

ANALISIS KETERLAMBATAN PEMBAYARAN UKT MAHASISWA MENGGUNAKAN METODE KLASIFIKASI NAÏVE BAYES DALAM PENDEKATAN *DATA SCIENCE*

I Gusti Ngurah Nyoman Bagiarta¹, Gusti Ngurah Mega Nata², I Nyoman Suraja Antaraja³

Sistem Informasi^{1,3} Manajemen Informatika²

Institut Teknologi dan Bisnis STIKOM Bali^{1, 2, 3}

mega@stikom-bali.ac.id^{1, 2, 3}

Abstrak

Biaya pendidikan mahasiswa atau Uang Kuliah Tunggal (UKT) merupakan komponen penting dalam keberlangsungan operasional perguruan tinggi, khususnya di Perguruan Tinggi Swasta (PTS). Namun, permasalahan keterlambatan pembayaran UKT masih sering terjadi, bahkan pada mahasiswa yang tidak tergolong dalam kategori prasejahtera. Permasalahan ini tidak hanya dipengaruhi oleh faktor kondisi keuangan mahasiswa, tetapi juga kemungkinan faktor lain seperti tingkat keseriusan dalam menjalani perkuliahan, kemampuan akademik mahasiswa, serta potensi penyalahgunaan dana UKT untuk keperluan di luar pendidikan. Oleh karena itu, diperlukan analisis lebih lanjut untuk memahami keterkaitan antara keterlambatan pembayaran UKT dengan karakteristik akademik mahasiswa. Penelitian ini bertujuan untuk menganalisis hubungan antara keterlambatan pembayaran UKT dengan data akademik mahasiswa melalui pendekatan *data mining*, khususnya dengan membangun model klasifikasi. Data yang digunakan dalam penelitian ini merupakan gabungan antara data transaksi pembayaran UKT dan data akademik mahasiswa reguler aktif. Variabel yang digunakan meliputi waktu perkuliahan, program studi, jumlah SKS kumulatif, Indeks Prestasi Semester (IPS), Indeks Prestasi Kumulatif (IPK), serta tingkat kehadiran mahasiswa pada setiap semester. Label klasifikasi yang digunakan adalah status pembayaran UKT, yaitu “*Lancar*” dan “*Menunda*”. Model klasifikasi dibangun menggunakan algoritma *Naïve Bayes Classifier* dan dievaluasi menggunakan *confusion matrix* untuk mengukur tingkat kinerja model. Dataset yang digunakan terdiri dari 12.935 data rekaman mahasiswa. Hasil penelitian menunjukkan bahwa model klasifikasi yang dibangun mampu memprediksi potensi keterlambatan pembayaran UKT dengan tingkat akurasi sebesar 59%. Tingkat akurasi tersebut menunjukkan adanya keterkaitan antara data akademik mahasiswa dengan perilaku pembayaran UKT, meskipun masih diperlukan pengembangan model lebih lanjut serta penambahan variabel lain untuk meningkatkan performa prediksi.

Kata kunci : Perguruan tinggi swasta, UKT, *Naïve Bayes*, *data science*

Abstract

Student education costs, or the Single Tuition Fee (UKT), are a crucial component in the operational continuity of higher education institutions, particularly in Private Universities (PTS). However, the phenomenon of late UKT payments still frequently occurs, even among students who are not classified as underprivileged. This problem is not only influenced by the financial condition of students, but also other possible factors such as the level of seriousness in studying, student academic ability, and the potential for misuse of UKT funds for non-educational purposes. Therefore, further analysis is needed to understand the relationship between late UKT payments and student academic characteristics. This study aims to analyze the relationship between late UKT payments and student academic data through a data mining approach, specifically by building a classification model. The data used in this study is a combination of UKT payment transaction data and academic data of active regular students. The variables used include lecture schedules, study programs, cumulative number of credits, Semester Grade Point Average (IPS), Cumulative Grade Point Average (IPK), and student attendance levels in each semester. The classification label used is the UKT payment status, namely Current and Postponed. The classification model was built using the Naïve Bayes Classifier algorithm and evaluated using a confusion matrix to measure the model's performance. The dataset used consisted of 12,935 student records. The results showed that the classification model was able to predict potential late tuition fees (UKT) payments with an accuracy of 59%. This accuracy level indicates a link between student academic data and UKT payment behavior, although further model development and the addition of other variables are needed to improve prediction performance.

Keywords : Tuition Delay, Naïve Bayes, academic performance, predictive model, data science.

I. PENDAHULUAN

Sumber pendanaan utama bagi perguruan tinggi swasta adalah pembayaran Uang Kuliah Tunggal (UKT) dari mahasiswa [1]. Pembayaran UKT dilakukan pada awal semester dan wajib diselesaikan sebelum batas waktu yang biasanya ditetapkan misal, sebelum pelaksanaan Ujian Tengah Semester (UTS). Namun, dalam praktiknya, banyak mahasiswa yang menunggak pembayaran hingga akhir semester. Permasalahan ini umumnya disebabkan oleh kondisi finansial mahasiswa. Namun, Perguruan tinggi kesulitan menilai kebenaran kondisi ekonomi mahasiswa. Berdasarkan data pendaftaran saat menjadi mahasiswa baru, dimana tidak sedikit mahasiswa yang menunda pembayaran semester adalah mahasiswa yang awalnya mendaftar lewat jalur ekonomi mampu. Diperguruan tinggi umumnya bagi calon mahasiswa yang tidak mampu akan diberikan beasiswa [2]. Jadi, Apakah mahasiswa kesulitan finansial atau faktor akademik lainnya yang membuat mahasiswa tidak melakukan pembayaran menjadi pertanyaan. Oleh karena itu, faktor keuangan tidak serta merta menjadi satu-satunya dasar dalam menganalisis keterlambatan pembayaran. Selain faktor ekonomi, terdapat pula faktor lain yang dapat menjadi penyebab keterlambatan pembayaran UKT. Faktor-faktor tersebut

dapat dikategorikan sebagai *outlier factors* [3], seperti kurangnya fokus mahasiswa dalam melanjutkan studi, atau perilaku tidak bertanggung jawab, misalnya menggunakan dana UKT yang telah diberikan oleh orang tua untuk keperluan pribadi lainnya [3].

Penelitian yang dilakukan bertujuan untuk mengetahui hubungan antara keterlambatan pembayaran UKT dengan data akademik dan aktivitas perkuliahan mahasiswa. Pendekatan yang digunakan yaitu dengan menganalisis data (*data science*), pada data pembayaran UKT dan data aktifisasi akademik. Metode analisis yang digunakan yaitu klasifikasi data berbasis *machine learning*. Atribut data yang digunakan dari tabel akademik yaitu Indeks Prestasi Kumulatif (IPK), Indeks Prestasi Semester (IPS), status pekerjaan, waktu kuliah yang dipilih, program studi, rata-rata kehadiran, serta jumlah SKS yang telah ditempuh. Atribut-atribut tersebut dijadikan dasar dalam membangun model prediksi keterlambatan pembayaran. Manfaat dari penelitian ini adalah memberikan informasi awal kepada pihak keuangan perguruan tinggi mengenai potensi mahasiswa yang berisiko menunggak pembayaran UKT, sehingga dapat dilakukan tindakan atau pendekatan yang tepat sesuai dengan permasalahan masing-masing mahasiswa. Data yang digunakan dalam penelitian ini adalah data mahasiswa reguler aktif dari salah satu perguruan tinggi swasta di Bali, mencakup lima program studi selama empat semester. Data akan diproses menggunakan algoritma *machine learning* yaitu Naïve Bayes Classifier [4], dan akan dibagi menjadi data pelatihan (*data training*) sebesar 60% dan data pengujian (*data testing*) sebesar 40%. Hasil utama dari penelitian ini adalah model klasifikasi yang dapat memprediksi kemungkinan keterlambatan pembayaran UKT oleh mahasiswa. Model ini akan dievaluasi berdasarkan tingkat akurasi dan reliabilitasnya menggunakan metode *confusion matrix* [5]. Penelitian terkait penggunaan algoritma Naïve Bayes dalam analisis data perguruan tinggi sebelumnya telah dilakukan pada beberapa studi, seperti dalam [6] dan [7]. Namun, penelitian-penelitian tersebut umumnya masih berfokus pada aspek akademik semata. Belum ditemukan penelitian yang secara khusus membahas hubungan antara data akademik mahasiswa dan keterlambatan pembayaran UKT. Penelitian pada paper [3] hanya memvisualisasikan transaksi pembayaran untuk menemukan *outlier*, tanpa mengkaji keterkaitan dengan data akademik. Oleh karena itu, penelitian ini penting dilakukan untuk mengisi celah tersebut dan memberikan kontribusi terhadap pengelolaan keuangan di perguruan tinggi swasta.

II. TINJAUAN PUSTAKA

1. Data Pembayaran Perguruan Tinggi

Data bersumber dari salah satu perguruan tinggi swasta. Bagian data yang diambil bersumber dari data loket pembayaran dan akademik. Data pembayaran yang digunakan berupa rekam pembayaran per mahasiswa per semester, jika pembayaran dilunasi lebih dari tanggal jatuh tempo maka dianggap mahasiswa terlambat pembayarannya. Keterangan pembayaran terlambat atau tidak terlambat digunakan sebagai label pembayaran mahasiswa pada data training. Data pembayaran Karakteristik data yang digunakan yaitu berupa data terstruktur. Data yang diolah berbentuk tabel yang disimpan dengan DBMS SQL Server. Pemahaman data (*data understanding*) didapat dari peraturan pembayaran, *manual book* dan diskusi langsung dengan petugas pengolahan data. Data yang digunakan merupakan data tiga angkatan mahasiswa reguler. Identitas dari setiap mahasiswa tidak digunakan untuk menjaga kerahasiaan data pribadi.

2. Algoritma Naïve Bayes

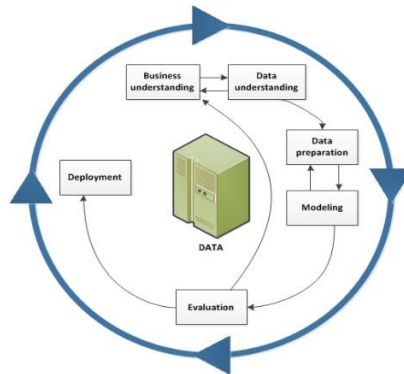
Algoritma Naïve Bayes adalah algoritma *Classifier* yang mengklasifikasikan probabilitas sederhana yang mengaplikasikan teorema bayes [5]. Tujuan dari algoritma Naïve Bayes yaitu untuk membuat komputer dapat memprediksi sesuatu berdasarkan data yang telah dikumpulkan dalam transaksi atau data entry. Algoritma Naïve Bayes dapat membangun model dari data kuantitatif atau data kategorikal. Jumlah data yang diperlukan untuk membangun model sejumlah kecil data training. Kelebihan dari algoritma Naïve Bayes yaitu relatif mudah untuk diimplementasikan karena tidak menggunakan optimasi numerik [4]. Algoritma ini menganggap semua faktor / kategori tidak saling berhubungan. Meskipun hubungan antar faktor / kategori tidak diperhitungkan namun hasil prediksi algoritma Naïve Bayes cukup akurat dan cepat. Cara kerja algoritma Naïve Bayes yang diterapkan untuk prediksi keterlambatan pembayaran UKT Mahasiswa yaitu pertama data keuangan dan data akademik digabungkan sebagai data training dan data testing. *History* data mahasiswa tersebut dihitung peluangnya dalam kelancaran pembayaran dan keterlambatan pembayaran UKT. Hasil perhitungan peluang tersebut adalah model prediksi yang digunakan untuk menentukan tingkat peluang lancar atau menunda pembayaran mahasiswa berikutnya.

3. Penelitian terkait

Pada paper [6] menyatakan pembiayaan pendidikan perguruan tinggi merupakan aspek penting bagi perguruan tinggi dan bagi mahasiswa. Namun, pada sistem pembayaran UKT pendidikan di Indonesia belum cukup bisa mengatasi persoalan pembiayaan UKT mahasiswa sehingga masih banyak mahasiswa yang belum bisa melakukan pembayaran. Sehingga perguruan tinggi akan kesulitan dalam pembiayaan operasionalnya. Dalam memprediksi keterlambatan pembayaran biaya kuliah (UKT) di perguruan tinggi seperti pada paper [1] telah menerapkan algoritma K-Nearest Neighbor. Data penelitian pada paper tersebut menggunakan sumber data dari keuangan dan juga data dari akademik. Data dari akademik sudah berisi pendapatan / jumlah pengasialan orang tua, tentunya data tersebut sudah memberikan gambaran kemampuan dari keluarga mahasiswa. Namun, data tersebut bersifat tetap padahal data pengasialan orang tua bisa saja berubah setiap tahunnya.

III. ANALISIS DAN PERANCANGAN

Metode pembangunan model menggunakan CRISP – DM. Tahapan – tahapan penelitian CRISP-DM terdiri dari enam bagian yaitu Pemahaman bisnis, pemahaman data yang dimiliki, persiapan data, Analisis Model, uji coba dan evaluasi, pengembangan aplikasi [7] [8]. Namun pada penelitian ini pengembangan aplikasi tidak dilakukan melainkan hanya menggunakan aplikasi weka. Jika digambarkan dalam sebuah *flowchart* maka seperti pada gambar berikut:



Gambar 1. Alur CRISP-DM (www.ibm.com)

1. Pemahaman Bisnis (*business understanding*)

Pembayaran UKT pada perguruan tinggi swasta yang menjadi objek penelitian dilakukan diawal semester, mahasiswa di ijinakan menunda pembayaran sampai 1 minggu sebelum UTS. Mahasiswa yang membayar lewat dari UTS sudah dikatakan terlambat dan dikenakan denda. Keterlambatan Pembayaran UKT selama ini tidak diketahui keterkaitannya dengan nilai dan kehadiran perkuliahan. Maka penelitian untuk mengetahui hubungan pembayaran dengan data akademik mahasiswa perlu dilakukan. Hasil dari pemodelan tersebut akan menjadi pengetahuan baru untuk membuat kebijakan.

2. Pemahaman Data (*data understanding*)

Data adalah hal penting dalam penelitian [9]. Karena data adalah salah satu faktor penting dalam keberhasilan peneliti maka dalam melaksanakan penelitian ini data yang akan digunakan bersumber dari data transaksi. Jumlah data yang akan digunakan yaitu keseluruhan data transaksi. Jenis data yang digunakan dalam penelitian ini adalah data yang bersumber dari data akademik dan data keuangan. Dari akademik yaitu data jumlah sks kumulatif, nilai Indeks Prestasi Semester (IPS) dan Indeks Prestasi Kumulatif (IPK) setiap semester yang mewakili tingkat kemampuan mahasiswa dalam bidang akademik. Keseriusan kuliah dari mahasiswa diambil dari data kehadiran mahasiswa yang diambil setiap satu semester, jika kehadiran mahasiswa dalam satu semester dibawah batas presensi kehadiran maka mahasiswa digolongkan “malas” kuliah sedangkan kebalikannya “rajin” kuliah. Data prodi yang digunakan yaitu 5 prodi, dimana kode 001 yaitu prodi Sistem Komputer, 002 manajemen informatika, 003 Sistem Informasi, 004 Teknologi Informasi, 005 Bisnis Digital. Data yang didapat kemudian digabung dalam satu tabel, Berikut adalah proses penggabungan data pada Query SQL.

	str_nim	bol_pagi_malam	str_jurusan	str_thn_ajaran	bol_semester	num_sks_kum	dec_ip_smt	dec_ip_kum	kehadiran
12920	210050002	0	005	2021/2022	0	18	3.83	3.83	Malas
12921	210050032	0	005	2021/2022	0	18	3.89	3.89	Malas
12922	210050034	0	005	2021/2022	0	18	3.89	3.89	Malas
12923	210050058	0	005	2021/2022	0	18	3.61	3.61	Malas
12924	210050078	0	005	2021/2022	0	18	3.78	3.78	Malas
12925	210050123	1	005	2021/2022	0	18	3.61	3.61	Malas
12926	210050138	1	005	2021/2022	0	4	2.50	2.50	Rajin
12927	210050150	0	005	2021/2022	0	18	3.44	3.44	Malas
12928	210050169	0	005	2021/2022	0	18	3.33	3.33	Malas
12929	210050177	0	005	2021/2022	0	18	3.67	3.67	Malas
12930	210050195	0	005	2021/2022	0	18	3.33	3.33	Rajin
12931	210050202	0	005	2021/2022	0	18	3.78	3.78	Malas
12932	210050203	1	005	2021/2022	0	34	2.75	2.41	Rajin
12933	210050212	0	005	2021/2022	0	18	0.00	0.00	Malas
12934	210050228	0	005	2021/2022	0	18	0.00	0.00	Malas
12935	210050229	0	005	2021/2022	0	18	0.00	0.00	Malas

Gambar 2. Penggabungan Data Training Dan Data Testing

Data yang dimiliki sudah berbentuk data terstruktur. Proses pengumpulan data dan penggabungan data dilakukan dengan bahasa SQL. Setiap mahasiswa akan dimasukkan dalam satu record per semesternya, sehingga akan terdapat 4 record untuk satu mahasiswa dalam 4 semester. Jumlah *record* yang digunakan yaitu 12.935

baris, jika dibagi 4 semester maka artinya terdapat 3.300 lebih mahasiswa aktif yang digunakan dalam analisis ini.

3. Persiapan data (*preprocessing data*)

Data preparation adalah proses mengubah data yang telah dimiliki menjadi data training dan data testing. Proses preparation terdiri dari pengapusan data yang tidak sesuai, mengubah data menjadi deskrit. Proses preparation dilakukan pada query sql dan aplikasi excel. Pada gambar 2 hasil query yang berupa tabel kemudian di sesuaikan dengan kebutuhan analisis dan algoritma yang digunakan. Algoritma Naïve Bayes merupakan algoritma yang probabilitas dengan nilai atribut diskrit [5]. Maka, data akademik dan data keuangan mahasiswa dideskritisasi. Data yang tidak digunakan dalam analisis dihapus, seperti NIM, tahun ajaran dan semester. sedangkan data yang masih berupa kode diubah menjadi data yang bisa dipahami, seperti pada kolom *bol_pagi_malam* dan *str_jurusan*. Berikut tabel penyesuaian isi kolom *bol_pagi_malam* dan *str_jurusan*

TABEL I
PENYESUAIAN KOLOM WAKTU KULIAH

Bol_pagi_malam	Penyesuaian
0	Pagi
1	Malam

Kolom *str_jurusan* pada gambar 2 merukan program studi yang dipilih oleh mahasiswa. Kolom yang berisi kode angka tersebut kemudian diubah menjadi hurup yang lebih mudah untuk dipahami. Berikut adalah penyesuaian isi kolom *str_jurusan* menjadi kolom *prodi*.

TABEL II
PENYESUAIAN KOLOM PROGRAM STUDI

Str_jurusan	Prodi	Keterangan
001	SK	Sistem Komputer
002	MI	Manajemen Informatika
003	SI	Sistem Informasi
004	TI	Teknologi Informasi
005	BD	Bisnis Digital

Setelah proses penyesuaian isi kolom *bol_pagi_malam* dan *str_jurusan* maka didapatlah tabel III berikut yang sudah mudah untuk dipahami. Tahap preparation berikutnya adalah mengubah nilai numerik menjadi ordinal untuk mengurangi dimensi data. Nilai numerik yang diubah menjadi ordinal yaitu nilai SKS, IPS dan nilai IPK.

TABEL III
DATA PREPARATION

NIM	Jadwal Kuliah	Prodi	SKS	IPS	IPK	Kehadiran
150010289	Malam	SK	143	3	3.09	Malas
150010289	Malam	SK	143	0	3.09	Malas
150030023	Pagi	SI	143	3.5	3.1	Rajin
150030023	Pagi	SI	145	4	3.11	Malas
150030030	Pagi	SI	147	1	3.72	Malas
150030030	Pagi	SI	147	0	3.72	Malas
150030031	Pagi	SI	143	2	2.74	Malas
150030031	Pagi	SI	131	2	2.62	Malas
150030031	Pagi	SI	139	1.67	2.64	Malas
150030329	Malam	SI	141	4	3.2	Malas
160010012	Pagi	SK	141	3.5	3.33	Rajin
160010046	Malam	SK	141	1	3.28	Malas

160010046	Malam	SK	141	1	3.28	Malas
160010046	Malam	SK	141	0	3.28	Malas
160010158	Pagi	SK	111	0	0.72	Malas
160010158	Pagi	SK	111	0.56	0.77	Malas
160010158	Pagi	SK	111	1.81	0.96	Malas
160010184	Malam	SK	151	0	2.91	Malas
160010202	Pagi	SK	141	0	3.17	Malas

Data yang didapat dari akademik dan data mahasiswa yang masih berupa angka kemudian di jadikan data diskrit agar mengurangi dimensi data. Seperti data IPS dan IPK yang berupa angka dikonversi menjadi peringkat akademik seperti tabel berikut.

TABEL IV
PREDIKAT AKADEMIK

INDEKS PRESTASI	PREDIKAT
4,00	Summa Cum Laude
3,80 -- 3,99	Magna Cum Laude
3,60 -- 3,79	Cum Laude
3,00 -- 3,59	Sangat Memuaskan
2,50 -- 2,99	Memuaskan
2,00 -- 2,49	Cukup
0-1,99	Rendah

Pada tabel 2. IPS dan IPK mahasiswa dikonversi menjadi predikat akademik. Nilai terendah IPK / IPS adalah 0 sedangkan nilai tertinggi adalah 4. Konversi nilai IPS / IPK menjadi predikat akademik berdasarkan standar predikat kelulusan. Pada tabel 3 di bawah ini adalah tabel pengelompokan jumlah SKS yang sudah diambil oleh mahasiswa. Setiap semester mahasiswa mengambil rata – rata 20 SKS, jadi, pengelompokan SKS kumulatif dibagi dalam rentang 20 SKS.

TABEL V
PENGELOMPOKAN SKS KUMULATIF

Kode	SKS Kumulatif
1	0-20
2	21-40
3	41-60
4	61-80
5	81-100
5	101-120
6	121-140
7	141-160
8	>=161

Pada tabel 1 belum terdapat label pembayaran lancar atau menunda. Label menunda dan pembayaran lancar didapat dari data keuangan. Berikut adalah penjelasan dari label *data training*.

TABEL VI
PENJELASAN LABEL

Label	Keterangan
Lancar	Mahasiswa tidak kena denda pada semester tersebut
Menunda	Mahasiswa Membayar denda 5% atau 10%

Data training yang sudah di rubah nilainya menjadi lebih mudah diproses (mengurangi dimensi) dan sudah ditambahkan labelnya kemudian dijadikan file dengan format CSV / ARFF. Berikut adalah *data training* yang sudah siap di unggah.

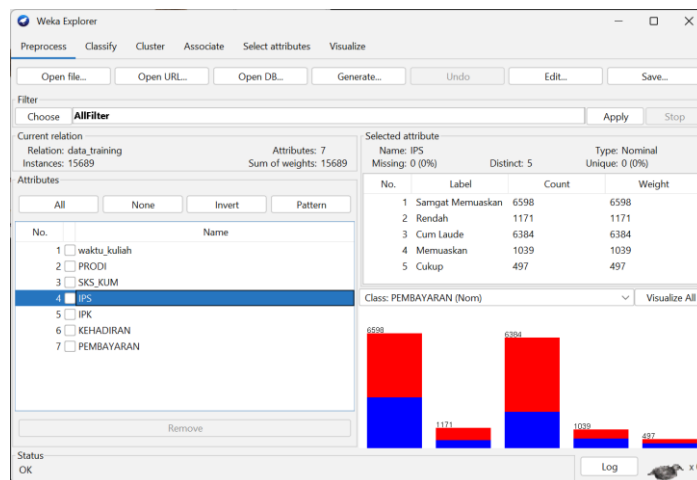
TABEL V
SAMPLE DATA TRAINING

waktu_kuliah	PRODI	SKS_KUM	IPS	IPK	KEHADIRAN	PEMBAYARAN
Malam	SK		7 Sangat Memuaskan	Sangat Memuaskan	Malas	MENUNDA
Malam	SK		7 Rendah	Sangat Memuaskan	Malas	MENUNDA
Malam	SK		8 Sangat Memuaskan	Sangat Memuaskan	Malas	MENUNDA
Pagi	SK		8 Cum Laude	Sangat Memuaskan	Malas	MENUNDA
Pagi	SI		8 Rendah	Cum Laude	Malas	MENUNDA
Pagi	SI		8 Rendah	Cum Laude	Malas	MENUNDA
Pagi	SI		8 Cum Laude	Sangat Memuaskan	Malas	LANCAR
Pagi	SI		8 Rendah	Sangat Memuaskan	Malas	MENUNDA
Pagi	SI		8 Rendah	Sangat Memuaskan	Malas	MENUNDA
Pagi	SI		5 Cum Laude	Cum Laude	Rajin	MENUNDA
Pagi	SI		7 Cum Laude	Cum Laude	Rajin	MENUNDA
Pagi	SI		8 Cum Laude	Cum Laude	Rajin	MENUNDA
Pagi	SI		8 Cum Laude	Cum Laude	Rajin	MENUNDA
Malam	SI		5 Sangat Memuaskan	Sangat Memuaskan	Malas	MENUNDA
Malam	SI		7 Sangat Memuaskan	Sangat Memuaskan	Malas	MENUNDA
Malam	SI		8 Cum Laude	Sangat Memuaskan	Malas	MENUNDA

Jumlah row yang berasal dijadikan data training yaitu 15.689 row. Dengan jumlah kolom yaitu 6 kolom, dimana kolom terakhir adalah label yang akan diprediksi.

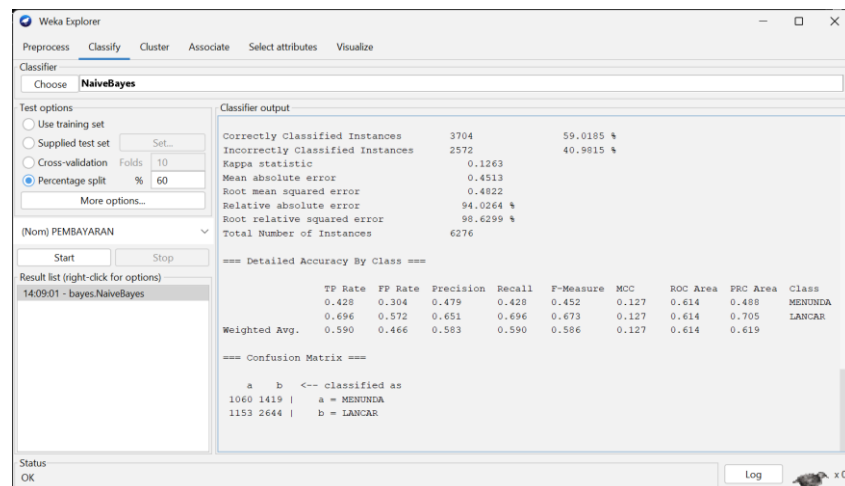
4. Pembangunan Model

Pembangunan model klasifikasi Naïve Bayes untuk prediksi mahasiswa yang akan menunda pembayaran menggunakan aplikasi WEKA. WEKA merupakan aplikasi *mesin learning* yang sudah memiliki algoritma Naïve Bayes yang sudah bisa digunakan untuk penelitian data mining [10]. *Data training* yang dimiliki pertama-tama format file nya dirubah menjadi format file ARFF / format CSV. file arff yang telah dibuat kemudian diupload ke weka dengan cara pilih menu explorer di weka kemudian open file. Berikut adalah gambar upload file arff ke weka.



Gambar 3. Upload File ARFF ke Weka

File yang sudah di upload ke weka kemudian dibuatkan model. Pembuatan model pada weka dengan cara pilih menu *classify* kemudian pilih model Naïve Bayes. *Test option* yang dipilih yaitu *percentage split 60%* Berikut adalah gambarannya.



Gambar 4. Pembuatan Model Classifikasi

5. Hasil dan Evaluasi

Pengujian model classifikasi Naïve Bayes yaitu dengan *precision*, *recall*, *f-measure* dan akurasi. Pembentukan *f-measure* dengan mengukur *True Positif*, *False Positif*, *precision*, dan *recall*. Berikut adalah nilai perhitingan dari *accuracy* dan *f-measure* dan confusion matrix. Nilai *correctly* 59% dan 40,9% tidak *correctly*, artinya ketepatan prediksi penundaan pembayaran berdasarkan nilai akademik dan kehadiran mahasiswa di kelas sangat berhubungan dengan akurasi 59%. Berikut adalah nilai hasil *accuracy*.

Classifier output		
Correctly Classified Instances	3704	59.0185 %
Incorrectly Classified Instances	2572	40.9815 %
Kappa statistic	0.1263	
Mean absolute error	0.4513	
Root mean squared error	0.4822	
Relative absolute error	94.0264 %	
Root relative squared error	98.6299 %	
Total Number of Instances	6276	

Gambar 5. Hasil Accuracy

Nilai *precision*, *recall* dan *f-measure* digunakan untuk mengukur dari masing–masing label. Pada hasil di bawah ini Label menunda memiliki presisi 0.479 dari maksimal 1. Sedangkan label Lancar 0.651. Jika nilai 0.5 atau 50% adalah nilai ragu – ragu, maka dari hasil tersebut disimpulkan jika sistem memprediksi menunda maka tingkat ketepatan masih diragukan karena di bawah 50% sedangkan jika model memprediksi Lancar maka tingkat ketepatan dapat dipercaya karena memiliki tingkat presisi di atas 50%.

=== Detailed Accuracy By Class ===								
	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area
	0.428	0.304	0.479	0.428	0.452	0.127	0.614	0.488
	0.696	0.572	0.651	0.696	0.673	0.127	0.614	0.705
Weighted Avg.	0.590	0.466	0.583	0.590	0.586	0.127	0.614	0.619
=== Confusion Matrix ===								
a	b	<-- classified as						
1060	1419	a = MENUNDA						
1153	2644	b = LANCAR						

Gambar 6. Perhitungan F-Measure

IV. KESIMPULAN

Data *training* telah dikumpulkan dari data akademik dan data keuangan mahasiswa. Jumlah data *training* yang digunakan model klasifikasi akan berpengaruh terhadap model yang dihasilkan. Pada penelitian ini menggunakan 60% model dan 40% data testing. Data bersumber dari nilai akademik dan kehadiran mahasiswa dalam perkuliahan, sedangkan data label / prediktif bersumber dari keuangan.

Hasil pengujian akurasi model klasifikasi 59% akurat, artinya hubungan data akademik dengan pembayaran memiliki saling keterkaitan sekitar 59%. Namun jika dilihat dari presisi perlabel prediksi MENUNDA memiliki nilai presisi dan

recall dibawah nilai tengah atau 0.5. sehingga model ini jika memprediksi menunda tidak selalu tepat. Sedangkan jika model prediksi Lancar maka hasil lebih bisa dipercaya, karena nilai presisi dan recall diatas nilai 0.5.

REFERENSI

- [1] Akhmad MR, Siswa TAY. Implementasi K-Nearest Neighbor Dalam Memprediksi Keterlambatan Pembayaran Biaya Kuliah Di Perguruan Tinggi. *Progresif J Ilm Komput.* 2022;18(2):185.
- [2] Putra IPW. Sistem pendukung keputusan beasiswa tidak mampu menggunakan metode dia pada itb stikom bali. 2025;07(02):134–42.
- [3] Ngurah G, Nata M. Deteksi Outlier Transaksi Menggunakan Visualisasi- Olap Pada Data Warehouse Perguruan Tinggi Swasta. *J Appl Intell Syst UDINUS.* 2016;1(2):77–89.
- [4] Rudy Hendrawan IN, Budhi Saputra IMA, Cahya Dewi GAP, Adi Pranata IGS, Wedasari NLN. Klasifikasi Lama Studi dan Predikat Kelulusan Mahasiswa menggunakan Metode Naïve Bayes. *J Eksplora Inform.* 2022;11(1):50–6.
- [5] Syukri Mustafa M, Rizky Ramadhan M, Thenata AP. Implementasi Data Mining untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier. *Citec J.* 2017;4(2):151–62.
- [6] Lorensius, Ping T. Model Pembiayaan Pendidikan di Perguruan Tinggi Pada Masa Covid-19. *Indones J Educ Humanit.* 2021;1(2):107–19.
- [7] Schröder C, Kruse F, Gómez JM. A systematic literature review on applying CRISP-DM process model. *Procedia Comput Sci.* 2021;181(2019):526–34.
- [8] Nata GNM, Anthony S, Yudiastra PP. Knowledge Discovery And Virtual Tour To Support Tourism Promotion. *IAIC Trans Sustain Digit Innov.* 2020;2(2):94–106.
- [9] Mega Nata GN, Ngurah Nyoman Bagiarta IG, Suraja Antarajaya IN, Effendy M. Frequent Pattern Untuk Katalog Digital. *Naratif J Nas Riset, Apl dan Tek Inform.* 2022;4(1):39–49.
- [10] Sunita B Aher, Lobo L M R J. Data Mining in Educational System using WEKA Walchand Institute of Technology Solapur. *IJCA Proc Int Conf.* 2011;(August):20–5.