Volume 5, Nomor 3, Juli 2021, Page 987-994 ISSN 2614-5278 (media cetak), ISSN 2548-8368 (media online) Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/mib DOI 10.30865/mib.v5i3.3030



Pengembangan Model Untuk Prediksi Tingkat Kelulusan Mahasiswa Tepat Waktu dengan Metode *Naïve Bayes*

M Riski Qisthiano, Tri Basuki Kurniawan*, Edi Surya Negara, Muhammad Akbar

Magister Teknik Informatika, Universitas Bina Darma, Palembang, Sumatera Selatan Email: ¹thiano72@gmail.com , ².*tribasukikurniawan@binadarma.ac.id, ³e.s.negara@binadarma.ac.id, ⁴akbartea@gmail.com Email Penulis Korespondensi: tribasukikurniawan@binadarma.ac.id

Abstrak—Banyak parameter yang mempengaruhi ketepatan waktu kelulusan mahasiswa, mulai dari minat mahasiswa terhadap jurusan tertentu, jenis kelas yang dipilih, hingga nilai setiap semester yang didapat. Hal ini merupakan faktor 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 yang ada 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 untuk attribut-attribut yang digunakan adalah Jurusan, Perguruan Tinggi, Jenis Kelas, Nilai IP Sementara dari semester 1 hinga 4, tahun lulus, dan angkatan kuliah. Lalu dari attribut dan model yang digunakan, peneliti menggunakan bahasa pemrograman Python 3 dan tools Jupyter Notebook untuk melakukan pemrosesan dataset yang sudah disiapkan. Selanjutnya pembagian dataset dibagi 70% untuk data training dan 30% sebagai data testing. Untuk menguji proses algoritma yang digunakan peneliti menggunakan K-Fold Validation Hasil penelitian ini merupakan akurasi dari model prediksi yang dilakukan, yang dimana hasil akurasi yang didapat dari bahasa pemrograman Python 3 dan algoritma Naïve Bayes adalah 0.8103.

Kata Kunci: Naïve Bayes; Python 3; Perguruan Tinggi; Mahasiswa; Alumni

Abstract—Many parameters affect the timeliness of student graduation, starting from the student's interest in certain majors, the type of class chosen, to the grades for each semester obtained. This is a determining factor in how students can graduate on time or not at the end of their education. So a model is needed to predict student graduation rates on time, using alumni data whose data is obtained from several universities in Palembang City. The model used is a Naïve Bayes algorithm which serves as a model for classification. The dataset used is alumni data that has been collected from several universities, while the attributes used are the Department, College, Class Type, Temporary IP Value from semester 1 to 4, graduation year, and college generation. Then from the attributes and models used, the researcher used the Python 3 programming language and the Jupyter Notebook tools to process the prepared dataset. Furthermore, the distribution of the dataset is divided by 70% for training data and 30% for testing data. To test the algorithmic process used by researchers using K-Fold Validation. The results of this study are the accuracy of the prediction model carried out, where the accuracy results obtained from the Python 3 programming language and the Naïve Bayes algorithm are 0.8103.

Keywords: Naïve Bayes; Python 3; University; Student; Alumni

1. PENDAHULUAN

Permasalahan yang terjadi pada sebuah lembaga pendidikan terutama pada Perguruan Tinggi baik itu negeri ataupun swasta ialah mencari prediksi tingkat kelulusan mahasiswa tepat waktu, karena dengan lulus tepat waktu merupakan salah satu penilaian dalam proses akreditasi Perguruan Tinggi, sehingga semakin banyak mahasiswa yang tepat waktu kelulusannya maka semakin baik penilaian yang didapat pada saat akreditasi, 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 prediski 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, jika lewat dari itu maka mahasiswa tersebut digolongkan mahasiswa tidak tepat waktu.

Penelitian yang berjudul Metode Naive Bayes Untuk Prediksi Kelulusan, pada penelitian ini menghasilkan nilai Presentase keakuratan menunjukkan kefektifan dataset Penerimaan Mahasiswa Baru yang diterapkan ke dalam metode *Naïve Bayes* Clasification yang mencapai 94 %. Lalu Nilai Presentase keakuratan menunjukkan kefektifan 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[1],

Penelitian yang berjudul Perbandingan Dan Analisis Metode Klasifikasi Untuk Menentukan Konsentrasi Jurusan, hasil penelitian ini menunjukan pemilihan terhadap konsentrasi jurusan hanya berdasarkan pada keinginan mahasiswa ataupun ikut-ikutan teman, untuk itu dirasa perlu adanya preferensi dalam membantu mahasiswa memilih konsentrasi. Metode yang digunakan yaitu C4.5 dan *Naïve Bayes* dengan menggunakan aplikasi *Rapid Miner* sebagai alat bantu untuk mengklasifikasikan penjurusan mahasiswa. Pada penelitian ini diketahui algoritma C4.5 memiliki tinggkat akurasi 48,06 % dan naïve bayes 42,79% [2].

M Riski Qisthiano, Copyright ©2021, MIB, Page 987 Submitted: 30/05/2021; Accepted: 01/07/2021; Published: 31/07/2021

Volume 5, Nomor 3, Juli 2021, Page 987-994ISSN 2614-5278 (media cetak), ISSN 2548-8368 (media online)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/mib
DOI 10.30865/mib.v5i3.3030



Sedangkan penelitian yang berjudul Penerapan Algoritma Naive 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[3].

Pada penelitian yang berjudul Komparasi Kinerja Algoritma C.45 Dan *Naive Bayes* Untuk Prediksi Kegiatan Penerimaanmahasiswa 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 Naive Bayes yang memiliki akurasi sebesar 87.24% [4].

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*, kelebihan *Naïve Bayes Classifier* dibandingkan algoritma lain adalah pada kemampuannya mengklasifikasi dokumen dengan kesederhanaan dan kecepatan komputasinya namun memiliki komputasi tinggi, metode *Naïve Bayes Classifier* juga memiliki kinerja yang baik terhadap pengklasifikasian data dokumen yang mengandung angka maupun teks [5].

Penelitian yang berjudul *The Combination of Naive Bayes and Particle Swarm Optimization Methods of Student's Graduation Prediction*, Hasilnya adalah metode *Naive 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 berdasarakan sampel acak pada mahasiswa UIGM Palembang[6].

Penelitian yang berjudul *Comparative Study of K-NN, Naive Bayes and Decision Tree Classification Techniques*, pada penelitian ini menghasilkan survei dan analisis tentang perbandingan antar data mining algoritma klasifikasi (Pohon keputusan, KNN, Bayesian), hal ini menunjukkan bahwa semua algoritma Decision Tree lebih akurat dan mereka memiliki tingkat kesalahan yang lebih sedikit dan itu adalah algoritma yang lebih mudah dibandingkan dengan K-NN dan Bayesian[7].

Penelitian yang berjudul *Analysis on Students Performance Using Naïve*, hasil peneltian ini ditunjukkan bahwa dengan 10 cross-fold-validation algoritma Naïve Bayes dapat digunakan untuk klasifikasi kinerja mahasiswa semester II tahap awal dengan akurasi 73,4% [8].

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[9].

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. Banyak penelitian telah dilakukan berkaitan dengan model *Naïve Bayes* tersebut. Kebanyakan dari penelitian tersebut mengarah kepada ilmu komputer atau informasi data mengenai statistika sehingga pada pembahasannya lebih ditekankan pada hasil prediksi untuk suatu model prediksi yang baru dalam menentukan prediksi tingkat kelulusan mahasiswa tepat waktu.

Pengembangan model baru pada model *Naive Bayes* dianggap cocok digunakan untuk klasifikasi prediksi tingkat kelulusan mahasiswa tepat waktu. Maka pada penelitian ini akan di implementasikan metode tersebut dan dicari hasil akurasi dengan model *Naive bayes* dan menggunakan bahasa pemrograman *Python 3* serta hasil pengelompokan dari tingkat akurasi prediksi tersebut, sehingga dapat menggali informasi mengenai hasil data mining pada prediksi tingkat kelulusan mahasiswa tepat waktu, dan menghasilkan prediksi data terhadap 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.

2. METODOLOGI PENELITIAN

2.1 Data Mining

Data mining adalah sebagai proses untuk mendapatkan informasi yang berguna dari gudang basis data yang besar. Data mining juga dapat diartikan sebagai pengekstrakan informasi baru yang diambil dari bongkahan data besar yang membantu dalam pengambilan keputusan. Data mining sering juga disebut dengan *Knowledge Discovery in Database* atau disingkat KDD, adalah kegiatan yang meliputi pengumpulan, pemakaian data histori untuk menelusuri data yang ada untuk membangun sebuah model agar dapat mengenali pola data yang lain berukuran besar[10]. Data mining merupakan proses penemuan pola - pola baru dari kumpulan - kumpulan data yang sangat besar, yang meliputi metode - metode irisan dari *artificial intelligence, machine learning, statistics*, dan *database system*. Data mining juga dapat mengekstraksi (mengambil intisari) pengetahuan dari sekumpulan data sehingga

M Riski Qisthiano, Copyright ©2021, MIB, Page 988 Submitted: 30/05/2021; Accepted: 01/07/2021; Published: 31/07/2021

Volume 5, Nomor 3, Juli 2021, Page 987-994

ISSN 2614-5278 (media cetak), ISSN 2548-8368 (media online) Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/mib DOI 10.30865/mib.v5i3.3030



didapatkan struktur yang dapat dimengerti manusia serta meliputi basis data dan manajemen data, prapemrosesan data, pertimbangan model dan inferensi, ukuran ketertarikan, pertimbangan kompleksitas, pasca pemrosesan terhadap struktur yang ditemukan, visualisasi, dan online updating[11].

2.2 Naïve Bayes

Naïve Bayes merupakan pengklasifikasian dengan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, 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. Probabilitas yang terlibat dalam memproduksi perkiraan akhir dihitung sebagai jumlah frekuensi dari "master" tabel keputusan. The Naive Bayes Classifier bekerja sangat baik dibanding dengan model Classifier lainnya[12]. Teorema bayes merupakan dasar aturan dari naive bayes Classifier berikut teorema bayes akan disajikan pada persamaan.

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

Keterangan:

x : Data dengan class yang belum diketahui

c : Hipotesis data merupakan suatu class spesifik

P(c/x): Probabilitas hipotesis berdasar kondisi (posteriori probability)

P(c): Probabilitas hipotesis (prior probability)

P(x/c): Probabilitas berdasarkan kondisi pada hipotesis

P(x): Probability c

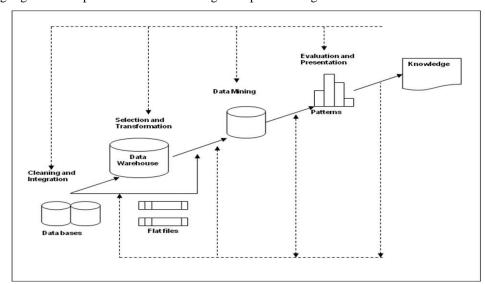
2.3 Klasifikasi

Klasifikasi adalah Suatu teknik dengan melihat pada kelakuan dan atribut dari kelompok yang telah didefinisikan. Teknik ini dapat memberikan klasifikasi pada data baru dengan memanipulasi data yang ada yang telah diklasifikasi dan dengan menggunakan hasilnya untuk memberikan sejumlah aturan [13].

Klasifikasi merupakan proses identifikasi kelompok dari suatu obyek berdasarkan kesamaan fitur tertentu, dimana setiap kelompok telah terbentuk melalui suatu proses tertentu. Keunikan setiap kelompok direpresentasikan dengan suatu vector fitur[14].

2.4 Knowledge Dicovery in Database (KDD)

Knowledge Discovery in Database adalah proses untuk menggali dan menganalisis sejumlah data dan mengestrak informasi dan pengetahuian yang berguna. Hasil pengetahuan yang diperoleh dalam proses tersebut dapat digunakan sebagai basis pengetahuan untuk keperluan pengambilan keputusan"[15]. Proses dalam KDD adalah proses yang digambarkan pada dan terdiri dari rangkaian proses sebagai berikut:



Gambar1. Tahapan Knowledge Discovery in Database

Data mining merupakan salah satu langkah dari proses *Knowledge Discovery from Data* atau lebih dikenal dengan singkatan KDD". Berikut langkah langkah dari KDD[16]: Tahap-tahap data mining ada 6 yaitu:

M Riski Qisthiano, Copyright ©2021, MIB, Page 989 Submitted: **30/05/2021**; Accepted: **01/07/2021**; Published: **31/07/2021**

Volume 5, Nomor 3, Juli 2021, Page 987-994

ISSN 2614-5278 (media cetak), ISSN 2548-8368 (media online) Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/mib DOI 10.30865/mib.v5i3.3030



a. Pembersihan data (data cleaning)

Pembersihan data merupakan proses menghilangkan noise dan data yang tidak konsisten atau data tidak relevan. Pada umumnya data yang diperoleh, baik dari database suatu perusahaan maupun hasil eksperimen, memiliki isian-isian yang tidak sempurna seperti data yang hilang, data yang tidak valid atau juga hanya sekedar salah ketik. Selain itu, ada juga atribut-atribut data yang tidak relevan dengan hipotesa data mining yang dimiliki. Data-data yang tidak relevan itu juga lebih baik dibuang. Pembersihan data juga akan mempengaruhi performasi dari teknik data mining karena data yang ditangani akan berkurang jumlah dan kompleksitasnya.

b. Integrasi data (data integration)

Integrasi data merupakan penggabungan data dari berbagai database ke dalam satu database baru. Tidak jarang data yang diperlukan untuk data mining tidak hanya berasal dari satu database tetapi juga berasal dari beberapa database atau file teks. Integrasi data dilakukan pada atribut-aribut yang mengidentifikasikan entitas-entitas yang unik seperti atribut nama, jenis produk, nomor pelanggan dan lainnya. Integrasi data perlu dilakukan secara cermat karena kesalahan pada integrasi data bisa menghasilkan hasil yang menyimpang dan bahkan menyesatkan pengambilan aksi nantinya. Sebagai contoh bila integrasi data berdasarkan jenis produk ternyata menggabungkan produk dari kategori yang berbeda maka akan didapatkan korelasi antar produk yang sebenarnya tidak ada.

c. Seleksi Data (Data Selection)

Data yang ada pada database sering kali tidak semuanya dipakai, oleh karena itu hanya data yang sesuai untuk dianalisis yang akan diambil dari database. Sebagai contoh, sebuah kasus yang meneliti faktor kecenderungan orang membeli dalam kasus market basket analysis, tidak perlu mengambil nama pelanggan, cukup dengan id pelanggan saja.

d. Transformasi data (Data Transformation)

Data diubah atau digabung ke dalam format yang sesuai untuk diproses dalam data mining. Beberapa metode data mining membutuhkan format data yang khusus sebelum bisa diaplikasikan. Sebagai contoh beberapa metode standar seperti analisis asosiasi dan clustering hanya bisa menerima input data kategorikal. Karenanya data berupa angka numerik yang berlanjut perlu dibagi-bagi menjadi beberapa interval. Proses ini sering disebut transformasi data.

e. Proses mining,

Merupakan suatu proses utama saat metode diterapkan untuk menemukan pengetahuan berharga dan tersembunyi dari data.

f. Evaluasi pola (pattern evaluation),

Untuk mengidentifikasi pola-pola menarik kedalam *knowledge-based* yang ditemukan. Dalam tahap ini hasil dari teknik data mining berupa pola-pola yang khas maupun model prediksi dievaluasi untuk menilai apakah hipotesa yang ada memang tercapai. Bila ternyata hasil yang diperoleh tidak sesuai hipotesa ada beberapa alternatif yang dapat diambil seperti menjadikannya umpan balik untuk memperbaiki proses data mining, mencoba metode data mining lain yang lebih sesuai, atau menerima hasil ini sebagai suatu hasil yang di luar dugaan yang mungkin bermanfaat.

g. Presentasi pengetahuan (knowledge presentation),

Merupakan visualisasi dan penyajian pengetahuan mengenai metode yang digunakan untuk memperoleh pengetahuan yang diperoleh pengguna. Tahap terakhir dari proses data mining adalah bagaimana memformulasikan keputusan atau aksi dari hasil analisis yang didapat. Ada kalanya hal ini harus melibatkan orang-orang yang tidak memahami data mining. Karenanya presentasi hasil data mining dalam bentuk pengetahuan yang bisa dipahami semua orang adalah satu tahapan yang diperlukan dalam proses data mining. Dalam presentasi ini, visualisasi juga bisa membantu mengkomunikasikan hasil data mining.

2.5 Python

Python merupakan salah satu alat yang direkomendasikan untuk hal tersebut. Dalam beberapa tahun terakhir, Dukungan perpustakaan Python yang ditingkatkan (terutama panda) telah membuatnya menjadi alternatif yang kuat untuk tugas analisis data. Dikombinasikan dengan kekuatan Python dalam pemrograman tujuan umum, itu adalah pilihan yang sangat baik sebagai bahasa tunggal untuk membangun aplikasi data-sentris[17].

Python merupakan salah satu bahasa pemrograman yang banyak digunakan oleh perusahaaan besar maupun para developer untuk mengembangkan berbagai macam aplikasi berbasis desktop, web dan mobile. Python diciptakan oleh Guido van Rossum di Belanda pada tahun 1990 dan namanya diambildari acara televisi kesukaan Guido Monty Python's Flying Circus. Van Rossum mengembangkan Python sebagaihobi, kemudian Python menjadi bahasa pemrograman yang dipakai secara luas dalam industri dan pendidikan karena sederhana, ringkas, sintak sintuitif dan memiliki pustaka yang luas[18].

M Riski Qisthiano, Copyright ©2021, MIB, Page 990 Submitted: **30/05/2021**; Accepted: **01/07/2021**; Published: **31/07/2021**

Volume 5, Nomor 3, Juli 2021, Page 987-994 ISSN 2614-5278 (media cetak), ISSN 2548-8368 (media online) Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/mib DOI 10.30865/mib.v5i3.3030



3. HASIL DAN PEMBAHASAN

3.1 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, *dataset* yang digunakan pada penelitian ini berjumlah 1739 yang berasal dari data alumni, setelah itu *dataset* dibagi menjadi data *training* yang berjumlah 70% dari keseluruhan data dan yang akan dijadikan data *testing* berjumlah 30% dari keseluruhan data. Berikut jumlah data yang sudah disusun pada gambar 1 berikut berdasarkan Perguruan Tinggi.

Out[15]:											
		JurusanAlumni	PT	JenisKelas	IP1	IP2	IP3	IP4	Angkatan	Tahun Lulus	Keterangan
	0	Ilmu Komunikasi	STISIPOL Candradimuka	Regular	3.43	3.33	3.10	3.05	2016	2020	Tepat Waktu
	1	Ilmu Komunikasi	STISIPOL Candradimuka	Regular	3.62	3.22	3.00	3.01	2016	2021	Tepat Waktu
	2	Ilmu Komunikasi	STISIPOL Candradimuka	Regular	3.57	3.54	2.63	3.21	2016	2021	Tidak Tepat Waktu
	3	Ilmu Komunikasi	STISIPOL Candradimuka	Regular	3.33	3.38	3.33	3.52	2016	2020	Tepat Waktu
	4	Ilmu Komunikasi	STISIPOL Candradimuka	Regular	3.57	2.25	2.89	3.33	2016	2021	Tidak Tepat Waktu
	5	Ilmu Komunikasi	STISIPOL Candradimuka	Regular	3.86	3.25	3.14	3.52	2016	2020	Tepat Waktu
	6	Ilmu Komunikasi	STISIPOL Candradimuka	Regular	4	4.00	3.88	3.75	2016	2020	Tepat Waktu
	7	Ilmu Komunikasi	STISIPOL Candradimuka	Regular	3.86	3.88	3.54	3.67	2016	2020	Tepat Waktu
	8	Ilmu Komunikasi	STISIPOL Candradimuka	Regular	3.33	3.29	3.19	3.33	2016	2020	Tepat Waktu
	9	Ilmu Komunikasi	STISIPOL Candradimuka	Regular	3.86	3.63	3.67	3.58	2016	2020	Tepat Waktu

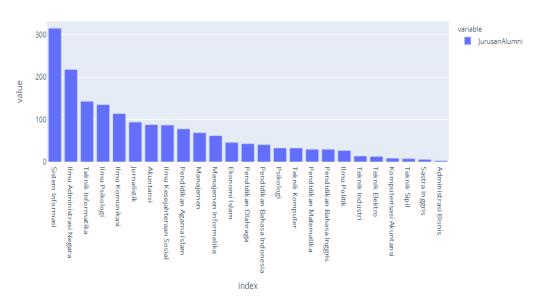
Gambar 2. Dataset

Dataset pada gambar 2 berisikan attribut data yang akan diolah pada proses klasifiksasi dengan algoritma Naïve Bayes, attribut tersebut berisikan jurusan, PT, jenis kelas, IP sementara dari semester 1 sampai 4, angkatan masuk, dan tahun lulus alumni tersebut.

3.2 Persebaran Distribusi Data

Berikut merupakan grafik distribusi data berdasarkan attribut jurusan, Perguruan Tinggi dan jenis kelas pada tiap mahasiswa, persebaran data ini dibagi kedalam 2 jenis distribusi, pertama distribusi sebelum data di proses *cleaning* dan kedua distribusi data setelah melewati proses *cleaning data*. Berikut gambar 3 yang berisikan distribusi data.





Gambar 3. Grafik Persebaran Data

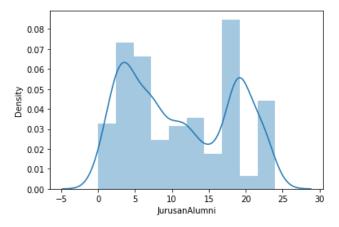
Berikut ini merupakan grafik-grafik distribusi data setelah dilakukan proses *cleaning* dan *transformasi data* dapat dilihat pada gambar grafik persebaran data berdasarkan jurusan seperti gambar 4 berikut ini.

M Riski Qisthiano, Copyright ©2021, MIB, Page 991 Submitted: 30/05/2021; Accepted: 01/07/2021; Published: 31/07/2021

Volume 5, Nomor 3, Juli 2021, Page 987-994

ISSN 2614-5278 (media cetak), ISSN 2548-8368 (media online) Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/mib DOI 10.30865/mib.v5i3.3030





Gambar 4. Grafik Persebaran Data Setelah Cleaning

Setelah melihat gambar 4 yang berisikan data persebaran pada *dataset*, maka proses selanjutnya adalah klasifikasi dengan model *Naïve Bayes* menggunakan bahasa pemrograman *Python 3* pada *tools jupyter notebook*.

3.3 Proses Klasifikasi Naïve Bayes Dengan Python 3

Pemilihan *algoritma Naïve Bayes* didasari oleh karena memiliki performa yang cukup tesis dan waktu pemrosesan yang cepat. Pada fase ini akan ditunjukan bagaimana melakukan prediksi terhadap kelulusan tepat waktu mahasiswa dengan algoritma *Naïve Bayes* pada *tools jupyter notebook* dan bahasa pemrograman *Python*. Sebelum melakukan proses klasifikasi pada penelitian ini adalah melakukan proses pemanggilan *library* dengan mengimport *modul (nampy, pandas, seaborn, matplotlib)*, berikut merupakan tahapan dari proses klasifikasi dengan *Naïve Bayes*.

- a. Pertama dataset yang telah disiapkan di import kedalam tools sebelum diolah dengan model Naïve Bayes.
- b. Setelah proses memasukan data, maka proses selanjutnya membaca *file* yang sudah diatur dengan ektensi *csv*.
- c. Langkah selanjutnya melakukan *preprocessing* atau merubah data sebelum diolah pada tahap selanjutnya, dikarenakan beberapa tipe data attribut masih berbentuk *string* atau *float* maka dilakukan *preprocessing* data yang berisi *source code* untuk merubah tipe attribut data ke *numberic*
- d. Setelah data masuk maka proses selanjutnya *cleaning data* dan *transformasi data* agar data dapat diolah dengan *model Naïve Bayes*.
- e. Setelah data training alumni telah siap untuk digunakan, maka proses selanjutnya adalah penggunaan *library* untuk melakukan prediksi serta menguji dan mengevaluasi model yang akan digunakan, berukut merupakan *source code* yang akan digunakan dalam memanggil *library-library* yang akan dipakai.
- f. Setelah memasukan *source code* untuk melakukan pemanggilan *library* maka proses selanjutnya adalah melakukan ekstraksi terhadap fitur yang akan digunakan dan melakukan klasifikasi sekaligus dengan satu kelas pipeline. pembagian data menjadi data *training* dan *data testing*, jumlah data *training* dibuat menjadi 70% dari keseluruhan data sedangkan *data testing* diambil 30% dari keseluruhan data, total data berjumlah 1739 *record* yang akan dipilih secara acak atau random *by system*.
- g. Setelah data siap maka proses selanjutnya memanggil model *Naïve Bayes* untuk di lihat hasil dari performa dan akurasi dari model yang dipakai.
- h. Setelah didapat hasil akurasi maka proses selanjutnya mencari data *K-Fold Validation*, pengujian ini dibagi menjadi 10 K untuk dicari hasil akurasi yang lebih baik.

3.4 Uji Data Dan Model

Pada tahap ini akan dilakukan uji data dan evaluasi model dari *source code* sebelumnya terhadap evaluasi model dengan *confussion matrix*. Serta melakukan pengujian terhadap *performa* dari *algoritma Naïve Bayes* pada bahasa pemrograman *Python 3* dan data alumni yang dginuakan, berikut merupakan data-data yang akan dihasilkan dari penelitian ini:

- a. Data hasil akurasi berdasakan data test terhadap data training,
- b. Data Confusion Matrix
- c. Hasil dari Precision, Recall, dan F1-Score

Terlihat pada hasil pengujian dari diatas dilakukan proses klasifikasi terhadap sebaran data testing. Pengujian dilakukan dengan melakukan klasifikasi terhadap data testing. Pada pengujian ini, penulis menggunakan nilai random_state yaitu 0 sampai 4. Hasil confussion matrix dari tiap sebaran data dapat dilihat pada tabel 1 berikut.

Volume 5, Nomor 3, Juli 2021, Page 987-994

ISSN 2614-5278 (media cetak), ISSN 2548-8368 (media online) Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/mib DOI 10.30865/mib.v5i3.3030



Tabel 1. Hasil Persebaran Confussion matrix

Algoritma Naïve Bayes					
_	_	Tepat Waktu	Tidak Tepat Waktu		
Actual Value	Tepat Waktu	220	49		
	Tidak Tepat Waktu	50	203		

Setelah nilai-nilai pada *Confussion Matrix* diketahui, maka proses selanjutnya adalah mengetahui atau menghitung dari nilai *precission, accuracy, recall* dan *F1-Score. Precission* merupakan ketepatan nilai antara permintaan pengguna pada respon *system*, sedangkan *accuracy* merupakan perbandingan antara informasi benar yang dijawab sistem dengan keseluruhan data, lalu *recall* merupakan ketepatan antara informasi yang sama dengan informasi yang pernah dipanggil sebelumnya, terakhir adalah *F1-Score* yaitu merupakan perbandingan rata-rata pada *precission* dan *recall* yang dibobotkan. Berikut hasil dari *source code* untuk melakukan pengujian model yang dapat dilihat pada tabel 2 berikut ini:

Tabel 2. Hasil Pengujian Model

Jenis Klasifikasi	Precision	Recall	F1-Score
Tepat Waktu	0.81	0.82	0.82
Tidak Tepat Waktu	0.81	0.80	0.80

Dari tabel diatas dapat dilihat bahwa hasil dari evaluasi model yang dapat dilihat nilai *precision* dan *recall* pada setiap kelas dapat terlihat kemampuan pemrosesan sistem dalam mencari tingkat ketepatan kelulusan tepat waktu terhadap mahasiswa antara Tepat Waktu yang memiliki nilai "81%" dan Tidak Tepat Waktu "81%". Tingkat keberhasilan dari pemrosesan sistem dalam memperoleh kembali informasi label Tepat Waktu adalah "82%", untuk label Tidak Tepat Waktu adalah "80%". Berikut ini gambar 5 yang merupakan hasil dari nilai *accuracy* dari *algoritma Naïve Bayes* yang sudah dijalankan.

```
In [27]: print("Nilai Akurasi Dengan Metode Naive Bayes Pada Data Testing", tree_data.score(xtest, ytest))

Nilai Akurasi Dengan Metode Naive Bayes Pada Data Testing 0.8103448275862069
```

Gambar 5. Hasil Akurasi Naïve Bayes

3.5 Hasil Pengujian Pengaruh K-FOLD

Nilai Akurasi Dengan Algoritma *Naive Bayes* Pada Data Testing 0.8103448275862069 atau 0.810. Setelah model di uji maka akan menunjukan hasil yang cukup baik. Sedangkan peneliti menggunakan *cross validation* untuk menguji performa dari model yang digunakan, berikut hasil yang didapat dengan nilai K = 10 dapat dilihat pada tabel 3.

Tabel 3. Hasil K-Fold Validation

No	K-Fold	Hasil
1	K-1	0.85714286
2	K-2	0.82857143
3	K-3	0.74285714
4	K-4	0.74285714
5	K-5	0.84285714
6	K-6	0.75714286
7	K-7	0.8115942
8	K-8	0.89855072
9	K-9	0.71014493
10	K-10	0.8115942

Dari tabel 3 di atas dapat terlihat bahwa hasil pengujian tertinggi terletak pada K=8, yang dimana berisi 0.898. Hasil pengujian ini memiliki hasil yang berbeda-beda sehingga dilakukan perhitungan rata-rata yang dimana hasil rata-ratanya adalah 0.800.

4. KESIMPULAN

Kesimpulan akhir dari penelitian yang dijalankan, pada penelitian yang berjudul Pengembangan Model Prediksi Kelulusan Mahasiswa Tepat Waktu, yang menggunakan sampel data alumni dari beberapa Perguruan Tinggi yang ada di Kota Palembang, berikut merupakan kesimpulan akhir dari penelitian ini.Penggunaan data uji menggunakan data alumni dari beberapa Perguruan Tinggi seperti UIN Raden Fatah Palembang, Universitas Bina Darma

M Riski Qisthiano, Copyright ©2021, MIB, Page 993 Submitted: 30/05/2021; Accepted: 01/07/2021; Published: 31/07/2021

Volume 5, Nomor 3, Juli 2021, Page 987-994

ISSN 2614-5278 (media cetak), ISSN 2548-8368 (media online) Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/mib DOI 10.30865/mib.v5i3.3030



Palembang, Universitas PGRI Palembang dan STISIPOL Candradimuka Palembang. Hasil data uji yang didapat memiliki tingkat Akurasi sebesar 0.810 dengan menggunakan *Algoritma Naïve Bayes*. Untuk nilai *percision* Kelas Tepat Waktu memiliki nilai "0.81" sedangkan Tidak Tepat Waktu sebesar "0.81". Untuk nilai *recall* Kelas Tepat Waktu memiliki nilai "0.82" sedangkang kelas Tidak Tepat Waktu sebesar "0.80".

REFERENCES

- [1] S. Syarli and A. A. Muin, "Metode Naive Bayes Untuk Prediksi Kelulusan (Studi Kasus: Data Mahasiswa Baru Perguruan Tinggi)," *J. Ilmu Komput.*, vol. 2, no. 1, pp. 22–26, 2018.
- [2] I. Hidayanti, T. B. Kurniawan, and A. Afriyudi, "Perbandingan Dan Analisis Metode Klasifikasi Untuk Menentukan Konsentrasi Jurusan," *J. Ilm. Inform. Glob.*, vol. 11, no. 1, pp. 16–21, 2020, doi: 10.36982/jig.v11i1.1067.
- [3] A. H. Mirza, "Application of Naive Bayes Classifier Algorithm in Determining New Student Admission Promotion Strategies," *J. Inf. Syst. Informatics*, vol. 1, no. 1, pp. 14–28, 2019, doi: 10.33557/journalisi.v1i1.2.
- [4] N. Yahya and A. Jananto, "Komparasi Kinerja Algoritma C.45 Dan Naive Bayes Untuk Prediksi Kegiatan Penerimaanmahasiswa Baru (Studi Kasus: Universitas Stikubank Semarang)," Pros. SENDI, no. 2014, pp. 978–979, 2019
- [5] D. N. Chandra, G. Indrawan, and I. N. Sukajaya, "Klasifikasi Berita Lokal Radar Malang Menggunakan Metode Naïve Bayes Dengan Fitur N-Gram," *J. Ilmu Komput. Indones.*, vol. 4, no. 2, 2019.
- [6] E. Purnamasari, D. P. Rini, and Sukemi, "The Combination of Naive Bayes and Particle Swarm Optimization Methods of Student's Graduation Prediction," J. Ilm. Tek. Elektro Komput. dan Inform., vol. 5, no. 2, 2019, doi: 10.26555/jiteki.v5i2.15317.
- [7] S. D. Jadhav and H. P. Channe, "Comparative Study of K-NN, Naive Bayes and Decision Tree Classification Techniques," *Int. J. Sci. Res.*, vol. 5, no. 1, pp. 1842–1845, 2016, doi: 10.21275/v5i1.nov153131.
- [8] M. Mokhairi, H. Nawang, and S. N. Wan, "Analysis on Students Performance Using Naïve," J. Theor. Appl. Inf. Technol., vol. 31, no. 16, pp. 3993–4000, 2017.
- [9] E. Sutoyo and A. Almaarif, "Educational Data Mining untuk Prediksi Kelulusan Mahasiswa Menggunakan Algoritme Naïve Bayes Classifier," J. RESTI (Rekayasa Sist. dan Teknol. Informasi), vol. 4, no. 1, pp. 95–101, 2020, doi: 10.29207/RESTI.V4I1.1502.
- [10] H. Prasetyo and W. Sutopo, "Perkembangan Keilmuan Teknik Industri Menuju Era," Semin. dan Konf. Nas. IDEC 2017, pp. 488–496, 2017.
- [11] J. Sumpena and N. Kurnia, "Analisis Prediksi Kelulusan Siswa PKBM Paket C dengan Metoda Algoritma Naïve Bayes," *Tedc*, vol. 13, no. 2, pp. 127–133, 2019.
- [12] D. L. Olson and D. Delen, Advanced Data Mining Techniques. Springer, 2008.
- [13] D. Aprilia, D. Aji Baskoro, L. Ambarwati, and I. W. S. Wicaksana, Belajar Data Mining Dengan Rapid Minner. 2013.
- [14] R. Y. Dillak, D. M. Pangestuty, and M. G. Bintiri, "Klasifikasi Jenis Musik Berdasarkan File Audio Menggunakan Jaringan Syaraf Tiruan Learning Vector Quantization," *Semin. Nas. Inform.*, vol. 2012, no. semnasIF, pp. 122–125, 2012.
- [15] D. A. Effendy, K. Kusrini, and S. Sudarmawan, "Classification of intrusion detection system (IDS) based on computer network," *Int. Conf. Inf. Technol. Inf. Syst. Electr. Eng.*, 2017.
- [16] H. Jiawei and K. Micheline, Data mining: concepts and techniques second edition. 2006.
- [17] M. L. Abbot and J. McKinney, Understanding and applying research design. New Jersey: John Wiley & Sons, Inc, 2013.
- [18] M. Romzi and B. Kurniawan, "Pembelajaran Pemrograman Python Dengan Pendekatan Logika Algoritma," *J. Tek. Inform. Mahakarya*, vol. 03, no. 2, pp. 37–44, 2020.

M Riski Qisthiano, Copyright ©2021, MIB, Page 994
Submitted: **30/05/2021**; Accepted: **01/07/2021**; Published: **31/07/2021**