

# KLASIFIKASI KETEPATAN WAKTU LULUS MAHASISWA JURUSAN MATEMATIKA UNIVERSITAS PATTIMURA MENGGUNAKAN METODE *NAÏVE BAYES CLASSIFIER*

Ilham Rizki Maziz Tomu<sup>1</sup>, Malon Stivo Noya Van Delsen<sup>2</sup>, Norisca Lewaherilla<sup>3</sup>

Program Studi Statistika, Jurusan Matematika, Fakultas MIPA<sup>1,2,3</sup>

Universitas Pattimura<sup>1,2,3</sup>

Email: ilhamtomu5@gmail.com<sup>1</sup>, marlonnvd@gmail.com<sup>2</sup>, lewaherillanorisca@gmail.com<sup>3</sup>

**Corresponding Author:** Norisca Lewaherilla email: lewaherillanorisca@gmail.com

**Abstrak.** Perguruan Tinggi memiliki peranan yang besar dalam mencetak sumber daya manusia yang kompeten, sehingga kualitas dan akreditasi Perguruan Tinggi menjadi hal yang perlu diperhatikan. Salah satu indikator yang mempengaruhi kualitas dan akreditasi Perguruan Tinggi adalah persentase mahasiswa lulus tepat waktu. Oleh karena itu, perlu adanya perhatian khusus dari Perguruan Tinggi mengenai ketepatan waktu lulus mahasiswa. Dengan melihat lima faktor yang mempengaruhi ketepatan waktu lulus mahasiswa yakni Usia Masuk Perguruan Tinggi, Jenis Kelamin, Akreditasi Asal SMA, Jalur Penerimaan dan IPK diharapkan dapat memberikan prediksi mahasiswa tersebut lulus tepat waktu atau tidak. Metode yang digunakan dalam mengklasifikasi ketepatan waktu lulus mahasiswa adalah metode *Naïve Bayes Classifier*, karena metode ini memiliki akurasi yang tinggi dan dapat bekerja lebih baik pada kasus di dunia nyata. Hasil klasifikasi ketepatan waktu lulus mahasiswa Jurusan Matematik Tahun Akademik 2015/2016 sampai 2019/2020 dengan menggunakan metode *Naïve Bayes Classifier* dengan *K-Fold Cross Validation* untuk  $K=5$  sampai  $K=10$  bantuan software Rapid Miner berhasil diklasifikasikan ketepatan waktu lulus mahasiswa dengan persentase keakuratan tertinggi yaitu untuk  $K=10$  sebesar 70% dan nilai error sebesar 30% serta nilai AUC 72%, yang artinya model pada metode *Naïve Bayes Classifier* diklasifikasikan dengan baik.

**Kata Kunci:** Ketepatan Waktu Lulus Mahasiswa, *Naïve Bayes Classifier*, *K-Fold Cross Validation*, Klasifikasi.

**Abstract.** Universities have a big role in producing competent human resources, so that the quality and accreditation of universities is something that needs to be considered. One indicator that influences the quality and accreditation of higher education is the percentage of students graduating on time. Therefore, there needs to be special attention from universities regarding the timeliness of student graduation. By looking at five factors that influence the timeliness of a student's graduation, namely College Entry Age, Gender, High School Accreditation, Admission Route and GPA, it is hoped that this can provide a prediction of whether the student will graduate on time or not. The method used to classify student graduation timeliness is the *Naïve Bayes Classifier* method, because this method has high accuracy and can work better in real world cases. The results of the classification of the accuracy of graduating students in the Mathematics Department for the 2015/2016 to 2019/2020 Academic Year using the *Naïve Bayes Classifier* method with *K-Fold Cross Validation* for  $K=5$  to  $K=10$  with the help of Rapid Miner software successfully classified the accuracy of students' graduation time with the highest percentage of accuracy is for  $K=10$  at 70% and the error value is 30% and an AUC value of 72%, which means the model in the *Naïve Bayes Classifier* method is classified good.

**Keywords:** Punctuality of Student Graduation, *Naïve Bayes Classifier*, *K-Fold Cross Validation*, Classification.

## A. Pendahuluan

Perguruan Tinggi adalah penyelenggara pendidikan dengan jenjang yang lebih tinggi dari pada pendidikan menengah di jalur pendidikan sekolah. Perguruan Tinggi diselenggarakan untuk mempersiapkan peserta didik menjadi anggota masyarakat yang memiliki kemampuan akademis dan profesional yang dapat menerapkan, mengembangkan, dan menciptakan ilmu pengetahuan, teknologi, dan kesenian (Kementrian Riset, Teknologi, 2015). Perguruan Tinggi memiliki peran besar



dalam mencetak sumber daya manusia yang kompeten dan siap menghadapi era persaingan global. Oleh karena itu, peningkatan kualitas Perguruan Tinggi sangat dibutuhkan.

Salah satu aspek pengukuran kualitas dalam evaluasi keberhasilan penyelenggaraan pendidikan tinggi adalah mahasiswa. Mahasiswa menjadi unsur penting karena mahasiswa nantinya akan langsung berhubungan dengan stakeholder. Perguruan tinggi dengan kualitas lulusan yang baik akan mendapat kepercayaan stakeholder dalam merekrut maupun menggunakan lulusannya. Terdapat beberapa unsur penilaian perguruan tinggi yang berhubungan dengan mahasiswa. Unsur penilaian tersebut antara lain nilai ujian masuk, nilai akademik, prestasi yang dicapai mahasiswa, kompetensi yang dimiliki mahasiswa, dan persentase mahasiswa lulus tepat waktu (Sabilla Imama & Eranti Putri, 2017). Salah satu aspek yang mengakibatkan tingginya persentase mahasiswa lulus tepat waktu ialah lama masa studi mahasiswa.

Lama masa studi merupakan waktu lama tempuh bagi mahasiswa untuk menyelesaikan studinya. Menurut Peraturan Kementerian Riset Teknologi dan Pendidikan Tinggi Nomor 44 Tahun 2015 tentang standar nasional pendidikan tinggi, masa dan beban belajar untuk program sarjana (S1) adalah maksimal 7 tahun akademik dan telah menempuh paling sedikit 144 SKS (Zakariyah & Z. Ismaini, 2015). Berdasarkan data Sistem Akademik (SIKAD) untuk Jurusan Matematika pada mahasiswa Tahun Akademik 2015/2016 sampai 2019/2020 yang berstatus lulus, diperoleh persentase mahasiswa yang lulus tepat waktu sebesar 42,4%. Masih rendahnya jumlah persentase mahasiswa yang lulus tepat waktu dapat mengurangi penilaian dari kualitas dan akreditasi Perguruan Tinggi. Oleh karena itu, diharapkan mahasiswa Jurusan Matematika dapat menyelesaikan studinya dalam kurun waktu 4 tahun atau 8 semester.

Metode yang akan digunakan pada penelitian ini adalah *Naïve Bayes Classifier* yang merupakan bagian dari metode klasifikasi. Klasifikasi merupakan suatu proses untuk memperoleh model yang mampu membedakan kelas data. Tujuannya adalah untuk mendapatkan perkiraan kelas dari suatu objek yang labelnya belum diketahui. Sementara *Naïve Bayes* merupakan metode klasifikasi statistik yang dapat dipakai untuk memprediksi probabilitas keanggotaan suatu kelas (Bustami, 2014). *Naïve Bayes Classifier* memiliki ciri utama yaitu pada asumsi independensi (*naïve*) antar fitur (variabel bebas), yang mana *Naïve Bayes Classifier* menganggap semua fitur (variabel bebas) dalam dataset adalah independen satu sama lain, tetapi pengujian independensi terhadap fitur tidak perlu dilakukan secara eksplisit (Haldi Widiyanto, 2019).

Penelitian mengenai lama studi mahasiswa telah dilakukan oleh (Widaningsih, 2019) dengan menggunakan empat teknik klasifikasi yakni, Algoritma C4.5, *Naïve Bayes*, *K-Nearest Neighbor*, dan *Support Vector Machine*. Hasil penelitian tersebut dari keempat algoritma tersebut diperoleh bahwa algoritma *Naïve Bayes* merupakan algoritma terbaik untuk memprediksi kelulusan mahasiswa yang tepat waktu dengan  $IPK \geq 3$  dengan nilai akurasi (76,79%), *error* (23,21%), dan nilai *Area Under Curve* (AUC) (0,850). Penelitian yang dilakukan oleh (Oktanisa & Afif Supianto, 2018) mengenai perbandingan teknik klasifikasi dalam *data mining* untuk *bank direct marketing*, dengan menggunakan sembilan teknik klasifikasi, yakni *Tree*, *Naïve bayes*, *Constant* dan *Stochastic Gradient Descent* (SGD), *Random Forest*, *K-Nearest Neighbor* (KNN), *CN-2 Rule Inducer*, *Adaptive Boosting*, dan *Support Vector Machine*. Hasil penelitian tersebut diperoleh empat model dengan nilai akurasi yang tinggi, yaitu *Tree*, *Naïve Bayes*, *Constant* dan SGD sebesar 97,2%. Selain itu, diperoleh juga kesimpulan bahwa *Naïve Bayes* merupakan model dengan hasil implementasi yang cepat dan mudah serta cenderung membuat asumsi yang sangat kuat yang dapat mempengaruhi distribusi data. Kemudian, pada penelitian (Sidik & Ema, 2019) mengenai prediksi tingkat kelulusan tepat waktu mahasiswa menggunakan *K-Nearest Neighbor* dan *Naïve Bayes* diperoleh hasil akurasi terbaik adalah dengan menggunakan metode *Naïve Bayes*, yakni sebesar 85%, sedangkan akurasi metode *K-Nearest Neighbor* diperoleh sebesar 68,89%.

Metode *Naïve Bayes Classifier* selain mudah diimplementasikan, sederhana, cepat dan memiliki akurasi yang tinggi, metode ini juga tidak membutuhkan data *training* yang besar dan dapat bekerja lebih baik pada kasus di dunia nyata yang kompleks, serta mempunyai kecepatan dan akurasi



yang tinggi dan nilai error yang rendah apabila data memiliki jumlah yang besar (Effrida et al., 2017). Penelitian ini difokuskan untuk mengklasifikasi ketepatan waktu lulus mahasiswa Jurusan Matematika Universitas Pattimura Tahun Akademik 2015/2016 sampai 2019/2020 menggunakan metode *Naïve Bayes Classifier*, yang oleh beberapa penelitian terdahulu merupakan teknik klasifikasi dengan nilai akurasi yang tinggi. Variabel independen yang digunakan, yakni Indeks Prestasi Kumulatif (IPK), jenis kelamin, jalur penerimaan PT (Raaihatul Mashfia, 2022), usia masuk PT, dan akreditasi asal Sekolah Menengah Atas (SMA).

## B. Metodologi Penelitian

Data yang digunakan ialah data mahasiswa Jurusan Matematika Tahun Akademik 2015/2016 sampai 2019/2020 yang lulus pada Tahun Akademik 2019/2020 sampai 2023/2024. Data bersumber dari Sistem Informasi Akademik (SIKAD) Universitas Pattimura untuk Jurusan Matematika, dengan jumlah data observasi sebanyak 224. Variabel yang digunakan dalam penelitian ini tercantum dalam Tabel 1.

**Tabel 1. Variabel penelitian**

Nama Variabel	Kategori
	<b>Variabel Dependen</b>
Ketepatan Waktu Lulus (Y)	0 = Lulus Tepat Waktu ( $\leq 48$ bulan) 1 = Lulus Tidak Tepat Waktu ( $>48$ bulan)
	<b>Variabel Independen</b>
Usia Masuk Perguruan Tinggi (X1)	1 = $\leq 18$ Tahun 2 = $>18$ Tahun
Jenis Kelamin (X2)	1 = Laki-laki 2 = Perempuan
Akreditasi Asal Sekolah Menengah Atas (SMA) (X3)	1 = A 2 = B 3 = Lainnya
Jalur Penerimaan (X4)	1 = SNMPTN 2 = SBMPTN 3 = Mandiri
Indeks Prestasi Mahasiswa (IPK) (X5)	1 = 2.50 – 3.00 (Memuaskan) 2 = 3.01 – 3.50 (Sangat Memuaskan) 3 = 3.51 – 4.00 (Dengan pujian)

Pada proses pengolahan data menggunakan bantuan *software* RapidMiner. Langkah-langkah analisis yang akan dilakukan adalah sebagai berikut:

1. Menganalisis data secara deskriptif untuk mengetahui gambaran umum mengenai data.
2. Membagi data menjadi *data training* dan *data testing* menggunakan *K-Fold Cross Validation*.
3. Pembuatan model klasifikasi dari *data training* dengan menggunakan algoritma *Naïve Bayes Classifier*, dengan tahapan-tahapan sebagai berikut:
  - a. Menghitung probabilitas prior pada *data training*, yakni probabilitas mahasiswa lulus tepat waktu dan tidak tepat waktu.
  - b. Menghitung probabilitas setiap variabel pada *data training* terhadap masing-masing kelas.
  - c. Menggunakan *data testing* yang berdasarkan probabilitas *data training*, hitung perkalian probabilitas prior dengan atribut pada masing-masing kelas.
  - d. Mencari nilai probabilitas maksimal pada masing-masing kelas.
4. Menghitung ketepatan klasifikasi dengan cara menghitung akurasi, sensitivitas dan spesifitas dalam proses pengklasifikasian.



### C. Hasil Penelitian dan Pembahasan

Berdasarkan data 224 mahasiswa yang digunakan, diperoleh sebanyak 95 mahasiswa yang lulus tepat waktu dan 129 mahasiswa lulus tidak tepat waktu.

#### 1. Analisis Deskriptif

Berdasarkan data yang digunakan, persentase jumlah mahasiswa yang lulus tepat waktu sebesar 42,4%, yang artinya selama periode tersebut sebanyak 57,6% mahasiswa menjalani masa studinya lebih dari empat tahun atau lulus tidak tepat waktu. Lebih lanjut, deskriptif data berdasarkan variabel yang digunakan telah tersaji pada Tabel 2.

**Tabel 2. Persentase ketepatan waktu lulus mahasiswa**

Variabel	Indikator	Tepat Waktu (%)	Tidak Tepat Waktu (%)
Ketepatan Waktu Lulus		42,40	57,60
Usia Masuk PT	≤ 18 tahun	57,70	42,30
	>18 tahun	42,90	57,10
Jenis Kelamin	Laki-laki	37,70	62,30
	Perempuan	44,50	55,50
Akreditasi Sekolah asal SMA	A	60,00	40,00
	B	43,10	56,90
	Lainnya	15,20	84,80
Jalur Penerimaan	SNMPTN	55,60	44,40
	SBMPTN	43,00	57,00
	Mandiri	35,80	64,10
Indeks Prestasi Kumulatif (IPK)	Memuaskan	100	0
	Sangat memuaskan	37,40	62,60
	Dengan Pujian	0,39	0,12

#### 2. Probabilitas Variabel Ketepatan Waktu Lulus Mahasiswa

Metode *Naïve Bayes Classifier* dalam penerapannya dilakukan pembagian data menjadi data training dan data testing. Data training ini menjadi data pembelajaran untuk model klasifikasi sedangkan data testing digunakan untuk memprediksi data sehingga dapat mengetahui performa metode yang sudah dilatih sebelumnya. Dalam penelitian ini pembagian data training dan data testing menggunakan *K-Fold Cross Validation* dengan  $K=5$  sampai  $K=10$ . *Naïve bayes* merupakan metode klasifikasi berdasarkan Teorema Bayes. *Naïve bayes* memprediksi kemungkinan masa depan berdasarkan pengalaman masa lalu. Persamaan (1) berikut merupakan prediksi Bayes (Pujianto et al., 2017):

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

Dimana:

- $X$  : data testing yang kelasnya belum diketahui
- $H$  : hipotesis data  $X$  yang merupakan suatu kelas yang lebih spesifik
- $P(H|X)$  : probabilitas hipotesis  $H$  berdasar kondisi  $X$  (*posteriori probability*)
- $P(X|H)$  : probabilitas hipotesis  $X$  berdasarkan kondisi  $H$  (*likelihood*)
- $P(H)$  : probabilitas hipotesis  $H$  (*prior probability*)
- $P(X)$  : probabilitas hipotesis  $X$  (*predictor prior probability*)

Selanjutnya dihitung probabilitas prior dari data training untuk masing-masing kelas ketepatan waktu lulus mahasiswa ( $Y$ ) dan probabilitas bersyarat untuk setiap variabel independen ( $X$ ), hasilnya ditunjukkan pada Tabel 3.



**Tabel 3. Probabilitas variabel ketepatan waktu lulus mahasiswa**

Variabel	Indikator	Tepat Waktu	Tidak Tepat Waktu
Ketepatan Waktu Lulus		0,424	0,576
Usia Masuk PT	≤ 18 tahun	0,84	0,84
	>18 tahun	0,16	0,16
Jenis Kelamin	Laki-laki	0,27	0,33
	Perempuan	0,73	0,67
Akreditasi Sekolah asal SMA	A	0,63	0,31
	B	0,26	0,26
	Lainnya	0,11	0,430
Jalur Penerimaan	SNMPTN	0,21	0,13
	SBMPTN	0,38	0,47
	Mandiri	0,31	0,40
Indeks Prestasi Kumulatif (IPK)	Memuaskan	0	0,12
	Sangat memuaskan	0,61	0,75
	Dengan Pujian	0,39	0,12

### 3. Klasifikasi Menggunakan Metode *Naïve Bayes Classifier*

Pembagian data training dan testing dilakukan dengan menggunakan *K-Fold Cross Validation* dengan nilai  $K$  yang digunakan adalah  $K = 5$  sampai  $K = 10$ . Pembagian ini memberikan keakuratan yang seimbang antara estimasi dan efisiensi komputasional dengan bantuan Rapid Miner, serta memberikan representasi yang baik dalam variasi dalam data. Misalkan, pada klasifikasi ini peneliti menampilkan perhitungan untuk  $K = 10$  sebagai gambaran *Naïve Bayes Classifier* bekerja dalam klasifikasi. Selanjutnya, *data training* dilatih untuk memperoleh model klasifikasi. Berikut ini adalah hasil klasifikasi yang diperoleh.

**Tabel 4. Hasil klasifikasi ketepatan waktu lulus dengan *Naïve Bayes Classifier***

Lulus Tepat Waktu	Tidak Lulus Tepat Waktu
95	129

Tabel 4 menggambarkan bahwa hasil klasifikasi pada 224 data untuk ketepatan waktu lulus mahasiswa Jurusan Matematika Tahun Akademik 2015/2016 sampai 2019/2020, yakni 95 mahasiswa yang lulus tepat waktu dan 129 mahasiswa yang lulus tidak tepat waktu. Hasil klasifikasi menunjukkan bahwa lebih banyak mahasiswa yang lulus tidak tepat waktu yakni 58%, dibandingkan dengan mahasiswa yang lulus tepat waktu yakni sebanyak 42%.

### 4. Ketepatan Klasifikasi

Pada proses klasifikasi diperlukan adanya evaluasi kinerja dari sistem klasifikasi. Evaluasi kinerja dari model yang sudah dibangun oleh algoritma klasifikasi *Naïve Bayes Classifier* dapat dilakukan dengan menjumlahkan data *testing* yang diprediksi benar atau salah oleh model tersebut. Evaluasi kinerja klasifikasi ini menggunakan *Confusion Matrix* yang berupa tabel pencatat hasil kerja klasifikasi (M. Fadly et al., 2017).

Tabel 5 merupakan *Confusion Matrix* yang berisi hasil klasifikasi ketepatan waktu lulus mahasiswa menggunakan metode *Naïve Bayes Classifier*.

**Tabel 5. *Confusion matrix* ketepatan waktu lulus**

Aktual	Prediksi	
	Tepat Waktu	Tidak Tepat Waktu
Tepat Waktu	57	29
Tidak Tepat Waktu	38	100

Pada Tabel 5 terlihat bahwa hasil klasifikasi ketepatan waktu lulus mahasiswa Jurusan Matematika Tahun Akademik 2015/2016 sampai 2019/2020 menggunakan metode *Naïve Bayes Classifier* adalah sebanyak 57 data mahasiswa yang lulus tepat waktu dan terklasifikasi benar lulus



tepat waktu, sebanyak 29 data mahasiswa yang lulus tepat waktu tetapi terklasifikasi salah yakni terklasifikasi lulus tidak tepat waktu, sebanyak 100 data mahasiswa yang lulus tidak tepat waktu dan terprediksi benar lulus tidak tepat waktu, dan terdapat 38 data mahasiswa yang lulus tidak tepat waktu yang terklasifikasi lulus tepat waktu.

Selanjutnya untuk mengetahui ketepatan klasifikasi dari algoritma *Naïve Bayes Classifier* digunakan nilai akurasi, sebagai perbandingan antara data yang terklasifikasi benar dengan keseluruhan data (Anam et al., 2017). Nilai akurasi dari algoritma *Naïve Bayes Classifier* dihitung dari hasil *Confusion Matrix* pada Tabel 5, yakni:

$$\text{Akurasi} = \frac{57+100}{57+100+38+29} = 0,70$$

Untuk mengetahui berapa persen data kategori positif yang terklasifikasikan dengan benar oleh metode, digunakan nilai sensitivitas (Fiqih et al., 2020) berdasarkan *confusion matrix* pada Tabel 5, yakni:

$$\text{Sensitivitas} = \frac{57}{57+29} = 0,66$$

Kemudian, untuk untuk membandingkan proporsi data negatif yang diprediksi benar (Eka Wulansari, 2015) digunakan nilai spesifitas berdasarkan *confusion matrix* pada Tabel 5 yakni:

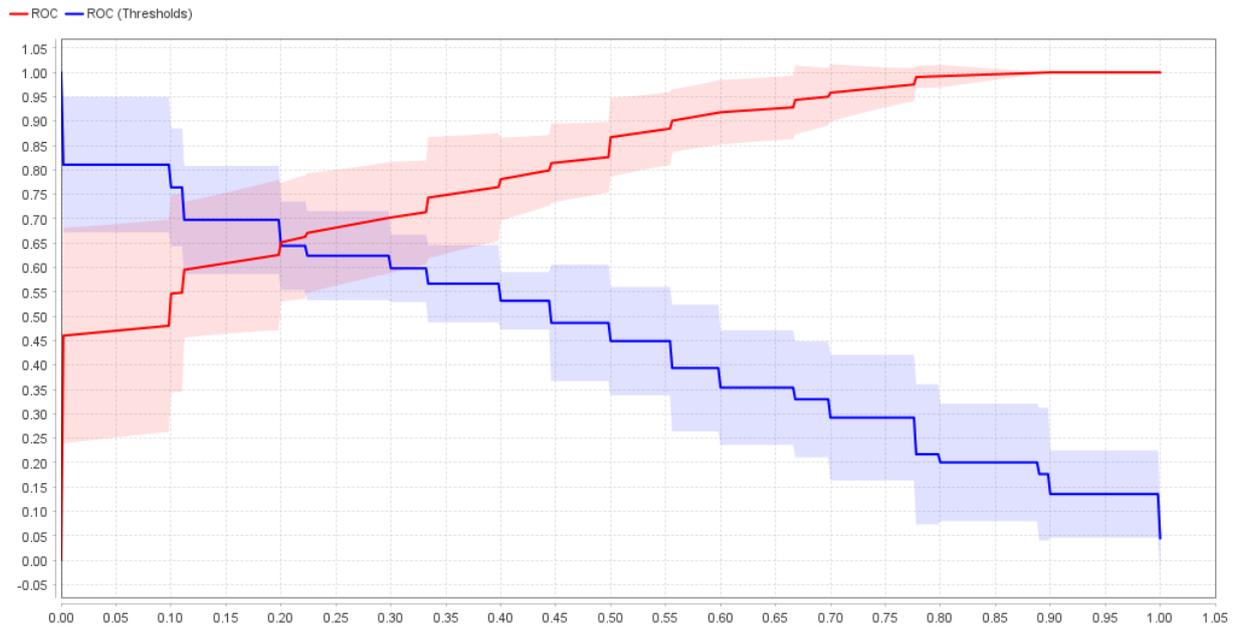
$$\text{Spesifitas} = \frac{100}{100+29} = 0,78$$

Berikutnya adalah pengujian metode *Naïve Bayes Classifier* dengan kurva *Receiving Operating Characteristic* (ROC) dan perhitungan *Area Under Curve* (AUC). Kurva *Receiving Operating Characteristic* (ROC) merupakan gambaran dari hubungan antara sensitivity dan specificity secara grafis (Arian R & Peter, 1998). Kurva ROC mengekspresikan *confusion matrix*, dimana pada grafiknya garis horizontal sebagai FP dan garis vertical sebagai TP (Triowali, 2016). Semakin dekat plot ROC ke sudut kiri atas, maka semakin tinggi pula akurasi dari keseluruhan tes (Mark H & Campbell, 1993). Area yang berada dibawah kurva ROC merupakan wilayah yang menunjukkan tingkat keakuratan dari model empirik dan dapat dihitung dengan *Area Under Curve* (AUC). AUC sendiri merupakan daerah berbentuk persegi yang nilainya selalu berada diantara 0 dan 1. Jika AUC yang dihasilkan  $< 0,5$ , maka model statistik yang dievaluasi memiliki tingkat keakuratan yang sangat rendah dan mengindikasikan bahwa model tersebut sangat buruk jika digunakan (Zou et al., 2007). Kurva ROC pada penelitian ini dapat dilihat pada Gambar 1, dimana terlihat bahwa garis yang berwarna merah merupakan kurva ROC dengan nilai 0,809. Hasil klasifikasi dikategorikan baik dikarenakan garis merah semakin dekat ke pojok kiri atas sedangkan untuk garis berwarna biru merupakan *Thresholds* (ambang). Pada kasus biner, nilai AUC dapat didekati dengan nilai *Balanced Accuracy* (Bekkar et al., 2013):

$$\begin{aligned} \text{AUC} &= \frac{1}{2}(\text{sensitifitas} + \text{spesifitas}) \\ \text{AUC} &= \frac{1}{2}(0,66 + 0,78) = 0,72 \end{aligned}$$



AUC: 0.809 +/- 0.066 (micro average: 0.809) (positive class: Tidak Tepat Waktu)



Gambar 1. Kurva ROC

Hasil perhitungan menggunakan metode *Naive Bayes Classifier* dalam mengklasifikasikan ketepatan waktu lulus mahasiswa Jurusan Matematika Tahun Akademik 2015/2016 sampai 2019/2020 memiliki tingkat akurasi yang baik yakni sebesar 0,70 atau 70%. Selanjutnya diperoleh nilai sensitivitas sebesar 0,66 atau 66%. Hal ini menunjukkan 66% keberhasilan metode *Naive Bayes Classifier* dalam menemukan kembali informasi pada klasifikasi ketepatan waktu lulus mahasiswa. Berikutnya diperoleh nilai spesifitas sebesar 78%. Hal ini menunjukkan 78% kemampuan metode *Naive Bayes Classifier* dalam membandingkan proporsi data negatif yang diklasifikasi benar. Kemudian, dengan menggunakan nilai AUC diperoleh hasil 72% yang artinya model pada metode *Naive Bayes Classifier* diklasifikasikan dengan baik.

##### 5. Hasil Klasifikasi dengan *K-Fold Cross Validation* untuk $K = 5$ sampai $K = 10$

Hasil klasifikasi metode *Naive Bayes Classifier* menggunakan *K-Fold Cross Validation* dengan nilai  $K = 5$  sampai  $K = 10$  dapat dilihat pada Tabel 6.

Tabel 6. Hasil pembagian data *training* dan data *testing* menggunakan *K-Fold Cross Validation*

<i>K-Fold Cross Validation</i>	Sensitifitas	Spesifitas	Akurasi
5	65,52%	76,74%	69,64%
6	63,74%	74,42%	68,82%
7	64,77%	75,97%	69,20%
8	64,77%	75,97%	69,20%
9	64,37%	75,97%	68,67%
10	66,28%	77,52%	70,06%

Tabel 6 ditampilkan untuk mengetahui pembagian dan hasil terbaik berdasarkan nilai akurasinya. Pembagian yang memiliki hasil akurasi terbaik adalah dengan  $K=10$  yakni sebesar 70,06%. Hasil klasifikasi diharapkan bermanfaat sebagai informasi kepada Jurusan Matematika Universitas Pattimura dalam meninjau ketepatan waktu lulus sesuai kurikulum yang berlaku berdasarkan indikator-indikator pada variabel penelitian ini. Evaluasi pembelajaran secara berkala dapat terus dilakukan sehingga strategi pembelajaran semakin ditingkatkan dalam upaya menambah jumlah mahasiswa lulus tepat waktu.



#### D. Kesimpulan

Ketepatan waktu lulus mahasiswa Jurusan Matematika Universitas Pattimura Tahun Akademik 2015/2016 sampai 2019/2020 dengan menggunakan metode *Naïve Bayes Classifier* dapat diklasifikasikan dengan persentase keakuratan tertinggi yaitu  $K=10$  sebesar 70% dan *error* sebesar 30%. Nilai AUC diperoleh 72% yang artinya model pada metode *Naïve Bayes Classifier* diklasifikasikan dengan baik. Penelitian lanjutan dapat dicoba dengan jumlah data yang besar sehingga memungkinkan hasil penelitian lebih baik lagi. Variabel yang dapat digunakan juga dapat ditambah, seperti kota tempat tinggal, dan kegiatan yang dilakukan selama di kampus.

## DAFTAR PUSTAKA

- Anam, S., Sugiman, S., & Sunarmi, S. (2017). Ketepatan Klasifikasi Dengan Menggunakan Metode Multivariate Adaptive Regression Spline (MARS) Pada Data Kelompok Rumah Tangga Kabupaten Cilacap. *Jurnal Matematika*, 6(1).
- Arian R, V. E., & Peter, M. (1998). Receiver Operating Characteristic (ROC) Analysis: Basic Principles and Application in Radiology. *European Journal of Radiology*, 27(2).
- Bekkar, M., Hassiba, K. D., & Taklit, A. A. (2013). Evaluation Measures for Models Assesment Over Imbalanced Data Sets. *Journal of Information Engginering and Application*, 3(10).
- Bustami. (2014). Penerapan Algoritma Naive Bayes Untuk Mengklasifikasikan Data Nasabah Asuransi. *Jurnal Informatika*, 8(1).
- Effrida, M., Ariwisanto Sianturi, F., & Manalu, R. (2017). Penerapan Algoritma Naive Bayes untuk Memprediksi Jumlah Produksi Barang Berdasarkan Data Persediaan dan Jumlah Pemesanan Pada CV. Papadan Mama Pastries. *Jurnal Mantik Penusa*, 1(2).
- Eka Wulansari, F. (2015). Analisa Data Mining Untuk Prediksi Penyakit Hepatitis dengan Menggunakan Metode Naive Bayes dan Support Vector Machine. *Jurnal Khatulistiwa Informatika*, 3(1).
- Fiqih, S., Zamharini, Z., & Apun Syaripudin, M. (2020). Prediksi Ketepatan Waktu Lulus Mahasiswa Menggunakan Algoritma C4.5 Pada Fakultas Dakwah dan Ilmu Komunikasi UIN Raden Intan Lampung. *Jurnal Ilmiah Matrik*, 22(1).
- Haldi Widiyanto, M. (2019). Algoritma Naive Bayes. *Binus Universsity*.
- Kementrian Riset, Teknologi, dan P. T. (2015). *Peraturan Menteri Riset, Teknologi, dan Pendidikan Tinggi Nomor 44 Tahun 2015 tentang Standar Nasional Pendidikan Tinggi*.
- Sabilla Imama, W., & Eranti Putri, T. (2017). Prediksi Ketepatan Waktu Lulus Mahasiswa dengan k-Nearest Neighbor dan Naive Bayes Classifier (Studi Kasus Prodi D3 Sistem Informasi Universitas Airlangga). *Jurnal Komputer Terapan*.
- Sidik, R., & Ema, U. (2019). Prediksi Tingkat Kelulusan Tepat Waktu dengan Metode Naive Bayes dan K-Nearst Neighbour. *Jurnal Informasi Dan Komputer*, 7(1).
- M. Fadly, R., D., D., & Nurma, I. (2017). Klasifikasi Untuk Diagnosa Diabetes Menggunakan Metode



Bayesian Regularization Neural Network (RBNN). *Jurnal Informatika*, 11(1).

Mark H, Z., & Campbell, G. (1993). Receiver Operating Characteristic (ROC) Plots: A Fundamental Evaluation Clinical Medicine. *Clinical Chemistry*, 39(4).

Oktanisa, I., & Afif Supianto, A. (2018). Perbandingan Teknik Klasifikasi dalam Data Mining untuk Bank Direct Marketing. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 5(5).

Triowali, R. (2016). Perbandingan Metode Naive Bayes Classifier dengan Metode Decision Tree (C4.5) untuk Menganalisa Kelancaran Pembiayaan (Study Kasus: KSPPS/BMT Al-Fadhila). *Jurnal TIM Darmajaya*, 2(1).

Widaningsih, S. (2019). Perbandingan Metode Data Mining Untuk Prediksi Nilai dan Waktu Kelulusan Mahasiswa Prodi Teknik Informatika Dengan Algoritma C4.5, Naive Bayes, Knn dan Svm. *Jurnal Tekno Insentif*, 13(1).

Zakariyah, & Z.Ismaini. (2015). Analisis Regresi Logistik Ordinal Pada Prestasi Belajar Lulusan Mahasiswa di ITS Berbasis SKEM. *Jurnal Sains Dan Seni Pomits*.

Zou, K. H., A.James, O., & Mauri, L. (2007). Receiver Operating Characteristic (ROC) Analysis For Evaluating Diagnostic Test and Predictive Models. *Circulation*, 115(5).

