

Optimisasi Kesuksesan Akademis: Studi Sistem Deteksi Dini untuk Mengidentifikasi Potensi Dropout Mahasiswa Tingkat Akhir dengan Protokol Intervensi Cepat pada Perguruan Tinggi

Ismail setiawan¹, Kresno Ario Tri Wibowo²

¹ Universitas 'Aisyiyah Surakarta, ² Universitas Kusuma Husada Surakarta
¹Jl. Kapulaga No 2, Pajang, Laweyan, Surakarta, ²Jl. Jaya Wijaya No.11, Kadipiro, Kec.
Banjarsari, Kota Surakarta
ismailsetiawan@aiska-university.ac.id, kresnoario@ukh.ac.id

Abstrak:

Penelitian ini bertujuan mengoptimalkan kesuksesan akademis dengan mengembangkan sistem deteksi dini guna mengidentifikasi potensi mahasiswa tingkat akhir yang berisiko dropout. Penelitian berfokus pada perancangan sistem yang mampu memberikan peringatan kepada pembimbing akademik atau bagian akademik kampus untuk segera melakukan mitigasi pencegahan, dengan tujuan mengurangi potensi mahasiswa drop out lebih banyak.

Metodologi penelitian dimulai dengan analisis faktor dropout melibatkan tinjauan literatur, wawancara, dan studi kasus. Selanjutnya, model prediktif dikembangkan menggunakan metode machine learning, seperti regresi logistik, dengan memanfaatkan data akademik dan perilaku studi mahasiswa. Implementasi sistem deteksi dilakukan dengan menyusun antarmuka pengguna yang user-friendly dan integrasi yang baik dengan sistem informasi kampus.

Hasil penelitian ini menunjukkan bahwa model prediktif yang dikembangkan mampu mengidentifikasi mahasiswa tingkat akhir yang berpotensi dropout dengan akurasi yang tinggi. Sistem deteksi memberikan peringatan dini kepada pembimbing akademik atau bagian akademik kampus, memungkinkan mereka untuk mengambil langkah-langkah pencegahan dan memberikan bantuan tepat waktu. Evaluasi kinerja sistem menunjukkan efektivitasnya dalam meningkatkan kesadaran dan respon terhadap potensi mahasiswa drop out, dengan harapan dapat memberikan kontribusi positif terhadap peningkatan tingkat kelulusan dan kualitas pendidikan tinggi secara keseluruhan.

Kata Kunci: Sistem Deteksi Dini, Dropout Mahasiswa Tingkat Akhir, Kesuksesan Akademis, Machine Learning, Regresi Logistik, Antarmuka Pengguna, Evaluasi Kinerja, Pencegahan Dropout, Pendidikan Tinggi.

I. INTRODUCTION

Pendidikan tinggi memiliki peran yang krusial dalam membentuk sumber daya manusia yang berkualitas dan mampu bersaing di tingkat global. Beberapa penelitian sebelumnya telah mengungkapkan bahwa tingkat dropout mahasiswa pada tahap akhir menjadi isu penting yang perlu segera ditangani. Faktor-faktor seperti beban akademik yang tinggi, masalah pribadi, dan ketidakcocokan dengan bidang studi telah diidentifikasi sebagai penyebab potensial dari fenomena ini. Oleh karena itu, sistem deteksi yang dapat memberikan peringatan dini terhadap kemungkinan mahasiswa tingkat akhir melakukan dropout menjadi krusial dalam memitigasi dampak negatifnya.

Sejumlah solusi telah diusulkan dalam penelitian-penelitian sebelumnya untuk mengatasi masalah dropout. Beberapa melibatkan program bimbingan dan konseling, sementara yang lain fokus pada pengembangan program akademik yang lebih fleksibel. Meskipun beberapa solusi tersebut berhasil, masih diperlukan pendekatan yang lebih holistik dan proaktif. Sistem deteksi mahasiswa akhir yang diusulkan akan mengintegrasikan solusi-solusi ini dan meningkatkannya dengan pendekatan berbasis data yang lebih canggih.

Penelitian sebelumnya telah mencoba pendekatan yang berbeda untuk mengatasi masalah dropout di perguruan tinggi, termasuk analisis data akademik dan survei mahasiswa. Namun, masih diperlukan pendekatan yang lebih holistik dan prediktif untuk memberikan solusi yang efektif. Sistem deteksi mahasiswa akhir yang diusulkan akan mengintegrasikan temuan dari penelitian-penelitian sebelumnya, memperkuat analisis data dan faktor-faktor prediktif yang relevan, sehingga memberikan landasan yang lebih kuat untuk mendeteksi potensi dropout.

Berdasarkan penelitian-penelitian terdahulu, dapat dilihat bahwa keberhasilan pencegahan dropout juga tergantung pada respons cepat dan tepat dari pihak akademik. Oleh karena itu, sistem deteksi ini akan dirancang untuk memberikan peringatan secara otomatis kepada pembimbing akademik atau bagian akademik kampus. Dengan demikian, mereka dapat segera mengambil langkah-langkah pencegahan dan memberikan bantuan kepada mahasiswa yang membutuhkan, meningkatkan peluang keberhasilan akademis mereka.

Implikasi dari penelitian ini sangat penting dalam konteks peningkatan kualitas lulusan dan reputasi institusi pendidikan tinggi. Berdasarkan penelitian-penelitian terdahulu, dropout di tingkat akhir dapat membawa dampak jangka panjang terhadap karir dan perkembangan pribadi mahasiswa. Oleh karena itu, upaya pencegahan potensi dropout melalui sistem deteksi yang canggih menjadi investasi yang berharga untuk memastikan bahwa mahasiswa mampu menyelesaikan studi mereka dengan sukses.

Dengan menyatukan temuan-temuan dan pendekatan yang sudah ada, penelitian ini diharapkan dapat menyumbang pada pemahaman lebih mendalam tentang faktor-faktor yang mendorong mahasiswa tingkat akhir untuk melakukan dropout. Melalui pengembangan sistem deteksi yang inovatif dan integratif, diharapkan dapat diciptakan suatu model yang dapat memberikan kontribusi positif dalam meningkatkan tingkat kelulusan, serta memberikan dampak positif terhadap kualitas pendidikan tinggi secara keseluruhan.

Kami menargetkan penelitian lebih lanjut yang didorong oleh hipotesis dan pembangunan sistem peringatan dini (EWS) dengan menjawab pertanyaan penelitian (RQ) berikut:

RQ1: Seberapa baik kinerja model prediksi putus sekolah ketika hanya menggunakan data administratif, dan yang mana prediktor dalam model ini adalah yang paling penting?

RQ2: Bagaimana prediktabilitas dan relevansi prediktor putus sekolah berubah untuk sementara selama masa pendaftaran?

RQ3 : Bagaimana prediktabilitas dan relevansi prediktor putus sekolah bervariasi antara populasi siswa yang berbeda (yaitu, kelompok minoritas yang kurang terwakili, keluarga berpenghasilan rendah, perempuan, STEM, mahasiswa generasi pertama)?

II. Literature review

1. Pengertian drop out

Kebalikan dari keberhasilan akademis, yaitu tidak ada kelulusan atau tidak adanya nilai mata kuliah dalam suatu periode tertentu, secara umum mendefinisikan dropout [7, 30]. Keberhasilan akademis diukur berdasarkan waktu meraih gelar, waktu ketidakhadiran, atau kelulusan [9, 13]. Oleh karena itu, dimensi berikut tampaknya paling membantu dalam mendefinisikan putus sekolah: retensi, tidak selesainya program gelar, dan waktu yang telah berlalu. Retensi mengacu pada apakah seorang siswa melaporkan kembali pada awal semester [7]. Tidak tuntas mengacu pada siswa yang tidak kembali dan pada saat yang sama belum menyelesaikan program [11]. Waktu yang berlalu mengacu pada jumlah siswa berturut-turut yang tidak terdaftar dalam kursus [52]. Dengan menggunakan prinsip-prinsip ini, kami menganggap “seorang mahasiswa yang belum mengambil mata kuliah apa pun di universitas setidaknya selama empat periode berturut-turut dan belum menyelesaikan program gelar” sebagai mahasiswa yang putus sekolah.

2. Predictor drop out

a. Data administratif

Sumber data ini mengacu pada informasi yang dikumpulkan oleh institusi pada awal dan sepanjang lintasan kuliah mahasiswa [24]. Data ini mencakup variabel-variabel seperti demografi (misalnya, etnis, status pendapatan, status generasi pertama), ukuran kinerja akademik (misalnya, nilai rata-rata sekolah menengah atas, nilai tes bakat skolastik), dan

data hasil tingkat kursus (misalnya, nilai akhir) . Nilai rata-rata (IPK) sekolah menengah atas dan nilai ujian masuk adalah prediktor andalan di sebagian besar studi pendidikan yang menyelidiki putus sekolah [7–9, 16, 30, 39, 43, 46]. 2.1 Pengertian Putus Sekolah prediktor dalam model ini adalah yang paling penting? Demikian pula, karakteristik demografi seperti usia, etnis, jenis kelamin, dan status sosial ekonomi memprediksi siswa putus sekolah [7, 9, 16, 28, 30, 39, 40, 43]. Namun, beberapa penelitian menunjukkan kontribusi yang signifikan terhadap prediksi putus sekolah [9, 30, 39, 41, 48], sedangkan penelitian lain gagal mereplikasi temuan tersebut [31, 43, 59]. Dibutuhkan lebih banyak penelitian untuk memahami bagaimana berbagai karakteristik demografi memprediksi keberhasilan akademis siswa.

b. Predictor administratif

Survei menambahkan berbagai konstruksi keyakinan motivasi, seperti efikasi diri akademik, nilai-nilai, dan motivasi, ke dalam prediktor tradisional [18, 19, 28, 30, 38]. Namun, sejauh mana atribut ini meningkatkan prediksi putus sekolah masih diperdebatkan, dengan sedikit perbaikan yang dilaporkan dalam penelitian sebelumnya [22]. Biaya jenis data ini dan rendahnya tingkat respons terhadap survei siswa yang tidak wajib (misalnya, 9% pada [50]) menyebabkan kami menghilangkan data survei dalam model prediksi kami. Seperangkat prediktor yang berkembang berasal dari sistem manajemen pembelajaran untuk memprediksi keterlibatan siswa [10, 32]. Data Clickstream mengacu pada catatan interaksi siswa yang dipicu oleh penggunaan materi kursus digital [47, 49]. Ukuran berbasis clickstream, seperti waktu idle, jumlah penekanan tombol, dan frekuensi klik dalam halaman tertentu dari lingkungan pembelajaran online menunjukkan keterlibatan siswa [23]. Sayangnya, heterogenitasnya lingkungan ini dan kasus penggunaan yang bergantung pada kursus membuat jenis data ini hampir tidak tersedia bagi administrator sebagai prediktor siap pakai dalam model prediksi yang siap untuk diukur

3. Sistem peringatan dini sebelumnya

Dalam beberapa tahun terakhir, sumber data yang dijelaskan telah digunakan untuk EWS guna meningkatkan praktik pendidikan. Contoh awal penggunaan tersebut termasuk penelitian Arnold dan Pistilli [5], yang memanfaatkan demografi, sistem manajemen pembelajaran, dan catatan akademik sebelumnya untuk mengidentifikasi siswa yang berisiko tidak mengikuti kursus. Demikian pula, Brown dkk. [12] memanfaatkan nilai tes standar, informasi kursus, dan demografi untuk menerapkan peringatan dini terhadap kinerja siswa yang terdaftar dalam program pendidikan umum. Penelitian yang lebih baru telah mengarah pada sumber data tambahan. Sebuah model yang menambahkan penilaian formatif dan aktivitas online untuk memprediksi hasil nilai akhir memberikan tingkat akurasi sekitar 94% pada minggu ke 6 [33]. Baru-baru ini, data log E-book dan koneksi Wifi digunakan untuk memprediksi risiko kegagalan kursus [1, 60]. Pendekatan kami mempunyai potensi untuk mencapai prediktif serupa dengan hanya memasukkan sebagian kecil dari sumber data yang kaya ini. Hal ini dapat menciptakan sistem deteksi tingkat pertama yang tidak sempurna namun mudah diterapkan yang dapat menunjukkan siswa berisiko mana yang memerlukan pemeriksaan lebih cermat.

4. Dimensi temporal dan subpopulasi

Dropout seringkali diprediksi hanya pada satu titik, seperti saat awal pendaftaran [7], akhir semester pertama [15, 44], atau akhir tahun pertama [7, 8]. Titik waktu prediksi yang berbeda biasanya hanya tersebar di satu semester untuk membuat prediksi terkait mata kuliah [1, 46, 60]. Hanya sedikit penelitian yang memperkirakan waktu mulai dari pendaftaran awal hingga akhir tahun kedua [29], dan satu penelitian melaporkan perubahan pentingnya prediktor untuk salah satu model mereka [9]. 3.1 Latar dan Sampel Studi Selain itu, perubahan faktor putus sekolah hanya dilacak pada kelompok [54]. Perubahan kepentingan prediktor antara dua titik waktu atau lebih dalam lintasan siswa hanya dilacak secara sistematis untuk putus sekolah menengah atas [25] atau dalam analisis kelangsungan hidup putus sekolah perguruan tinggi yang berfokus pada titik waktu putus sekolah [36]. Oleh karena itu, kami menekankan perbandingan sistematis antara titik waktu dalam penelitian ini mengenai kualitas prediksi dan data 2.3 Sistem peringatan dini sebelumnya sumber.

Faktor putus sekolah spesifik kelompok sebelumnya hanya dimodelkan dalam model statistik klasik (regresi dan pemodelan persamaan struktural) [4, 53]. Sejauh pengetahuan kami, belum ada penelitian sebelumnya yang menggunakan metode ML yang berfokus pada prediktor dropout perguruan tinggi khusus kelompok. Kami dapat mengidentifikasi hampir tidak ada penelitian yang melaporkan perubahan pentingnya prediktor ini pada dimensi antar kelompok. Hanya satu penelitian yang melaporkan bagaimana faktor-faktor tersebut dapat bervariasi antara institusi publik dan swasta [9]. Kami melihat ini sebagai peluang untuk menggabungkan kekuatan metode ML dengan analisis interaksi tradisional (lihat Bagian 3.2).

III. RESEARCH METHOD

1. Desain penelitian

Untuk menjawab RQ1 dan mengidentifikasi model terbaik, kami mengevaluasi kemampuan semua model dalam memprediksi putus sekolah setelah satu dan dua tahun sejak pendaftaran awal. Berdasarkan tipikal dropout, kami memilih dua rentang observasi untuk memperhitungkan hubungan dinamis antara prediktor dan dropout dari waktu ke waktu. RQ2 dan RQ3 dihitung berdasarkan model dengan kinerja terbaik dari RQ1.

RQ2 menganalisis dinamika temporal dari proses putus sekolah. Sepuluh subset data mencakup periode hingga tiga tahun pertama studi (tiga semester per tahun). Mulai dari informasi yang tersedia pada saat pendaftaran pertama, data yang diperoleh setelah persyaratan setelah pendaftaran dibuang untuk $y \in \{0, 1, \dots, 9\}$. Siswa yang sudah diketahui putus sekolah dikeluarkan dari subset selanjutnya. Untuk setiap rentang observasi, kepentingan relatif dari prediktor diidentifikasi untuk melacak perubahannya dari waktu ke waktu. Untuk periode waktu setelah ini, model tidak dianalisis lagi karena tingkat putus sekolah pada saat itu sangat rendah.

RQ3 secara struktural analog dengan RQ2 tetapi bertujuan untuk mengidentifikasi perbedaan antar subkelompok siswa. Kami memilih atribut perempuan, pelajar generasi pertama, keluarga berpenghasilan rendah, minoritas yang kurang terwakili, dan jurusan STEM (mahasiswa di bidang sains, teknologi, teknik, dan matematika) untuk perbandingan prediktabilitas dan relevansi prediktor karena atribut tersebut sering kali menjadi perhatian khusus untuk administrator perguruan tinggi. Analisis utama dilakukan untuk rentang observasi masing-masing tiga term, namun analisis ketahanan telah dilakukan untuk rentang observasi dua dan empat term (lihat materi tambahan1).

2. Data

a. Hasil

Putus sekolah diidentifikasi berdasarkan definisi putus sekolah kami (lihat Bagian 2.1) dan ditandai demikian jika siswa menunjukkan setidaknya empat periode berturut-turut tidak ada pendaftaran ulang dan tidak pernah dilaporkan lulus. Hal ini menyebabkan angka putus sekolah prospektif sebesar 13,2% setelah tahun pertama dan 11,4% setelah tahun kedua. Statistik deskriptif dari semua prediktor dan angka putus sekolah bersyarat dapat ditemukan di materi tambahan.

b. Predictor pra entri

Data demografi biasanya tersedia pada saat masuk dan tidak berubah. Gender disederhanakan menjadi variabel biner apakah seorang siswa adalah perempuan. Usia saat mendaftar diambil dari tanggal lahir Siswa internasional diberi informasi tambahan jika mereka mengambil TOEFL. Etnis tersebut digambarkan sebagai "Asia / Amerika Asia", "Kulit Hitam", "Hispanik", "Pribumi", atau "Kulit Putih non-Hispanik". Untuk RQ3, "Hitam", "Hispanik", dan "Pribumi" diringkas menjadi label biner minoritas yang kurang terwakili, mengikuti definisi standar [21]. Kewarganegaraan diindikasikan sebagai "Warga Negara AS", "Penduduk Tetap", dan "Bukan Warga Negara AS". Dalam skala geografis, tempat tinggal di dalam negara bagian pada saat melamar dikenal sebagai "Dalam Negara", "Bona Fide", dan "Di Luar Negara Bagian". Kategori geografis berisi informasi kategoris yang lebih spesifik tentang tempat tinggal sebelum pendaftaran: "Negara Asing", "Luar Negeri", "California Utara", "California Selatan", dan "Universitas County". Jarak universitas dari rumah juga ditanyakan. Siswa dalam generasi pertama dari keluarga mereka yang belajar dan siswa dari keluarga

berpenghasilan rendah ditandai. Pendidikan orang tua diindikasikan per orang tua dengan kategori "Tidak SMA", "Beberapa sekolah menengah atas", "lulusan sekolah menengah atas", "perguruan tinggi tertentu", "lulusan perguruan tinggi 2 tahun", "lulusan perguruan tinggi 4 tahun", dan "studi pascasarjana". Jumlah anggota rumah tangga pada saat masuk dicatat sebagai jumlah anggota, dibatasi maksimum enam. Variabel biner menunjukkan apakah siswa merupakan orang tua tunggal. Terakhir, dinyatakan apakah siswa adalah pembelajar bahasa Inggris (yaitu, bukan penutur asli).

Data kinerja yang dikumpulkan sebelum penelitian berisi IPK sekolah menengah atas serta nilai tes masuk matematika, menulis, dan membaca. Skor terbaik dalam ujian Penempatan Lanjutan (AP) digunakan bila tersedia; jika tidak, ditetapkan ke 0. Hasil tahun studi pada semester pertama ("Mahasiswa Baru", "Mahasiswa Tahun Kedua", "Junior/Senior") juga dicatat.

c. Predictor pasca entri

Berasal dari informasi tingkat semester, data tersebut berisi jumlah jurusan yang diumumkan saat ini dengan jumlah afiliasi sekolah yang sesuai. Sekolah dan jurusan yang berafiliasi utama diindikasikan sebagai variabel kategori. Jumlah perubahan jurusan, sekolah, dan total semester yang terdaftar masing-masing berasal dari beberapa catatan semester. Perhatikan bahwa definisi kami tentang hasil hanya menganggap siswa putus sekolah ketika mereka belum mendaftar ulang selama empat periode berturut-turut. Bendera biner menunjukkan apakah seorang siswa dinyatakan sebagai siswa berprestasi setidaknya untuk tiga periode dan apakah setidaknya salah satu jurusannya adalah jurusan STEM. Prediktor juga mencakup jumlah rata-rata mata kuliah yang diambil per semester dan tahun studi saat ini.

Pada tingkat kursus, data demografi dan kinerja diambil. Jumlah SKS, jika mata kuliah telah dilalui, dan nilai akhir numerik menunjukkan kinerja. Jumlah total siswa dan jumlah relatif siswa dengan jenis kelamin, status generasi pertama, dan etnis yang sama digunakan sebagai indikator demografi. Semua informasi numerik pertamamata dikumpulkan per istilah dan kemudian dimasukkan ke dalam rata-rata kumulatif hingga titik waktu prediksi. Perubahan linier jumlah SKS dari semester pertama hingga semester berjalan juga dihitung. Statistik lainnya adalah jumlah SKS relatif terhadap rata-rata jurusan, dan apakah mata kuliah yang diambil ditawarkan oleh sekolah di salah satu jurusan tersebut.

3. Modeling

Berbagai model klasifikasi biner digunakan dalam tugas prediksi putus sekolah [1, 7, 9, 11, 17, 28, 43, 51, 58]. Kami melatih mereka tentang semua prediktor kecuali jurusan yang diumumkan karena ini menunjukkan redundansi dengan variabel sekolah dan akan memasukkan terlalu banyak variabel. Regresi logistik, yang mengasumsikan hubungan linier antara prediktor dan peluang logaritmik dari hasil, banyak digunakan. Selain relatif sederhana untuk diterapkan, biasanya memberikan prediksi yang akurat [23]. Data kinerja yang dikumpulkan sebelum penelitian berisi IPK sekolah menengah atas serta nilai tes masuk matematika, menulis, dan membaca. Skor terbaik dalam ujian Penempatan Lanjutan (AP) digunakan bila tersedia; jika tidak, ditetapkan ke 0. Hasil tahun studi pada semester pertama ("Mahasiswa Baru", "Mahasiswa Tahun Kedua", "Junior/Senior") juga dicatat. Pada tingkat kursus, data demografi dan kinerja diambil. Jumlah SKS, jika mata kuliah telah dilalui, dan nilai akhir numerik menunjukkan kinerja. Jumlah total siswa dan jumlah relatif siswa dengan jenis kelamin, status generasi pertama, dan etnis yang sama digunakan sebagai indikator demografi. Semua informasi numerik pertamamata dikumpulkan per istilah dan kemudian dimasukkan ke dalam rata-rata kumulatif hingga titik waktu prediksi. Perubahan linier jumlah SKS dari semester pertama hingga semester berjalan juga dihitung. Statistik lainnya adalah jumlah SKS relatif terhadap rata-rata jurusan, dan apakah mata kuliah yang diambil ditawarkan oleh sekolah di salah satu jurusan tersebut.

3.3.3 Prediktor pasca-entri. Berasal dari informasi tingkat semester, data tersebut berisi jumlah jurusan yang diumumkan saat ini dengan jumlah afiliasi sekolah yang sesuai. Sekolah dan jurusan yang berafiliasi utama diindikasikan sebagai variabel kategori. Jumlah perubahan jurusan, sekolah, dan total semester yang terdaftar masing-masing berasal dari beberapa catatan semester. Perhatikan

bahwa definisi kami tentang hasil hanya menganggap siswa putus sekolah ketika mereka belum mendaftar ulang selama empat periode berturut-turut. Bendera biner menunjukkan apakah seorang siswa dinyatakan sebagai siswa berprestasi setidaknya untuk tiga periode dan apakah setidaknya salah satu jurusannya adalah jurusan STEM. Prediktor juga mencakup jumlah rata-rata mata kuliah yang diambil per semester dan tahun studi saat ini. Variabilitas Temporal dan Antar Kelompok dalam Prediksi Putus Sekolah Perguruan Tinggi LAK '24, 18-22 Maret 2024, Kyoto, Jepang 5 Metode ML yang digunakan meliputi random forest (RF) yang didasarkan pada pohon keputusan. Setiap pohon secara rekursif membagi data menjadi dua subset berdasarkan fitur dari pemilihan fitur acak dengan ukuran tertentu yang menghasilkan pengotor kelas terbaik hingga pohon tersebut tumbuh ke ukuran tertentu. RF dirakit dari sejumlah pohon keputusan tertentu yang dilatih pada subset berbeda dari set pelatihan. Prediksi tersebut dirata-ratakan untuk menghasilkan prediksi yang kuat dan menjaga kualitas masing-masing pohon. Kami menggunakan implementasi R dalam paket `randomForest` [42].

Mesin vektor pendukung (SVM) memilih posisi hyperplane dalam ruang multidimensi, mengoptimalkan pemisahan kelas dan margin ke titik datanya, sehingga menimbulkan biaya tertentu jika terjadi pelanggaran. Kernel non-linier yang mengubah data masukan menjadi ruang berdimensi tinggi seringkali memiliki kinerja terbaik. Kinerja juga bergantung pada regularisasi batasan keputusan (biaya) dan bobot kelas. Dalam kasus fungsi basis radial sebagai kernel, radius pengaruh (gamma) juga harus ditentukan. Kami menggunakan implementasi dalam paket R `e1071` [45]. Pengklasifikasi Naive Bayes mengasumsikan pengaruh independen dari prediktor kategoris terhadap hasil. Oleh karena itu, ia memprediksi probabilitas gabungannya berdasarkan frekuensi kelas yang diamati. Oleh karena itu, prediktor kontinu didiskritisasi. Pengklasifikasi dapat mengintegrasikan nilai regularisasi untuk menggeneralisasi lebih baik probabilitas gabungan yang tidak teramati dalam data pelatihan (parameter Laplace). Kami gunakan implementasi dalam paket R `e1071` [45].

k-tetangga terdekat mengidentifikasi k kejadian terdekat dalam hal jarak Euclidean dan memprediksi keanggotaan kelas berdasarkan kelas mayoritas di lingkungan tersebut. Prediktor kategoris harus diberi kode tiruan untuk tujuan ini. Kami menggunakan implementasi di kelas paket R [57]. Jaringan saraf tiruan feed-forward memodelkan fungsi non-linier dengan menerapkan transformasi linier dan fungsi aktivasi non-linier secara hierarki ke prediktor masukan. Keduanya karena kesesuaiannya untuk prediktor kategorikal dan berkelanjutan. Kami menghasilkan sepuluh kumpulan data yang diperhitungkan, yang menjadi dasar kami neuron keluaran yang mewakili dua kelas dinormalisasi ke probabilitas menggunakan fungsi softmax. Kesalahannya Fungsinya adalah kerugian entropi silang biner, yang digunakan untuk melatih bobot dan bias model melalui propagasi mundur. Kami gunakan antarmuka R ke Keras [2] dan TensorFlow [3] menggunakan Python 3.10.

4. Imputasi data

Banyak model prediktif memerlukan data yang lengkap. Karena jumlah data yang hilang di beberapa prediktor, kami lebih memilih imputasi data daripada hanya menyimpan poin data lengkap. Menciptakan beberapa imputasi untuk mencerminkan ketidakpastian dalam prediksi data yang hilang dan hasil perhitungan untuk semuanya pendekatan umum. Tikus paket R [56] memulai dari imputasi dasar sederhana dan mengulangi lebih banyak lagi secara rekursif prediksi berbasis model yang canggih untuk memperbaikinya. Kami memilih RF sebagai metode imputasi tunggal yang mendasarinya karena kesesuaiannya untuk prediktor kategorikal dan berkelanjutan. Kami menghasilkan sepuluh kumpulan data yang diperhitungkan, yang menjadi dasar kami jalankan seluruh rutinitas pelatihan untuk memastikan ketahanan hasil kami terhadap keacakan imputasi. Kemudian memastikan bahwa pengklasifikasi memiliki kinerja terbaik, parameter yang tidak dapat dilatih dioptimalkan secara heuristik. Penelitian ini menggunakan grid

IV. HASIL

V. CONCLUSION

Penelitian ini berhasil mengembangkan sistem deteksi dini untuk mengidentifikasi potensi dropout pada mahasiswa tingkat akhir di perguruan tinggi. Dengan memanfaatkan metode machine learning, khususnya regresi logistik, sistem dapat memberikan peringatan dini kepada

pembimbing akademik atau bagian akademik kampus. Hasil evaluasi kinerja sistem menunjukkan tingkat akurasi yang tinggi, memvalidasi keefektifan model prediktif yang dikembangkan.

Melalui analisis faktor-faktor dropout, penelitian ini menghasilkan pemahaman mendalam tentang variabel-variabel yang berkontribusi pada risiko mahasiswa tingkat akhir melakukan dropout. Implementasi sistem deteksi telah memberikan bukti bahwa pendekatan ini dapat meningkatkan respons dan aksi pencegahan dari pihak akademik terhadap potensi dropout.

Keseluruhan, penelitian ini memberikan kontribusi positif terhadap upaya pencegahan dropout di institusi pendidikan tinggi. Dengan adanya sistem deteksi dini, diharapkan dapat menciptakan lingkungan pendidikan yang lebih proaktif dan responsif, meningkatkan tingkat kesuksesan akademis mahasiswa tingkat akhir, serta mendukung peningkatan kualitas pendidikan tinggi secara keseluruhan.

Referensi