Digital Transformation Technology (Digitech) | e-ISSN: 2807-9000
Volume 4, **Number** 2, September 2024
https://doi.org/10.47709/digitech.v4i2.4963

Penerapan Algoritma Naive Bayes dalam Memprediksi Lulus Tepat Waktu Mahasiswa

M. Julkarnain[1*], Mar'i Yustiardin[2]
[1,2,] Universitas Teknologi Sumbawa, Indonesia
[1] m.julkarnain@uts.ac.id , [2]yustiardin004@gmail.com

open access
Histori Artikel:
Diajukan: 16 November 2024
Disetujui: 30 November 2024
Dipublikasi: 3 Desember 2024
Kata Kunci:
kelulusan mahasiswa, tepat waktu, data mining, *naïve bayes*, prediksi
Digital Transformation Technology (Digitech) *is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0).*

Abstrak
Pendidikan tinggi memainkan peran penting dalam pengembangan sumber daya manusia yang berkualitas. Salah satu indikator keberhasilan pendidikan tinggi di Indonesia adalah tingkat kelulusan mahasiswa tepat waktu. Namun, Program Studi Informatika di Universitas Teknologi Sumbawa (UTS) menghadapi tantangan dengan sejumlah mahasiswa yang tidak lulus tepat waktu setiap tahunnya. Penelitian ini bertujuan untuk memprediksi kemungkinan mahasiswa lulus tepat waktu dengan menggunakan algoritma Naïve Bayes. Data akademik mahasiswa yang digunakan dalam penelitian ini bersumber dari Program Studi Informatika UTS. Metode *Naïve Bayes* dipilih karena kemampuannya dalam menangani data besar dan memberikan prediksi yang akurat berdasarkan probabilitas. Hasil penelitian menunjukkan bahwa algoritma ini mampu memberikan prediksi yang signifikan dalam menentukan mahasiswa yang berpotensi lulus tepat waktu yang memiliki akurasi 94,31%, *precision* 91%, *recall* 95%, dan *F-1 Score* 93%, sehingga menghasilkan model yang dapat memprediksi kelulusan tepat waktu mahasiswa dengan mutakhir. Dengan demikian, hasil penelitian ini diharapkan dapat membantu pihak universitas dalam mengambil keputusan yang lebih baik terkait dengan manajemen mahasiswa, serta meningkatkan efisiensi dan efektivitas proses pendidikan di UTS.

PENDAHULUAN

Pendidikan tinggi merupakan salah satu faktor penting dalam mengembangkan sumber daya manusia yang berkualitas. Di Indonesia, salah satu indikator keberhasilan pendidikan tinggi adalah tingkat kelulusan mahasiswa tepat waktu. Lulus tepat waktu menjadi harapan setiap mahasiswa dan institusi pendidikan, karena berkaitan dengan efisiensi waktu dan biaya Pendidikan. Selain itu, kelulusan tepat waktu juga dapat meningkatkan daya saing lulusan dalam memasuki dunia kerja atau melanjutkan studi ke jenjang yang lebih tinggi (Priyatman et al., 2019)

Universitas Teknologi Sumbawa (UTS) menjadi salah satu lembaga pendidikan swasta terkenal di Sumbawa. Kampus UTS teletak di Jalan Raya Olat Maras Batu Alang, Pernek, Kec. Moyo Hulu, Kabupaten Sumbawa, Nusa Tenggara Barat 84371. Kampus UTS memiliki 8 fakultas pada tahun 2024, salah satunya Fakultas Rekayasa Sistem (FRS). FRS UTS terdapat 5 program studi seperti Teknik Mesin, Teknik Elekro, Teknik Industri, Teknik Sistem Energi, dan Informatika. Program Studi Informatika UTS merupakan salah satu program studi yang mencetak lulusan di bidang teknologi informasi. Tercatat sebanyak 1.178 mahasiswa terdaftar pada program studi informatika UTS hingga semester genap tahun ajaran 2023/2024. Setiap tahunnya, terdapat mahasiswa yang lulus tepat waktu dan ada pula yang tidak. Berdasarkan hasil pengamatan data dan wawancara dengan sekretaris program studi informatika UTS bahwa mahasiswa yang tidak lulus tepat waktu pada tahun 2017 tercatat 50% mahasiswa dari 26 mahasiswa terdaftar. Selanjutnya, tahun 2018 tercatat 45,8% dari 72 mahasiswa terdaftar. Tahun 2019 sebanyak 27,2% dari 99 mahasiswa terdaftar. Tahun 2020 sebanyak 69,3% dari 101 mahasiswa terdaftar. Tahun 2021 sebanyak 63,1% dari 133 mahasiswa terdaftar. Tahun 2022 sebanyak 48,9% dari 147 mahasiswa terdaftar. Tahun 2023 sebanyak 60% dari 125 mahasiswa terdaftar.

Berdasarkan data tersebut, Program Studi perlu terus berupaya melakukan analisis mendalam dan evaluasi berkelanjutan terhadap permasalahan yang dihadapi mahasiswa yang belum dapat menyelesaikan studi tepat waktu atau yang mengalami status non-aktif. Upaya ini dapat dilakukan dengan cara memahami karakteristik dan latar belakang mahasiswa-mahasiswa tersebut, serta menyusun kebijakan dan regulasi yang dapat mendorong mereka untuk mengikuti kurikulum dengan lebih baik. Apabila kemampuan untuk memprediksi waktu kelulusan mahasiswa dapat diwujudkan, maka manajemen mahasiswa akan menjadi lebih efektif dan efisien dalam memitigasi risiko jumlah mahasiswa yang terkena sanksi non-aktif. Salah satu pendekatan yang dapat digunakan untuk menghasilkan prediksi tersebut adalah dengan memanfaatkan teknik data mining.

Data mining adalah teknik dalam kemajuan digital berfungsi meringankan perusahaan atau universitas mencari data informasi penting di gudang data. Data mining sendiri dibagi menjadi dua tujuan utama yaitu

predictive dan *descriptive*. Data mining merupakan cabang dari bidang pembelajaran mesin atau *machine learning*. Tujuan utama data mining untuk memprediksi nilai untuk setiap kasus dalam dataset baru dan untuk mengidentifikasi hubungan antara variabel input dan variabel target. Sementara itu, model deskriptif dikembangkan untuk mengidentifikasi pola atau hubungan dalam data yang dapat memberikan pemahaman tentang data tersebut (Han et al., 2022). Data mining memiliki beberapa fungsi yaitu *classification clustering, associasion, regression, forecasting* dan *sequence analysis* (saleh & Nasari, 2018).

Beberapa algoritma pembelajaran mesin sering digunakan untuk memprediksi, antara lain *Support Vector Machine (SVM), Decision Tree, Maximum Entropy Classification,* dan *Naïve Baye*s (NB) (Pradnyana & Darmawiguna, 2022). Proses prediksi ketepatan kelulusan mahasiswa akan dilakukan dengan memanfaatkan teknologi *Visual Studio Code* sebagai lingkungan pengembangan terpadu (*Integrated Development Environment/IDE*). Implementasi komputasi akan memanfaatkan *Jupyter Notebook*, sebuah aplikasi web sumber terbuka yang memungkinkan penulisan dan eksekusi kode secara interaktif. Bahasa pemrograman yang akan digunakan adalah *Python* versi 3.11, yang merupakan versi terbaru pada saat penelitian ini dilakukan. *Python* dipilih karena fleksibilitas dan ketersediaannya dalam berbagai pustaka yang mendukung pemrosesan data serta *machine learning*.

Dalam berbagai kasus memprediksi, metode *Naive Bayes* merupakan metode yang paling sering digunakan. *Naive Bayes* adalah metode sederhana yang dikembangkan berdasarkan aturan Bayes dengan melihat kondisi yang ada dan peluang untuk setiap kondisi. Menjadikan informasi sebagai solusi pendukung keputusan dalam dunia perusahaan dan pendidikan (Nugraha & Hairani, 2019). Dengan menggunakan metode ini dapat diketahui secara detail data yang diperlukan dalam penelitian. *Naive Bayes classifier* adalah metode klasifikasi statistik yang menggunakan teorema Bayes dan asumsi independensi fitur. Metode ini menghitung probabilitas suatu data masuk ke dalam kategori tertentu berdasarkan distribusi nilai fitur pada data latih. Meskipun sederhana, *Naive Bayes classifier* mampu memberikan performa yang baik dalam berbagai kasus klasifikasi, termasuk mengidentifikasi fitur-fitur yang paling berpengaruh terhadap hasil prediksi dalam bentuk nilai probabilitas. Berdasarkan uraian diatas, maka penulis tertarik untuk melakukan penelitian yang bertujuan untuk memprediksi kemungkinan mahasiswa lulus tepat waktu dengan menggunakan algoritma Naïve Bayes.

STUDI LITERATUR

Penelitian yang dilakukan oleh Kemal Refta Diska dan Khairi Budayawan menggunakan algoritma *Naive Bayes Classifier* bertujuan memprediksi kelulusan di program studi Pendidikan Teknik Informatika. Kekurangan sistem atau algoritma terletak pada jumlah data yang digunakan yang masih sedikit, serta hanya menggunakan dua atribut. Kelebihan sistem atau algoritma ini adalah sistem telah diaplikasikan dan akurasi yang didapatkan cukup tinggi, yaitu 82,6% (Diska & Budayawan, 2023). Penelitian ini sama dalam penggunaan algoritma dan kasus yang diangkat, namun sistem yang digunakan tidak diaplikasikan, serta atribut yang digunakan penulis lebih banyak sehingga menghasilkan akurasi yang lebih tinggi. Selain itu, penelitian lain juga memilih algoritma *Naïve Bayes Classifier* untuk memprediksi kelulusan mahasiswa tepat waktu, dengan kelemahan algoritma pada penggunaan aplikasi yang berbeda sehingga hasil perbandingan sangat signifikan (Setiyani et al., 2020). Kelebihan algoritma ini adalah jumlah data yang digunakan banyak serta perbandingan versi *Rapidminer* dan literatur yang dilakukan merupakan bentuk dukungan bagi peneliti lain yang meneliti topik ini. Perbedaan penelitian ini membandingkan hasil penelitian terdahulu berdasarkan jumlah data, jumlah atribut, ada tidaknya metode validasi, dan penggunaan aplikasi *Rapidminer* dari berbagai versi.

Ada pula penelitian dalam memprediksi kelulusan mahasiswa yang lain, namun algoritma yang digunakan adalah *Support Vector Machine*. Pada algoritma ini, jumlah atribut yang digunakan sedikit, sehingga hasilnya kurang memuaskan, namun jumlah data yang digunakan banyak, tingkat akurasi sistem tinggi, dan penggunaan *library SK-Learn* yang mempermudah proses transformasi data (Haryatmi & Hervianti, 2021). Persamaan dengan penelitian ini adalah kasus yang diangkat sama dan penggunaan *library SK-Learn*. Penelitian lain dengan judul "Klasterisasi Menggunakan Algoritma *K-Means Clustering* untuk Memprediksi Waktu Kelulusan Mahasiswa" bertujuan memprediksi waktu kelulusan mahasiswa. Kekurangan algoritma terletak pada tidak dihasilkannya akurasi sehingga hasil prediksinya tidak dapat dinyatakan benar. Kelebihannya adalah adanya perhitungan manual dalam penelitian ini (Priyatman et al., 2019). Persamaan dengan penelitian penulis adalah topik yang diangkat sama dan beberapa atribut yang digunakan juga sama. Perbedaan dengan penelitian penulis adalah kasus yang diangkat oleh peneliti mencakup ruang lingkup yang lebih luas, serta algoritma yang digunakan berbeda.

Algoritma *Naive Bayes* merupakan salah satu algoritma yang digunakan dalam teknik pengklasifikasian data (Apridiansyah et al., 2021). *Naive Bayes* adalah *classifier* yang menggunakan metode probabilistik dan statistik yang diperkenalkan oleh ilmuwan Inggris Thomas Bayes yang memprediksi kemungkinan masa depan berdasarkan pengalaman masa lalu dengan cara yang dikenal sebagai teorema Bayes. Klausa tersebut diikuti oleh "naïve", yang mengasumsikan bahwa kondisi antar atribut tidak bergantung satu sama lain. Algoritma Naïve Bayes merupakan algoritma data mining yang dapat digunakan untuk melakukan klasifikasi, dan merupakan algoritma yang akan memberikan hasil kemungkinan yang akan terjadi di masa akan datang dari hasil klasifikasi

[11]. Naïve Bayes berdasar pada Teorema Bayes dengan rumus sebagai berikut (Hutagulung et al., 2023).

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

Keterangan:
X = Data dengan class yang belum diketahui
H = Hipotesis data X merupakan suatu kelas spesifik
$P(X|H)$ = Probabilitas hipotesis H berdasarkan kondisi X
$P(H)$ = Probabilitas hipotesis H

Salah satu bahasa pemrograman yang banyak digunakan dalam *machine learning* adalah *Python*. Bahasa pemrograman *Python* adalah bahasa pemrograman yang mengalami perkembangan dan popularitas yang meningkat. Bahasa pemrograman ini memiliki banyak kelebihan yang menjadikannya sebagai alternatif yang sesuai untuk berbagai keperluan. Phyton merupakan pemrograman yang dinamis sehingga mudah dibaca. *Python* dapat digunakan untuk berbagai keperluan, seperti *machine learning* dapat digunakan untuk pengembangan algoritma *machine learning*, seperti klasifikasi, regresi, dan *clustering* (Mulyadien & Enri, 2022). *Python* adalah bahasa pemrograman tingkat tinggi, diinterpretasi, dan serba guna. Filosofi desainnya menekankan keterbacaan kode dengan penggunaan indentasi yang signifikan. *Python* adalah dinamis-diketik dan dilengkapi dengan pengumpul sampah. *Python* mendukung beberapa paradigma pemrograman, termasuk terstruktur, berorientasi objek dan pemrograman fungsional (Pandey, 2020).

Django adalah sebuah *framework* berbasis *Python* yang digunakan untuk membangun aplikasi web. Framework Django sangat tepat digunakan untuk membangun sebuah model visualisasi data perusahaan yang menjadi tujuan dalam penelitian. Django memiliki arsitektur MVT (*Model View Template*) yang memudahkan pengembangan aplikasi web secara fleksibel. Model digunakan untuk berinteraksi dengan basis data, View menangani logika aplikasi, dan Template menangani bagian depan atau tampilan aplikasi. Arsitektur ini memungkinkan pembagian tugas yang jelas (Vamsi et al., 2021). Django memiliki arsitektur MVT yang memudahkan pembagian tugas pengembangan, sudah menyediakan fitur administrasi basis data secara internal, menggunakan ORM yang mempermudah interaksi dengan basis data, serta dapat mengintegrasikan basis data, antarmuka, dan logika aplikasi menjadi satu kesatuan (Novaliendry & Hakim, 2022). Dengan keunggulan-keunggulan tersebut, Framework Django dianggap mampu memenuhi kebutuhan pengelolaan data alumni secara elektronik dan berbasis web.

Dalam menulis kode program *Python* dibutuhkan suatu code editor yang disebut Integrated Development Environment atau IDE. Salah satu IDE yang cukup populer terutama di kalangan data scientist untuk menulis kode *Python* adalah Jupyter Notebook. Ada beberapa cara dalam melakukan instalasi Jupyter Notebook (Aribowo, 2021). Salah satu cara yang paling mudah adalah melalui instalasi Anaconda, yaitu suatu platform yang mendistribusikan bahasa *Python* khususnya untuk bidang *scientific computing* seperti *data mining, machine learning, predictive analysis* dan lain sebagainya. Selanjutnya, *scikit-learn (SK-Learn)* merupakan salah satu library *machine learning* yang paling populer dan banyak digunakan dalam komunitas *Python*. Library ini menyediakan implementasi yang efisien untuk berbagai algoritma *machine learning*, termasuk klasifikasi, regresi, clustering, dan pengurangan dimensi. *SK-Learn* dibangun di atas *NumPy, SciPy*, dan *Matplotlib*, yang merupakan beberapa library paling populer untuk komputasi ilmiah dan numerik di *Python* (Aryyantomo & Ratama, 2023). *Scikit-learn* berfungsi untuk membantu melakukan *processing* data ataupun melakukan *training data* untuk kebutuhan *machine-learning*. Fitur yang dapat digunakan didalam *Scikitlearn* antara lain, seperti *Clustering*, *Classification, Regression, Dimensionality, Model selection, Preprocessing* Data, dan *Reduction* (Darmawan & Dianta, 2023).

Cross-Validation adalah teknik yang digunakan dalam pembelajaran mesin dan statistika untuk mengevaluasi kinerja model pada data yang belum pernah dilihat sebelumnya. Teknik ini melibatkan pembagian dataset menjadi subset data latih dan data validasi, dimana model dilatih pada data latih dan dievaluasi pada data validasi. *Cross-Validation* membantu mengatasi masalah *overfitting* dan *underfitting* dalam pemodelan dengan memberikan perkiraan kinerja model yang lebih akurat pada data yang belum pernah dilihat sebelumnya (Diska & Budayawan, 2023). Pengujian algoritma juga perlu dilakukan untuk mengevaluasi hasil penelitian menggunakan beberapa hyperparameter, antara lain *Confusion matrix, accuracy, precision, recall, F1-Score* dan *ROC (Receiver Operating Characteristic*). *Confusion matrix* adalah tabel yang menyatakan klasifikasi jumlah data uji yang benar dan jumlah data uji yang salah (Normawati & Prayogi, 2021).

Tabel 1. *Confusion Matrix*

		Kelas Prediksi	
		1	0
Kelas Sebenarnya	1	TP	FN
	0	FP	TN

Keterangan:
TP (*True Positive*) = jumlah dokumen dari kelas 1 yang benar diklasifikasikan sebagai kelas 1
TN (*True Negative*) = jumlah dokumen dari kelas 0 yang benar diklasifikasikan sebagai kelas 0
FP (*False Positive*) = jumlah dokumen dari kelas 0 yang salah diklasifikasikan sebagai kelas 1
FN (*False Negative*) = jumlah dokumen dari kelas 1 yang salah diklasifikasikan sebagai kelas 0.

Akurasi (*accuracy*) yaitu menunjukkan tingkat akurasi model dalam mengklasifikasikan data dengan benar. *Precision* yaitu menunjukkan akurasi antara data yang diminta dengan hasil prediksi model. *Recall* menunjukkan keberhasilan model dalam menemukan kembali sebuah informasi. Rumus yang digunakan untuk menentukan akurasi, presisi dan *recall* adalah sebagai berikut (Raihan et al., 2021).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (2)$$

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

$$Recall = \frac{TP}{TP + FN} \quad (4)$$

F1-score adalah metrik evaluasi yang digunakan dalam pembelajaran mesin dan data mining untuk mengukur kinerja model klasifikasi biner. *F1-score* merupakan rata-rata harmonik dari presisi (*precision*) dan *recall*. F1-score memiliki nilai antara 0 dan 1, dengan nilai 1 menunjukkan kinerja terbaik. Nilai F1-score yang tinggi mengindikasikan bahwa model dapat memprediksi sampel positif dengan *presicion* dan *recall* yang baik. Presisi mengukur seberapa banyak sampel positif yang diidentifikasi dengan benar, sedangkan *recall* mengukur seberapa banyak sampel positif yang berhasil ditemukan (Raihan et al., 2021).

$$F1 = 2\frac{(Precison \times Recall)}{(Precison + Recall)} \quad (5)$$

Hasil perhitungan divisualisasikan dengan kurva ROC (*Receiver Operating Characteristic*) atau AUC (*Area Under Curve*). ROC memiliki tingkat nilai diagnosa sebagai berikut (Wahyono & Nugroho, 2019).

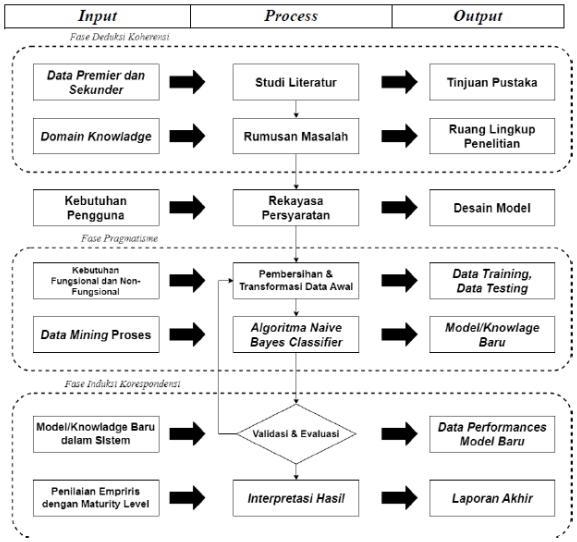
Tabel 2. Kriteria Akurasi

No.	Rentang Nilai	Kriteria
1.	0.90 – 1.00	*excellent classification*
2.	0.80 – 0.90	*good classification*
3.	0.70 – 0.80	*fair classification*
4.	0.60 – 0.70	*poor classification*
5.	0.50 – 0.60	*failure*

Setelah memahami dan menimbang dari kelebihan dan kekurangan pada penelitian yang telah ada, maka pada penelitian ini akan menggunakan algoritma *Naive Bayes* dalam memprediksi lulus tepat waktu mahasiswa, menggunakan *Jupyter Notebook* sebagai *Integrated Development Environment* (IDE) yang berbasis bahasa pemrograman *Python* dalam memprediksi lulus tepat waktu mahasiswa, menggunakan *Library SK-Learn* dalam mempermudah transformasi data mahasiswa, menggunakan *Cross-Validation* dalam memvalidasi hasil prediksi lulus tepat waktu mahasiswa, menggunakan ROC (*Receiver Operating Characteristic*) dalam memvisualisasikan hasil prediksi lulus tepat waktu mahasiswa dan atribut yang akan digunakan adalah gender, indeks prestasi kumulatif, indeks prestasi semester 1 sampai 8, jumlah SKS kumulatif, status beasiswa mahasiswa, status organisasi dari mahasiswa.

METODE

Penelitian ini merupakan penelitian kualitatif, dengan alur dari penelitian seperti yang ditunjukkan pada gambar berikut.



Gambar 1. Alur Penelitian

Penelitian ini dimulai dengan mengumpulkan data sekunder dan pengetahuan domain untuk memahami kebutuhan pengguna serta mendefinisikan kebutuhan fungsional dan non-fungsional. Proses penelitian melibatkan studi literatur untuk merumuskan masalah, rekayasa persyaratan, pembersihan dan transformasi data awal, penerapan algoritma *Naive Bayes* Classifier, validasi, evaluasi, dan interpretasi hasil. Output penelitian mencakup tinjauan pustaka, ruang lingkup penelitian, desain model, data training dan testing, model atau pengetahuan baru, kinerja model baru, dan laporan akhir. Alur penelitian ini dirancang untuk menghasilkan model yang dapat diimplementasikan dalam sistem berdasarkan analisis kebutuhan dan data yang diperoleh.

Tahap pembersihan data (*data cleaning*) diperlukan untuk membuang data awal yang tidak konsisten dan *noise.* Dalam proses ini, beberapa data yang *noise* atau kosong dan atribut yang tidak berpengaruh dalam perhitungan terkait prediksi lulus tepat waktu mahasiswa akan dihapus, yaitu nama mahasiswa dan identitas lain yang bersifat pribadi. Atribut yang masih dipakai yaitu jenis kelamin, status mahasiswa, IPS per semester, IPK, status kelulusan dan lain sebagainya. Selanjutnya seleksi data yaitu menyeleksi atribut yang tidak berpengaruh dalam proses data mining, yaitu IP Semester 5 sampai Semester 8.

Pada tahap transformasi data (*data transformation*), dilakukan pra-pemrosesan data yang sebelumnya telah dibersihkan untuk mempersiapkan data agar sesuai dengan format yang dibutuhkan dalam proses data mining. Dari data asli, terdapat beberapa atribut yang perlu ditransformasi agar dapat diolah dengan baik dalam perhitungan data mining. Sebagai contoh, atribut status mahasiswa yang awalnya berbentuk teks, diubah menjadi representasi numerik dengan mengkodekan 'non-aktif' menjadi 0 dan 'aktif' menjadi 1. Transformasi serupa juga dilakukan pada atribut lain yang relevan. Setelah proses transformasi data selesai, data kemudian dibagi menjadi dua subset, yaitu data latih dan data uji, yang akan digunakan dalam tahap pemodelan selanjutnya.

Tahapan penerapan algoritma *Naive Bayes* menggunakan *Python* ialah mengimplementasikan algoritma *Naive Bayes* dengan menggunakan bahasa pemrogram *Python*, data latih dan data uji yang diperoleh dari data akademik mahasiswa Program Studi Informatika UTS. Model ini dilatih untuk memprediksi status kelulusan mahasiswa, yaitu lulus tepat waktu atau tidak lulus tepat waktu, berdasarkan fitur-fitur yang relevan seperti nilai mata kuliah, IPK, dan faktor-faktor lain yang mempengaruhi kelulusan. Tahap selanjutnya melakukan validasi terhadap model yang telah dihasilkan dari penerapan algoritma *Naive Bayes* pada tahap sebelumnya. Validasi dilakukan dengan menggunakan teknik *cross-validation*, dimana data latih dibagi menjadi dua bagian, yaitu 80% digunakan sebagai data latih dan 20% sisanya digunakan sebagai data validasi. Hal ini bertujuan untuk mengevaluasi kinerja model secara lebih komprehensif dan memastikan bahwa model yang dibangun bekerja dengan baik serta dapat memberikan prediksi yang akurat.

Tahap terakhir adalah evaluasi dengan melakukan evaluasi kinerja model secara keseluruhan menggunakan data uji yang terpisah dari data latih dan validasi. Pada tahap ini, berbagai matrik evaluasi akan diterapkan untuk mengukur performa model *Naive Bayes* dalam memprediksi status kelulusan mahasiswa, meliputi akurasi, presisi, *recall*, *F1-score*, serta kurva ROC dan nilai AUC untuk mengevaluasi kinerja model secara komprehensif dan membandingkannya dengan baseline atau model lain. Akurasi mengukur ketepatan prediksi keseluruhan, sementara presisi, *recall*, dan *F1-score* memberikan gambaran lebih rinci untuk masing-masing kelas. Hasil evaluasi akan dianalisis mendalam untuk mengetahui kekuatan dan kelemahan model, memberikan wawasan tentang faktor-faktor yang memengaruhi kelulusan tepat waktu, serta menyusun rekomendasi yang tepat bagi Program Studi Informatika UTS untuk meningkatkan kelulusan tepat waktu mahasiswa.

Metode pengumpulan data yang digunakan ialah observasi, wawancara dan studi pustaka. Observasi dilakukan dengan cara terjun langsung pada pihak terkait untuk mendapatkan data-data atau informasi yang diperlukan untuk membangun sistem untuk memprediksi lulus tepat waktu mahasiswa program studi informatika UTS, yaitu mengamati data akademik mahasiswa program studi informatika UTS yang berupa Indeks Prestasi Semester 1 sampai 4, IPK mahasiswa di setiap tahunnya, dan jumlah SKS mahasiswa. Wawancara langsung dilakukan pada sekretaris Program Studi Informatika UTS yang bertujuan untuk mendapatkan data-data atau informasi terkait data mahasiswa yaitu adanya keluhan terkait banyaknya mahasiswa yang tidak dapat menyelesaikan studi tepat waktu atau sesuai dengan masa studi yang telah ditetapkan. Beberapa faktor yang disebutkan antara lain nilai mata kuliah yang kurang memuaskan, IPK yang rendah, kurangnya motivasi dalam menyelesaikan studi, masalah ekonomi, serta faktor-faktor lain yang dapat menghambat proses studi mahasiswa. Studi pustaka bertujuan mengumpulkan sumber sebagai referensi untuk membangun sistem untuk memprediksi lulus tepat waktu mahasiswa Program Studi Informatika UTS. Adapun sumber yang berkaitan dengan pengelolaan data mahasiswa adalah informasi dari sekretaris Program Studi Informatika UTS dan berbagai jurnal yang terkait.

HASIL

Data awal yang digunakan merupakan data mahasiswa Program Studi Informatika jenjang Starta 1 yang bersumber dari Sekretariat Proogram Studi Informatika Universitas Teknologi Sumbawa mulai dari angkatan 2013 sampai dengan angkatan 2019 yang berjumlah 703 data mahasiswa. Atribut yang digunakan adalah jenis kelamin, Indeks Prestasi Kumulatif (IPK), Indeks Prestasi Semester 1 sampai semester 8, total SKS sampai semester 8, biaya kuliah, dan keterangan (Lulus Tepat Waktu/Tidak Lulus Tepat Waktu). Berikut adalah data awal mahasiswa Program Studi Informatika Universitas Teknologi Sumbawa angkatan 2013 hingga 2019.

Tabel 3. Data Awal Mahasiswa Program Studi Informatika UTS

ID	Status	IPK	IPS I	IPS II	IPS III	IPS IV	IPS V	IPS VI	IPS VII	IPS VIII	Total SKS	Biaya Kuliah	Jenis kelamin
13002	Tidak Tepat Waktu	2,74	3.00	2.03	1.94	2.34	2.58	3.22	2.28	3.00	144	NaN	L
13003	Tepat Waktu	3,43	3.65	3.53	3.52	2.93	3.58	4.00	3.10	3.50	144	NaN	L
13004	Tepat Waktu	3,56	3.60	3.33	3.19	3.58	4.00	4.00	3.48	3.50	144	NaN	P
13005	Tepat Waktu	3,25	3.45	3.55	3.24	3.02	3.00	3.80	2.95	3.25	144	NaN	L
13006	Tidak Tepat Waktu	2,96	2.93	2.48	2.47	2.78	3.11	3.65	3.15	3.50	131	NaN	P
...													
19124	Tepat Waktu	3,42	3.15	3.11	3.25	3.50	3.65	3.44	0.00	4.00	154	NaN	P
19125	Tepat Waktu	3,27	3.26	2.93	3.00	3.44	2.97	3.93	0.00	3.50	145	NaN	L

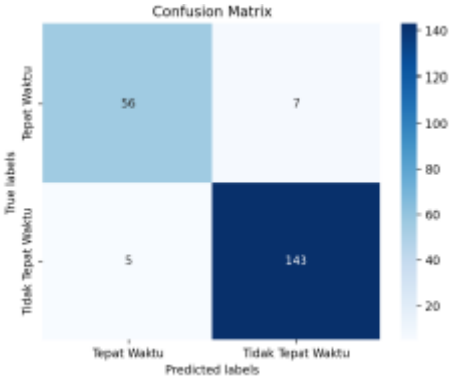
Proses pembersihan data dan atribut yang tidak berpengaruh dalam perhitungan yaitu untuk seluruh atribut yang tidak bernilai (NaN) telah diubah menggunakan fitur dari *SK-Learn* yaitu dengan fitur *interpolate()*, sehingga atribut biaya kuliah telah diubah menjadi bernilai Nol (0). Proses seleksi data dilakukan dengan cara memilih atribut yang akan digunakan dalam proses data mining. Hasil dari proses ini untuk seluruh atribut IP semester 5 hingga semseter 8 telah dihapus karena akan berpengaruh besar pada model *Naive Bayes* nanti. Selanjutnya, proses transfomasi data dilakukan agar *dataset* dapat diproses kedalam data mining. Hasil dari proses transformasi data adalah sebagai berikut.

Tabel 4. Dataset Setelah Proses Transfomasi Data

ID	Status	IPK	IPS I	IPS II	IPS III	IPS IV	Biaya Kuliah	Jenis kelamin
13002	0	2,74	3.00	2.03	1.94	2.34	0	6
13003	1	3,43	3.65	3.53	3.52	2.93	0	6
13004	1	3,56	3.60	3.33	3.19	3.58	0	9
13005	1	3,25	3.45	3.55	3.24	3.02	0	6
13006	0	2,96	2.93	2.48	2.47	2.78	0	9
...								
19124	1	3,42	3.15	3.11	3.25	3.50	0	9
19125	1	3,27	3.26	2.93	3.00	3.44	0	6

Setelah penerapan algoritma Naïve Bayes, model perlu dievaluasi untuk memantau hasil kerja model Naïve Bayes. Adapun teknik yang digunakan dalam memvalidasi model yaitu teknik *cross-validation*. Dalam teknik ini, data akan dibagi menjadi dua yakni data *training* dan data validasi dengan perbandingan 70% untuk data training

dan 30% sisanya akan dijadikan data validasi yang selanjutnya hasil dari data validasi akan di evaluasi menggunakan *Confusion Matrix* yang meliputi *Accuracy Model, Precision Model, Recall*, dan *F1-Score.* Berikut adalah tampilan *Confusion Matrix* dan *Classification Report* dari data validasi.



Gambar 2. Confusion Matrix Data Mahasiswa

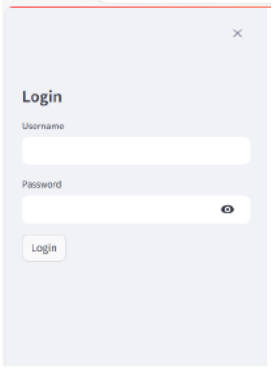
Berdasarkan Gambar 2 dapat diketahui bahwa terdapat 211 data mahasiswa yang di uji. Dari 211 data mahasiswa, model *Naive Bayes* dapat memprediksi 56 data mahasiswa "lulus tepat waktu" dengan benar. Namun, 7 data mahasiswa yang sebenarnya "lulus tepat waktu" diprediksi "tidak lulus tepat" waktu oleh model *Naive Bayes*. Sedangkan, model *Naive Bayes* memprediksi 143 data mahasiswa yang "tidak lulus tepat waktu" dengan benar. Namun, 5 data mahasiswa yang sebenarnya "tidak lulus tepat waktu" diprediksi "lulus tepat waktu" oleh model. Berdasarkan hasil dari confusion matrix maka didapatkanlah hasil dari Classification Report. Berikut adalah hasil dari classification report.

Tabel 5. *Classification Report* Model *Naive Bayes*

	Hasil Uji Coba			
	Precision	*Recall*	*F1-Score*	*Support*
Tepat Waktu	0.91	0.95	0.93	63
Tidak Tepat Waktu	0.97	0.95	0.96	148
Accuracy			0.94	211
Macro avg	0.94	0.95	0.94	211
Weighted avg	0.95	0.95	0.95	211

Hasil pengujian untuk kelas tepat waktu yaitu presisi mendapatkan skor 0.91, *recall* mendapatkan skor 0.95, *F1-score* mendapatkan skor 0.93 dengan support 63 data mahasiswa. Sedangkan untuk kelas tidak tepat waktu yaitu untuk presisi mendapatkan skor 0.97, *recall* medapatkan skor 0.95, *F1-score* mendapatkan skor 0.96 dengan support 148 data mahasiswa. Untuk akurasi model *Naive Bayes* sendiri mendapatkan skor 0.94, maka dapat disimpulkan bahwa model *Naive Bayes* untuk memprediksi lulus tepat waktu mahasiswa ini masuk dalam golongan *excellent classification*.

Selanjutnya penerapan aplikasi prediksi kelulusan tepat waktu mahasiswa informatika UTS. Berikut adalah tampilan halaman login aplikasi prediksi kelulusan tepat waktu mahasiswa informatika UTS.



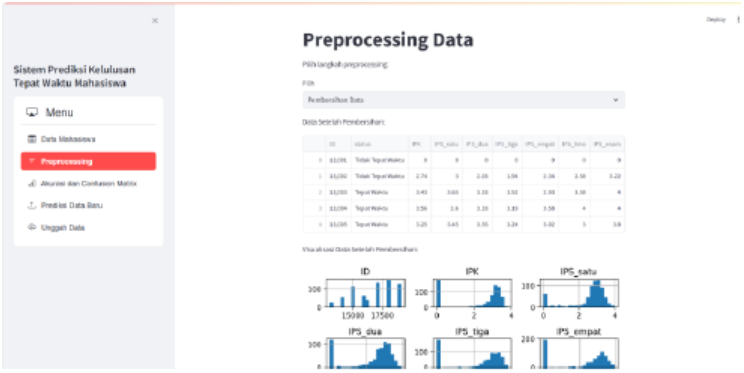
Gambar 3. Halaman Login Aplikasi

Halaman ini adalah langkah awal yang harus dilalui oleh admin atau pengguna untuk mengakses fitur-fitur dalam aplikasi prediksi kelulusan tepat waktu mahasiswa informatika UTS. Setelah berhasil login, pengguna akan diarahkan ke dashboard utama atau halaman lain yang relevan dengan peran dan fungsionalitas mereka dalam aplikasi. Tampilan halaman Menu dan Data Mahasiswa aplikasi prediksi kelulusan tepat waktu mahasiswa informatika UTS disajikan pada gambar berikut.



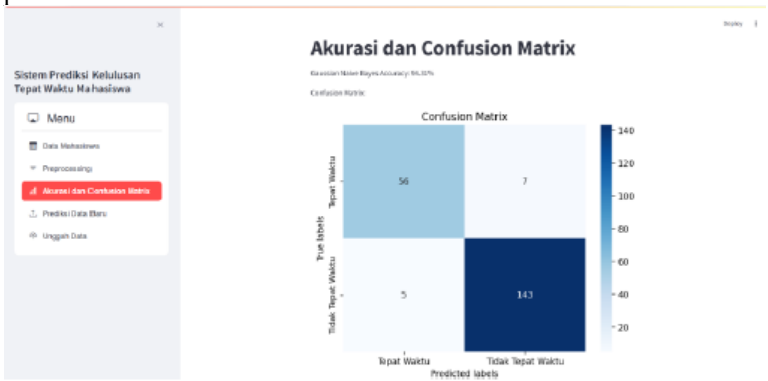
Gambar 4. Halaman Menu Utama dan Data Mahasiswa

Halaman ini menampilkan data mahasiswa sebelum dikelola. Ada indikator halaman yang menunjukkan bahwa data ditampilkan dalam beberapa halaman. Pengguna bisa mengatur nomor halaman yang ingin dilihat. Tampilan ini memudahkan pengguna untuk melihat data akademik mahasiswa secara menyeluruh dan memahami status kelulusan mereka, yang merupakan bagian penting dalam proses prediksi kelulusan tepat waktu. Kemudian, tampilan halaman Preprocessing aplikasi prediksi kelulusan tepat waktu mahasiswa informatika UTS adalah sebagai berikut.



Gambar 5. Halaman Preprocessing Data Mahasiswa

Pada gambar diatas menampilkan data mahasiswa yang telah mengalami proses pembersihan data dan transformasi data. Adanya indikator “Pilih” menunujukkan bahwa data ditampilkan dalam dua proses yaknik pembersihan data dan transformasi data. Berikut adalah tampilan halaman Akurasi dan Confussion Matrix aplikasi prediksi kelulusan tepat waktu mahasiswa informatika UTS.



Gambar 6. Halaman Akurasi dan Confussion Matrix

Berdasarkan Gambar 4.5 menunjukkan bahwa halaman Akurasi dan Confussion Matrix ini menampilkan

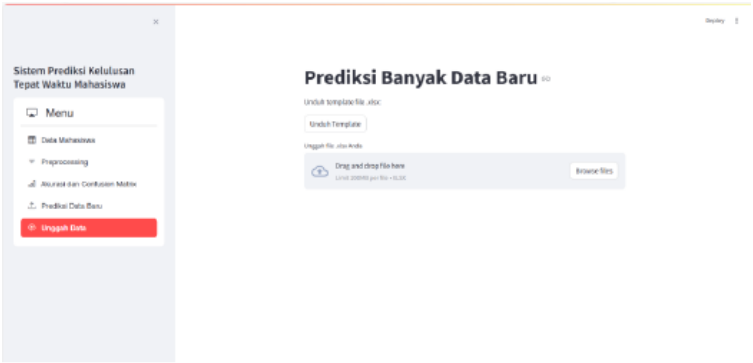
gambar confusion matrix dari model *Naive Bayes*. Selain itu, halaman ini juga menampilkan tabel classification report yang menunjukkan kinerja dari model *Naive Bayes* dengan spesifikasi *Accuracy, Precision, Recall, F-1 Score,* dan *Support*.

Tampilan halaman Prediksi Data Baru aplikasi prediksi kelulusan tepat waktu mahasiswa informatika UTS merupakan halaman untuk meng-*input* data baru yang nantinya akan di prediksi oleh model *Naive Bayes,* yang ditunjukkan pada gambar berikut.



Gambar 7. Halaman Prediksi Data Baru

Sedangkan halaman Unggah Data (Prediksi banyak data) merupakan halaman untuk dapat memprediksi data dalam jumlah banyak. Adanya fitur "Unduh Template" akan mempermudah dalam pengisian data baru dengan jumlah yang banyak. Setelah mengisi data yang formatnya berupa ".xlsx", data perlu diunggah untuk diprediksi oleh model *Naive Bayes* dan menghasilkan kolom baru dalam tabel dengan nama field "Status" yang menjelaskan keterangan data mahasiswa yang "Tepat Waktu" atau "Tidak Tepat Waktu". Tampilan halaman Unggah Data (Prediksi banyak data) aplikasi prediksi kelulusan tepat waktu mahasiswa informatika UTS ditunjukkan pada gambar berikut.



Gambar 8. Halaman Unggah Data

Tahap terakhir adalah pengujian aplikasi untuk memastikan apakah sistem sudah sesuai atau tidak dengan hasil yang diinginkan. Tahap pengujian sistem ini merupakan hal yang sangat penting untuk melihat apakah sistem perangkat lunak sudah sesuai dengan standar atau tidak. Pada tahap pengujian ini dilakukan dengan cara pengujian *blackbox testing*. Pengujian ini dilakukan oleh Herfandi, M.Kom dan Sekretaris Program Studi Informatika.

Tabel 6.Pengujian Aplikasi

Menu Pengujian	Aksi Pengujian	Hasil yang Diharapkan
Login	*Login*	Berhasil
Dasboard	Menampilkan Data Awal sebelum *Preprocessing* Data	Berhasil
Pembersihan Data	Menampilkan Dataset setelah pembersihan data serta visualisasi data setelah pembersihan	Berhasil
Transfomasi Data	Menampilkan Dataset setelah transformasi data serta visualisasi data setelah transformasi	Berhasil
Evaluasi Model	Menampilkan Akurasi, *Confusion Matrix,* dan *Classification Report*	Berhasil
Prediksi Data baru (Individu)	Menampilkan hasil Prediksi dari model secara induvidu (*Input*)	Berhasil
Download Template	Mengunduh *Template* yang di sediakan oleh program	Berhasil

Unggah Template	Mengunggah template data yang telah diisi sesuai *field* yang ada	Berhasil
Prediksi Data dalam Jumlah Banyak	Menampilkan Tabel data baru yang merupakan hasil prediksi dari mode	Berhasil

PEMBAHASAN

Algoritma yang digunakan model prediksi adalah algoritma *Naive Bayes* yang memiliki akurasi 94,31%, *precision* 91%, *recall* 95%, dan *F-1 Score* 93%, sehingga menghasilkan model yang dapat memprediksi kelulusan tepat waktu mahasiswa dengan baik, namun masih perlu untuk dilakukan perbaikan lebih lanjut lagi di masa mendatang dengan memperhatikan komponen biaya kuliah karena beberapa mahasiswa yang nonaktif diduga karena masalah biaya kuliah namun data siakad di universitas belum akurat untuk merincikan hal tersebut sehingga dalam model ini diabaikan. Beberapa jenis beasiswa juga mungkin mempengaruhi untuk dilakukan uji lebih lanjut.

KESIMPULAN

Berdasarkan dari hasil pembahasan sebelumnya terdapat permasalahan yang dihadapi oleh mahasiswa yang belum dapat menyelesaikan studi tepat waktu sehingga mahasiswa non-aktif. Adanya aplikasi prediksi lulus tepat waktu mahasiswa pada Program Studi Informatika UTS ini diharapkan dapat menjadi salah satu upaya Program Studi Informatika untuk memahami karakteristik dan latar belakang mahasiswa terkait sehingga mengurangi risiko jumlah mahasiswa yang terkena sanksi non-aktif. Aplikasi ini dikembangkan menggunakan bahasa pemrograman *Python* beserta support dari *scikit-learn* yang sangat membantu dalam proses awal pembersihan data hingga proses evaluasi model yang terdiri dari 703 data mahasiswa yang terdiri dari Indeks Prestasi Kumulatif (IPK), Indeks Prestasi Semester (IPS) 1 sampai semester 8, dan total SKS. Pada penelitian ini telah dibangun aplikasi menggunakan bahasa pemrograman *Python*. Algoritma yang digunakan model prediksi adalah algoritma *Naive Bayes* yang memiliki akurasi 94,31%, *precision* 91%, *recall* 95%, dan *F-1 Score* 93%, sehingga menghasilkan model yang dapat memprediksi kelulusan tepat waktu mahasiswa dengan mutakhir. Saran dari penulis dalam pengembangan aplikasi ataupun model agar lebih sempurna yaitu perbanyak dataset mahasiswa yang dibuat agar hasil prediksi lebih akurat dan menambahkan atribut yang berkaitan dengan prediksi kelulusan mahasiswa, seperti alamat, lulusan sekolah, dan sebagainya, serta aplikasi prediksi lulus tepat waktu mahasiswa ini diharapkan bisa dikembangkan ke arah sistem pengambilan keputusan mahasiswa.

REFERENSI

Priyatman, H., Sajid, F., & Haldivany, D. (2019). Klasterisasi Menggunakan Algoritma K-Means Clustering untuk Memprediksi Waktu Kelulusan Mahasiswa. *Jurnal Edukasi Dan Penelitian Informatika (JEPIN)*, *5*(1), 62. https://doi.org/10.26418/jp.v5i1.29611

Han, J., Pei, J., & Tong, H. (2022). *Data mining: concepts and techniques.* Morgan Kaufmann, United States

Saleh, A., & Nasari, F. (2018). Penerapan Equal-Width Interval Discretization Dalam Metode Naive Bayes Untuk Meningkatkan Akurasi Prediksi Pemilihan Jurusan Siswa (Studi Kasus: Mas Pab 2 Helvetia, Medan). *Masy. Telemat. Dan Inf. J. Penelit. Teknol. Inf. dan Komun*, *8*(1), 1. http://dx.doi.org/10.17933/mti.v9i1.113

Pradnyana, G. A., & Darmawiguna, I. G. M. (2022). Web-Based System for Bali Tourism Sentiment Analysis during The Covid-19 Pandemic using Django Web Framework and Naive Bayes Method. *Proceedings of the 4th International Conference on Innovative Research Across Disciplines (ICIRAD 2021)*, *613*(Icirad), 316–320. https://doi.org/10.2991/assehr.k.211222.050

Nugraha, G. S., & Hairani, H. (2019). Aplikasi Pemetaan Kualitas Pendidikan di Indonesia Menggunakan Metode K-Means. *Jurnal MATRIK*, *17*(2), 13–23. https://doi.org/10.30812/matrik.v17i2.84

Diska, K. R., & Budayawan, K. (2023). Sistem Informasi Prediksi Kelulusan Menggunakan Metode Naive Bayes Classifer (Studi Kasus: Prodi Pendidikan Teknik Informatika). *Jurnal Pendidikan Tambusai*, 7(1), 936–943. 37. https://doi.org/10.31004/jptam.v7i1.5375

Setiyani, L., Wahidin, M., Awaludin, D., & Purwani, S. (2020). Analisis Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode Data Mining Naïve Bayes: Systematic Review. *Faktor Exacta*, *13*(1), 35. https://doi.org/10.30998/faktorexacta.v13i1.5548

Haryatmi, E., & Hervianti, S. P. (2021). Penerapan Algoritma Support Vector Machine Untuk Model Prediksi Kelulusan Mahasiswa Tepat Waktu. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, *5*(2), 386–392. https://doi.org/10.29207/resti.v5i2.3007

Apridiansyah, Y., David, N., Veronika, M., & Putra, E. D. (2021). Prediksi Kelulusan Mahasiswa Fakultas Teknik Informatika Universitas Muhammadiyah Bengkulu Menggunakan Metode Naive Bayes. JSAI : *Journal Scientific and Applied Informatics*, 4(2), 236–247. https://doi.org/10.36085

Gunawan, M., Zarlis, M., & Roslina, R. (2021). Analisis Komparasi Algoritma Naïve Bayes dan K-Nearest Neighbor Untuk Memprediksi Kelulusan Mahasiswa Tepat Waktu. *Jurnal Media Informatika Budidarma*, 5(2), 513. https://doi.org/10.30865/mib.v5i2.2925

Hutagalung, N. A., P, H., Yusuf, M. A., Frastio, & M Ali Akbar. (2023). Mengimplementasikan Metode Naive Bayes Untuk Memprediksi Tingkat Kelulusan Mahasiswa Tepat Waktu. *INNOVATIVE: Journal Of Social Science Research*, 3(2), 9363–9372. https://j-innovative.org/index.php/Innovative

Mulyadien, M. K., & Enri, U. (2022). Algoritma K-Means Untuk Pengelompokan Bantuan Langsung Tunai (BLT). *Jurnal Ilmiah Wahana Pendidikan*, 8(12), 198–210. https://doi.org/10.5281

Pandey, R. (2020). *What is Python? Javatpoint.* Javapoint. https://www.javatpoint.com/python-tutorial

Vamsi, K. M., Lokesh, P., Reddy, K. N., & Swetha, P. (2021). Visualization of Real World Enterprise Data using Python Django Framework. IOP Conference Series: Materials Science and Engineering, 1042(1), 012019. https://doi.org/10.1088/1757-899x/1042/1/012019

Novaliendry, D., & Hakim, N. U. (2022). Development of A Tracer Study Information System In Senior High School with Devops Method Using Python Application and Django Framework. Jurnal Teknologi Informasi Dan Pendidikan, 15(1), 96–104. https://doi.org/10.24036/jtip.v15i1.586

Aribowo, B. (2021). *Belajar Python dari Nol via Jupyter Notebook*. Diandra Kreatif, Sleman.

Arvyantomo, M., & Ratama, N. (2023). Analisis Sentimen Masyarakat Indonesia terhadap Invasi Russia di Ukraina menggunakan Metode Naive Bayes pada Media Sosial Facebook. *LOGIC: Jurnal Ilmu Komputer*, 1(4), 705–717. https://www.journal.mediapublikasi.id/index.php/logic/article/view/2142

Darmawan, E. M. Z., & Dianta, A. F. (2023). Implementasi Optimasi Hyperparameter GridSearchCV Pada Sistem Prediksi Serangan Jantung Menggunakan SVM. *Teknologi: Jurnal Ilmiah Sistem Informasi*, 13(1), 8–15. https://doi.org/10.26594/teknologi.v13i1.3098

Normawati, D., & Prayogi, S. A. (2021). Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter. *Jurnal Sains Komputer & Informatika (J-SAKTI),* 5(2), 697–711. http://dx.doi.org/10.30645/j-sakti.v5i2.369

Raihan, M., Allaam, R., & Wibowo, A. T. (2021). Klasifikasi Genus Tanaman Anggrek Menggunakan Metode Convolutional Neural Network (CNN). *E-Proceeding of Engineering*, *8*(2), 1153. https://openlibrarypublications.telkomuniversity.ac.id/index.php/engineering/article/view/14708

Wahyono, W., & Nugroho, A. (2019). Penerapan Algoritma C4.5 Untuk Prediksi Tingkat Kompetensi Karyawan Pt Multistrada Arah Sarana. *Joutica*, *3*(1), 145. https://doi.org/10.30736/jti.v3i1.201