

## Prediksi Piutang Biaya Pendidikan Mahasiswa Tak Tertagih menggunakan Algoritma Naïve Bayes di Institut Teknologi dan Bisnis Muhammadiyah Wakatobi

Sry Faslia Hamka<sup>1</sup>, Kusrini<sup>2</sup>, and Kusnawi<sup>3</sup>

<sup>1</sup>Program Studi Magister Teknik Informatika, Universitas Amikom Yogyakarta, Jl. Ring Road Utara, Ngringin, Condongcatur, Kec. Depok, Kabupaten Sleman, Daerah Istimewa Yogyakarta 55281  
e-mail: <sup>1</sup>sryfasliairmawan@students.amikom.ac.id

<sup>2,3</sup> Program Studi Magister Teknik Informatika, Universitas Amikom Yogyakarta, Jl. Ring Road Utara, Ngringin, Condongcatur, Kec. Depok, Kabupaten Sleman, Daerah Istimewa Yogyakarta 55281  
e-mail: <sup>2</sup>kusrini@amikom.ac.id, <sup>3</sup>khusnawi@amikom.ac.id

Submitted Date: 2023-06-13  
Revised Date: 2023-06-24

Reviewed Date: 2023-06-23  
Accepted Date: 2023-06-29

### Abstract

*The Muhammadiyah Wakatobi Institute of Technology and Business (ITBMW) determines the tuition fee that must be paid by students within a certain period or done in installments. However, more and more students are in arrears because there are still many students who have not paid tuition fees and the installment system that is applied. As a result, the higher the amount of student receivables, the more likely it is that student tuition fees are uncollected. This study aims to predict uncollectible student tuition receivables at ITBMW using a classification method, namely the Naïve Bayes algorithm. The data that will be utilized consists of ITBMW student information obtained from PDDikti for 2020/2021, 2021/2022, and 2022/2023 periods. In addition, internal data from the ITBMW Financial Administration Bureau will also be used for the 2021, 2022, and 2023 fiscal years. Data processing is carried out to obtain optimal prediction results by evaluating the performance of the algorithm to obtain the best results. Supporting attributes used in the available datasets are NIM, student name, status, university, study program, level, village/village address, sub-district address, guardian's education, guardian's job, guardian's income, description, amount of single tuition receivables ( UKT) of students, age of student UKT receivables, amount of receivables for education costs (BPP), age of BPP receivables, the status of study program receivables, level of study, address, education of father/mother/guardian, occupation of father/mother/guardian, the income of father/mother/trustee, the amount of UKT receivables, the age of UKT receivables, the amount of DPP receivables, and the age of DPP receivables. The targets and objectives of this data processing are student receivables with collectible and uncollectible status, using two experiments, namely the proportion of training data and data testing 80:20 and 90:10. Of the two experimental combinations, the proportion of training data and testing data is 80:20 which shows a high level of accuracy, namely 92.31%, which is the best accuracy rate compared to the proportion of 90:10 which produces an accuracy rate of 88.46%.*

*Keywords: Predictions; Receivables; Tuition Fees; Bad Debts; Naïve Bayes*

### Abstrak

Institut Teknologi dan Bisnis Muhammadiyah Wakatobi (ITBMW) menetapkan biaya pendidikan yang wajib dibayarkan oleh mahasiswa dalam jangka waktu tertentu atau dilakukan dengan cara mengangsur. Akan tetapi malah semakin banyak mahasiswa yang menunggak karena masih banyak mahasiswa yang belum membayar biaya pendidikan dan sistem angsuran yang diterapkan. Akibatnya, semakin tinggi jumlah piutang mahasiswa, semakin besar kemungkinan bahwa biaya pendidikan mahasiswa tak tertagih. Penelitian ini bertujuan untuk memprediksi piutang biaya pendidikan mahasiswa



tak tertagih di ITBMW menggunakan metode klasifikasi yaitu algoritma Naïve Bayes. Data yang akan dimanfaatkan terdiri dari informasi mahasiswa ITBMW yang didapatkan dari PDDikti selama periode 2020/2021, 2021/2022, dan 2022/2023 selain itu juga akan digunakan data internal Biro Administrasi Keuangan ITBMW untuk tahun anggaran 2021, 2022 dan 2023. Pengolahan data dilakukan untuk memperoleh hasil prediksi yang optimal dengan mengevaluasi kinerja algoritma sehingga memperoleh hasil yang terbaik. Atribut pendukung yang digunakan pada dataset yang tersedia yaitu: NIM, nama mahasiswa, status, perguruan tinggi, program studi, jenjang, alamat kelurahan/desa, alamat kecamatan, pendidikan wali, pekerjaan wali, penghasilan wali, keterangan, jumlah piutang uang kuliah tunggal (UKT) mahasiswa, umur piutang UKT mahasiswa, jumlah piutang biaya sarana dan prasarana pembangunan (BPP), umur piutang BPP, status piutang, program studi, jenjang studi, alamat, pendidikan ayah/ibu/wali, pekerjaan ayah/ibu/wali, penghasilan ayah/ibu/wali, jumlah piutang UKT, umur piutang UKT, jumlah piutang DPP, dan umur piutang DPP. Target dan sasaran dari pengolahan data ini adalah piutang mahasiswa dengan status tertagih dan tidak tertagih, dengan menggunakan dua percobaan yaitu dengan data proporsi data training dan data testing 80:20 dan 90:10. Dari dua kombinasi percobaan tersebut proporsi data training dan data testing 80:20 menunjukkan tingkat akurasi yang tinggi yaitu 92,31% merupakan tingkat akurasi yang terbaik dibandingkan dengan proporsi 90:10 yang menghasilkan tingkat akurasi 88,46%.

Kata Kunci : Prediksi; Piutang; Biaya Pendidikan; Piutang Tak Tertagih; Naïve Bayes

## 1. Pendahuluan

Prediksi merupakan proses yang dilakukan secara sistematis untuk memperkirakan hal yang paling mungkin terjadi. Hal ini didasarkan pada informasi yang tersedia saat ini dan sebelumnya. Tujuan utama dari prediksi adalah untuk meminimalisir kesalahan antara hasil perkiraan dengan kenyataan yang sebenarnya (Rahmayanti et al., 2022).

Piutang merupakan instrumen yang krusial dan memerlukan perhatian yang serius dalam mengelola perusahaan. Kinerja suatu perusahaan dapat dipengaruhi oleh besarnya nilai piutang yang dimilikinya. Apabila nilai piutang terlalu besar, maka dapat menjadi ancaman bagi kelangsungan hidup perusahaan. Ketika melakukan penagihan, perusahaan seringkali menghadapi kendala, salah satunya adalah keterlambatan pembayaran (Munandar et al., 2018). Beberapa macam risiko kerugian piutang meliputi Risiko ketidakmampuan pelunasan seluruh tagihan (piutang); Risiko tunggakan pembayaran sebagian piutang; Risiko pembayaran piutang yang terlambat; Risiko modal tidak terinvestasikan dalam piutang. Pengendalian piutang yang efektif sangat diperlukan agar dapat mengoptimalkan penagihan piutang tepat waktu (Aimbu et al., 2021).

Institut Teknologi dan Bisnis Muhammadiyah Wakatobi (ITBMW) didirikan pada tahun 2020 dan terletak di bawah naungan persyarikatan Muhammadiyah, setelah mendapatkan izin pendirian dari Menteri

Pendidikan dan Kebudayaan melalui Keputusan Nomor 942/M/2020. ITBMW adalah institusi pendidikan tinggi yang berorientasi pada pelayanan masyarakat dan tidak memiliki tujuan komersial. Institusi ini menitikberatkan pada pemberian kesempatan kepada masyarakat kurang mampu untuk mengakses pendidikan tinggi. Hal ini didukung dengan biaya pendidikannya yang masih terjangkau dan dalam pembayarannya bisa dilakukan dengan sistem mengangsur. Akan tetapi masih banyak mahasiswa yang menunggak atau tidak tepat waktu dalam pembayaran biaya pendidikan sehingga mengakibatkan semakin tinggi jumlah piutang mahasiswa dan semakin besar pula kemungkinan bahwa biaya pendidikan mahasiswa tak ter tagihkan.

Dalam konteks perguruan tinggi, memprediksi piutang biaya pendidikan mahasiswa yang belum terbayar sangatlah penting. Hal ini dapat membantu perguruan tinggi untuk mengambil tindakan yang sesuai dalam mengurangi risiko keterlambatan pembayaran atau piutang yang tak tertagih. Selain itu, prediksi ini juga membantu perguruan tinggi dalam perencanaan keuangan jangka panjang.

Model klasifikasi statistik yang dikenal sebagai Algoritma Naïve Bayes memiliki kemampuan untuk memprediksi probabilitas keanggotaan suatu kelas. Model ini berguna untuk memprediksi kelas dari data berdasarkan nilai-nilai yang terdapat pada fitur-fiturnya (Dewi, 2019).

Penelitian ini bertujuan untuk menggunakan algoritma Naïve Bayes dalam memprediksi piutang biaya pendidikan mahasiswa yang tidak tertagih di Institut Teknologi dan Bisnis Muhammadiyah Wakatobi. Dataset yang akan diolah terdiri dari informasi mahasiswa ITBMW yang didapatkan dari PDDikti selama periode 2020/2021, 2021/2022, dan 2022/2023 selain itu juga akan digunakan data internal Biro Administrasi Keuangan ITBMW untuk tahun anggaran 2021, 2022, dan 2023. Atribut pendukung yang digunakan pada dataset yang tersedia yaitu: NIM, nama mahasiswa, program studi, jenjang studi, alamat, pendidikan ayah/ibu/wali, pekerjaan ayah/ibu/wali, penghasilan ayah/ibu/wali, jumlah piutang UKT, umur piutang UKT, jumlah piutang DPP, dan umur piutang DPP. Target dan sasaran dari pengolahan data ini adalah piutang mahasiswa di Institut Teknologi dan Bisnis Muhammadiyah Wakatobi dengan status tertagih dan tidak tertagih.

Software atau Tool yang digunakan sebagai alat untuk melakukan pengujian metode adalah Rapidminer. Pengolahan data dilakukan untuk memperoleh hasil prediksi yang optimal dengan mengevaluasi kinerja algoritma sehingga memperoleh hasil yang terbaik.

Beberapa penelitian terdahulu yang telah menggunakan algoritma Naïve Bayes dalam pengolahan data untuk mendukung penelitian ini adalah Abdi Rahim Damanik dan kawan-kawan (2021) dengan penelitiannya menggunakan metode Naïve Bayes pada Algoritma Data Mining untuk memklasifikasi tingkat kepuasan mahasiswa terhadap pembelajaran daring. Ada empat parameter yang akan dipertimbangkan, yaitu Komunikasi, Suasana Pembelajaran di kelas, Evaluasi oleh Mahasiswa, dan Pemaparan Materi. Setelah melakukan pengujian akhir dengan menggunakan 80 informasi training dan 30 informasi testing, diperoleh hasil yang sangat baik dengan tingkat akurasi mencapai 100%, serta nilai precision dan recall sebesar 100% (Damanik et al., 2021).

Weko Susanto dan Luthfi Indriyani (2019) melakukan penelitian dengan mengambil studi kasus dengan dataset keanggotaan koperasi keluarga guru di Jakarta menggunakan algoritma Naive Bayes untuk memprediksi apakah kredit akan lancar atau macet serta mengevaluasi kinerja metode naive bayes. Hasil menunjukkan bahwa metode naive bayes memperoleh tingkat akurasi sebesar 84,00% yang termasuk dalam klasifikasi

baik. Selain itu, algoritma naive bayes juga mampu menangani perbedaan range data yang signifikan (Susanto & Indriyani, 2019).

Siti Lestari dan kawan-kawannya (2020) melakukan pengujian menggunakan data training yang terdiri dari 438 sampel dan data testing sebanyak 146 sampel yang berasal dari database nasabah pada tahun 2015. Dari hasil pengujian, terbukti bahwa Algoritma Naïve Bayes dapat digunakan untuk memprediksi kelayakan pemberian pinjaman di Koperasi Anugerah Bintang Cemerlang. Tingkat akurasi tertinggi yang dicapai mencapai 78,08%, sedangkan presentase kesalahan sebesar 21,92%. (Lestari et al., 2020).

Lian Siliyani dan kawan-kawan (2020) melakukan penelitian menggunakan data dari mahasiswa yang telah mengisi kuesioner. Totalnya terdapat 217 responden yang berasal dari mahasiswa semester ganjil di STIKES Ahmad Dahlan Cirebon pada Tahun Akademik 2019/2020. Dalam penilaian ini, terdapat beberapa indikator yang digunakan, antara lain tangible (bukti langsung), reliability (keandalan), responsiveness (daya tangkap), assurance (jaminan), dan empathy (empati). Selanjutnya, data tersebut dibagi menjadi dua bagian, yakni data training sebesar 80% dan data testing sebesar 20%. Dari hasil pengujian, ditemukan bahwa tingkat akurasi mencapai 97,73%, nilai presisi mencapai 100,00%, dan nilai recall mencapai 66,67% (Siliyani et al., 2020).

Wahyu Ananda dan kawan-kawan (2021) melakukan penelitian yang bertujuan untuk melakukan prediksi terhadap jumlah produksi panen kelapa sawit serta memberikan rekomendasi kepada PTPN IV Dolok Sinubah untuk meningkatkan produksi panen sawit setiap tahunnya. Dalam penelitian ini, terdapat total 7 rekaman yang diprediksi meningkat dan 5 rekaman yang diprediksi menurun. Dengan demikian, nilai akurasi yang diperoleh adalah 100% (Ananda et al., 2021).

Data mining adalah bidang ilmu yang memungkinkan prediksi yang akurat tentang keadaan di masa depan dengan meningkatkan efisiensi dalam mencari informasi dari basis data yang besar dan mengidentifikasi pola-pola yang sebelumnya tidak terdeteksi (Aeni & Asy'ari, 2020).

Data mining merupakan proses yang digunakan untuk mengungkap pengetahuan yang tersembunyi dalam basis data, yang juga dikenal

sebagai *Knowledge Discovery in Database* (KDD). Pengetahuan yang ditemukan melalui data mining dapat berupa pola data atau hubungan yang valid antara data. Dalam konteks ini, Sumpena dan Kurnia (2019) menjelaskan langkah-langkah yang terlibat dalam data mining yaitu:

- a. Proses Data Cleaning adalah menghapus data yang tidak relevan, tidak konsisten atau mengandung noise.
- b. Proses Integrasi Data adalah menggabungkan data dari beberapa sumber basis data ke dalam satu basis data yang baru.
- c. Proses Data Selection merupakan proses memilih data yang relevan dan cocok untuk dianalisis dari basis data, karena tidak semua data dalam basis data diperlukan.
- d. Proses Transformasi Data (*Data Transformation*) adalah mengubah atau menggabungkan data ke dalam format yang dapat diolah dengan baik dalam proses data mining.
- e. Proses Mining merupakan tahap dalam mengaplikasikan metode-metode tertentu untuk menemukan informasi penting dan tersembunyi dari data. Proses ini menjadi tahap utama dalam data mining. Metode-metode yang digunakan dalam proses ini bervariasi tergantung pada jenis pengelompokan data mining yang diterapkan.
- f. Proses Evaluasi Pola (*Pattern Evaluation*) merupakan proses untuk mengidentifikasi pola-pola yang menarik dan relevan dari Knowledge Base yang telah ditemukan.
- g. Persentasi Pengetahuan (*Knowledge Presentation*) melibatkan visualisasi dan penyajian informasi tentang metode yang digunakan untuk memperoleh pengetahuan yang telah ditemukan kepada pengguna.

Prediksi atau peramalan merupakan suatu proses terstruktur untuk memperkirakan kemungkinan terjadinya suatu kejadian di masa depan dengan memanfaatkan informasi yang diperoleh dari masa lalu dan saat ini. Tujuan dari prediksi adalah untuk mengurangi kesalahan dalam perkiraan atau selisih antara hasil prediksi dengan kenyataan yang terjadi. Ada juga definisi lain yang menyatakan bahwa prediksi atau peramalan melibatkan seni dan ilmu dalam memprediksi peristiwa masa depan (Zulfauzi & Alamsyah, 2020).

Piutang merupakan klaim yang diberikan oleh pihak yang ingin menerima pembayaran

kepada pihak yang memiliki kewajiban untuk membayar dengan bentuk pembayaran umumnya berupa uang. Piutang dapat dikelompokkan menjadi 3 kategori berdasarkan jangka waktu pembayarannya, yaitu jangka pendek (1 sampai 2 bulan); jangka menengah (6-12 bulan); dan jangka panjang (dengan jangka waktu pembayaran yang mencapai beberapa tahun, tergantung pada jumlah piutang dan persyaratan dari pihak yang memberikan piutang)

Resiko piutang yang tidak tertagih merujuk pada kemungkinan kerugian yang timbul akibat transaksi penjualan yang dilakukan secara kredit. Hal ini dikenal sebagai resiko kerugian piutang. Resiko kerugian piutang dapat terdiri dari beberapa bentuk, termasuk: 1) risiko ketidakmampuan pelanggan untuk membayar seluruh tagihan piutang; 2) risiko ketidakmampuan pelanggan untuk membayar sebagian tagihan piutang; 3) risiko keterlambatan pelunasan piutang; 4) risiko ketidakmampuan modal yang diinvestasikan dalam piutang (Aimbu et al., 2021).

Biaya pendidikan adalah jumlah uang yang harus dibayarkan oleh peserta didik ke lembaga pendidikan tempat mereka belajar. Selain itu, ada juga biaya lain seperti peralatan, transportasi, makanan, akomodasi, dan sebagainya. Namun, yang harus dibayarkan oleh peserta didik kepada lembaga pendidikan adalah biaya pendidikan itu sendiri. Besar biaya pendidikan yang harus dibayarkan oleh peserta didik tergantung pada ketentuan yang berlaku di lembaga pendidikan tersebut (Hendri, 2021).

Pengklasifikasi Naive Bayes merupakan sebuah algoritma klasifikasi probabilistik yang didasarkan pada teorema Bayes. Algoritma ini mengasumsikan bahwa setiap fitur memberikan kontribusi yang saling independen dan memiliki bobot yang sama terhadap kelas target. Dalam anggapan bahwa setiap fitur adalah independen dan tidak saling berinteraksi, setiap fitur akan memberikan kontribusi secara mandiri dan sama-sama pada probabilitas sampel menjadi bagian dari kelas tertentu. Pengklasifikasi Naive Bayes mudah dilaksanakan dan dapat bekerja dengan cepat pada kumpulan data yang besar dan memiliki dimensi yang tinggi. Algoritma ini dapat digunakan dalam aplikasi waktu nyata dan tidak terpengaruh oleh noise. Dalam pengklasifikasi Naive Bayes, probabilitas kelas dan probabilitas bersyarat dihitung dengan memproses set data pelatihan. Untuk menentukan probabilitas,

frekuensi setiap nilai fitur yang terkait dengan kelas tertentu dibagi dengan frekuensi instance yang memiliki kelas tersebut. Algoritma klasifikasi Naive Bayes akan memberikan hasil terbaik ketika fitur-fitur yang saling berkorelasi dihapus, karena jika tidak, fitur-fitur tersebut akan terpilih dua kali dalam model dan menghasilkan penekanan yang berlebihan pada pentingnya fitur-fitur tersebut (Al Hafiz et al., 2019).

Teorema Bayes dalam bentuk persamaan matematika adalah sebagai berikut (Novianti, 2019).

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

Keterangan:

- X : Data class yang belum diketahui.
- H : Hipotesis data class spesifik.
- P(H|X) : Probabilitas hipotesis H berdasar kondisi X (posteriori probability).
- P(H) : Probabilitas hipotesis H (prior probability).
- P(X|H) : Probabilitas X berdasarkan kondisi H (like hood probability).
- P(X) : Probabilitas X (evidence probability).

*Confusion matrix* merupakan teknik yang sering dipakai dalam menghitung akurasi pada pembelajaran data mining, yang menghasilkan empat output yaitu *recall*, *precision*, akurasi, dan tingkat kesalahan. (Dewi, 2019).

Cara kerja *confusion matrix* adalah dengan mengevaluasi performa model klasifikasi berdasarkan jumlah prediksi yang benar dan salah pada objek yang diuji. *Confusion matrix* memuat nilai-nilai berikut ini (Novianti, 2019) :

- a. *True positif* (TP) adalah nilai positif yang terdeteksi dengan benar, baik dalam prediksi maupun nilai aktualnya.
- b. *False positif* (FP) adalah nilai yang diprediksi positif, namun nilai aktualnya negatif.
- c. *False negative* (FN) adalah nilai yang diprediksi negatif, namun nilai aktualnya positif.
- d. *True negative* (TN) adalah nilai negatif yang terdeteksi dengan benar, baik dalam prediksi maupun nilai aktualnya.
- e. Berikut rumus menghitung rasio *accuracy*, *precision*, *recall*, dan *f1-score*:

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

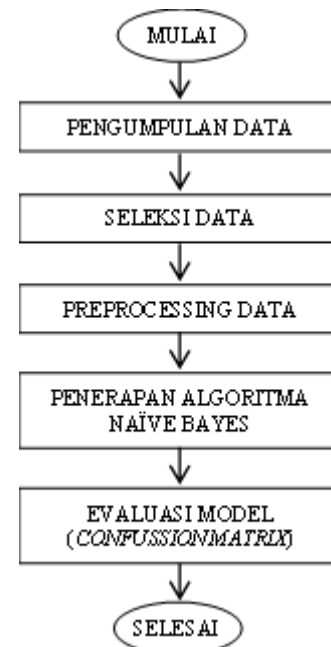
$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$F1 - Score = \frac{2 \times recall \times precision}{recall + precision}$$

Rapid Miner adalah sebuah perangkat lunak sumber terbuka yang dapat digunakan untuk melaksanakan proses data mining. RapidMiner menyediakan berbagai fitur, termasuk *Extraction*, *Transformation*, *Loading* (ETL), *preprocessing data*, visualisasi, pemodelan, dan evaluasi. RapidMiner menawarkan antarmuka pengguna grafis (GUI) yang memungkinkan pengguna untuk membuat file XML (Extensible Markup Language) yang menggambarkan proses analitis yang akan dijalankan pada data. File XML ini dapat dibaca oleh RapidMiner untuk menjalankan analisis secara otomatis (Sumpena & Kurnia, 2019).

## 2. Tahapan Penelitian

Terdapat langkah-langkah penelitian yang dapat dilihat pada Gambar 1, yang menggambarkan alur proses penelitian:



Gambar 1 Alur proses penelitian

Urutan langkah-langkah dalam penelitian ini secara spesifik dapat dijelaskan sebagai berikut:

## 1. Pengumpulan Data

Diawali dengan melakukan Pengumpulan Data yang terdiri dari informasi mahasiswa ITBMW yang didapatkan dari PDDikti selama periode 2020/2021, 2021/2022, dan 2022/2023 di import ke bentuk xls. Data Mahasiswa tersebut terdiri dari 12 Atribut dan setiap Atribut terdiri dari beberapa nilai. Tabel 1 merupakan tampilan data informasi mahasiswa yang diambil dari PDDikti.

Pada Tabel 1 menunjukkan beberapa data mahasiswa dari Pangkalan Data Pendidikan Tinggi melalui situs <https://pddikti-admin.kemdikbud.go.id/admin/> untuk periode Tahun Akademik 2020/2021, 2021/2022 dan 2022/2023 berjumlah 3.283 record data.

Selain itu juga untuk data piutang mahasiswa akan digunakan data internal Biro Administrasi Keuangan ITBMW untuk Tahun Anggaran 2021, 2022, dan 2023. Data yang telah diimport ke bentuk xls. Data Piutang Mahasiswa

Tabel 1 Data beberapa Mahasiswa dari PDDikti

NIM	Nama Mahasiswa	Status	Perguruan Tinggi	Program Studi	Jenjang	Alamat Kelurahan / Desa	Alamat Kecamatan	Pendidikan Wali	Pekerjaan Wali	Penghasilan Wali	Ket
220542470001	NINIANTI	Aktif	Institut Teknologi dan Bisnis Muhammadiyah Wakatobi	Ilmu Perikanan	S1	Pada Raya Makmur	Wangi-Wangi	SD Sederajat	Petani	Kurang dari Rp. 500.000	Valid
220542470002	ANITA	Aktif	Institut Teknologi dan Bisnis Muhammadiyah Wakatobi	Ilmu Perikanan	S1	Wanci	Wangi-Wangi	SD Sederajat	Nelayan	Kurang dari Rp. 1.000.000	Valid
220542470003	BINTANG SUDARNO	Aktif	Institut Teknologi dan Bisnis Muhammadiyah Wakatobi	Ilmu Perikanan	S1	Mola Selatan	Wangi-Wangi Selatan	SMA Sederajat	Nelayan	Rp. 1.000.000 - Rp. 1.999.999	Valid
220542470004	MA'RUF	Aktif	Institut Teknologi dan Bisnis Muhammadiyah Wakatobi	Ilmu Perikanan	S1	Pada Raya Makmur	Wangi-Wangi	Putus SD	Buruh Harian Lepas	Kurang dari Rp. 500.000	Valid
220542470005	YADIR	Aktif	Institut Teknologi dan Bisnis Muhammadiyah Wakatobi	Ilmu Perikanan	S1	Mola Nelayan Bakti	Wangi-Wangi Selatan	SMA Sederajat	Nelayan	Kurang dari Rp. 1.000.000	Valid
220542470006	MERI HANDAYANI	Aktif	Institut Teknologi dan Bisnis Muhammadiyah Wakatobi	Ilmu Perikanan	S1	Mola Samaturu	Wangi-Wangi Selatan	D4/S1	Wiraswasta	Rp. 1.000.000 - Rp. 1.999.999	Valid
220542470007	WAHDANIA	Aktif	Institut Teknologi dan Bisnis Muhammadiyah Wakatobi	Ilmu Perikanan	S1	Mola Samaturu	Wangi-Wangi Selatan	SMA Sederajat	Nelayan	Kurang dari Rp. 1.000.000	Valid
220542470008	KASMADIN	Aktif	Institut Teknologi dan Bisnis Muhammadiyah Wakatobi	Ilmu Perikanan	S1						Valid
220542470009	FITRIA A	Aktif	Institut Teknologi dan Bisnis Muhammadiyah Wakatobi	Ilmu Perikanan	S1	Mola Bahari	Wangi-Wangi Selatan	SMP Sederajat	Nelayan	Kurang dari Rp. 1.000.000	Valid
220542470010	AHMAD AMIRUDIN	Aktif	Institut Teknologi dan Bisnis Muhammadiyah Wakatobi	Ilmu Perikanan	S1						Valid

tersebut terdiri dari Data Piutang Biaya Sarana dan Prasarana Pembangunan terdiri dari 7 Atribut dan Data Piutang Uang Kuliah Tunggal terdiri dari 5 Atribut. Tabel 2 merupakan tampilan Data Piutang Biaya Sarana dan Prasarana Pembangunan.

Tabel 2 Data Piutang BPP

NIM	Nama Mahasiswa	Pembayaran DPP Uang Pendaftn, Reg Ulang, Infak				
		Pendaftar awal masuk	Registras i Ulang	Infak	BPS	Total
220572010006	Fitriani	-	-	-	1,000,000	1,000,000
220572010003	Bambang	-	-	-	-	-
220572010008	Jamalu	-	500,000	300,000	2,000,000	2,800,000
220572010024	Wa Ewi	-	-	-	1,300,000	1,300,000
220572010026	Yuni	-	-	-	1,000,000	1,000,000
220542470003	Bintang S	-	-	300,000	2,000,000	2,300,000
220942020001	Rostiana	-	-	300,000	2,000,000	2,300,000
220942020002	Ani M	-	-	-	1,800,000	1,800,000
220572010017	Nurul Y	-	500,000	300,000	2,000,000	2,800,000
220572010021	Suhartono	-	-	300,000	2,000,000	2,300,000

Pada Tabel 2 menunjukkan beberapa data piutang mahasiswa untuk biaya sarana dan prasarana pembangunan untuk Tahun Anggaran 2020, 2021 dan 2022 berjumlah 5.211 record data.

Tabel 3 merupakan tampilan Data Piutang Uang Kuliah Tunggal. Tabel ini menampilkan informasi terkait piutang yang terkait dengan uang kuliah tunggal.

Tabel 3. Data Piutang UKT

NIM	Nama Mahasiswa	Jumlah Piutang	Penerimaan Piutang	Saldo Tahun
220572010006	Fitriani	5,000,000	4,000,000	1,000,000
220572010003	Bambang	5,000,000	4,000,000	1,000,000
220572010008	Jamalu	5,000,000	2,000,000	3,000,000
220572010024	Wa Ewi	5,000,000	5,000,000	-
220572010026	Yuni	5,000,000	4,500,000	500,000
220542470003	Bintang Sudarno	5,000,000	4,900,000	100,000
220942020001	Rostiana	5,000,000	5,000,000	-
220942020002	Ani Mazayarah	5,000,000	5,000,000	-
220572010017	Nurul Yakini	5,000,000	1,000,000	4,000,000
220572010021	Suhartono	5,000,000	3,000,000	2,000,000

Pada Tabel 3, terdapat beberapa data piutang Uang Kuliah Tunggal (UKT) mahasiswa untuk Tahun Anggaran 2020, 2021, dan 2022. Jumlah total record data dalam tabel ini adalah sebanyak 1.572.

## 2. Seleksi Data

Langkah selanjutnya adalah melakukan Seleksi Data agar hasil yang optimal, tepat, dan akurat, maka dilakukan penggabungan, pengolahan dan pembersihan data dengan mempertimbangkan kriteria yang telah ditetapkan selama proses klasifikasi sehingga hanya menjadi 5.717 record data.

### 3. Preprocessing Data

Proses yang harus dilakukan pada data yang diperoleh adalah memperbaikinya, *Preprocessing data* menjadi langkah berikutnya. *Preprocessing data* adalah melakukan pembersihan, menambahkan, dan menyusun data agar sesuai dengan kebutuhan yang ditentukan. Data awal yang diperoleh dari proses seleksi data berjumlah 5.717 record data setelah dilakukan *preprocessing data* dengan membersihkan data *null*, menghapus data yang tidak relevan maka diperoleh 17 Atribut dengan total 4.418 record data.

### 4. Penerapan algoritma naïve bayes

Dalam metode pengujian ini, digunakan pendekatan menggunakan Naïve Bayes. Naïve Bayes adalah suatu metode klasifikasi yang berbasis pada probabilitas dan statistik, yang awalnya dikembangkan oleh seorang ilmuwan Inggris bernama Thomas Bayes. Metode ini bertujuan untuk memprediksi peluang peristiwa di masa depan berdasarkan pengalaman yang terjadi di masa lalu, dan populer dengan sebutan Teorema Bayes. Proses ini menggunakan Rapidminer dengan algoritma naïve bayes.

### 5. Evaluasi Model

Hasil ini akan disajikan dalam format yang mudah dipahami oleh pembaca. Dalam proses ini, melibatkan identifikasi pola dan informasi yang ditemukan, serta menggunakan Confusion Matrix untuk mengevaluasi model klasifikasi yang digunakan dalam memprediksi objek yang benar dan salah. Prediksi hasil akan dibandingkan dengan kelas sebenarnya dari data tersebut. Confusion Matrix digunakan sebagai alat evaluasi kinerja model berdasarkan kemampuan prediksi. Salah satu parameter evaluasi yang digunakan adalah akurasi, yang mengukur sejauh mana aturan klasifikasi yang dihasilkan tepat dalam mengklasifikasikan Data Testing berdasarkan atribut yang ada dalam Kelasnya.

## 3. Metode Penelitian

### 1. Teknik Pengumpulan Data

Dalam penelitian ini, pengumpulan data dilakukan dengan menggunakan dua sumber data utama, yaitu pangkalan data Dikti dan data piutang mahasiswa yang diperoleh dari biro keuangan. Data dari pangkalan data Dikti digunakan untuk mendapatkan informasi terkait dengan data akademik dan profil mahasiswa, sedangkan data piutang mahasiswa dari biro

keuangan digunakan untuk memperoleh informasi tentang status piutang mahasiswa.

### 2. Teknik Analisis Data

Teknik analisis data yang digunakan dalam penelitian ini adalah analisis data yang berasal dari sumber sekunder berupa data yang telah dikumpulkan dilakukan seleksi data untuk menggabungkan, mengolah dan membersihkan data dengan mempertimbangkan kriteria yang telah ditetapkan selama proses klasifikasi, kemudian dilakukan teknik *preprocessing data* untuk menghilangkan data yang tidak relevan, data kosong, data ganda dan anomali data.

### 3. Teknik Pengolahan Data

Pada tahap pengolahan data, dilakukan menggunakan aplikasi Rapidminer dengan metode Klasifikasi Naïve Bayes. Data set yang diperoleh dibagi menjadi dua bagian, di mana bagian pertama digunakan sebagai data training dan bagian kedua digunakan sebagai data testing. Penelitian ini melakukan dua percobaan pengujian yaitu percobaan pertama dengan proporsi data training dan data testing adalah 80:20 dan percobaan kedua dengan proporsi data training dan data testing 90:10.

### 4. Model yang diusulkan

Model yang diusulkan adalah menggunakan algoritma Naïve Bayes untuk memprediksi piutang biaya pendidikan mahasiswa tak tertagih.

## 4. Hasil dan Pembahasan

Metode klasifikasi yang digunakan dalam tahap ini adalah Naive Bayes. Sebelum melanjutkan ke tahap pengujian, terlebih dahulu dilakukan tahap preprocessing data. Pada saat pengujian klasifikasi, hasil penelitian yang telah diuji dievaluasi, digunakan dua kali percobaan yaitu percobaan pertama dengan proporsi data training dan data testing 80:20 dan percobaan kedua dengan proporsi data training dan data testing 90:10, kemudian dilakukan tahap evaluasi menggunakan *Confusion Matrix*.

Hasil prepossessing data setelah digabungkan dari dua sumber data utama menjadi 17 Atribut dengan total 4.418 record data. Data hasil preprocessing per variable dapat dilihat dari Tabel 4 berikut ini

Tabel 4 Jumlah Atribut yang digunakan

No	Atribut	Jumlah
----	---------	--------

1	NIM	260
2	Nama Mahasiswa	260
3	Status	1
4	Perguruan Tinggi	1
5	Program Studi	3
6	Jenjang	1
7	Alamat Kelurahan/ Desa	50
8	Alamat Kecamatan	13
9	Pendidikan Wali	8
10	Pekerjaan Wali	15
11	Penghasilan Wali	8
12	Keterangan	1
13	Jumlah Piutang UKT Mahasiswa	31
14	Umur Piutang UKT Mahasiswa	3
15	Jumlah Piutang BPP	24
16	Umur Piutang BPP	3
17	Status Piutang	2

Berdasarkan Tabel 4 terdapat 17 Atribut yang digunakan yaitu NIM terdapat 260 nilai, Nama Mahasiswa terdapat 260 nilai, Status terdapat 1 nilai, Perguruan Tinggi terdapat 1 nilai, Program Studi 3 nilai, jenjang 1 nilai, alamat kelurahan/desa terdapat 50 nilai, Alamat Kecamatan 13 nilai, Pendidikan Wali 8 nilai, Pekerjaan Wali 15 nilai, keterangan 1 nilai, Jumlah Piutang UKT Mahasiswa 31 nilai, Umur Piutang UKT Mahasiswa 3 nilai, Jumlah Piutang BPP 24 nilai, umur piutang BPP terdapat 3 nilai dan Status piutang memiliki 2 nilai.

Tabel 5 Atribut dan Tipe data

No	Atribut	Tipe Data
1	NIM	Real
2	Nama Mahasiswa	Polynomial
3	Status	Binomial
4	Perguruan Tinggi	Binomial
5	Program Studi	Polynomial
6	Jenjang	Binomial
7	Alamat Kelurahan/ Desa	Polynomial
8	Alamat Kecamatan	Polynomial
9	Pendidikan Wali	Polynomial
10	Pekerjaan Wali	Polynomial
11	Penghasilan Wali	Polynomial
12	Keterangan	Binomial
13	Jumlah Piutang UKT Mahasiswa	Integer
14	Umur Piutang UKT Mahasiswa	Polynomial
15	Jumlah Piutang BPP	Integer
16	Umur Piutang BPP	Polynomial
17	Status Piutang	Polynomial

Dari hasil percobaan pertama pengujian dengan proporsi data training dan data testing sebesar 80:20, dapat dilihat pada Gambar 2.

	true Piutang tertagih	true Piutang tak tertagih	class precision
pred Piutang tertagih	29	2	93.55%
pred Piutang tak tertagih	2	19	90.48%
class recall	93.55%	90.48%	

Gambar 2 Hasil Akurasi pada proporsi 80:20

Dari tabel yang terdapat pada Gambar 2, dapat dijelaskan bahwa akurasi mencapai 92,31%. Hal ini menunjukkan bahwa piutang mahasiswa yang telah tertagih diprediksi dengan tepat sebanyak 29 data, sedangkan 2 data diprediksi tidak tepat sebagai piutang tertagih dan sebaliknya, piutang mahasiswa yang tidak tertagih diprediksi dengan tepat sebanyak 19 data, namun terdapat 2 data yang diprediksi tidak tepat sebagai piutang tidak tertagih. Adapun bentuk tabel dari *confusion matrix*.

Tabel 6 Tabel *Confusion matrix* Proporsi 80:20

	True Positif	True Negatif
False positif	29	2
false negatif	2	19

Adapun perhitungan dari tabel *confusion matrix*:

Kita dapat menghitung tingkat akurasi dan persamaan berikut:

$$\begin{aligned}
 &= \frac{TP + TN}{FP + FN + TP + TN} \times 100\% \\
 &= \frac{29 + 19}{29 + 19 + 2 + 2} \times 100\% \\
 &= \frac{48}{52} \times 100\% \\
 &= 92,31\%
 \end{aligned}$$

Dan selanjutnya adalah nilai *precision*

$$\begin{aligned}
 &= \frac{TP}{FP + TP} \times 100\% \\
 &= \frac{29}{29 + 2} \times 100\% \\
 &= 0,9355 \times 100\% = 93,55\%
 \end{aligned}$$

Yang terakhir adalah *Recall*

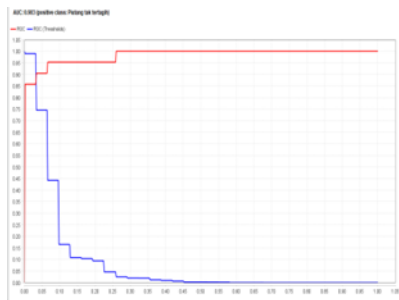
$$\begin{aligned}
 &= \frac{TP}{FN + TP} \times 100\% \\
 &= \frac{29}{29 + 2} \times 100\% \\
 &= \frac{29}{31} \times 100\% = 93,55\%
 \end{aligned}$$

Dari perhitungan tabel *confusion matrix* tersebut maka dapat kita lihat hasil untuk nilai



accurasi sebesar 92,31%, berarti algoritma tersebut telah berhasil mengklasifikasikan piutang tak tertagih biaya pendidikan mahasiswa, untuk selanjutnya nilai Precision sebesar 93,55% hal ini menunjukkan bahwa sistem berhasil memberikan informasi yang akurat. Selain itu, nilai recall sebesar 93,55% menunjukkan bahwa sistem juga berhasil memberikan informasi yang benar.

Pada gambar 3, terdapat Grafik ROC yang menunjukkan bahwa nilai AUC (Area Under Curve) adalah 0,983%. Hal ini mengindikasikan bahwa klasifikasi tersebut merupakan diagnosis terbaik.



Gambar 3 Grafik ROC dan nilai AUC Proporsi 80:20

Hasil pengujian percobaan kedua dengan data proporsi data training dan data testing 90:10 dapat dilihat pada gambar 3 yaitu:

accuracy: 88.46%

	true Piutang tertagih	true Piutang tak tertagih	class precision
pred. Piutang tertagih	14	1	93.33%
pred. Piutang tak tertagih	2	9	81.82%
class recall	87.50%	90.00%	

Gambar 4 Hasil Akurasi pada proporsi 90:10

Dari tabel yang terdapat pada Gambar 4, dapat dijelaskan bahwa akurasi mencapai 88,46%. Hal ini menunjukkan bahwa piutang mahasiswa yang telah tertagih diprediksi dengan tepat sebanyak 14 data, namun terdapat 1 data yang diprediksi tidak tepat sebagai piutang tertagih. Sedangkan, piutang mahasiswa yang tidak tertagih diprediksi dengan tepat sebanyak 9 data, namun terdapat 2 data yang diprediksi tidak tepat sebagai piutang tidak tertagih.

Adapun bentuk tabel dari *confusion matrix*.

Tabel 6 Tabel Confusion matrix Proporsi 90:10

	True Positif	True Negatif
False positif	14	1
false negatif	2	9

Adapun perhitungan dari tabel *confusion matrix*:

Kita dapat menghitung tingkat akurasi dan persamaan berikut:

$$\begin{aligned}
 &= \frac{TP + TN}{FP + FN + TP + TN} \times 100\% \\
 &= \frac{14 + 9}{14 + 9 + 2 + 1} \times 100\% \\
 &= \frac{23}{26} \times 100\% \\
 &= 88,46\%
 \end{aligned}$$

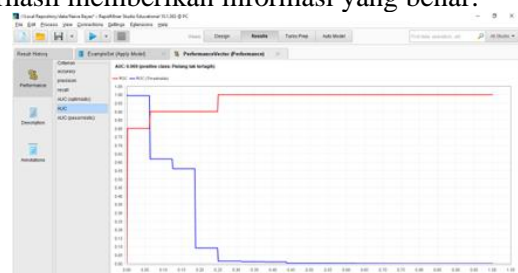
Dan selanjutnya adalah nilai *precision*.

$$\begin{aligned}
 &= \frac{TP}{FP + TP} \times 100\% \\
 &= \frac{14}{14 + 1} \times 100\% \\
 &= 0,933 \times 100\% = 93,33\%
 \end{aligned}$$

Yang terakhir adalah *Recall*.

$$\begin{aligned}
 &= \frac{TP}{FN + TP} \times 100\% \\
 &= \frac{14}{14 + 2} \times 100\% \\
 &= \frac{14}{16} \times 100\% = 87,5\%
 \end{aligned}$$

Dari perhitungan tabel *confusion matrix* tersebut maka dapat kita lihat hasil untuk nilai accurasi sebesar 88,46%, berarti algoritma tersebut telah berhasil mengklasifikasikan piutang tak tertagih biaya pendidikan mahasiswa, untuk selanjutnya nilai *Precision* sebesar 93,33% ini berarti sistem telah berhasil memberikan informasi yang tepat, sedangkan yang terakhir adalah nilai dari *recall* sebesar 87,5 yang artinya sistem telah berhasil memberikan informasi yang benar.



Gambar 5 Grafik ROC dan nilai AUC Proporsi 90:10

Pada gambar 5 yaitu Grafik ROC menjelaskan bahwa nilai untuk AUC adalah sebesar 0,969 % artinya bahwa klasifikasi tersebut merupakan diagnose terbaik.

## 5. Kesimpulan

Dalam kesimpulan ini menunjukkan hasil klasifikasi dengan menggunakan dua percobaan yaitu dengan data proporsi data training dan data testing 80:20 dan 90:10. Dari dua kombinasi percobaan tersebut proporsi data training dan data testing 80:20 menunjukkan tingkat akurasi yang tinggi yaitu 92,31% merupakan tingkat akurasi yang terbaik dibandingkan dengan proporsi 90:10 yang menghasilkan tingkat akurasi 88,46%. Dari nilai akurasi tersebut dapat disimpulkan bahwa piutang biaya pendidikan mahasiswa tak tertagih dapat diprediksi dengan menggunakan algoritma naïve bayes dengan status piutang tertagih merupakan status piutang mahasiswa yang lebih unggul dibandingkan status piutang tak tertagih dengan demikian prediksi status piutang mahasiswa lebih banyak adalah piutang tertagih.

## References

- Aeni, K., & Asy'ari, M. F. (2020). Prediksi Kepuasan Layanan Akademik Menggunakan Algoritma Naïve Bayes. *Jurnal Teknik Informatika dan Sistem Informasi*, 7(3), 601-609. <https://doi.org/10.35957/jatisi.v7i3.603>
- Aimbu, G., Karamoy, H., & Tangkuman, S. J. (2021). Analisis Pengendalian Piutang untuk meminimalkan risiko piutang tak tertagih pada PT Samudera Mandiri Sentosa. *Jurnal Riset Akuntansi*, 16(4), 366-374.
- Ananda, W., Safir, M., & Fauzan, M. (2021). Prediksi Jumlah Hasil Panen Sawit menggunakan Algoritma Naïve Bayes. *TIN: Terapan Informatika Nusantara*, 1(10), 513-519.
- Al Hafiz, D. M., Amaly, K., Jonathan, J., Rachmatullah, M. T., & Rosidi. (2021). Sistem Prediksi Penyakit Jantung menggunakan Metode Naïve Bayes. *Jurnal Rekayasa Elektro Sriwijaya*, 2(2), 151-157. DOI: <https://doi.org/10.36706/jres.v2i2.29>
- Arifin, T., Syalwah, S. (2020). Prediksi Keberhasilan Immunotherapy pada Penyakit Kutil dengan Menggunakan Naïve Bayes. *Jurnal Responsif: Riset Sains dan Informatika*, 2(1), 38-43. <https://doi.org/10.51977/jti.v2i1.177>
- Damanik, A. R., Sumijan., & Nurcahyo, G. W. (2021). Prediksi Tingkat Kepuasan dalam Pembelajaran Daring menggunakan Algoritma Naïve Bayes. *Jurnal Sistem Informasi dan Teknologi*, 3(2), 88-94. DOI: <https://doi.org/10.37034/jsisfotek.v3i3.49>
- Dewi, S. (2019). Komparasi Metode Algoritma Data Mining pada Prediksi Uji Kelayakan Credit Approval pada Calon Nasabah Kredit Perbankan. *Jurnal Khatulistiwa Informatika*, VII(1). DOI: <https://doi.org/10.31294/jki.v7i1.5744.g3253>
- Hendri, H. (2021). Implementasi Data Mining dengan Metode C4.5 untuk Prediksi Mahasiswa Penerima Beasiswa. *Indonesian Journal of Computer Science*, 10(2), 312-321. DOI: <https://doi.org/10.33022/ijcs.v10i2.3013>
- Lestari, S., Akmaludin., & Badrul, M. (2020). Implementasi Klasifikasi Naïve Bayes untuk prediksi kelayakan pemberian pinjaman pada Koperasi Anugrah Bintang Cemerlang. *Jurnal PROSISKO*, 7(1). DOI: <https://doi.org/10.30656/prosisko.v7i1.2129>
- Munandar, A., Huda, N., & Muhajirin. (2018). Analisis Piutang tak tertagih pada PT Astra Internasional Tbk. *Jurnal Manajemen dan Keuangan*, 7(2). DOI: <https://doi.org/10.33059/jmk.v7i2.859>
- Rahmayanti, A., Rusdiana, L., & Suratno. (2022). Perbandingan Metode Algoritma C4.5 dan Naïve Bayes untuk memprediksi kelulusan mahasiswa. *Walisongo Journal of Information Technology*, 4(1), 11-22. DOI: <http://dx.doi.org/10.21580/wjit.2022.4.1.9654>
- Siliyani, L., Junizar, I. A., Nuraeni, U., Tohidi, E., & Ali, I. (2020). Penerapan Algoritma Naïve Bayes untuk Mengetahui Kepuasan Mahasiswa Terhadap Layanan Administrasi Keuangan. *KOPERTIP: Jurnal Ilmiah Manajemen Informatika dan Komputer*, 4(3). DOI: <https://doi.org/10.32485/kopertip.v4i3.122>
- Sumpena, J., & Kurnia, N. H. (2019). Analisis Prediksi Kelulusan Siswa PKBM Paket C dengan Metode Algoritma Naïve Bayes. *Jurnal TEDC*, 13(2), 127-123.
- Susanto, W., & Indriyani, L. (2019). Analisis Penerapan Naïve Bayes untuk memprediksi Risiko Kredit Anggota Koperasi Keluarga Guru. *Jurnal Informatika*, 6(2), 262-270. DOI: <https://doi.org/10.31294/ji.v6i2.5724>
- Utami, D. Y., Nurlalah, E., Hasan, F. N. (2021). Comparison of Neural Network Algorithms Naïve Bayes and Logistic Regression to Find The Highest Accuracy in Diabetes. *Journal of Informatics and Telecommunication Engineering*, 5(1). DOI: 10.31289/jite.v5i1.5201
- Zulfauzi., & Alamsyah, N. (2020). Penerapan Algoritma Naïve Bayes untuk Memprediksi Penerimaan Mahasiswa Baru Studi Kasus Universitas Bina Insan Fakultas Komputer. *Jurnal Teknologi Informasi Mura*, 12(2). DOI:

