

PREDIKSI TINGKAT KELULUSAN TEPAT WAKTU DENGAN METODE NAÏVE BAYES DAN K-NEAREST NEIGHBOR

Sidik Rahmatullah¹, Ema Utami²

¹) Teknik Informatika ²) Magister Teknik Informatika

¹) STMIK Dian Cipta Cendikia Kotabumi, ²) Universitas AMIKOM Yogyakarta

Jl. Negara No.03 Candimas Kotabumi Lampung Utara

E-mail : sidik@dcc.ac.id, emma@nrar.net

ABSTRAK

Lulusan adalah status yang dicapai mahasiswa setelah menyelesaikan proses pendidikan sesuai dengan persyaratan kelulusan yang ditetapkan oleh program studi. Sebagai salah satu keluaran langsung dari proses pendidikan yang dilakukan oleh program studi, lulusan yang bermutu memiliki ciri penguasaan kompetensi akademik termasuk *hard skills* dan *soft skills* sebagaimana dinyatakan dalam sasaran mutu serta dibuktikan dengan kinerja lulusan di masyarakat sesuai dengan profesi dan bidang ilmu. Program studi yang bermutu memiliki sistem pengelolaan lulusan yang baik sehingga mampu menjadikannya sebagai *human capital* bagi program studi yang bersangkutan.

Penelitian ini menggunakan metode data mining yang digunakan untuk memprediksi tingkat kelulusan mahasiswa menggunakan dua metode yaitu *Naive Bayes* dan *K-Nearest Neighbor*. Hasil dari penelitian ini dapat memprediksi mahasiswa tepat lulus atau terlambat. Uji coba dilakukan dengan menggunakan data lulusan mahasiswa S1 Sistem informasi STMIK Dian Cipta Cendikia Kotabumi sebanyak 600 data untuk training dan 180 data untuk testing. Hasil uji coba menunjukkan bahwa dengan menggunakan *Naive Bayes* menghasilkan akurasi sebesar 85%, sedangkan menggunakan algoritma *K-nearest neighbor* menghasilkan akurasi sebesar 68.89 %.

Kata kunci: *Naive Bayes*, *K-Nearest Neighbor*, Prediksi Ketepatan Waktu Lulus

Abstracts

Graduates is status reached students after completing the process of education in accordance with the requirements of graduation set by the study program. As one of the output is to directly from the process of education that have been undertaken by by the study program , a graduate of a who are fully dedicated rewarded with lofty mansions in academic pt batubara bukit kendi total competence including a hard skills the and soft skills the as claimed in the target group quality and shall be proven with the performance of a graduate of a in the midst of society in accordance with for a profession and the field of science. Course of study quality having management systems good graduates order to be able to make them a human capital for successor program study concerned.

This research uses the data mining used to measure graduation rates students using two method is naive bayes and k-nearest neighbor. The result of this research to predict students just pass or late. The trial was done using data from information system undergraduates stmik dian cipta cendikia kotabumi about 600 data for training and 180 data for testing The results of the tryouts shows that by the use of naive bayes produce of accuracy of the numbers bayes as much as 85 % , while using an algorithm k-nearest neighbor produce of accuracy of the numbers as much as 68.89 %.

Keywords: naive bayes, k-nearest neighbor, timeliness of students graduation prediction

1. PENDAHULUAN

Dalam dunia pendidikan khususnya perguruan tinggi tidak lepas dengan adanya data yang menumpuk baik itu data mahasiswa dan juga data alumni, setiap tahun perguruan tinggi mengadakan penerimaan mahasiswa baru sehingga jumlah mahasiswa diperguruan tinggi akan semakin bertambah banyak. Sehingga perlu diperhatikan antara rasio mahasiswa yang ada dengan rasio dosen tetap pada program studi sesuai dengan surat edaran dari Kementerian Riset Teknologi dan Pendidikan Tinggi[1]

Berdasarkan [2] Lulusan adalah status yang dicapai mahasiswa setelah menyelesaikan proses pendidikan sesuai dengan persyaratan kelulusan yang ditetapkan oleh program studi. Sebagai salah satu keluaran langsung dari proses pendidikan yang dilakukan oleh program studi, lulusan yang bermutu memiliki ciri penguasaan kompetensi akademik termasuk *hard skills* dan *soft skills* sebagaimana dinyatakan dalam sasaran mutu serta dibuktikan dengan kinerja lulusan di masyarakat sesuai dengan profesi dan bidang ilmu. Program studi yang bermutu memiliki sistem pengelolaan lulusan yang baik sehingga mampu menjadikannya sebagai *human capital* bagi program studi yang bersangkutan.

Perguruan tinggi seharusnya memiliki dan menerapkan kebijakan untuk mengelola dan mengevaluasi standar tersebut yaitu kelulusan mahasiswa yang mencerminkan kinerja perguruan tinggi dalam peningkatan mutu. Salah satu upaya yang dilakukan untuk mengelola kelulusan mahasiswa adalah tercapainya masa studi mahasiswa sesuai dengan lama studi yang telah terjadwal. Namun beberapa perguruan tinggi mengalami kesulitan bagaimana mahasiswa dapat mencapai masa studi tepat waktu sesuai dengan jadwal yang telah ditetapkan perguruan tinggi[3]

Berdasarkan hasil penelitian dari [4] *Naïve Bayes* merupakan algoritma yang dapat digunakan untuk prediksi masa studi mahasiswa yang bisa dikategorikan tepat waktu, terlambat dan tidak lulus. Dengan menggunakan data mining khususnya klasifikasi untuk prediksi dengan algoritma *naïve bayes* dapat dilakukan prediksi terhadap ketepatan waktu studi dari mahasiswa berdasarkan data *training* yang ada. Pengujian yang dipakai yaitu *k-fold cross validation 10-fold*. Hasil pengujian didapat nilai akurasi rata-rata sebesar 85.17 % sedangkan nilai akurasi tertinggi sebesar 88.96 %.

Pada Penelitian [5] salah satu aspek pengukuran kualitas dalam evaluasi keberhasilan penyelenggaraan pendidikan tinggi adalah ketepatan lulus mahasiswa. Jumlah prosentase mahasiswa yang lulus tepat waktu menjadi indikator keberhasilan pelaksanaan proses belajar mengajar di

suatu program studi. Penelitian ini menawarkan penggunaan metode data mining untuk memprediksi waktu lulus mahasiswa menggunakan dua metode yaitu *k-nearest neighbor* dan *naïve bayes classifier*. Hasil penelitian ini berupa system yang dapat memprediksi ketepatan waktu lulus. Uji coba dilakukan dengan menggunakan data lulusan mahasiswa D3 Sistem informasi Universitas Airlangga. Hasil uji coba menunjukkan bahwa metode *k-Nearest neighbor* menghasilkan akurasi lebih tinggi dibandingkan dengan *naïve bayes classifier*. Akurasi tertinggi diperoleh dengan menggunakan metode *k-nearest neighbor* yaitu sebesar 98.7% oleh karena itu dapat disimpulkan bahwa system yang dibangun pada penelitian ini mampu memprediksi ketepatan waktu lulus dengan akurasi cukup tinggi.

Selanjutnya penelitian [6] menjelaskan Sistem Prediksi merupakan salah satu metode yang tepat untuk membentuk pola-pola yang mungkin dapat memberikan indikasi yang bermanfaat bagi pengelola perguruan tinggi sejak dini, sehingga pengelola dapat melakukan langkah-langkah positif dalam meminimalisasi masalah yang ada. Pada sistem prediksi ini dapat digunakan algoritma *Nearest-Neighbor classifier* dan *Naïve Bayes classifier* dalam mengukur tingkat kelulusan kemungkinan mahasiswa lulus tepat waktu, tidak tepat waktu atau *drop-out*.

Dari uraian masalah diatas, maka penulis akan memprediksi tingkat kelulusan mahasiswa yang akan dapat membantu pihak akademik dalam mengelola dan mengevaluasi mutu dan kualitas yang berkaitan dengan kelulusan mahasiswa di lingkungan STMIK Dian Cipta Cendikia Kotabumi. Apabila masa studi dapat diprediksi dan hasil dari prediksi tersebut tidak sesuai atau melebihi masa studi yang seharusnya, maka pihak akademik dapat melakukan pencegahan dan membantu mahasiswa untuk dapat mengatasi kendala dalam mencapai kelulusan tepat waktu.

2.METODE PENELITIAN

2.1 Metode Pengumpulan Data

Dalam metode pengumpulan data ini, digunakan metode deskriptif yang bertujuan mengumpulkan data, menyajikan data dan membahas analisis data yang diperoleh.

a. Observasi

Melakukan pengamatan secara langsung terhadap kondisi yang ada di STMIK Dian Cipta Cendikia Kotabumi

b. Wawancara

Melakukan wawancara secara langsung terhadap pihak terkait yang ada di Perguruan Tinggi STMIK Dian Cipta Cendikia Kotabumi.

c. Studi Pustaka

Dengan membaca buku-buku, literatur-literatur, melakukan pencarian melalui berbagai media terhadap apa saja yang berhubungan data mining maupun kelulusan mahasiswa sehingga lebih menambah referensi yang dibutuhkan.

2.2 Metode Analisis Data

Tahap analisis data merupakan tahap preprocessing data terkait data yang akan diuji atau dievaluasi menggunakan machine learning. Tahap ini terdiri dari, pemilihan atribut atau fitur yang mendukung penelitian, data cleaning, dan menghasilkan data set berupa data training dan data testing yang nantinya akan dievaluasi.

Dalam penelitian ini akan dilakukan analisis data Preprocessing data yang dilakukan pada penelitian ini meliputi :

1. Penentuan data set

Pengelolaan Dataset yang didapatkan dari data bagian akademik data tersebut dalam bentuk Microsoft excel yang nantinya dapat diproses menggunakan rapidminer.

2. Penentuan Atribut

Dalam penentuan atribut dilakukan untuk menentukan perlu atau tidaknya atribut digunakan dalam proses klasifikasi.

3. Pembagian Dataset

Dataset keseluruhan dibagi menjadi dua bagian yakni data training dan data testing. Data training berisi 600 data sedangkan data testing berisi 180 data.

4. Pemodelan

Pembuatan model dari pelatihan data latih pada penelitian yakni menggunakan *k-nearest neighbor* dan *naïve bayes*. Masing-masing algoritma mempunyai tahapan algoritma klasifikasi masing-masing. Pembuatan model dibuat dengan mengikuti algoritma masing-masing. Aplikasi yang digunakan untuk melakukan pemodelan yaitu Rapidminer.

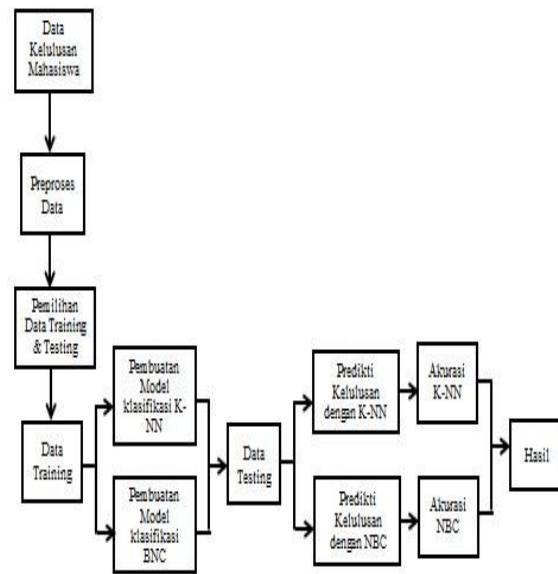
5. Evaluasi dan Hasil

Di dalam tahap evaluasi dan perbandingan dilakukan pengujian data uji sesuai masing-masing algoritma klasifikasi yaitu *k-nearest neighbor* dan *naïve bayes*. Untuk pengujian data uji menggunakan algoritma knn maka dilakukan penentuan nilai k berdasarkan percobaan nilai k pada angka tertentu. Setelah itu, menghitung jarak terdekat antara data uji dengan data latih sesuai jumlah k yang telah didapatkan sebelumnya. Kemudian, mengurutkan jarak yang paling minimum sehingga data uji bisa dikategorikan ke dalam kelas tertentu.

2.3 Alur Penelitian

alur penelitian pertama adalah diambil data kelulusan mahasiswa prodi Sistem Informasi berikut atribut yang akan digunakan. Kemudian data di masukkan dan dilakukan praproses terlebih dahulu

agar data tersebut dapat digunakan ke dalam metode klasifikasi yaitu *K-Nearest Neighbor* dan *Naïve Bayes Classifier*. Setelah itu dilakukan pemilihan data training dan data testing secara acak. Data training merupakan data yang digunakan untuk membuat model pengklasifikasi baik pada metode *k-Nearest Neighbor* dan *Naïve Bayes Classifier*. Model pengklasifikasi yang dihasilkan oleh data training selanjutnya digunakan untuk memprediksi kelas dari data testing, sehingga pada data testing label kelas tidak digunakan. Hasil prediksi berupa label kelas akan dibandingkan dengan label kelas sebenarnya untuk dihitung akurasi metode baik *k-nearest neighbor* maupun *Naïve bayes classifier*.



Gambar 1. Alur Penelitian

3.HASIL DAN PEMBAHASAN

3.1 Dataset

Sebelum melakukan proses perhitungan algoritma, terlebih dahulu dilakukan pengumpulan data. Masalah yang akan diolah datanya dalam pembahasan ini adalah memprediksi tingkat kelulusan mahasiswa prodi Sistem Informasi di STMIK Dian Cipta Cendikia Kotabumi. Data lulusan yang digunakan adalah 600 lulusan yaitu angkatan 2012 – 2013. Atribut-atribut data yang digunakan adalah NPM, Jenis Kelamin, IPS 1-5, Konsentrasi, asal tinggal, jumlah sks, status pekerjaan, status kelulusan.

Adapun data Training yang digunakan adalah sebanyak 600 data kelulusan pada program studi S1 Sistem Informasi angkatan 2012 – 2013.

Tabel 1. Cuplikan Data Training Kelulusan

No	Npm	Jenis_Kelamin	IPK_S1	IPK_S2	IPK_S3	IPK_S4	IPK_S5	Jumlah_SKS	Konsentrasi	Status_Kerja	Asal_Tinggal	Status_Lulus
1	13071011	Laki	3.11	3.6	3.4	1.8	1	10	EIS	Bekerja	Luar	Tepat Waktu
2	13071015	Perempuan	3.21	3.26	3.42	3	3.36	18	EIS	Bekerja	Luar	Tepat Waktu
3	13071020	Laki	3.14	3.52	3.17	3.25	2.45	10	EDS	Bekerja	Luar	Tepat Waktu
4	13071025	Laki	2.96	3.35	3.33	3.18	1.45	90	EDS	Tidak Bekerja	Luar	Terlambat
5	13071030	Perempuan	3	3.17	2.82	3.27	3.35	10	EIS	Tidak Bekerja	Kotabumi	Tepat Waktu
6	13071034	Laki	2.91	2.76	2.67	1.17	2.18	10	EDS	Tidak Bekerja	Luar	Tepat Waktu
7	13071044	Perempuan	3.5	3.3	3.33	3.5	3.39	18	EIS	Tidak Bekerja	Kotabumi	Tepat Waktu
8	13071057	Perempuan	3.36	3.33	3.25	3.5	3.55	18	EIS	Bekerja	Luar	Tepat Waktu
9	13071065	Laki	3.04	2.26	2.67	2.82	1.91	100	EDS	Tidak Bekerja	Luar	Tepat Waktu
10	13071073	Perempuan	3.23	3.62	3.08	3.42	3.45	18	EIS	Tidak Bekerja	Luar	Tepat Waktu
11	13071071	Perempuan	1.5	1.7	2.08	2.33	3.3	80	EIS	Tidak Bekerja	Kotabumi	Tepat Waktu
12	13071078	Perempuan	2.33	2.3	1.83	3.17	3.39	10	EIS	Bekerja	Kotabumi	Tepat Waktu
13	13071085	Perempuan	3.36	3.37	2.5	3.08	3.55	18	EIS	Bekerja	Luar	Tepat Waktu
14	13071092	Laki	3.61	3.83	3.67	3.73	2.64	18	EIS	Tidak Bekerja	Luar	Tepat Waktu
15	13071102	Perempuan	2.92	2.9	2.18	3.09	3.35	188	EIS	Tidak Bekerja	Kotabumi	Tepat Waktu
16	13071104	Perempuan	3.61	3.65	3.83	3.55	3.91	18	EIS	Bekerja	Luar	Tepat Waktu
17	13071116	Laki	3.79	3.57	3.5	1.57	2.5	18	EIS	Bekerja	Luar	Tepat Waktu
18	13071117	Laki	2.59	2.95	3.83	3.67	3.82	18	EIS	Bekerja	Luar	Tepat Waktu
19	13071122	Perempuan	3.43	3.43	3.67	3.64	3.82	18	EIS	Bekerja	Luar	Tepat Waktu
20	13071140	Perempuan	3.43	3.17	3.58	3.55	2.6	18	EIS	Tidak Bekerja	Luar	Tepat Waktu
21	13071146	Laki	3.5	2.86	2.42	3	2.36	100	EDS	Tidak Bekerja	Luar	Tepat Waktu
22	13071154	Laki	2.25	2.22	2.08	3.17	3.74	10	EIS	Bekerja	Kotabumi	Tepat Waktu
23	13071162	Perempuan	3	3.65	3.25	3.42	3.57	18	EIS	Tidak Bekerja	Kotabumi	Tepat Waktu
24	13071169	Perempuan	3.17	2.61	2.58	3.25	3.57	18	EIS	Tidak Bekerja	Kotabumi	Tepat Waktu
25	13071453	Laki	2.76	1.96	2.5	3.08	3.39	100	EIS	Tidak Bekerja	Kotabumi	Tepat Waktu
26	13071552	Perempuan	3.32	3.52	3.28	2.82	3.52	18	EIS	Tidak Bekerja	Kotabumi	Tepat Waktu
27	13071163	Perempuan	3.25	2.78	2.91	3.64	2.65	10	EIS	Tidak Bekerja	Kotabumi	Tepat Waktu
28	13071192	Laki	2.74	3.04	3.09	1.91	3	48	EIS	Tidak Bekerja	Luar	Tepat Waktu
29	13071195	Laki	2.92	3.67	2.75	2.83	2.55	18	EIS	Bekerja	Luar	Tepat Waktu
30	13071200	Perempuan	3.5	3.43	2.55	2.9	3.35	18	EDS	Tidak Bekerja	Kotabumi	Tepat Waktu

Sedangkan untuk menguji menggunakan data testing sebanyak 180 data mahasiswa seperti yang tertera pada tabel 2. dibawah ini.

Tabel 2. Cuplikan Data Testing Mahasiswa

No	Npm	Jenis_Kelamin	IPK_S1	IPK_S2	IPK_S3	IPK_S4	IPK_S5	Jumlah_SKS	Konsentrasi	Status_Kerja	Asal_Tinggal	Status_Lulus
1	16071001	Perempuan	3.1	3.33	3.42	3.6	3.4	18	EIS	Tidak Bekerja	Kotabumi	
2	16071002	Perempuan	3	3.25	3.42	3.6	3.3	18	EIS	Tidak Bekerja	Kotabumi	
3	16071003	Perempuan	2.95	3.25	3.25	3.8	3	18	EIS	Tidak Bekerja	Kotabumi	
4	16071004	Perempuan	3.25	3.33	3.67	3.8	3.25	18	EIS	Tidak Bekerja	Kotabumi	
5	16071005	Perempuan	3	2.92	3.33	3.5	3	18	EDS	Tidak Bekerja	Luar	
6	16071006	Perempuan	3.1	3	3.5	3.7	3.3	18	EDS	Tidak Bekerja	Kotabumi	
7	16071007	Laki	2.8	1.15	1.24	0	0	80	EDS	Tidak Bekerja	Luar	
8	16071008	Laki	3	3.08	2.75	3.5	2.95	18	EDS	Tidak Bekerja	Luar	
9	16071009	Laki	2.5	1	1.5	1.8	2.85	10	EIS	Tidak Bekerja	Luar	
10	16071010	Perempuan	2.95	3.08	3	3.8	3.4	18	EIS	Tidak Bekerja	Kotabumi	
11	16071011	Perempuan	3.1	3.17	3.42	3.6	3.2	18	EIS	Tidak Bekerja	Luar	
12	16071012	Perempuan	2.75	2.58	1.75	1.1	2.85	10	EIS	Tidak Bekerja	Luar	
13	16071013	Laki	3.2	2.92	2.67	3.5	3.25	18	EDS	Tidak Bekerja	Luar	
14	16071014	Laki	1.95	2.25	1.33	1	2.75	10	EDS	Tidak Bekerja	Kotabumi	
15	16071015	Perempuan	3	3.25	3.33	3.8	3.25	18	EDS	Tidak Bekerja	Kotabumi	
16	16071016	Perempuan	3.1	3.42	3.5	3.6	3.25	18	EDS	Tidak Bekerja	Luar	
17	16071017	Perempuan	3.1	3.17	3.42	3.9	3.3	18	EIS	Tidak Bekerja	Luar	
18	16071018	Laki	2.95	2.67	3.33	2.7	3	10	EIS	Tidak Bekerja	Luar	
19	16071019	Laki	2	2	1	2	0	10	EDS	Tidak Bekerja	Luar	
20	16071020	Laki	2	1.5	0	0	0	60	EDS	Tidak Bekerja	Luar	

3.2 Perhitungan Manual Algoritma Naive Bayes Classifier

Langkah-langkah perhitungan manual algoritma Naive Bayes Classifier dan pendukungnya dapat dilihat dibawah ini :

1. Menentukan Prior Probability (P)

Berdasarkan data training kelulusan yang digunakan maka akan dicari kelas kelulusan tepat waktu dan terlambat seperti pada Tabel 3

Tabel 3. Data Kelulusan Mahasiswa

Kelulusan	Jumlah Kejadian
Tepat Waktu	354
Terlambat	246
Jumlah	600

Menghitung jumlah kelas/label

$$P(Y = \text{Tepat Waktu})$$

$$= 354/600 = 0,59$$

$$P(Y = \text{Terlambat})$$

$$= 246/600 = 0,41$$

2. Menentukan Probabilitas kemunculan setiap nilai untuk atribut Jenis Kelamin (X1), Konsentrasi (X8), Status Pekerjaan (X9), Asal Tinggal (X10)

a. Menentukan probabilitas kemunculan setiap nilai untuk atribut Jenis Kelamin (X1)

Dalam menentukan probabilitas kemunculan atribut jenis kelamin berdasarkan data training maka diperoleh hasil kemunculan seperti yang terlihat pada tabel 4

Tabel 4. Data Jenis Kelamin

Jenis Kelamin X1	Jumlah Kejadian	
	Tepat Waktu	Terlambat
Laki	191	139
Perempuan	163	107
Jumlah	354	246

$$P(\text{Jenis Kelamin} = \text{Laki} | Y = \text{Tepat Waktu})$$

$$= 191/354 = 0.540$$

$$P(\text{Jenis Kelamin} = \text{Laki} | Y = \text{Terlambat})$$

$$= 139/246 = 0.565$$

$$P(\text{Jenis Kelamin} = \text{Perempuan} | Y = \text{Tepat Waktu})$$

$$= 163/354 = 0.460$$

$$P(\text{Jenis Kelamin} = \text{Perempuan} | Y = \text{Terlambat})$$

$$= 107/246 = 0.435$$

b. Menentukan probabilitas kemunculan setiap nilai untuk atribut Konsentrasi (X8)

Dari data training diatas di peroleh data konsentrasi seperti pada tabel 5

Tabel 5. Data Konsentrasi

Konsentrasi X8	Jumlah Kejadian	
	Tepat Waktu	Terlambat
EIS	218	142
EDS	136	104
Jumlah	354	246

$$P(\text{Konsentrasi} = \text{EIS} | Y = \text{Tepat Waktu})$$

$$= 218/354 = 0.616$$

$$P(\text{Konsentrasi} = \text{EIS} | Y = \text{Terlambat})$$

$$= 142/246 = 0.577$$

$$P(\text{Konsentrasi} = \text{EDS} | Y = \text{Tepat Waktu})$$

$$= 136/354 = 0.384$$

$$P(\text{Konsentrasi} = \text{EDS} | Y = \text{Terlambat})$$

$$= 104/246 = 0.423$$

- c. Menentukan probabilitas kemunculan setiap nilai untuk atribut Status Pekerjaan (X9)

Dari data training diatas di peroleh data Status Pekerjaan seperti pada tabel 6

Tabel 6. Data Status Pekerjaan

Status Pekerjaan X9	Jumlah Kejadian	
	Tepat Waktu	Terlambat
Bekerja	133	96
Tidak Bekerja	221	150
Jumlah	354	246

$$P(\text{Status Pekerjaan} = \text{Bekerja} | Y=\text{Tepat Waktu}) = 133/354 = 0.376$$

$$P(\text{Status Pekerjaan} = \text{Bekerja} | Y=\text{Terlambat}) = 96/246 = 0.390$$

$$P(\text{Status Pekerjaan} = \text{Tidak Bekerja} | Y=\text{Tepat Waktu}) = 221/354 = 0.624$$

$$P(\text{Status Pekerjaan} = \text{Tidak Bekerja} | Y=\text{Terlambat}) = 150/246 = 0.610$$

- d. Menentukan probabilitas kemunculan setiap nilai untuk atribut Asal Tinggal (X9)

Dari data training diatas di peroleh data Asal Tinggal seperti pada tabel 7

Tabel 7. Data Kemunculan Asal Tinggal

Asal Tinggal X9	Jumlah Kejadian	
	Tepat Waktu	Terlambat
Kotabumi	265	206
Luar	89	40
Jumlah	354	246

$$P(\text{Asal Tinggal} = \text{Kotabumi} | Y=\text{Tepat Waktu}) = 265/354 = 0.749$$

$$P(\text{Asal Tinggal} = \text{Kotabumi} | Y=\text{Terlambat}) = 206/246 = 0.837$$

$$P(\text{Asal Tinggal} = \text{Luar} | Y=\text{Tepat Waktu}) = 89/354 = 0.251$$

$$P(\text{Asal Tinggal} = \text{Luar} | Y=\text{Terlambat}) = 40/246 = 0.163$$

- e. Menghitung probabilitas untuk setiap kelas berdasarkan atribut dari kasus diatas akan dilakukan perhitungan dengan nilai probabilitas dari variabel yang bersifat kontinu yaitu IP_S1 (X2), IP_S2(X3), IP_S3(X4), IP_S4(X5), IP_S5(X6), Jumlah SKS (X7)

Berikut ini adalah perhitungan untuk mencari data mean dan standar deviasi untuk IPK_S1 sampai dengan IPK_S5 dan Jumlah SKS.

Tabel 8. Mean dan Standar Deviasi untuk setiap IPK Semester dengan masing-masing kategori

Nama Atribut	Standar Deviasi		Mean	
	Tepat Waktu	Terlambat	Tepat Waktu	Terlambat
IPK_S1 (X2)	0.432762	0.495487	2.812062	2.711667
IPK_S2 (X3)	0.331813	0.369557	2.948701	2.842683
IPK_S3 (X4)	0.288127	0.340476	2.875650	2.796829
IPK_S4 (X5)	0.324597	0.409733	2.883755	2.770121
IPK_S5 (X6)	0.395623	0.469013	2.921722	2.792071
Jumlah SKS	10.6158	15.2483	111.870	110.524

- 3. Menghitung Probabilitas dari setiap atribut Contoh Kasus Perhitungan menggunakan algoritma Naïve Bayes

- 1) Pengujian data testing mahasiswa Kasus 1 NPM : 16071001
 - X1=Perempuan X4=3.42 X7=118
 - X2 = 3.10 X5=3.60 X8=EIS
 - X3 = 3.33 X6=3.40
 - X9=Tidak Bekerja X10 = Kotabumi

$$P(X1 = \text{Perempuan} | \text{Tepat Waktu}) = 163/354 = 0.460$$

$$P(X1 = \text{Perempuan} | \text{Terlambat}) = 107/246 = 0.435$$

$$P(X2=3.10 | \text{Tepat Waktu}) = \frac{1}{\sqrt{2\pi} (0.432762)} e^{-\frac{(3.10 - 2.812062)^2}{2(0.432762)^2}} = 0.606564 \times 0.801439 = 0.486124$$

$$P(X2=3.10 | \text{Terlambat}) = \frac{1}{\sqrt{2\pi} (0.495487)} e^{-\frac{(3.10 - 2.711667)^2}{2(0.495487)^2}} = 0.566897 \times 0.735558 = 0.416985$$

$$P(X3=3.33 | \text{Tepat Waktu}) = \frac{1}{\sqrt{2\pi} (0.331813)} e^{-\frac{(3.33 - 2.948701)^2}{2(0.331813)^2}} = 0.692745 \times 0.516715 = 0.357952$$

$$P(X3=3.33 | \text{Terlambat}) = \frac{1}{\sqrt{2\pi} (0.369557)} e^{-\frac{(3.33 - 2.842683)^2}{2(0.369557)^2}} = 0.656418 \times 0.419188 = 0.275162$$

$$P(X4=3.42 | \text{Tepat Waktu}) = \frac{1}{\sqrt{2\pi} (0.288127)} e^{-\frac{(3.42 - 2.875650)^2}{2(0.288127)^2}} = 0.743410 \times 0.16785 = 0.124781$$

$$P(X4=3.42 | \text{Terlambat}) = \frac{1}{\sqrt{2\pi} (0.340476)} e^{-\frac{(3.42 - 2.796829)^2}{2(0.340476)^2}} = 0.6833875 \times 0.1877308 = 0.128096$$

$$P(X5=3.60|Tepat Waktu) = \frac{1}{\sqrt{2\pi (0.324595)}} e^{-\frac{(3.60-2.883757)^2}{2(0.324595)^2}} = 0.70041 \times 0.087641 = 0.061384$$

$$P(X5=3.60|Terlambat) = \frac{1}{\sqrt{2\pi (0.409735)}} e^{-\frac{(3.92-2.770127)^2}{2(0.3246)^2}} = 0.623403 \times 0.128580 = 0.080162$$

$$P(X6=3.40|Tepat Waktu) = \frac{1}{\sqrt{2\pi (0.395623)}} e^{-\frac{(3.65-2.921723)^2}{2(0.395623)^2}} = 0.634424 \times 0.481548 = 0.305505$$

$$P(X6=3.40|Terlambat) = \frac{1}{\sqrt{2\pi (0.469015)}} e^{-\frac{(3.65-2.792073)^2}{2(0.469015)^2}} = 0.582676 \times 0.431691 = 0.251536$$

$$P(X7=118|Tepat Waktu) = \frac{1}{\sqrt{2\pi (10.615882)}} e^{-\frac{(118-111.8701)^2}{2(10.615882)^2}} = 0.122474 \times 0.846443 = 0.103667$$

$$P(X7=118|Terlambat) = \frac{1}{\sqrt{2\pi (15.248333)}} e^{-\frac{(118-110.5244)^2}{2(15.248333)^2}} = 0.10219 \times 0.886764 = 0.090619$$

$$P(\text{Konsentrasi} = \text{EIS} | Y=\text{Tepat Waktu}) = 218/354 = 0.616$$

$$P(\text{Konsentrasi} = \text{EIS} | Y=\text{Terlambat}) = 142/246 = 0.577$$

$$P(\text{Status Pekerjaan} = \text{Tidak Bekerja} | Y=\text{Tepat Waktu}) = 221/354 = 0.624$$

$$P(\text{Status Pekerjaan} = \text{Tidak Bekerja} | Y=\text{Terlambat}) = 150/246 = 0.610$$

$$P(\text{Asal Tinggal} = \text{Kotabumi} | Y=\text{Tepat Waktu}) = 265/354 = 0.749$$

$$P(\text{Asal Tinggal} = \text{Kotabumi} | Y=\text{Terlambat}) = 206/246 = 0.837$$

4. Membandingkan hasil kelas Tepat Waktu dan Terlambat

Likelihood Tepat Waktu = P(Perempuan|Tepat Waktu) * P(IPK_S1|Tepat Waktu) * P(IPK_S2|Tepat Waktu) * P(IPK_S3|Tepat Waktu) * P(IPK_S4|Tepat Waktu) * P(IPK_S5|Tepat Waktu) * P(Jumlah SKS|Tepat Waktu) * P(Konsentrasi|Tepat Waktu) * P(Status Pekerjaan|Tepat Waktu) * P(Asal Tinggal|Tepat Waktu)

$$\text{Jadi : } 0.460 * 0.486124 * 0.357952 * 0.124781 * 0.061384 * 0.305505 * 0.103667 * 0.616 * 0.624 * 0.749$$

$$= 0.00000559039$$

Likelihood Terlambat = P(Perempuan|Terlambat) * P(IPK_S1|Terlambat) * P(IPK_S2|Terlambat) * P(IPK_S3|Terlambat) * P(IPK_S4|Terlambat) * P(IPK_S5|Terlambat) * P(Jumlah SKS|Terlambat) * P(Konsentrasi|Terlambat) * P(Status Pekerjaan|Terlambat) * P(Asal Tinggal|Terlambat)

Jadi :

$$0.727 * 0.416985 * 0.275162 * 0.128096 * 0.080162 * 0.251536 * 0.090619 * 0.577 * 0.609 * 0.609 = 0.00000344154$$

Jadi dari contoh kasus diatas bisa diprediksi bahwa mahasiswa tersebut diprediksi hasilnya adalah "Tepat Waktu" karena nilai likelihood Status lulus Tepat waktu lebih besar dibandingkan dengan Terlambat

3.3 Contoh Kasus Perhitungan Manual Algoritma K-Nearest Neighbor

Langkah-langkah perhitungan manual algoritma K-Nearest Neighbor pendukungnya dapat di lihat dibawah ini :

1. Pemberian Nilai Bobot untuk Tiap Kriteria
 Dalam pemberian bobot kriteria, dimana bobot antara satu kriteria dengan kriteria yang lain dapat didefinisikan dengan nilai yang berbeda. Berikut ini adalah contoh pemberian bobot pada kriteria indek prestasi semester (IPKS) 1 – 5 dan jumlah SKS dimana nilai bobot digunakan untuk menghitung nilai kriteria.

Tabel 9. Nilai IPK Semester (IPS)

No	Kriteria	Nilai	Bobot
1	IPKS >=3.5	100 = 1	30%=0.3
2	IPKS >=3 dan <3.5	80 = 0.8	
3	IPKS >=2 dan <3	60 = 0.6	
4	IPKS <2	40 = 0.4	

Nilai Kriteria IPKS adalah

$$\text{Nilai Kriteria} = \frac{\sum(100 \times 0.3) + (80 \times 0.3) + (60 \times 0.3) + (40 \times 0.3)}{N}$$

$$\text{Nilai Kriteria} = \frac{30+24+18+12}{4}$$

$$\text{Nilai Kriteria} = \frac{84}{4}$$

$$\text{Nilai Kriteria} = 21$$

Tabel 10. Nilai Atribut Jumlah SKS

No	Kriteria	Nilai	Bobot
1	<=120 dan >=115	100 = 1	20%=0.2
2	<115 dan >=100	80 = 0.8	
3	<= 60 dan <100	60 = 0.6	
4	>60	40 = 0.4	

Nilai Kriteria Jumlah SKS adalah

$$\text{Nilai Kriteria} = \frac{\sum(100 \times 0.2) + (80 \times 0.2) + (60 \times 0.2) + (40 \times 0.2)}{n}$$

$$\text{Nilai Kriteria} = \frac{20+16+12+8}{4}$$

$$\text{Nilai Kriteria} = \frac{56}{4}$$

$$\text{Nilai Kriteria} = 14$$

Perhitungan yang sama dilakukan untuk semua kriteria dalam menentukan nilai dari kriteria-kriteria. Pada tabel 11 menunjukkan bobot untuk tiap kriteria yang digunakan.

Tabel 11. Bobot kriteria

No	Atribut	Bobot
1	IPKS	0.3
2	Jenis Kelamin	0.1
3	Jumlah SKS	0.2
4	Konsentrasi	0.1
5	Status Pekerjaan	0.2
6	Asal Tinggal	0.1

2. Menghitung Nilai Kedekatan Antar Kriteria
Selanjutnya menentukan nilai kedekatan untuk semua kriteria, dimana nilai kedekatan adalah nilai yang berada pada jarak nilai antara 0 dan 1. Nilai 0 berarti kedua kasus mutlak tidak mirip, dan untuk nilai 1 artinya kasus mirip secara mutlak. Nilai kedekatan akan digunakan untuk menghitung jarak terdekat antara data training dan data testing. Berikut ini adalah tabel nilai kedekatan untuk kriteria IPK S :

Tabel 12. Nilai Kedekatan Kriteria IPK S

IPKS	>=35	>=3 dan <3.5	>=2 dan <3	<2 dan =<1
>=35	1	0.8	0.6	0.4
>=3 dan <3.5	0.8	1	0.75	0.5
>=2 dan <3	0.6	0.75	1	0.667
<2 dan =<1	0.4	0.5	0.667	1

Selanjutnya untuk menentukan Nilai kedekatan untuk atribut Jumlah SKS

Tabel 13. Nilai Kedekatan Jumlah SKS

Jumlah SKS	<=120 dan >=115	<115 dan >=100	<= 60 dan <100	>60
<=120 dan >=115	1	0.8	0.6	0.4
<115 dan >=100	0.8	1	0.75	0.5
<= 60 dan <100	0.6	0.75	1	0.667
>60	0.4	0.5	0.667	1

Berikut ini adalah Nilai kedekatan untuk Jenis Kelamin, Konsentrasi, Status Pekerjaan dan Asal Tinggal.

Pada masing-masing penentuan nilai kedekatan antar kriteria diperoleh nilai seperti pada Tabel 14, tabel 15, Tabel 16 dan Tabel 17.

Tabel 14. Nilai Kedekatan Jenis Kelamin

Jenis Kelamin	Laki	Perempuan
Laki	1	0
Perempuan	0	1

Tabel 15. Nilai Kedekatan Konsentrasi

Konsentrasi	EIS	EDS
EIS	1	0
EDS	0	1

Tabel 16. Nilai Kedekatan Status Pekerjaan

Status Pekerjaan	Bekerja	Tidak Bekerja
Bekerja	1	0
Tidak Bekerja	0	1

Tabel 17. Nilai Kedekatan Asal Tinggal

Status Pekerjaan	Kotabumi	Luar
Kotabumi	1	0
Luar	0	1

3. Perhitungan fungsi *similarity*

Selanjutnya dilakukan perbandingan antara data testing dan data training yang telah dipersiapkan sebelumnya dengan menggunakan rumus, dimana data testing ke-1 dibandingkan dengan kasus yang telah terjadi sebelumnya.

Data testing ke-1 NPM (16071002), dihitung nilainya berdasarkan tahapan-tahapan yang telah diulas sebelumnya. Berikut ini adalah contoh perhitungan dari tabel hasil perhitungan dengan kasus yang baru yaitu :

Menghitung kasus ke 1

- Kedekatan jenis kelamin = 0
- Bobot kedekatan jenis kelamin = 0.1
- Kedekatan nilai IPK_S1 = 1
- Kedekatan nilai IPK_S2 = 0.8
- Kedekatan nilai IPK_S3 = 1
- Kedekatan nilai IPK_S4 = 0.4
- Kedekatan nilai IPK_S5 = 0.5
- Bobot kedekatan nilai IPKS = 0.3
- Kedekatan jumlah sks = 0.8
- Bobot kedekatan jumlah sks = 0.2
- Kedekatan konsentrasi = 1
- Bobot kedekatan konsentrasi = 0.1
- Kedekatan status pekerjaan = 0
- Bobot kedekatan status pekerjaan = 0.2
- Kedekatan asal tinggal = 0
- Bobot kedekatan asal tinggal = 0.1

Menghitung (Kasus 1)

Jarak =

$$\frac{(a*b)+(c*h)+(d*h)+(e*h)+(f*h)+(g*h)+(i*j)+(k*1)+(m*n)+(o*p)}{b+(5*h)+j+l+n+q}$$

Jarak =

$$\frac{(0*0.1)+(1*0.3)+(0.8*0.3)+(1*0.3)+(0.4*0.3)+(0.5*0.3)+(0.8*0.2)+(1*0.1)+(0*0.2)+(0*0.1)}{0.1+(5*0.3)+0.2+0.1+0.2+0.1}$$

Jarak = $\frac{1.37}{2.2}$
Jarak = 0.622

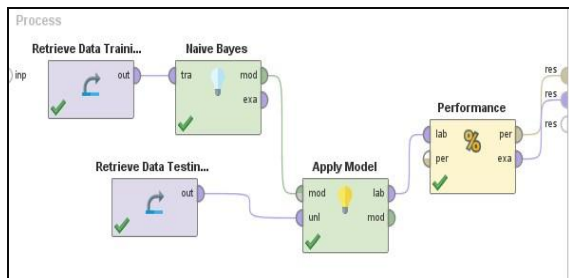
3.4. Pengujian Model Algoritma Data Mining

Dalam penyelesaian penelitian tentang prediksi kelulusan mahasiswa pada perguruan tinggi STMIK Dian Cipta Cendikia Kotabumi, metode yang digunakan adalah *Naive Bayes* dan *K-nearest neighbor* dengan 11 parameter yaitu 10 atribut dan 1 sebagai label yang terdiri dari Jenis Kelamin, IPK Semester 1 – IPK Semester 5, Jumlah SKS, Konsentrasi, Status Pekerjaan, Asal Tinggal dan label Tepat Waktu dan Terlambat.

3.4.1 Pengujian Model Algoritma Naive Bayes

1. Proses Pengujian Model dengan Rapidminer
Pengujian model *Naive Bayes* dengan menggunakan rapidminer akan menguji akurasi algoritma *Naive Bayes*, seberapa besar akurasi yang akan dihasilkan, hasil dari pengujian akurasi ini akan dibandingkan dengan metode lainnya.

Berikut ini adalah tahapan-tahapan yang dilakukan dalam pengujian dengan menggunakan *performance* untuk menghasilkan tingkat akurasi dan dari metode yang digunakan. Pengujian model *Naive Bayes* dengan menggunakan data training dan data testing seperti gambar 2



Gambar 2. Proses Pemodelan Algoritma *Naive Bayes*

Setelah dilakukan proses pemodelan dengan algoritma *naive bayes* maka akan diperoleh prediksi dibawah ini :

	Status_Lulus	prediction(Status_Lulus)	confidence(Tepat...)	confidence(Terlambat)	Npm	Nama	Jenis_K...
1	Tepat Waktu	Tepat Waktu	0.671	0.329	16071001	Arisa Agustini	Perempu...
2	Tepat Waktu	Tepat Waktu	0.678	0.322	16071002	Ardiana Saffri	Perempu...
3	Tepat Waktu	Tepat Waktu	0.569	0.431	16071003	Dea Puspita	Perempu...
4	Tepat Waktu	Terlambat	0.451	0.549	16071004	Desi Restiana	Perempu...
5	Tepat Waktu	Tepat Waktu	0.755	0.245	16071005	Devil Oktasari	Perempu...
6	Tepat Waktu	Tepat Waktu	0.549	0.451	16071006	Dissmal Pus...	Perempu...
7	Terlambat	Terlambat	0	1	16071007	Doni Pratama	Laki
8	Tepat Waktu	Tepat Waktu	0.739	0.261	16071008	Fachri Pawiga	Laki
9	Tepat Waktu	Terlambat	0.000	1.000	16071009	Firhan Hudab...	Laki
10	Tepat Waktu	Tepat Waktu	0.619	0.381	16071010	Indriyani	Perempu...
11	Tepat Waktu	Tepat Waktu	0.782	0.218	16071011	Jenni Astina P.	Perempu...

Gambar 3. Proses Hasil Pemodelan algoritma *Naive Bayes*

2. Hasil Analisa Pengujian Model

Setelah proses pengujian model dengan data training sebanyak 600 data dan data testing sebanyak 180 data akan menghasilkan *output accuracy* sebesar 85.00% *precision* sebesar 44.90% dan *recall* sebesar 100% seperti yang terlihat pada gambar 4 dibawah ini.

Criterion	Table View	Plot View	
accuracy	accuracy: 85.00%		
precision			
recall			
AUC (optimistic)	true Tepat Waktu	true Terlambat	class precision
AUC	pred. Tepat Waktu	0	100.00%
AUC (pessimistic)	pred. Terlambat	27	44.90%
class recall	82.91%	100.00%	

Gambar 4. Hasil *Confusion Matrix Naive Bayes*

Selanjutnya lebih jelasnya untuk hasil *accuracy*, *precision*, *recall* terlihat pada gambar 5

PerformanceVector	
PerformanceVector:	
accuracy:	85.00%
ConfusionMatrix:	
True:	Tepat Waktu Terlambat
Tepat Waktu:	131 0
Terlambat:	27 22
precision:	44.90% (positive class: Terlambat)
ConfusionMatrix:	
True:	Tepat Waktu Terlambat
Tepat Waktu:	131 0
Terlambat:	27 22
recall:	100.00% (positive class: Terlambat)
ConfusionMatrix:	
True:	Tepat Waktu Terlambat
Tepat Waktu:	131 0
Terlambat:	27 22
AUC (optimistic):	0.999 (positive class: Terlambat)
AUC:	0.999 (positive class: Terlambat)
AUC (pessimistic):	0.999 (positive class: Terlambat)

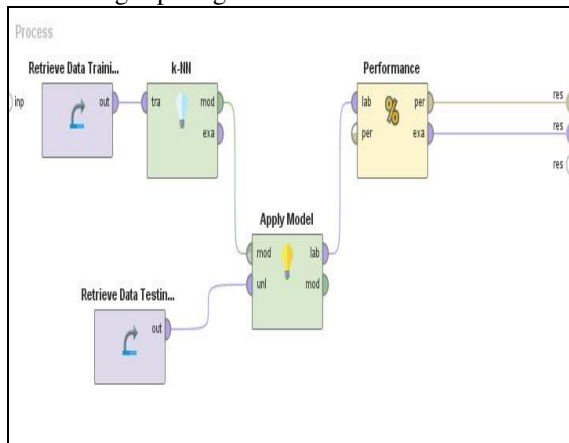
Gambar 5. Hasil *Performance Vektor Naive Bayes*

3.4.2 Pengujian Model Algoritma K-nearest neighbor

1. Proses Pengujian Model dengan Rapidminer
Pengujian model *K-Nearest Neighbor* dengan menggunakan rapidminer akan menguji akurasi algoritma *K-Nearest Neighbor*, seberapa besar akurasi yang akan dihasilkan, hasil dari pengujian akurasi ini akan dibandingkan dengan metode lainnya.

Berikut ini adalah tahapan-tahapan yang dilakukan dalam pengujian dengan menggunakan *performance* untuk menghasilkan tingkat akurasi dan dari metode yang digunakan. Pengujian model *K-Nearest*

Neighbor dengan menggunakan data training dan data testing seperti gambar 6



Gambar 6. Proses pengujian Algoritma K-Nearest Neighbor

Setelah dilakukan proses pemodelan dengan algoritma K-Nearest Neighbor maka akan diperoleh prediksi pada gambar 7 dibawah ini :

Ro...	Status_Lulus	prediction(S...	confidence(Tepat W...	confidence(Terlambat)	Npm	Nama	Jenis_Kela...
1	Tepat Waktu	Tepat Waktu	0.788	0.212	16071001	Anisa Agustin...	Perempuan
2	Tepat Waktu	Tepat Waktu	0.605	0.395	16071002	Andiana Safitri	Perempuan
3	Tepat Waktu	Tepat Waktu	0.596	0.404	16071003	Dea Puspiita	Perempuan
4	Tepat Waktu	Tepat Waktu	0.604	0.396	16071004	Desi Resitiana	Perempuan
5	Tepat Waktu	Tepat Waktu	0.595	0.405	16071005	Devil Oktasari	Perempuan
6	Tepat Waktu	Tepat Waktu	0.800	0.200	16071006	Diasmai Pusp...	Perempuan
7	Terlambat	Terlambat	0.187	0.813	16071007	Doni Pralama	Laki
8	Tepat Waktu	Tepat Waktu	0.599	0.401	16071008	Fachri Pawiga	Laki
9	Tepat Waktu	Tepat Waktu	0.601	0.399	16071009	Fitnah Hulaba...	Laki
10	Tepat Waktu	Tepat Waktu	0.597	0.403	16071010	Indriyani	Perempuan
11	Tepat Waktu	Tepat Waktu	1	0	16071011	Jeri Astina Putri	Perempuan

Gambar 7. Proses Hasil Pemodelan Algoritma K-Nearest Neighbor

2. Hasil Analisa Pengujian Model

Setelah proses pengujian model, akan menghasilkan output accuracy seperti gambar 8 dibawah ini

Criterion	Table View	Plot View
accuracy	accuracy: 68.89%	
precision		
recall		
AUC (optimistic)	true Tepat Waktu	true Terlambat
AUC	pred. Tepat Waktu	class precision
AUC (pessimistic)	112	10
	46	12
	70.89%	54.55%

Gambar 8. Hasil Confusion Matrix K-Nearest Neighbor

Dari gambar 8 memberikan informasi bahwa hasil accuracy dari metode klasifikasi K-Nearest Neighbor adalah sebesar 68.89%. Untuk nilai precision adalah sebesar 20.69% dan nilai recall adalah sebesar 54.55%.

```

PerformanceVector
PerformanceVector:
accuracy: 68.89%
ConfusionMatrix:
True:  Tepat Waktu    Terlambat
Tepat Waktu:  112     10
Terlambat:    46      12
precision: 20.69% (positive class: Terlambat)
ConfusionMatrix:
True:  Tepat Waktu    Terlambat
Tepat Waktu:  112     10
Terlambat:    46      12
recall: 54.55% (positive class: Terlambat)
ConfusionMatrix:
True:  Tepat Waktu    Terlambat
Tepat Waktu:  112     10
Terlambat:    46      12
AUC (optimistic): 0.646 (positive class: Terlambat)
AUC: 0.646 (positive class: Terlambat)
AUC (pessimistic): 0.646 (positive class: Terlambat)
    
```

Gambar 9. Hasil Performance Vektor K-Nearest Neighbor

Kesimpulan

Berdasarkan pembahasan yang telah diuraikan, yaitu tentang Prediksi Tingkat Kelulusan Mahasiswa Dengan Metode *Naïve Bayes* dan *K-Nearest Neighbor* untuk memprediksi tingkat kelulusan Tepat Waktu mahasiswa, maka dapat disimpulkan beberapa hal sebagai berikut:

1. Hasil pengujian pada perhitungan manual naïve bayes pada data testing ke-1 di prediksi hasilnya adalah tepat waktu
2. Hasil pengujian menggunakan naïve bayes menggunakan rapidminer dengan jumlah training 600 dan data testing 180 di dapat nilai akurasi sebesar 85 %
3. Hasil pengujian manual K-nearest neighbor pada testing ke-1 diprediksi hasilnya adalah tepat waktu.
4. Hasil pengujian menggunakan k-nearest neighbor menggunakan rapidminer dengan jumlah training 600 dan data testing 180 di dapat nilai akurasi sebesar 68.89 %

5.1. Saran

Saran yang ditambahkan untuk pengembangan dan penelitian selanjutnya adalah sebagai berikut :

1. Hasil penelitian ini perlu diimplementasikan menjadi perangkat lunak yang dapat digunakan oleh pihak STMIK Dian Cipta Cendikia Kotabumi untuk dapat memprediksi ketepatan waktu lulus mahasiswa.
2. Untuk penelitian mendatang, pemilihan algoritma klasifikasi selain naïve bayes dan k-nearest neighbor seperti C.45, support vector machine serta penggunaan teknik optimasi.
3. Selain atribut yang digunakan oleh penulis dalam melakukan penelitian, akan lebih baik menambahkan atribut selain yang penulis gunakan.

DAFTAR PUSTAKA

- [1] Alannuary, Adis, 2015, *Edaran Dikti Tentang Nisbah Rasio Dosen/Mahasiswa Di Atas 100*, <http://www.kopertis4.or.id/2015/08/edaran-dikti-tentang-nisbah-rasio-dosenmahasiswa-di-atas-100>
- [2] BAN-PT, B. A., 2010, *AKREDITASI PROGRAM STUDI DIPLOMA*. Jakarta: BAN-PT.
- [3] Windarti, M., 2016, *Prediksi Masa Studi Mahasiswa Menggunakan Kombinasi Algoritma Bayesian Network Dan K-Nearest Neighbors* (Doctoral dissertation, UAJY).

- [4] Lumenta, A.S. and Jacobus, A., 2017. Prediksi Masa Studi Mahasiswa dengan Menggunakan Algoritma Naïve Bayes. *JURNAL TEKNIK INFORMATIKA UNIVERSITAS SAM RATULANGI*, 11(1).

- [5] Novianti, A.G. and Prasetyo, D., PENERAPAN ALGORITMA K-NEAREST NEIGHBOR (K-NN) UNTUK PREDIKSI WAKTU KELULUSAN MAHASISWA.

- [6] Ginting, S.L.B., Zarman, W. and Darmawan, A., 2014. Teknik Data Mining Untuk Memprediksi Masa Studi Mahasiswa Menggunakan Algoritma K-Nearest Neighborhood. *Jurnal Teknik Komputer Unikom*, 3(2), pp.29-34.