

PENERAPAN MODEL KLASIFIKASI NAÏVE BAYES UNTUK PREDIKSI MAHASISWA TEPAT WAKTU

M Riski Qisthiano

Institut Teknologi dan Bisnis Nasional (ITBN)
Jl. Cahaya Berlian No.81 Pangkalan Balai, Kec. Banyuasin 3, Kab. Banyuasin
thiano72@gmail.com

ABSTRAK

Banyak teknik pengukuran dalam prediksi mahasiswa tepat waktu, kriteria yang digunakan mulai dari jurusan, jenis kelas dan nilai semester mahasiswa yang merupakan salah satu penentu bagaimana mahasiswa bisa lulus tepat waktu atau tidak di akhir pendidikannya. Maka dibutuhkan suatu model untuk melakukan prediksi tingkat kelulusan mahasiswa tepat waktu, dengan menggunakan data alumni yang datanya didapat dari beberapa Perguruan Tinggi di Kota Palembang. Model yang digunakan ini menggunakan algoritma *Naïve Bayes* yang berfungsi sebagai salah satu model untuk melakukan klasifikasi. Dataset yang digunakan adalah data alumni yang sudah dikumpulkan dari beberapa Perguruan Tinggi, sedangkan kriteria dan atribut yang digunakan adalah jurusan, Perguruan Tinggi setiap mahasiswa, jenis kelas pilihan dan nilai setiap semester mahasiswa yang diambil dari semester 1 sampai dengan 4, tahun lulus setiap mahasiswa, dan angkatan pada saat mahasiswa masuk. Lalu dari atribut dan model yang digunakan, peneliti menggunakan alat bantu untuk mengelola data yaitu *Rapidminer* untuk melakukan pemrosesan dataset yang sudah disiapkan. Selanjutnya pengujian menggunakan 5 kali uji validasi dengan membagi dataset kedalam training dan testing. Hasil penelitian ini merupakan akurasi dari model prediksi yang dilakukan, yang dimana hasil akurasi yang didapat dari alat bantu *Rapidminer* dan algoritma *Naïve Bayes* adalah 0.8264.

Kata Kunci: *Naïve Bayes*, *Rapidminer*, Perguruan Tinggi, Mahasiswa, Alumni

ABSTRACT

There are many measurement techniques in predicting students on time, the criteria used ranging from majors, class types and student semester grades are one of the determinants of how students can graduate on time or not at the end of their education. So we need a model to predict student graduation rates on time, using alumni data whose data is obtained from several universities in the city of Palembang. The model used is using the Naïve Bayes algorithm which serves as one of the models for classifying. The dataset used is alumni data that has been collected from several universities, while the criteria and attributes used are majors, colleges for each student, type of class choice and grades for each semester of students taken from semester 1 to 4, year of graduation for each student, and the class at the time the student enters. Then from the attributes and models used, the researcher uses a tool to manage data, namely Rapidminer to process the datasets that have been prepared. Furthermore, the examiner used 5 validation tests by dividing the dataset into training and testing. The results of this study are the accuracy of the prediction model carried out, where the accuracy results obtained from the Rapidminer tool and the Naïve Bayes algorithm are 0.8264.

Key Word: *Naïve Bayes*, *Rapidminer*, University, Student, Alumni.

PENDAHULUAN

Permasalahan yang terjadi dalam suatu Perguruan Tinggi ialah bagaimana menentukan setiap mahasiswa itu kedalam kategori lulusan tepat waktu atau tidak tepat waktu. Kelulusan mahasiswa tepat waktu merupakan salah satu penilaian dalam proses akreditasi perguruan tinggi (Rohmawan, 2018). Maka untuk menjalankan penelitian ini dibutuhkan beberapa dataset lengkap mengenai data kelulusan mahasiswa, seperti data detail mahasiswa mulai dari NIM, nama, jurusan, fakultas, tahun masuk kuliah dan tahun kelulusan mahasiswa. Dalam hal ini sampel data yang diambil dari beberapa

sumber Perguruan Tinggi terkemuka di kota Palembang baik itu Universitas, Institut, ataupun Sekolah Tinggi sebagai penentu dalam hal prediksi tingkat kelulusan mahasiswa tepat waktu, normalnya mahasiswa yang dapat lulus tepat waktu adalah mahasiswa yang masa pendidikannya mampu ditempu selama empat (4) tahun masa studi, jikalewat dari itu maka mahasiswa tersebut digolongkan mahasiswa tidak tepat waktu.

Penelitian yang berjudul Metode *Naïve Bayes* Untuk Prediksi Kelulusan, pada penelitian ini menghasilkan nilai Presentase keakuratan

menunjukkan keefektifan dataset Penerimaan Mahasiswa Baru yang diterapkan ke dalam metode *Naïve Bayes Clasification* yang mencapai 94 %. Lalu Nilai Presentase keakuratan menunjukkan keefektifan dataset Penerimaan Mahasiswa Baru yang diterapkan ke dalam metode *Naïve Bayes Clasification*. Impelementasi *Naïve Bayes* menggunakan aplikasi WEKA dapat menelusuri karakteristik atribut dari dataset dengan luaran Pilihan Lulus. Pengelompokkan Pilihan Lulus dilakukan berdasarkan atribut terpilih yaitu Prodi, Pilihan Pertama, Pilihan Kedua dan Nilai Rata-rata(Syarli & Muin, 2018).

Sedangkan penelitian yang berjudul Penerapan Algoritma *Naïve Bayes Classifier* Dalam Menentukan Strategi Promosi Penerimaan Mahasiswa Baru, Berdasarkan data hasil dan pembahasan diatas dapat kita jadikan sebagai acuan untuk menentukan strategi promosi pada Universitas Bina Darma Palembang dengan melihat data atribut dari Alamat, Asal Sekolah dan Program Studi. Dimana dari data Penerimaan Mahasiswa Baru (PMB) dari tahun 2014-2016 dengan jumlah data 6495 yang registrasi 6121 dan tidak registrasi 374(Mirza, 2019).

Pada penelitian yang berjudul Komparasi Kinerja Algoritma C.45 Dan *Naïve Bayes* Untuk Prediksi Kegiatan Penerimaan mahasiswa Baru, pada penelitian ini didapatkan hasil tertinggi yaitu pengujian dengan data training sebanyak 2133 dan data testing sebanyak 533 yaitu algoritma C45 memiliki nilai akurasi lebih tinggi sebesar 88.74% dibandingkan algoritma *Naïve Bayes* yang memiliki akurasi sebesar 87.24%(Yahya & Jananto, 2019).

Solusi yang ditawarkan berdasarkan permasalahan diatas adalah dengan memprediksi akurasi tepat waktu atau tidak tingkat kelulusan mahasiswa dengan model *Naïve Bayes*. Model yang diambil dalam penelitian ini adalah *Naïve Bayes*, memiliki beberapa kelebihan yaitu, cepat dalam proses perhitungan, algoritma yang sederhana dan akurasi yang tinggi. Namun probabilitas pada *Naïve Bayes Classifier* tidak bisa mengukur seberapa besar tingkat keakuratan sebuah prediksi(Muhamad et al., 2017).

Penelitian yang berjudul *The Combination of Naïve Bayes and Particle Swarm Optimization Methods of Student's Graduation Prediction*, Hasilnya adalah metode *Naïve Bayes* memiliki nilai akurasi yang lebih tinggi jika digabungkan dengan *Particle Swarm Optimization*. Dengan demikian universitas dapat lebih mudah memprediksi apakah siswa lulus tepat waktu untuk yang akan datang periode kelulusan. Nilai akurasi peneltian ini mencapai 100% dari 10 sampel yang digunakan berdasarkan sampel acak pada mahasiswa UIGM Palembang(Purnamasari et al., 2019).

Penelitian yang berjudul *Educational Data Mining* untuk Prediksi Kelulusan Mahasiswa Menggunakan *Algoritme Naïve Bayes Classifier*, Hasil penelitian ini membuktikan bahwa NBC berhasil diterapkan untuk memprediksi kelulusan mahasiswa. Prediksi dari kelulusan mahasiswa tersebut mampu menghasilkan akurasi 73,725%, presisi 0,742, recall 0,736 dan F-measure 0,735(Sutoyo & Almaarif, 2020).

Berdasarkan penelitian-penelitian diatas maka salah satu fungsi penerapan data mining dengan metode *Naïve Bayes* dalam Perguruan Tinggi adalah untuk memprediksi tingkat kelulusan mahasiswa tepat waktu yang dalam hal ini studi kasus diambil dari beberapa Universitas, Institut ataupun Sekolah Tinggi yang ada di Kota Palembang. *Naïve Bayes* merupakan pengklasifikasian dengan metode probabilitas dan statistik yaitu memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya sehingga dikenal sebagai Teorema Bayes. *Naïve Bayes* untuk setiap kelas keputusan, menghitung probabilitas dengan syarat bahwa kelas keputusan adalah benar, mengingat vektor informasi obyek. Algoritma ini mengasumsikan bahwa atribut obyek adalah independen(Olson & Delen, 2008).

Metode *Naïve Bayes* juga termasuk dalam golongan klasifikasi yang dimana Klasifikasi merupakan proses identifikasi kelompok dari suatu obyek berdasarkan kesamaan fitur tertentu, dimana setiap kelompok telah terbentuk melalui suatu proses tertentu(Dillak et al., 2012). Alat bantu yang digunakan dalam memproses data ini adalah *Rapidminer* yang dimana *Rapidminer* merupakan

perangkat lunak yang dapat diakses oleh siapa saja dan bersifat terbuka (*open source*). *Rapidminer* ini dijadikan sebuah solusi untuk menganalisa terhadap *data processing*. Pada *Rapidminer* ini digunakan berbagai teknik seperti teknik deskriptif dan prediksi. *Rapidminer* menggunakan bahasa Java untuk pengoperasian nya(Sari et al., 2020).

Maka pada penelitian ini akan di implementasikan metode tersebut dan dicari hasil akurasi dengan model *Naïve Bayes* dan menggunakan alat bantu *Rapidminer* serta hasil pengelompokan dari tingkat akurasi prediksi tersebut, sehingga dapat menggali informasi mengenai hasil klasifikasi terhadap prediksi tingkat kelulusan mahasiswa tepat waktu, terhadap data mahasiswa yang masih aktif menjalankan studi serta dapat mengukur hasil dari model ini dalam hal data mining pada prediksi tingkat kelulusan mahasiswa tepat waktu khususnya di beberapa Universitas, Institut ataupun Sekolah Tinggi yang ada di Kota Palembang.

METODE PENELITIAN

Desain Penelitian

Adapun desain penelitian ini terdiri dari beberapa tahapan hingga mendapatkan hasil yang akurat. Penelitian ini menggunakan pendekatan *Naïve Bayes* sebagai pengembangan model baru terhadap tingkat kelulusan mahasiswa tepat waktu. Tools yang digunakan adalah program komputer Jupyter Notebook untuk mengolah data-data yang ada sedangkan alat bantu yang digunakan adalah *Rapidminer*. Pendekatan penelitian ini secara garis besar terdiri dari lima tahap, antara lain:

1. Mulai atau persiapan. Pada tahap persiapan penelitian, penulis melakukan pengamatan (observasi) terlebih dahulu pada objek secara langsung yaitu dengan mengumpulkan *dataset*.
2. Peneliti selanjutnya melakukan pemrosesan data yang sudah dikumpulkan dengan metode penelitian KDD.
3. Hasil dan Pembahasan. Pada tahapan ini membahas mengenai hasil pengenalan pola angka dengan pendekatan *Naïve Bayes* dalam implemetasinya.
4. Kesimpulan. Tahapan ini, penulis menarik kesimpulan dari hasil dan pembahasan yang telah dilakukan pada tahapan sebelumnya.

Data Penelitian

Untuk menunjang penelitian ini maka diperlukannya data penelitian. Data penelitian terdiri dari data sample, teknik pengumpulan data jenis dan sumber data beserta analisis data

Teknik Pengumpulan Data

Metode pengumpulan data tesis ini yang digunakan terdiri dari dua macam teknik pengumpulan data:

1. Data diambil dari Beberapa Universitas, Institut ataupun Sekolah Tinggi.
2. Studi Literatur (*Literature Research*). Melakukan penelitian dengan cara mempelajari buku, jurnal, dan bacaan lainnya yang berkaitan dengan judul penelitian

HASIL DAN PEMBAHASAN

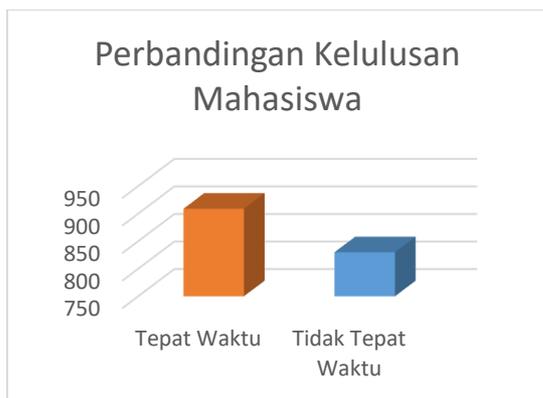
Pengumpulan Data

Setelah melakukan analisis data pada bab sebelumnya, maka pada bab ini merupakan hasil dan pembahasan dari penelitian yang berjudul Pengembangan Model untuk prediksi tingkat kelulusan mahasiswa tepat waktu. Dengan menggunakan data alumni maka akan dapat dicari tahu seberapa besar tingkat tepat waktu mahasiswa yang masih menjalani studi untuk lulus tepat waktu, dari hasil pengumpulan data maka sudah di dapat beberapa sampel data dari beberapa Universitas ataupun institusi pendidikan, data alumni.

Tabel 1. Tabel Jumlah Sampel Berdasarkan PT

No	Perguruan Tinggi	Jumlah Data
1	UIN Raden Fatah Palembang	513
2	Universitas Bina Darma Palembang	700
3	Universitas PGRI Palembang	90
4	Stisipol Candradimuka Palembang	436

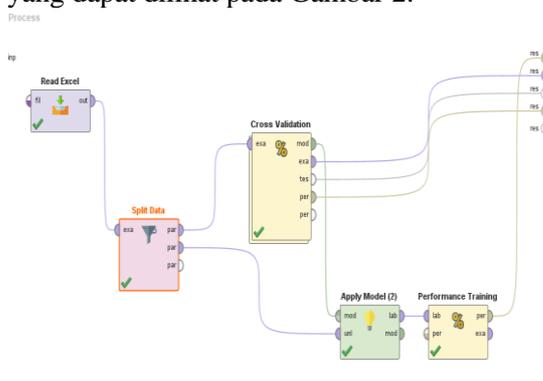
Jumlah data alumni yang telah dikumpulkan berjumlah 1739 mahasiswa dari berbagai Perguruan Tinggi yang ada di Kota Palembang, untuk data mahasiswa yang lulus Tepat Waktu berjumlah 909 sedangkan yang tidak tepat waktu berjumlah 830, berikut merupakan grafik perbandingan dari data kelulusan mahasiswa



Gambar 1. Grafik Label Data

Pengujian Dengan Rapidminer

Berikut ini uji coba dengan menggunakan tools yang berbeda untuk melihat perbandingan, tools yang digunakan kali ini adalah *Rapidminer*, berikut merupakan hasil uji data dengan bantuan tools *Rapidminer* yang dapat dilihat pada Gambar 2.



Gambar 2. Pengujian Dengan Rapidminer

Hasil Pengujian Dengan Rapidminer

Berikut ini merupakan hasil dari uji coba dengan menggunakan tools yang *Rapidminer* yang dimana hasil tersebut merupakan hasil uji data dengan bantuan tools *Rapidminer* yang dapat dilihat pada Tabel 2.

Tabel 2. Tabel Hasil Pengujian

	Tepat Waktu	Tidak Tepat Waktu	Precision
Pred. Tepat Waktu	382	78	83.04%
Pred. Tidak Tepat Waktu	73	337	82.20%
Recall	83.96%	81.20%	

Hasil Uji Validasi

Berikut ini merupakan hasil dari uji coba Penerapan Model Klasifikasi *Naïve Bayes* Untuk Prediksi Mahasiswa Tepat Waktu dengan menggunakan tools *Rapidminer* dengan 1739 record data dengan melakukan 5

kali ujicoba *confussion matrix* dimana hasil tersebut akan dapat dilihat pada Tabel 3.

Tabel 3. Tabel Hasil Uji Validasi

Training	Testing	Pembagian Data	Hasil Presentase
1565	174	90:10	79.31%
1391	348	80:20	79.02%
1217	522	70:30	81.80%
1043	696	60:40	81.47%
870	870	50:50	82.64%

Berdasarkan hasil pengujian diatas didapat hasil akurasi tertinggi yaitu dengan menggunakan 50% data training dan 50% data testing dari total 1739 record data yang menghasilkan 82.64% akurasi terhadap Penerapan Model Klasifikasi *Naïve Bayes* Untuk Prediksi Mahasiswa Tepat Waktu Berikut ini perhitungan manual terhadap hasil pengujian validasi.

- Hasil *Recall*

$$\text{Tepat Waktu} = \frac{382}{382+73} = \frac{382}{455} = 0.8396 \text{ (83.96\%)}$$

$$\text{Tidak Tepat Waktu} = \frac{337}{78+337} = \frac{337}{415} = 0.8120 \text{ (81.20\%)}$$
- Hasil *Precision*

$$\text{Tepat Waktu} = \frac{382}{382+78} = \frac{382}{460} = 0.8304 \text{ (83.04\%)}$$

$$\text{Tidak Tepat Waktu} = \frac{337}{73+337} = \frac{337}{410} = 0.8220 \text{ (82.20\%)}$$
- Hasil Akurasi

$$\text{Accuracy} = \frac{382+337}{382+78+73+337} = \frac{719}{870} = 0.8264 \text{ (82.64\%)}$$

SIMPULAN DAN SARAN

Kesimpulan akhir dari penelitian yang dijalankan, pada penelitian yang berjudul Penerapan Model Klasifikasi *Naïve Bayes* Untuk Prediksi Mahasiswa Tepat Waktu, yang menggunakan sampel data alumni dari beberapa Perguruan Tinggi yang ada di Kota Palembang. Hasil data uji yang didapat memiliki tingkat Akurasi sebesar 0.8264 atau sebesar 82.64% dengan pembagian data 50% data training dan 50% data test yang digunakan dari 1739 record data dengan menggunakan Algoritma *Naïve Bayes*.

DAFTAR PUSTAKA

Dillak, R. Y., Pangestuty, D. M., & Bintiri, M. G. (2012). *Klasifikasi Jenis Musik*

- Berdasarkan File Audio Menggunakan Jaringan Syaraf Tiruan Learning Vector Quantization. *Seminar Nasional Informatika, 2012(semnasIF)*, 122–125.
- Mirza, A. H. (2019). Application of *Naïve Bayes* Classifier Algorithm in Determining New Student Admission Promotion Strategies. *Journal of Information Systems and Informatics*, 1(1), 14–28. <https://doi.org/10.33557/journalisi.v1i1.2>
- Muhamad, H., Prasojo, C. A., Sugianto, N. A., Surtiningsih, L., & Cholissodin, I. (2017). Optimasi *Naïve Bayes* Classifier Dengan Menggunakan Particle Swarm Optimization Pada Data Iris. *Jurnal Teknologi Informasi Dan Ilmu Komputer (JTIIK)*, 4(September), 180–184.
- Olson, D. L., & Delen, D. (2008). *Advanced Data Mining Techniques*. Springer.
- Purnamasari, E., Rini, D. P., & Sukemi. (2019). The Combination of *Naïve Bayes* and Particle Swarm Optimization Methods of Student's Graduation Prediction. *Jurnal Ilmiah Teknik Elektro Komputer Dan Informatika (JITEKI)*, 5(2). <https://doi.org/10.26555/jiteki.v5i2.15317>
- Rohmawan, E. P. (2018). Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode Decision Treed dan Artificial Neural Network. *Jurnal Ilmiah MATRIK*, 20(1), 21–30.
- Sari, Y. R., Sudewa, A., Lestari, D. A., & Jaya, T. I. (2020). Penerapan Algoritma K-Means Untuk Clustering Data Kemiskinan Provinsi Banten Menggunakan Rapidminer. *CESS (Journal of Computer Engineering, System and Science)*, 5(2), 192. <https://doi.org/10.24114/cess.v5i2.18519>
- Sutoyo, E., & Almaarif, A. (2020). Educational Data Mining untuk Prediksi Kelulusan Mahasiswa Menggunakan Algoritme *Naïve Bayes* Classifier. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 4(1), 95–101. <https://doi.org/10.29207/RESTI.V4I1.1502>
- Syarli, S., & Muin, A. A. (2018). Metode *Naïve Bayes* Untuk Prediksi Kelulusan (Studi Kasus: Data Mahasiswa Baru Perguruan Tinggi). *Jurnal Ilmu Komputer*, 2(1), 22–26. <https://fikom-unasman.ac.id/ejournal/index.php/jikom/article/view/26/17>
- Yahya, N., & Jananto, A. (2019). Komparasi Kinerja Algoritma C.45 Dan *Naïve Bayes* Untuk Prediksi Kegiatan Penerimaan Mahasiswa Baru. *Prosiding SENDI, 2014*, 978–979. <https://www.unisbank.ac.id/ojs/index.php/sendu/article/view/7389/2369>