabilitas yang mendekati 1. Penghitungan manual melalui microsoft excel dan WEKA juga memiliki keselarasan dengan mendeteksi ada 2(dua) data yang invalid (kesalahan prediksi) dan 8(delapan) data valid (prediksi tepat). Pengujian menggunakan *confusion matrix* mendapatkan hasil dengan *recall* sebesar 80%, *accuracy* sebesar 80% dan *precision* sebesar 80%.

DAFTAR REFERENSI

- [1] STMIK Primakara, Buku Pedoman Akademik. 2014.
- [2] Sembiring, M. A., Sibuea, M. F. L., & Sapta, A. Analisa Kinerja Algoritma C4.5 Dalam Memprediksi Hasil Belajar, *Jssr*. 2018; 43(07): 73–79.
- [3] Saleh, A. Implementasi metode klasifikasi naive bayes dalam memprediksi besarnya penggunaan listrik rumah tangga. *Creative Information Technology Journal*. 2015; 2(3): 207-217.
- [4] Masripah, S. Komparasi Algoritma Klasifikasi Data Mining untuk Evaluasi Pemberian Kredit. *Bina Insani ICT Journal*. 2016; 3(1): 187-193.
- [5] Hasan, M. Prediksi Tingkat Kelancaran Pembayaran Kredit Bank Menggunakan Algoritma Naive Bayes Berbasis Forward Selection. *ILKOM Jurnal Ilmiah*. 2017; 9(3): 317-324.
- [6] Wati, M., & Hadi, A. Implementasi Algoritma Naive Bayesian Dalam Penentuan Penerima Program Bantuan Pemerintah. *JTRISTE*. 2016; 3(1): 22-26.
- [7] Fadlan, C., Ningsih, S., & Windarto, A. P. Penerapan Metode Naive Bayes Dalam Klasifikasi Kelayakan Keluarga Penerima Beras Rastra. *JUTIM (Jurnal Teknik Informatika Musirawas)*. 2018; 3(1): 1-8.
- [8] Saputra, R. A., & Ayuningtias, S. Penerapan Algoritma Naive Bayes Untuk Penentuan Calon Penerima Beasiswa Pada Smk Pasim Plus Sukabumi. *Swabumi*. 2016; 4(2): 114-120.
- [9] Bahar, B., & Pratama, D. W. Penerapan Teorema Bayes Dalam Sistem Pakar Untuk Konsultasi Siswa Bermasalah. *Jutisi: Jurnal Ilmiah Teknik Informatika dan Sistem Informasi*. 2018; 6(2): 1529-1540.
- [10] Nofriansyah, D. *Konsep Data Mining VS Sistem Pendukung Keputusan*. 2014.
- [11] Larose, D.T. *Discovering Knowledge in Data: An Introduction to Data Mining*. Canada: John Wiley & Sons, Inc, 2014.
- [12] Jiawei Han, M.K. *Data Mining Concepts and Techniques*, II. San Francisco: Elsevier, Inc.
- [13] Deng, H. Sun, Y., Chang and Han, J. Probabilistic Models for Classification, in *Data Classification: Algorithms and Applications*, 1 ed., Minnesota, Chapman and Hall/CRC, 2014
- [14] Dahri, D., Agus, F., & Khairina, D. M. Metode Naive Bayes Untuk Penentuan Penerima Beasiswa Bidikmisi Universitas Mulawarman. *Jurnal Informatika Mulawarman*. 2016; 11(2): 30-36.
- [15] Rohman, A. Model Algoritma K-Nearest Neighbor (K-Nn) Untuk Prediksi Kelulusan Mahasiswa. *Neo Teknika*. 2015; 1(1): 1-9
- [16] Kusumawati, D., Winarno, W. W., & Arief, M. R. Prediksi Kelulusan Mahasiswa Menggunakan Metode Neural Network dan Particle Swarm Optimization. *SEMNASTEKNOMEDIA ONLINE*. 2015; 3(1): 3-8.
- [17] Adinugroho, S., Sari, Y.A. *Implementasi Data Mining Menggunakan WEKA*, I. Malang: UB Press, 2018.