

KLASIFIKASI KEPUTUSAN PENDAFTARAN ULANG MAHASISWA BARU UHW PERBANAS MENGGUNAKAN ALGORITMA NAIVE BAYES

CLASSIFICATION OF RE-REGISTRATION DECISIONS FOR NEW STUDENTS AT UHW PERBANAS USING THE NAIVE BAYES ALGORITHM

Syifa Karima^{1*}, Moch. Anang Karyawan¹

*E-mail: syifakarima@gmail.com

¹Sistem Informasi, Fakultas Teknik dan Desain, Universitas Hayam Wuruk Perbanas

Abstrak

Pemanfaatan data historis penerimaan mahasiswa baru pada perguruan tinggi belum dilakukan secara optimal sehingga informasi yang dihasilkan belum mampu memberikan gambaran terkait keputusan daftar ulang mahasiswa baru. Kondisi tersebut menyebabkan proses evaluasi penerimaan mahasiswa masih dilakukan secara terbatas berdasarkan jumlah pendaftar dan mahasiswa yang melakukan registrasi ulang. Penelitian ini bertujuan untuk mengklasifikasikan keputusan daftar ulang mahasiswa baru menggunakan algoritma Naive Bayes berdasarkan data pendaftaran mahasiswa program studi sarjana UHW Perbanas periode 2023 hingga 2025. Penelitian menggunakan pendekatan data mining dengan kerangka CRISP-DM yang terdiri atas business understanding, data understanding, data preparation, modeling, dan evaluation. Proses pengolahan data dilakukan menggunakan *KNIME* Analytics Platform melalui tahapan preprocessing data, pemodelan klasifikasi, dan evaluasi menggunakan confusion matrix dengan pengukuran accuracy, precision, recall, dan f1-score. Hasil pengujian menunjukkan bahwa algoritma Naive Bayes menghasilkan nilai accuracy sebesar 74,67% dengan performa klasifikasi yang cukup baik pada data keputusan daftar ulang mahasiswa baru. Hasil penelitian menunjukkan bahwa data historis penerimaan mahasiswa dapat dimanfaatkan untuk membantu proses evaluasi dan pengambilan keputusan pada penerimaan mahasiswa baru secara lebih objektif dan berbasis data.

Kata kunci: data mining, naive bayes, klasifikasi, mahasiswa baru, daftar ulang.

Abstract

*Historical data on new student admissions at universities has not been optimally utilized, resulting in insufficient information to provide a clear picture of new student re-enrollment decisions. This situation has resulted in the student admission evaluation process being limited to the number of applicants and re-registered students. This study aims to classify new student re-enrollment decisions using the Naive Bayes algorithm based on undergraduate student registration data from UHW Perbanas for the 2023-2025 period. The study uses a data mining approach with the CRISP-DM framework, which consists of business understanding, data understanding, data preparation, modeling, and evaluation. Data processing was carried out using the *KNIME* Analytics Platform through the stages of data preprocessing, classification modeling, and evaluation using a confusion matrix with measurements of accuracy, precision, recall, and f1-score. The test results show that the Naive Bayes algorithm produces an accuracy value of 74.67% with fairly good classification performance on new student re-enrollment decision data. The results of this study indicate that historical student admission data can be*

utilized to assist the evaluation process and decision-making on new student admissions in a more objective and data-driven manner.

Keywords: *data mining, naive bayes, classification, new students, re-enrollment.*

1. PENDAHULUAN

Peningkatan kebutuhan masyarakat terhadap pendidikan tinggi menyebabkan persaingan antar perguruan tinggi dalam memperoleh calon mahasiswa baru semakin meningkat[1]. Perguruan tinggi tidak hanya bersaing dalam menyediakan kualitas akademik, tetapi juga dalam menarik minat calon mahasiswa melalui berbagai strategi penerimaan mahasiswa baru. Kondisi tersebut dipengaruhi oleh semakin banyaknya pilihan perguruan tinggi dan kemudahan akses informasi yang memungkinkan calon mahasiswa membandingkan berbagai institusi sebelum menentukan pilihan akhir[2]. Keputusan calon mahasiswa dipengaruhi berbagai faktor seperti biaya pendidikan, program studi, fasilitas kampus, reputasi institusi, dan peluang karir setelah lulus. Situasi ini menuntut perguruan tinggi memahami pola perilaku calon mahasiswa secara lebih mendalam agar proses penerimaan mahasiswa baru dapat direncanakan secara efektif dan terarah.

Kompleksitas keputusan calon mahasiswa tidak hanya terjadi pada tahap pemilihan perguruan tinggi, tetapi juga pada tahap keputusan daftar ulang setelah dinyatakan diterima. Sebagian calon mahasiswa masih menunggu hasil seleksi dari perguruan tinggi lain sebelum melakukan registrasi akhir. Kondisi tersebut menyebabkan jumlah calon mahasiswa yang melakukan pendaftaran awal tidak selalu sama dengan jumlah mahasiswa yang benar-benar melakukan daftar ulang. Ketidakpastian tersebut memengaruhi proses perencanaan penerimaan mahasiswa baru, khususnya dalam penentuan kuota, strategi promosi, dan evaluasi efektivitas jalur penerimaan. Perguruan tinggi membutuhkan pendekatan yang mampu memberikan gambaran mengenai kecenderungan keputusan daftar ulang calon mahasiswa berdasarkan data historis yang dimiliki. Informasi tersebut dapat membantu perguruan tinggi dalam menyusun strategi penerimaan mahasiswa secara lebih terukur[3]. Kondisi serupa juga terjadi pada proses penerimaan mahasiswa baru di Universitas Hayam Wuruk Perbanas Surabaya (UHW Perbanas). UHW Perbanas telah menerapkan sistem penerimaan mahasiswa baru berbasis digital yang menyimpan data pendaftaran calon mahasiswa secara terstruktur pada setiap periode penerimaan. Data tersebut memuat berbagai informasi yang berkaitan dengan proses pendaftaran mahasiswa baru, seperti waktu pendaftaran, gelombang pendaftaran, jalur penerimaan, program studi, dan kondisi ekonomi calon mahasiswa. Pemanfaatan data historis selama ini masih terbatas pada kebutuhan administratif dan pelaporan jumlah pendaftar sehingga informasi yang tersimpan belum dimanfaatkan secara optimal untuk analisis lebih lanjut. Kondisi tersebut menyebabkan data pendaftaran belum mampu memberikan gambaran yang jelas terkait kecenderungan keputusan daftar ulang mahasiswa baru. Data historis penerimaan mahasiswa sebenarnya memiliki potensi untuk digunakan dalam mengenali pola perilaku calon mahasiswa selama proses penerimaan berlangsung. Pemanfaatan data secara optimal dapat membantu UHW Perbanas menyusun strategi penerimaan mahasiswa baru secara lebih terukur dan sistematis.

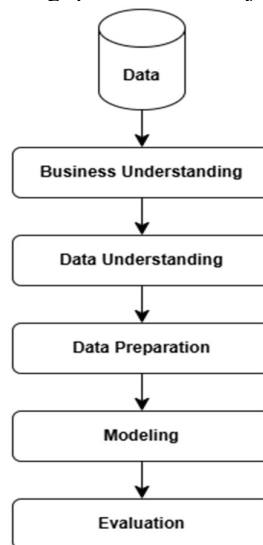
Kebutuhan terhadap pemanfaatan data historis mendorong penggunaan pendekatan data mining dalam proses analisis data penerimaan mahasiswa baru. Data mining digunakan untuk menemukan pola, hubungan, dan informasi penting dari kumpulan data dalam jumlah besar melalui proses analisis tertentu. Salah satu teknik data mining yang banyak digunakan untuk kebutuhan prediksi dan klasifikasi adalah algoritma Naive Bayes. Algoritma ini menggunakan pendekatan probabilitas berdasarkan Teorema Bayes dengan asumsi independensi antar atribut[4]. Karakteristik tersebut membuat Naive Bayes mampu melakukan klasifikasi secara sederhana, cepat, dan efisien pada berbagai jenis dataset. Algoritma Naive Bayes juga mampu

menghasilkan performa yang stabil meskipun menggunakan data dengan atribut yang cukup beragam[5]. Penelitian di bidang pendidikan menunjukkan bahwa metode Naive Bayes dapat digunakan untuk memprediksi perilaku akademik maupun kecenderungan keputusan mahasiswa berdasarkan data historis yang tersedia[6].

Penelitian ini bertujuan mengklasifikasikan keputusan daftar ulang mahasiswa baru menggunakan algoritma Naive Bayes berdasarkan data pendaftaran mahasiswa program studi sarjana UHW Perbanas periode 2023–2025. Pengolahan data dilakukan menggunakan *KNIME* melalui tahapan *preprocessing*, validasi data, pemodelan klasifikasi, dan evaluasi performa model. Pengukuran performa menggunakan confusion matrix dengan metrik *accuracy*, *precision*, *recall*, dan *f1-score*. Hasil penelitian diharapkan dapat memberikan gambaran kecenderungan keputusan daftar ulang mahasiswa baru sebagai pendukung evaluasi proses penerimaan mahasiswa di UHW Perbanas.

2. METODOLOGI

Penelitian ini menggunakan pendekatan data mining untuk mengklasifikasikan keputusan daftar ulang mahasiswa baru di UHW Perbanas menggunakan algoritma Naive Bayes. Proses penelitian disusun menggunakan kerangka CRISP-DM (*Cross Industry Standard Process for Data Mining*) yang terdiri atas *business understanding*, *data understanding*, *data preparation*, *modeling*, dan *evaluation*. Kerangka tersebut digunakan sebagai acuan dalam proses pengolahan dan analisis data penelitian. Alur metodologi penelitian ditunjukkan pada Gambar 1.



Gambar 1. Tahapan Penelitian[7]

Pada tahap *business understanding* dilakukan untuk mengidentifikasi permasalahan terkait keputusan daftar ulang mahasiswa baru pada proses penerimaan mahasiswa di UHW Perbanas. Penelitian difokuskan pada pembangunan model klasifikasi untuk mengetahui kecenderungan keputusan daftar ulang berdasarkan data historis pendaftaran mahasiswa.

Tahap *data understanding* dilakukan dengan memahami karakteristik dan kualitas dataset sebelum proses pemodelan dilakukan. Proses ini meliputi pemeriksaan struktur data, distribusi atribut, missing value, dan inkonsistensi data untuk memastikan kualitas data sesuai dengan kebutuhan penelitian. Hasil pemeriksaan data digunakan sebagai dasar dalam menentukan proses *preprocessing* dan transformasi data pada tahap berikutnya.

Penelitian menggunakan data historis penerimaan mahasiswa baru program studi sarjana UHW Perbanas periode 2023 hingga 2025 yang diperoleh dari sistem penerimaan mahasiswa

berbasis digital. Data penelitian terdiri atas atribut tanggal pendaftaran, gelombang pendaftaran, jalur pendaftaran, program studi, pendapatan orang tua, dan status daftar ulang. Status digunakan sebagai label klasifikasi dalam penelitian. Ringkasan spesifikasi dataset penelitian ditunjukkan pada Tabel 1.

Table 1. Spesifikasi Dataset Penelitian

Atribut	Tipe Data	Keterangan
Tanggal Pendaftaran	Date	Tanggal calon mahasiswa melakukan pendaftaran
Gelombang Pendaftaran	String	Gelombang penerimaan mahasiswa baru
Jalur Pendaftaran	String	Jalur penerimaan mahasiswa baru
Program Studi	String	Program studi yang dipilih calon mahasiswa
Pendapatan Orang Tua	Integer	Pendapatan orang tua/wali calon mahasiswa
Status	String	Status keputusan daftar ulang mahasiswa (Ya/Tidak)

Tahap *data preparation* meliputi proses cleaning data, transformasi data kategorik, dan penyesuaian format atribut sebelum proses klasifikasi dilakukan. Tahapan ini dilakukan agar dataset siap digunakan pada proses klasifikasi. Proses *preprocessing* juga dilakukan untuk mengurangi kemungkinan kesalahan klasifikasi yang disebabkan oleh data yang tidak konsisten atau tidak lengkap.

Tahap *modeling* dilakukan menggunakan algoritma Naive Bayes untuk membangun model klasifikasi keputusan daftar ulang mahasiswa baru. Proses validasi model dilakukan menggunakan *k-fold cross validation* untuk mengukur performa model terhadap data pelatihan dan data pengujian. Algoritma Naive Bayes menggunakan pendekatan probabilitas berdasarkan Teorema Bayes untuk menentukan probabilitas suatu data terhadap kelas tertentu. Persamaan algoritma Naive Bayes ditunjukkan pada Persamaan 1.

$$P(C|X) = \frac{P(C) \cdot P(X)}{P(X)} \dots \dots \dots (1)$$

dimana $P(C|X)$ merupakan probabilitas hipotesis berdasarkan data, $P(X|C)$ merupakan probabilitas data terhadap hipotesis, $P(C)$ merupakan probabilitas awal hipotesis, dan $P(X)$ merupakan probabilitas data.

Tahap *evaluation* dilakukan menggunakan confusion matrix untuk mengetahui performa model klasifikasi yang dihasilkan. Evaluasi model menggunakan *accuracy*, *precision*, *recall*, dan *f1-score* untuk mengukur kemampuan algoritma Naive Bayes dalam mengklasifikasikan keputusan daftar ulang mahasiswa baru. *Accuracy* digunakan untuk mengukur ketepatan klasifikasi model. *Precision*, *recall*, dan *f1-score* digunakan untuk mengevaluasi kualitas hasil klasifikasi pada setiap kelas.

Seluruh proses pengolahan data dilakukan menggunakan *KNIME Analytics Platform* mulai dari preprocessing data, proses klasifikasi, hingga evaluasi model. Hasil analisis digunakan untuk menggambarkan kecenderungan keputusan daftar ulang mahasiswa baru berdasarkan data penerimaan mahasiswa di UHW Perbanas.

3. HASIL DAN PEMBAHASAN

Proses pengolahan data dan pengujian model klasifikasi dilakukan menggunakan *KNIME Analytics Platform* melalui tahapan *preprocessing* data, pemodelan, dan evaluasi performa model. Hasil pengujian digunakan untuk mengetahui kemampuan model dalam mengenali pola keputusan daftar ulang berdasarkan atribut pendaftaran mahasiswa.

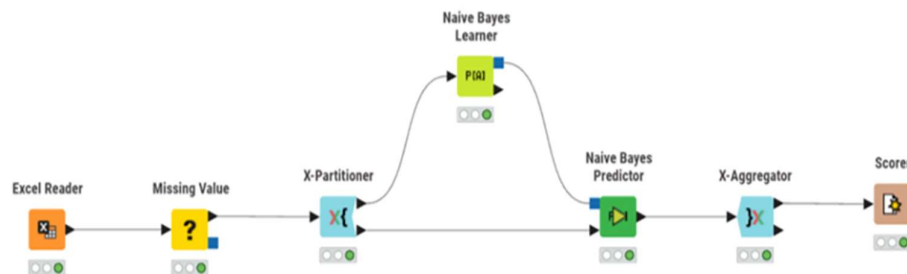
3.1 Hasil

Penelitian menggunakan data historis penerimaan mahasiswa baru program sarjana UHW Perbanas periode 2023–2025 sebanyak 1.528 data. Dataset memuat atribut tahun pendaftaran, bulan pendaftaran, jalur pendaftaran, program studi, pendapatan orang tua, dan status daftar ulang

sebagai label klasifikasi. Sebelum proses pemodelan, data melalui tahap preprocessing sehingga seluruh atribut siap digunakan pada proses klasifikasi.

Tahap *preprocessing* dilakukan melalui *cleaning* data, transformasi atribut kategorikal, dan penyesuaian format atribut. Proses transformasi dilakukan pada atribut tanggal pendaftaran dengan memecah data menjadi atribut tahun dan bulan untuk mempermudah analisis pola waktu pendaftaran mahasiswa. Atribut gelombang pendaftaran tidak digunakan pada tahap pemodelan karena memiliki keterkaitan informasi dengan waktu pendaftaran. Hasil *preprocessing* menghasilkan dataset yang terdiri atas atribut tahun, bulan, jalur, program studi, pendapatan orang tua, dan status sehingga data menjadi lebih terstruktur untuk proses klasifikasi.

Dataset hasil *preprocessing* digunakan pada tahap pengujian model menggunakan metode *10-fold cross validation*. Metode tersebut digunakan untuk mengukur performa model pada data pelatihan dan data pengujian secara bergantian sehingga hasil evaluasi menjadi lebih stabil. Proses klasifikasi dilakukan menggunakan *workflow* pada *KNIME* yang terdiri atas tahapan *preprocessing* data, penerapan model Naive Bayes, dan evaluasi hasil klasifikasi menggunakan confusion matrix. *Workflow* proses klasifikasi ditunjukkan pada Gambar 2.



Gambar 2. Workflow klasifikasi pada KNIME

Hasil pengujian model dievaluasi menggunakan confusion matrix untuk mengetahui kemampuan klasifikasi pada masing-masing kelas. Evaluasi dilakukan dengan membandingkan hasil prediksi model terhadap data aktual sehingga dapat diketahui jumlah data yang berhasil diklasifikasikan secara benar maupun data yang masih mengalami kesalahan klasifikasi. Hasil confusion matrix ditunjukkan pada Gambar 3.

RowID	Tidak Daftar Ulang Number (Integer)	Daftar Ulang Number (Integer)
Tidak Daftar Ulang	785	246
Daftar Ulang	141	356

Gambar 3. Confusion matrix hasil klasifikasi

Hasil confusion matrix menunjukkan bahwa sebagian besar data berhasil diklasifikasikan dengan baik pada kedua kelas. Beberapa kesalahan klasifikasi masih ditemukan pada data yang memiliki karakteristik atribut yang serupa sehingga model mengalami kesulitan dalam membedakan pola antar kelas. Kondisi tersebut menunjukkan bahwa beberapa atribut pada data pendaftaran masih memiliki hubungan yang cukup dekat antara mahasiswa yang melakukan daftar ulang dan yang tidak melakukan daftar ulang. Hasil klasifikasi diukur menggunakan *accuracy*, *precision*, *recall*, dan *f1-score* untuk mengetahui kualitas performa model secara lebih detail. Pengukuran dilakukan pada masing-masing kelas untuk melihat kemampuan model dalam mengenali pola keputusan daftar ulang mahasiswa. Hasil evaluasi model ditunjukkan pada Tabel 2.

Table 2. Hasil Evaluasi Model

Kelas	<i>Precision (%)</i>	<i>Recall (%)</i>	<i>F1-Score (%)</i>
-------	----------------------	-------------------	---------------------

Tidak Daftar Ulang	84,77	76,14	80,22
Daftar Ulang	59,14	71,63	64,79

Hasil evaluasi model menunjukkan nilai *accuracy* sebesar 74,67%. Nilai *precision* tertinggi diperoleh pada kelas Tidak Daftar Ulang sebesar 84,77%, sedangkan nilai *recall* tertinggi pada kelas Tidak Daftar Ulang mencapai 76,14%, dan nilai *f1-score* 80,22%. Pada kelas Daftar Ulang, model memperoleh *precision* sebesar 59,14%, *recall* 71,63%, dan *f1-score* 64,79%, sehingga secara keseluruhan model menunjukkan kemampuan klasifikasi yang cukup baik dalam memprediksi keputusan daftar ulang mahasiswa.

3.2 Pembahasan

Hasil pengujian menunjukkan bahwa algoritma *Naive Bayes* mampu digunakan untuk mengklasifikasikan keputusan daftar ulang mahasiswa berdasarkan data historis penerimaan mahasiswa baru di UHW Perbanas. Nilai *accuracy* sebesar 74,67% menunjukkan bahwa model memiliki kemampuan yang cukup baik dalam mengenali pola keputusan daftar ulang berdasarkan atribut yang digunakan.

Performa model pada kelas Tidak Daftar Ulang lebih tinggi dibandingkan kelas Daftar Ulang. Nilai *precision* sebesar 84,77% menunjukkan bahwa sebagian besar mahasiswa yang diprediksi tidak melakukan daftar ulang sesuai dengan kondisi sebenarnya. Sebaliknya, nilai *precision* pada kelas Daftar Ulang masih lebih rendah karena terdapat beberapa data dengan karakteristik atribut yang mirip sehingga menyebabkan model menghasilkan kesalahan klasifikasi. Meskipun demikian, nilai *recall* sebesar 71,63% menunjukkan bahwa sebagian besar mahasiswa yang benar-benar melakukan daftar ulang masih dapat dikenali oleh model.

Proses *preprocessing* turut mendukung pembentukan model klasifikasi. Transformasi atribut tanggal pendaftaran menjadi atribut tahun dan bulan serta penghapusan atribut gelombang pendaftaran membantu mengurangi redundansi informasi. Selain itu, atribut jalur pendaftaran, program studi, dan pendapatan orang tua memberikan informasi yang berkontribusi dalam pembentukan probabilitas klasifikasi sehingga model mampu mengenali pola keputusan daftar ulang secara lebih efektif.

Hasil penelitian menunjukkan bahwa pendekatan data mining menggunakan algoritma *Naive Bayes* dapat dimanfaatkan sebagai pendukung analisis penerimaan mahasiswa baru. Informasi yang dihasilkan dapat membantu UHW Perbanas mengidentifikasi calon mahasiswa yang berpotensi tidak melakukan daftar ulang sehingga strategi komunikasi, promosi, maupun tindak lanjut kepada calon mahasiswa dapat dilakukan secara lebih terarah.

4. KESIMPULAN DAN SARAN

Penelitian ini menunjukkan bahwa algoritma *Naive Bayes* dapat digunakan untuk mengklasifikasikan keputusan daftar ulang mahasiswa baru berdasarkan data penerimaan mahasiswa UHW Perbanas. Model dibangun menggunakan atribut tahun pendaftaran, bulan pendaftaran, jalur pendaftaran, program studi, dan pendapatan orang tua, serta menghasilkan *accuracy* sebesar 74,67% dengan nilai *precision*, *recall*, dan *f1-score* yang menunjukkan performa model berada pada kategori cukup baik. Hasil tersebut menunjukkan bahwa data historis penerimaan mahasiswa memiliki pola yang dapat dimanfaatkan untuk mendukung proses prediksi keputusan daftar ulang mahasiswa.

Penelitian ini juga menunjukkan bahwa proses *preprocessing* dan penggunaan *KNIME* membantu pengolahan data dilakukan secara sistematis dan terstruktur. Hasil klasifikasi yang diperoleh masih dapat dikembangkan lebih lanjut melalui penambahan jumlah data maupun atribut lain yang berkaitan dengan keputusan daftar ulang mahasiswa. Penggunaan metode klasifikasi lain juga dapat dilakukan sebagai perbandingan untuk mengetahui performa model yang lebih optimal pada data penerimaan mahasiswa baru. Pengembangan tersebut diharapkan

mampu meningkatkan kualitas hasil klasifikasi sehingga dapat mendukung proses evaluasi dan pengambilan keputusan pada penerimaan mahasiswa baru secara lebih efektif.

5. DAFTAR RUJUKAN

- [1] W. U. Putri, R. Alexandro, and A. Rahmadianor, “Minat Siswa SMA Swasta untuk Melanjutkan Ke Perguruan Tinggi Universitas Palangka Raya,” *J. Pendidik. Ilmu Pengetah. Sos.*, vol. 12, no. 1, pp. 57–61, 2020.
- [2] R. Pardiyo and H. D. Puspita, “Studi Tentang Faktor Dominan dalam Pemilihan Perguruan Tinggi Swasta,” vol. V, no. 2, pp. 279–289, 2021.
- [3] T. Mesra, Fitra, R. Faisal, and F. Alfiyad, “Faktor-Faktor Yang Mempengaruhi Pemilihan Perguruan Tinggi Swasta Di Kota Dumai,” *Integr. J. Ilm. Tek. Ind.*, vol. 10, no. 02, pp. 181–192, 2025.
- [4] A. Al-Kasidmi and Megawaty Byah Ayu, “Penerapan Algoritma Naïve Bayes untuk Prediksi Minat Siswa Melanjutkan Studi ke Perguruan Tinggi,” *DINAMIK*, vol. 31, no. 1, pp. 313–322, 2026, doi: <https://doi.org/10.35315/dinamik.v31i1.10416>.
- [5] A. Ainunrochman, M. N. Hayati, and A. M. A. Striya, “Perbandingan Klasifikasi Analisis Diskriminan Fisher Dan Metode Naive Bayes,” *J. Apl. Stat. Komputasi Stat.*, vol. 11, no. 2, pp. 37–48, 2019, doi: <https://doi.org/10.34123/jurnalasks.v11i2.156>.
- [6] R. Pikriyansah, F. R. Umbara, and P. N. Sabrina, “Klasifikasi Daftar Ulang Calon Mahasiswa Baru Dengan Menggunakan Metode Klasifikasi Naive Bayes,” *Informatics Digit. Expert*, vol. 4, no. 2, pp. 70–74, 2022.
- [7] K. A. Pratama, G. A. Pradnyana, and I. K. R. Arthana, “Pengembangan Sistem Cerdas Untuk Prediksi Daftar Kembali Mahasiswa Baru Dengan Metode Naive Bayes (Studi Kasus: Universitas Pendidikan Ganesha),” *SINTECH (Science Inf. Technol. J.)*, vol. 3, no. 1, pp. 22–34, 2020.