

Sistem Prediksi Lama Studi Kuliah Menggunakan Metode Naive Bayes

Muqorobin^{*1}, Moch Bagoes Pakarti²

^{1,2} Institut Teknologi Bisnis AAS Indonesia

^{1,2}Informatika ITB AAS Indonesia

E-mail: ^{*1}robbyaullah@gmail.com, ²8a9oes@gmail.com

Abstrak

Institut Teknologi Bisnis AAS Indonesia merupakan salah satu lembaga pendidikan perguruan tinggi yang ada di Jawa Tengah. Permasalahan lama studi setiap mahasiswa biasanya disebabkan atau dipengaruhi oleh banyak faktor, sehingga perlu dibuat sistem prediksi kelulusan mahasiswa untuk mengetahui faktor apa saja yang dapat mempengaruhi kelulusan mahasiswa. Sistem yang dirancang adalah sistem prediksi tingkat kelulusan mahasiswa pada tepat atau terlambatnya studi menggunakan metode data mining dengan algoritma naïve bayes. Kriteria yang digunakan dalam memprediksi tingkat kelulusan adalah Prodi, Gender, Asal Sekolah, Rata-rata SKS dan Asdos. Sistem diolah menggunakan Microsoft Excel dan Pengujian Sistem Menggunakan Tools RapidMiner. Dalam perhitungannya menggunakan Data sampel untuk data training berjumlah 13 data, untuk testing berjumlah 5 data. Hasil Penelitian dalam Uji sistem telah memiliki tingkat akurasi 80%, recall 50% dan presisi 100%.

Kata Kunci— *Prediksi, Lama Studi, Naïve Bayes*

Abstract

AAS Indonesia Institute of Business Technology is one of the higher education institutions in Central Java. The problem of length of study for each student is usually caused or influenced by many factors, so it is necessary to create a student graduation system to find out what factors can affect student graduation. The system designed is a predictive system for student graduation rates that are right or late in studies using data mining methods with naïve Bayes algorithms. The criteria used in predicting the graduation rate are Study Program, Gender, School Origin, Average Credits and Asdos. The system is obtained using Microsoft Excel and System Testing is using RapidMiner Tools. In the calculations using sample data for training data, which amounted to 13 data, for testing as many as 5 data. The results of the research in the system test have 80% accuracy, 50% recall and 100% precision.

Keywords— *Prediction, Length of Study, Naïve Bayes*

1. PENDAHULUAN

Institut Teknologi Bisnis AAS Indonesia merupakan salah satu Perguruan Tinggi yang berperan penting dalam menjalankan tri darma perguruan tinggi. Teknologi komputasi dan media penyimpanan telah memungkinkan manusia untuk mengumpulkan dan menyimpan data dari berbagai sumber dengan jangkauan yang amat luas. Meskipun teknologi basis data modern telah menghasilkan media penyimpanan yang besar, Teknologi untuk membantu menganalisis, memahami, atau bahkan memvisualisasikan data belum banyak tersedia. Hal inilah yang melatarbelakangi dikembangkannya

konsep data mining. Lama studi dari mahasiswa sangatlah penting bagi mahasiswa itu sendiri dan tentunya bagi program studi dalam rangka untuk menentukan langkah-langkah yang diperlukan agar perjalanan studi mahasiswa menjadi lancar. Permasalahan lama studi setiap mahasiswa biasanya disebabkan atau dipengaruhi oleh banyak faktor, sehingga perlu dibuat sistem prediksi kelulusan mahasiswa untuk mengetahui faktor apa saja yang dapat mempengaruhi kelulusan mahasiswa.

Ada beberapa penelitian yang telah diteliti sebelumnya tentang prediksi kelulusan mahasiswa diantaranya menurut Huda[1] dalam penelitiannya “aplikasi data mining untuk menampilkan tingkat kelulusan mahasiswa dengan studi kasus FMIPA Universitas Diponegoro” dengan menggunakan metode data mining dengan algoritma apriori menyebutkan bahwa data proses masuk, asal sekolah, kota asal dan program studi menjadi pertimbangan dalam menentukan tingkat kelulusan. Sedangkan menurut Meinanda [2] dalam penelitiannya tentang prediksi masa studi sarjana dengan menggunakan metode artificial neural network menyebutkan bahwa masa studi seorang sarjana dipengaruhi oleh IPK, jumlah mata kuliah yang diambil, jumlah mata kuliah yang mengulang, jumlah mata kuliah tertentu.

Salah satu metode yang dapat digunakan untuk memprediksi tingkat kelulusan mahasiswa adalah metode klasifikasi Naïve Bayes. Beberapa penelitian tentang metode klasifikasi Naive Bayes diantaranya adalah, “Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classification” yang memiliki tingkat keberhasilan 70%, sedangkan algoritma Naive Bayes untuk perkiraan waktu studi mahasiswa memiliki tingkat keberhasilan 80%. Berdasarkan kajian penelitian tingkat kelulusan mahasiswa menggunakan metode Naive Bayes memiliki akurasi tinggi dibandingkan metode-metode lainnya. Untuk itu peneliti menggunakan metode Naïve Bayes untuk meneliti kelulusan mahasiswa dengan kriteria yang digunakan diantaranya nama, nim, jurusan, gener, asal sekolah, nilai rata-rata sks dan asisten ahli, tahun lulus, kategori kelulusan (tepat waktu, tidak tepat waktu). Tujuan yang ingin dicapai dalam penelitian ini adalah mengetahui prediksi lama masa studi dari setiap mahasiswa yang telah menempuh kuliah hingga minimal semester IV dengan menggunakan dasar data kelulusan sebelumnya.

Adapun beberapa manfaat yang diharapkan pada penelitian ini adalah dengan mengetahui prediksi dari lama masa studi, maka dapat diambil manfaat bagi mahasiswa khususnya untuk memperhatikan perjalanan studinya dan umumnya bagi program studi untuk melakukan tindakan preventif berkaitan dengan lama studi setiap mahasiswa tersebut.

2. METODE PENELITIAN

Menurut Gartner Group, data mining adalah suatu proses menemukan hubungan yang berarti, pola, dan kecenderungan dengan memeriksa dalam sekumpulan besar data yang tersimpan dalam penyimpanan dengan menggunakan teknik pengenalan pola seperti teknik statistik dan matematika. Keluaran dari data mining ini bisa dipakai untuk memperbaiki pengambilan keputusan di masa depan[3].

Penelitian ini menggunakan teknik dalam metode mining dengan beberapa tahapan penelitian sebagai berikut:

a. Identifikasi Masalah.

Mengidentifikasi masalah-masalah yang relevan dengan penelitian ini yaitu bagaimana cara mencari perkiraan waktu studi mahasiswa.

b. Pengumpulan Data

Menyiapkan data-data yang dibutuhkan, yaitu data berkaitan dengan identitas pribadi dari mahasiswa dan nilai matakuliah yang telah ditempuh oleh mahasiswa hingga menyelesaikan seluruh sks yang telah ditentukan dan telah dinyatakan lulus. Data nilai matakuliah yang telah ditempuh hanya diambil sampai dengan semester IV dan kemudian dihitung indeks prestasinya (IPK) secara kumulatif sampai dengan semester IV. Data akademik mahasiswa di dapat dari bagian akademik jurusan akuntansi dan ekonomi syariah dan manajemen pajak di ITB AAS Indonesia. Dan data kelulusan di dapat dari Koordinator TA. Data yang di ambil merupakan data mahasiswa tahun 2020 sampai 2021 yang sudah dinyatakan lulus, yang nantinya data tersebut merupakan data training. Data mahasiswa tahun 2020 sampai 2021 yang diasumsikan belum lulus merupakan data target[4].

Metode yang digunakan dalam pengambilan data tersebut adalah secara random atau acak dari setiap tahun mahasiswa. Sumber Data: data yang digunakan dalam penelitian ini terdiri dari beberapa sumber data, yaitu data asli (master) Mahasiswa, data mata kuliah, dan data Kelulusan.

Setelah seluruh data terkumpul, selanjutnya dilakukan tahapan proses data mining,yaitu[5].

1. Pembersihan data (cleaning)

Setelah data di peroleh, maka akan ditemukan data-data yang tidak lengkap. Untuk itu diperlukan tahapan pembersihan data yaitu membuang data-data yang tidak lengkap.

2. Integrasi Data (data integration).

Integrasi data yaitu penggabungan data.

3. Seleksi Data.

Seleksi data, yaitu menyeleksi data-data yang akan digunakan.

4. Transformasi data.

Transformasi data yaitu perubahan data atau penggabungan data ke dalam format yang sesuai untuk diproses.

5. Proses Mining.

Proses mining yaitu merupakan proses utama saat metode diterapkan untuk menemukan hasil prediksi kelulusan mahasiswa.

c. Konversi Data

Untuk mempermudah dalam proses mining selanjutnya maka dilakukan proses konversi kedalam bentuk yang dapat diolah dengan alat bantu data mining yang ada. Selanjutnya dengan menggunakan alat bantu Microsoft Excel yang berguna untuk membantu proses perhitungan Algoritma Naïve Bayes. Sehingga

dapat menentukan kelas (prediksi) dari sebuah record data baru dengan berdasar data training tersebut[6].

d. Analisa

1. Analisa permasalahan.

Berkaitan dengan mengidentifikasi kebutuhan dalam suatu penelitian. Akan dijelaskan mengenai permasalahan yang terjadi dan bagaimana tahap penyelesaian masalah tersebut.

2. Analisa kebutuhan data

Akan di jelaskan mengenai data-data apa saja yang diperlukan dalam pembuatan sistem ini.

3. Analisa metode naïve bayes

Teorema Bayes memiliki bentuk umum sebagai berikut [4]:

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

Keterangan rumus:

X = Data dengan class yang belum diketahui

C_i = Hipotesis data X merupakan suatu class spesifik

P(C_i|X) = Probabilitas hipotesis C_i berdasarkan kondisi x (posteriori prob.)

P(C_i) = Probabilitas hipotesis C_i (prior prob.)

P(X|C_i) = Probabilitas X berdasarkan kondisi tersebut

P(X) = Probabilitas dari X

Untuk menentukan atau memprediksi pilihan kelas, digunakan peluang maksimal dari seluruh kelas dengan fungsi:

$$\text{argmax } C_i = \frac{P(X|C_i) P(C_i)}{P(X)}$$

Karena nilai P(X) konstan untuk semua kelas, maka P(X) dapat diabaikan sehingga menghasilkan fungsi:

$$F_c(X) = \text{argmax } C_i = P(X|C_i) P(C_i)$$

Keterangan rumus:

F_c(X) = output hasil klasifikasi naive bayes argmax

C_i = nilai mak dari kelas

P(C_i) = peluang dari kategori yang diberikan

P(X|C_i) = kategori pada suatu class.

Setelah kedua peluang diatas dicari nilainya masing-masing, maka kemudian dua nilai peluang tersebut dipilih yang mempunyai nilai yang lebih besar menjadi kesimpulan kelas yang dicari dari data yang diprediksi[7].

Adapun tahapan algoritma Naive bayes dalam sistem adalah:

- a. Mulai
- b. Baca data training
 1. Hitung $P(C_i)$ untuk setiap kelas
 2. Hitung $P(X|C_i)$ untuk setiap kriteria dan setiap kelas
 3. Cari $P(X|C_i)$ yang paling besar menjadi kesimpulan
- c. Tampilkan hasil prediksi
- d. Selesai

e. Perancangan Model

Model dikembangkan dengan menggunakan Tools Microsoft Excel yang berguna untuk melakukan proses perhitungan Algoritma Naïve Bayes dalam melakukan sistem prediksi lama studi di ITB AAS Indonesia[8].

f. Implementasi dan Pengujian

1. Implementasi

Pada proses implementasi ini akan dilakukan pembuatan modul-modul yang telah dirancang dalam tahap perancangan kedalam bahasa pemrograman. Implementasi sistem akan dilakukan dengan spesifikasi sebagai berikut:

Perangkat Lunak yang digunakan sebagai berikut :

- Sistem operasi : Windows 7 Ultimate
- Bahasa pemrograman : PHP versi 5.3.8
- Tools perancangan : Notepad ++ versi 5.8.2
- Web server : Apache Web Server Version 2.2.21
- Database : MySQL Database Version 5.0.8

Perangkat keras yang akan digunakan dalam pembuatan dan penerapan aplikasi ini:

- Processor : Intel Pentium
- Memori (RAM) : 2.00 GB

2. Pengujian

Pengujian merupakan tahapan dimana aplikasi akan dijalankan, tahap pengujian diperlukan untuk menjadi ukuran bahwa sistem dapat dijalankan sesuai dengan tujuan.

Tahap pengujiannya yaitu:

a. Pengujian black box, pengujian ini berfokus pada tingkah laku aplikasi yang telah dirancang. Pengujian ini bertujuan untuk memastikan sistem yang telah dibangun sesuai dengan analisa dan perancangan.

b. Pengujian tingkat akurasi sistem yang mana terdiri dari tingkat akurasi sistem dalam data training dan tingkat akurasi sistem di luar data training. Akurasi diukur menggunakan rumus akurasi dan laju eror di bawah ini [9].

$$\text{Akurasi} = \frac{\text{jumlah data yang diprediksi secara benar}}{\text{Jumlah data yang diprediksi}} \times 100\%$$

g. Kesimpulan dan Saran

Dalam tahapan ini dapat ditentukan kesimpulan terhadap hasil pengujian yang telah dilakukan untuk mengetahui apakah implementasi sistem yang telah dilakukan dapat beroperasi dengan baik dan sesuai dengan tujuan yang diinginkan serta memberikan saran-saran untuk penyempurnaan dan pengembangan penelitian selanjutnya[10].

3. HASIL DAN PEMBAHASAN

Pada bab ini akan dijelaskan hasil dari analisis masalah, analisis data, hasil analisis metode klasifikasi Naïve Bayes untuk prediksi kelulusan mahasiswa, dan perancangan sistem serta hasil implementasi, terakhir pengujian dan analisis hasil pengujian.

3.1 Analisis Masalah

Setiap perguruan tinggi baik negeri maupun swasta tentunya memiliki kebijakan kurikulum yang berbeda. Di ITB AAS Indonesia khususnya pada 3 program studi tingkat kelulusan masih terbilang rendah hal ini disebabkan oleh beberapa faktor, yang diantaranya: kurangnya strategi mahasiswa dalam belajar, sistem pengajaran yang kurang baik dan adanya perbedaan kemampuan setiap mahasiswa. Dari beberapa faktor tersebut menyebabkan banyak mahasiswa yang menempuh masa studi lebih dari lima tahun.

Untuk mengatasi permasalahan tersebut maka penulis membuat solusi mengimplementasikan data mining untuk memprediksi tingkat kelulusan mahasiswa. Dengan memanfaatkan sistem data mining khususnya algoritma klasifikasi Naïve Bayes dibuat sebuah perancangan aplikasi untuk memprediksi masa studi mahasiswa berdasarkan data pribadi mahasiswa dan data nilai (IPK) mahasiswa, dengan tiga kategori yaitu cepat, sedang dan lambat. Dengan adanya aplikasi ini diharapkan dapat dimanfaatkan dosen dan staff jurusan untuk memprediksi kelulusan mahasiswa dalam masa studinya.

3.2. Analisis Data

Data yang digunakan dalam penelitian ini terdiri dari beberapa sumber data, yaitu data asli (master) Mahasiswa, data mata kuliah, dan data Kelulusan. Setelah seluruh data terkumpul, selanjutnya dilakukan proses tahapan data mining yaitu:

1. Pembersihan data (cleaning).

Setelah data di peroleh, maka akan ditemukan data-data yang tidak lengkap. Untuk itu diperlukan tahapan pembersihan data yaitu membuang data-data yang tidak lengkap.

2. Integrasi Data (data integration).

Setelah data di Cleaning, maka proses selanjutnya adalah integrasi data yaitu penggabungan data asli (master) Mahasiswa, data mata kuliah, dan data Kelulusan.

3. Seleksi Data.

Proses selanjutnya adalah seleksi data, yaitu menyeleksi data-data yang akan digunakan. Data yang digunakan adalah No, Jurusan, Gender, Asal Sekolah, Rerata SKS dan Asisten. Untuk Label yaitu Tepat dan Terlambat. Hasilnya dapat dilihat pada tabel 1.

Langkah Perhitungan Algoritma Naïve Bayes

- a. Menentukan Dataset
- b. Menghitung Query **Label** Tepat/Terlambat)
- c. Menghitung Query Variabel **Jurusan** berdasarkan kelas Tepat/Terlambat
- d. Menghitung Query Variabel **Gender** berdasarkan kelas Tepat/Terlambat
- e. Menghitung Query Variabel **Asal Sekolah** berdasarkan kelas Tepat/Terlambat
- f. Menghitung Query Variabel Rata SKS berdasarkan kelas Tepat/Terlambat
- g. Menghitung Query Variabel Asisten dosen berdasarkan kelas Tepat/Terlambat
- h. Melakukan Uji Prediksi
- i. Evaluasi Sistem Model
- j. Uji Model RapidMiner

Berdasarkan Tahapan Implementasi dari Algoritma Naïve Bayes

1. Menentukan Dataset

Dataset yang dipergunakan dalam penelitian ini hanya menggunakan data sampel sebanyak 13 data training seperti ditunjukkan pada tabel 1.

Tabel 1. Hasil Seleksi Data

NO	JURUSAN	GENDER	ASAL_SEKOLAH	RERATA_SKS	ASISTEN	LAMA_STUDI
1	AKUNTANSI	PRIA	SURAKARTA	SKS > 18	TIDAK	TERLAMBAT
2	AKUNTANSI	PRIA	SURAKARTA	SKS > 18	TIDAK	TERLAMBAT
3	AKUNTANSI	PRIA	SURAKARTA	SKS > 18	TIDAK	TERLAMBAT
4	EKONOMI	WANITA	SURAKARTA	SKS > 18	TIDAK	TERLAMBAT
5	EKONOMI	WANITA	LUAR	SKS > 18	TIDAK	TERLAMBAT
6	EKONOMI	PRIA	LUAR	SKS > 18	TIDAK	TERLAMBAT
7	EKONOMI	PRIA	SURAKARTA	SKS > 18	YA	TEPAT
8	AKUNTANSI	PRIA	SURAKARTA	SKS <= 18	TIDAK	TERLAMBAT
9	AKUNTANSI	WANITA	SURAKARTA	SKS <= 18	TIDAK	TERLAMBAT
10	EKONOMI	PRIA	SURAKARTA	SKS > 18	YA	TEPAT
11	LAIN	PRIA	SURAKARTA	SKS > 18	TIDAK	TERLAMBAT
12	EKONOMI	PRIA	LUAR	SKS <= 18	TIDAK	TERLAMBAT
13	EKONOMI	WANITA	SURAKARTA	SKS <= 18	TIDAK	TEPAT

2. Menghitung Query **Label** (Tepat/Terlambat)

Dalam menentukan perhitungan nilai label Tepat/Terlambat menggunakan Rumus di Microsoft Excel

$$\text{Tepat} = \text{COUNTIF}(H4:H16;K5)/13 = 23\%$$

$$\text{Terlambat} = =\text{COUNTIF}(H4:H16;L5)/13 = 77\%$$

3. Menghitung Query Variabel **Jurusan** berdasarkan kelas Tepat/Terlambat

Hasil Perhitungan nilai Tepat/Terlambat pada Variabel Jurusan dapat dilihat pada tabel 2.

Tabel 2. Query Variabel Jurusan

P(J=↓ ...	TEPAT	TERLAMBAT
EKONOMI	100%	40%
AKUNTANSI	0%	50%
LAIN	0%	10%

Rumus Perhitungan

Tepat = COUNTIFS(\$C\$4:\$C\$16;\$J6;\$H\$4:\$H\$16;K\$5)/COUNTIF(\$H\$4:\$H\$16;K\$5)

Terlambat = COUNTIFS(\$C\$4:\$C\$16;\$J6;\$H\$4:\$H\$16;L\$5)/COUNTIF(\$H\$4:\$H\$16;L\$5)

4. Menghitung Query Variabel **Gender** berdasarkan kelas Tepat/Terlambat
Hasil Perhitungan nilai Tepat/Terlambat pada Variabel Jurusan dapat dilihat pada tabel 3.

Tabel 3. Query Variabel Gender

P(G=↓ ...	TEPAT	TERLAMBAT
PRIA	67%	70%
WANITA	33%	30%

Rumus Perhitungan

Tepat = COUNTIFS(\$D\$4:\$D\$16;\$J11;\$H\$4:\$H\$16;K\$5)/COUNTIF(\$H\$4:\$H\$16;K\$5)

Terlambat = COUNTIFS(\$D\$4:\$D\$16;\$J11;\$H\$4:\$H\$16;L\$5)/COUNTIF(\$H\$4:\$H\$16;L\$5)

5. Menghitung Query Variabel **Asal Sekolah** berdasarkan kelas Tepat/Terlambat
Hasil Perhitungan nilai Tepat/Terlambat pada Variabel Asal Sekolah dapat dilihat pada tabel 4.

Tabel 4. Query Variabel Asal Sekolah

P(A=↓ ...	TEPAT	TERLAMBAT
SURAKARTA	100%	70%
LUAR	0%	30%

Rumus Perhitungan

Tepat = COUNTIFS(\$E\$4:\$E\$16;\$J15;\$H\$4:\$H\$16;K\$5)/COUNTIF(\$H\$4:\$H\$16;K\$5)

Terlambat = COUNTIFS(\$E\$4:\$E\$16;\$J15;\$H\$4:\$H\$16;L\$5)/COUNTIF(\$H\$4:\$H\$16;L\$5)

6. Menghitung Query Variabel Rata SKS berdasarkan kelas Tepat/Terlambat
Hasil Perhitungan nilai Tepat/Terlambat pada Variabel Rerata SKS dapat dilihat pada tabel 5.

Tabel 5. Query Variabel Rerata SKS

P(R=↓ ...	TEPAT	TERLAMBAT
SKS > 18	67%	70%
SKS <= 18	33%	30%

Rumus Perhitungan

Tepat= COUNTIFS(\$F\$4:\$F\$16;\$N6;\$H\$4:\$H\$16;O\$5)/COUNTIF(\$H\$4:\$H\$16;K\$5)

Terlambat = COUNTIFS(\$F\$4:\$F\$16;\$N6;\$H\$4:\$H\$16;P\$5)/COUNTIF(\$H\$4:\$H\$16;L\$5)

- Menghitunga Query Variabel Asisten dosen berdasarkan kelas Tepat/Terlambat Hasil Perhitungan nilai Tepat/Terlambat pada Variabel Asal Sekolah dapat dilihat pada tabel 6.

Tabel 6. Query Variabel Asal Sekolah

P(ASI=↓ ...	TEPAT	TERLAMBAT
YA	67%	0%
TIDAK	33%	100%

Rumus Perhitungan

Tepat = COUNTIFS(\$G\$4:\$G\$16;\$N11;\$H\$4:\$H\$16;O\$5)/COUNTIF(\$H\$4:\$H\$16;K\$5)

Terlambat = COUNTIFS(\$G\$4:\$G\$16;\$N11;\$H\$4:\$H\$16;P\$5)/COUNTIF(\$H\$4:\$H\$16;L\$5)

- Melakukan Uji Prediksi
Dalam melakukan Uji Prediksi disini peneliti menggunakan 5 data testing secara acak seperti ditunjukkan pada tabel 7.

Tabel 7. Uji Prediksi

NO	JURUSAN	GENDER	ASAL_SEKOLAH	RERATA_SKS	ASISTEN	LAMA_STUDI DATA RIIL
1	IPS	WANITA	SURAKARTA	SKS <= 18	TIDAK	TERLAMBAT
2	IPA	PRIA	SURAKARTA	SKS > 18	YA	TEPAT
3	LAIN	PRIA	SURAKARTA	SKS > 18	TIDAK	TERLAMBAT
4	IPA	PRIA	LUAR	SKS <= 18	TIDAK	TERLAMBAT
5	IPA	WANITA	SURAKARTA	SKS <= 18	TIDAK	TEPAT

Berdasarkan hasil perhitungan dengan Algoritma Naïve Bayes maka diperoleh hasil prediksi seperti ditunjukkan pada tabel 8.

Tabel 8. Hasil Prediksi Metode Naïve Bayes

NO	JURUSAN	GENDER	ASAL_SEKOLAH	RERATA_SKS	ASISTEN	LAMA_STUDI PREDIKSI
1	IPS	WANITA	SURAKARTA	SKS <= 18	TIDAK	TERLAMBAT
2	IPA	PRIA	SURAKARTA	SKS > 18	YA	TEPAT
3	LAIN	PRIA	SURAKARTA	SKS > 18	TIDAK	TERLAMBAT
4	IPA	PRIA	LUAR	SKS <= 18	TIDAK	TERLAMBAT
5	IPA	WANITA	SURAKARTA	SKS <= 18	TIDAK	TERLAMBAT

- Evaluasi Sistem Model

Tahap terakhir yakni melakukan uji evaluasi sistem dengan membandingkan antara hasil label data riil dengan hasil label data prediksi. Adapun proses perbandingnya dapat dilihat pada tabel 9.

Tabel 9. Perbandingan Uji Testing Algoritma Naïve Bayes

NO	LAMA STUDI	Prediction	CLASS PREDICTION	TEPAT	TERLAMBAT
1	TERLAMBAT	Prediction	TERLAMBAT	0,0%	2,4%
2	TEPAT		TEPAT	6,8%	0,0%
3	TERLAMBAT		TERLAMBAT	0,0%	2,6%
4	TERLAMBAT		TERLAMBAT	0,0%	1,9%
5	TEPAT		TERLAMBAT	0,9%	1,9%

Perhitungan Nilai Akurasi Model

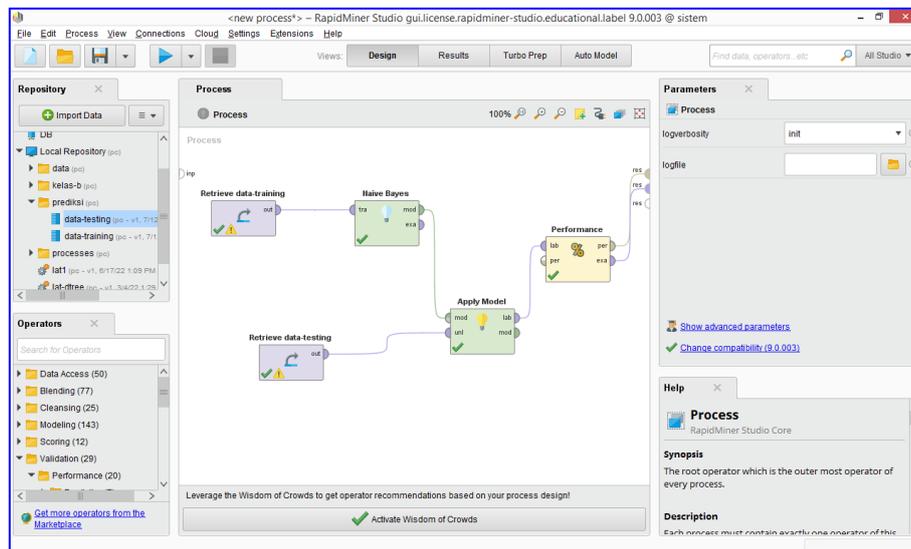
Dalam melakukan uji akurasi model dilakukan dengan metode Confusion Matrik. Adapun hasil rekap perhitungannya dapat dilihat pada tabel 10.

Tabel 10. Hasil Rekap Uji Confusion Matrik

Confusion Matrik	CLASS		AKURASI	
	Predicted ↓	TEPAT		TERLAMBAT
	TEPAT	1		0
TERLAMBAT	1	3		

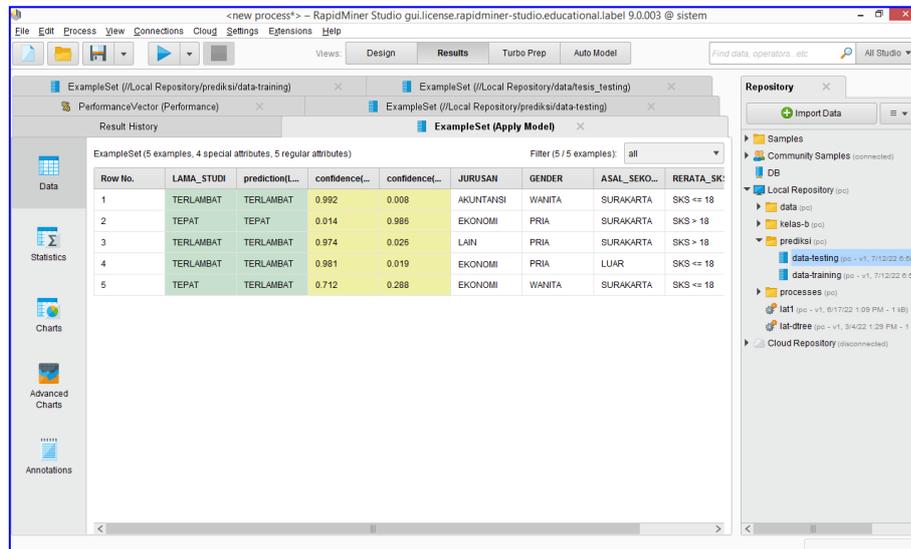
10. Uji Model RapidMiner

Dalam Implementasi pada pengujian sistem, peneliti juga menggunakan Tools RapidMiner yang berguna untuk melihat desain model dan nilai akurasi model yang telah dibuat. Adapun desain model implementasi sistem prediksi dapat dilihat pada gambar 1.



Gambar 1. Uji Evaluasi Model Prediksi

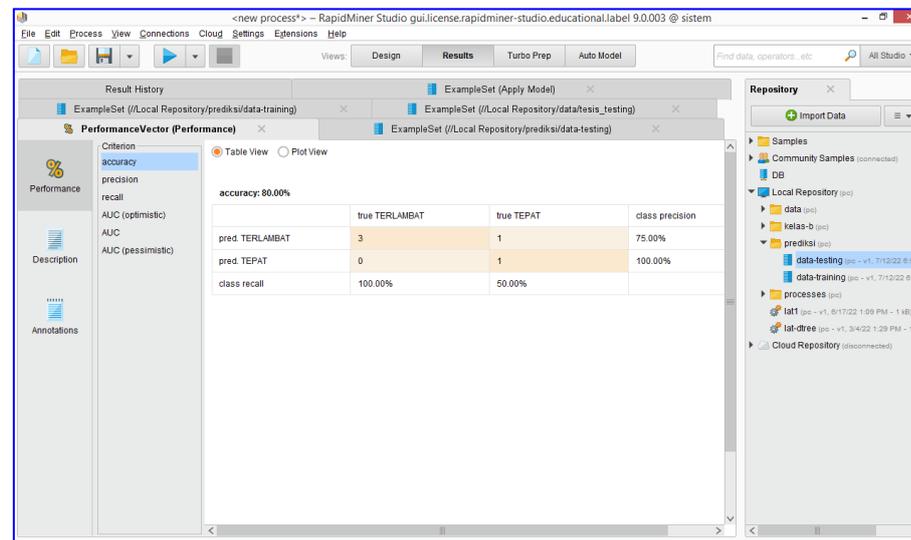
Dalam Implementasi Pada RapidMiner dilakukan bertujuan untuk menguji model yang telah dibuat, serta untuk melihat nilai prediksi dan akurasi model. Berdasarkan hasil penerapan dari rapidminer maka diperoleh nilai prediksi yang sama dengan perhitungan manual di Microsoft Excel seperti ditunjukkan pada gambar 2.



Row No.	LAMA_STUDI	prediction(...)	confidence(...)	confidence(...)	JURUSAN	GENDER	ASAL_SEKO...	RERATA_SK...
1	TERLAMBAT	TERLAMBAT	0.992	0.008	AKUNTANSI	WANITA	SURAKARTA	SKS > 18
2	TEPAT	TEPAT	0.014	0.986	EKONOMI	PRIA	SURAKARTA	SKS > 18
3	TERLAMBAT	TERLAMBAT	0.974	0.026	LAIN	PRIA	SURAKARTA	SKS > 18
4	TERLAMBAT	TERLAMBAT	0.981	0.019	EKONOMI	PRIA	LUAR	SKS <= 18
5	TEPAT	TERLAMBAT	0.712	0.288	EKONOMI	WANITA	SURAKARTA	SKS <= 18

Gambar 2. Hasil Prediksi Pada Tools RapidMiner

Hasil Prediksi dari Implemtasi dari menggunakan Tools Rapid Miner juga telah diperoleh hasil yang sama sebesar 80%, seperti ditunjukkan pada gambar 3.



Criterion	accuracy	precision	recall
AUC (optimistic)			
AUC			
AUC (pessimistic)			

	true TERLAMBAT	true TEPAT	class precision
pred TERLAMBAT	3	1	75.00%
pred TEPAT	0	1	100.00%
class recall	100.00%	50.00%	

Gambar 3. Hasil Akurasi Pada Tools RapidMiner

4. KESIMPULAN

Penyebaran virus corona yang semakin luas menyebabkan proses pembelajaran diperkuliahan harus dilakukan secara online daring. Peran teknolog sistem informasi menjadi hal utama dalam mesolusi proses pembelajaran secara online. Berdasarkan hasil penelitian yang telah dilakukan selama penelitian tentang Analisis Peran Teknologi Sistem Informasi dalam Pembelajaran Kuliah dimasa Pandemi Virus Corona menunjukkan media online yang digunakan dalam kuliah daring seperti zoom, edlink, skype, Google Classroom, Google Meet, Whatshapp. Hasil penelitian menunjukkan bahwa Media Google Classroom paling banyak diminati yaitu 55,9% sebagai share materi dan tugas. Sedangkan untuk media Video Conferences yang paling banyak diminati yaitu Google Meet : 70,6%. Hasil penilaian peran teknologi sistem informasi berada di point 4 (Baik) sebanyak : 44,1 %. Hal ini menunjukkan bahwa peran teknologi sistem informasi berupa Google Classroom & Google Meet menjadi solusi utama dalam proses pembelajaran perkuliahan secara daring.

Berdasarkan hasil dan analisis sistem prediksi tingkat kelulusan mahasiswa menggunakan metode data mining dengan algoritma naïve bayes di 3Jurusan ITB AAS Indonesia ini, dapat diambil kesimpulan sebagai berikut:

1. Sistem prediksi tingkat kelulusan mahasiswa menggunakan metode data mining dengan algoritma naïve bayes ini berhasil dirancang dan diimplementasikan dapat digunakan untuk memprediksi dari masa studi atau ketepatan masa studi dari mahasiswa dengan data training dan data testing yang telah diperoleh.
2. Berdasarkan proses perhitungan uji medel dengan menggunakan metode Confusion Matrik pada 13 data training dan 5 data testing telah diperoleh hasil Akurasi 80%, recall 50% dan presisi 100%.
3. Model yang telah dikembangkan ini dapat membantu bagian pusat unit informasi dalam memberikan informasi perkiraan mahasiswa akan tepat atau terlambat kuliahnya, sehinggga kampus dapat lebih cepat dalam menangani pada sistem yang diprediksi akan terlambat.

5. SARAN

Penelitian ini dilakukan berdasarkan hasil data sampel dari ITB AAS Indonesia Surakarta serta hanya di implementasikan Microsoft Excel dan Tools RapidMiner serta belum dikembangkan kedalam suatu bentuk Aplikasi. Maka bagi peneliti selanjutnya dapat mengembangkan penelitini degnan membuat aplikasi sistem prediksi sehingga hal tersebut dapat memberikan kontribusi lebih bagi kampus dan perguruan teinggi yang lainnya.

CAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada semua pihak yang telah ikut serta dalam membantu proses penelitian ini dengan judul Sistem Prediksi Lama Studi Kuliah Menggunakan Metode Naive Bayes. Semoga penelitian ini dapat bermanfaat bagi pembaca dan peneltian selanjutnya untuk mengembangkan penelitian ini lebih luas.

DAFTAR PUSTAKA

- [1] Huda, N.M. 2010. "Aplikasi data mining untuk menampilkan tingkat kelulusan mahasiswa dengan studi kasus FMIPA Universitas Diponegoro", Skripsi, Program Studi Teknik Informatika Jurusan MIPA UNDIP, Semarang.
- [2] Meinanda, M.H., Anisa M .Muhandri, N.Suryadi, K. 2009. "Prediksi masa studi sarjana dengan artificial neural network", Internet working Indonesia Journal, Vol.1No.2, pp. 31-35.
- [3] Santosa, Budi, 2007, "Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis", Graha Ilmu, Yogyakarta.
- [4] Kusrini, Luthfi, E.T. 2009. "Algoritma Data Mining", Andi Offset. Surabaya.
- [5] Kusrini, K., Luthfi, E. T., Muqorobin, M., & Abdullah, R. W. (2019, November). Comparison of Naive Bayes and K-NN Method on Tuition Fee Payment Overdue Prediction. In 2019 4th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE) (pp. 125-130). IEEE.
- [6] Han,J., Kamber, M., & Pei, J. 2012. Data Mining Concepts and Techniques (3rd ed.). Morgan Kaufmann Publishers.
- [7] Muqorobin, M., Kusrini, K., & Luthfi, E. T. (2019). Optimasi Metode Naive Bayes Dengan Feature Selection Information Gain Untuk Prediksi Keterlambatan Pembayaran Spp Sekolah. Jurnal Ilmiah SINUS, 17(1), 1-14.
- [8] Moonallika, P. S. C., Fredlina, K. Q., & Sudiatmika, I. K. (2020). Penerapan Data Mining Untuk Memprediksi Kelulusan Mahasiswa Menggunakan Algoritma Naive Bayes Classifier (Studi Kasus STMIK Primakara). Progresif: Jurnal Ilmiah Komputer, 16(1), 47-56.
- [9] Salmu, S., & Solichin, A. (2017, April). Prediksi Tingkat Kelulusan Mahasiswa Tepat Waktu Menggunakan Naive Bayes: Studi Kasus UIN Syarif Hidayatullah Jakarta. In Prosiding Seminar Nasional Multidisiplin Ilmu (Vol. 22).
- [10] Dwiwana, D., & Muqorobin, M. (2021). Analysis of Adi Soemarmo Solo Airport Parking Payment System. International Journal of Computer and Information System (IJCIS), 2(1), 1-3.