

Pengembangan Sistem Informasi Prediksi Risiko Dropout Mahasiswa Berbasis Web Menggunakan Algoritme CatBoost

DOI: <http://dx.doi.org/10.35889/progresif.v22i2.3656>

Creative Commons License 4.0 (CC BY –NC)



Tania Aurellia^{1*}, Genrawan Hoendarto², Thommy Willay³

^{1,3}Sistem Informasi, Universitas Widya Dharma Pontianak, Pontianak, Indonesia

²Informatika, Universitas Widya Dharma Pontianak, Pontianak, Indonesia

*e-mail *Corresponding Author*: 22412910_tania_aurelia@widyadharma.ac.id

Abstract

The problem addressed was the limited use of dropout prediction models, which had generally focused on algorithm performance and had not been integrated into an academic monitoring system design. This study aimed to develop a web-based prototype information system for predicting student dropout risk using the CatBoost algorithm to support academic monitoring. The development method used was Prototyping, while the prediction model was built from the Predict Students Dropout and Academic Success dataset, which was reduced to 3,399 records with two classes, namely dropout and graduate. The results showed that the developed prototype included manual prediction, CSV import, prediction monitoring, and role-based reporting features. The CatBoost model achieved 90.29% accuracy, 85.33% precision, 88.76% recall, and 87.01% F1-score. These findings indicated that the prototype had the potential to serve as a basis for developing an early detection system for students at risk of dropout.

Keywords: *Academic monitoring; CatBoost; Dropout prediction; System prototype; Web-based information system*

Abstrak

Permasalahan yang diangkat adalah keterbatasan pemanfaatan model prediksi dropout yang umumnya masih berfokus pada performa algoritme dan belum terintegrasi ke dalam rancangan sistem pemantauan akademik. Penelitian ini bertujuan mengembangkan prototype sistem informasi prediksi risiko dropout mahasiswa berbasis web menggunakan algoritme CatBoost untuk mendukung pemantauan akademik. Metode pengembangan yang digunakan adalah *Prototyping*, sedangkan model prediksi dibangun dari dataset *Predict Students Dropout and Academic Success* yang diseleksi menjadi 3.399 data dengan dua kelas, yaitu dropout dan graduate. Hasil penelitian menunjukkan bahwa prototype yang dikembangkan memuat fitur prediksi manual, impor file CSV, monitoring hasil prediksi, dan laporan berbasis peran pengguna. Model CatBoost memperoleh accuracy 90,29%, precision 85,33%, recall 88,76%, dan F1-score 87,01%. Temuan ini menunjukkan bahwa prototype tersebut berpotensi menjadi dasar pengembangan sistem deteksi dini mahasiswa berisiko dropout.

Kata kunci: *Pemantauan akademik; CatBoost; Prediksi dropout; Prototype sistem; Sistem informasi berbasis web*

1. Pendahuluan

Transformasi digital di perguruan tinggi telah mengubah fungsi sistem informasi akademik dari yang semula administratif menjadi sarana analitik untuk mendukung pengambilan keputusan berbasis data. Dalam konteks ini, *educational data mining* dan *learning analytics* semakin dimanfaatkan untuk memahami performa, keterlibatan, dan risiko akademik mahasiswa. Pemanfaatan analisis data membantu pengelola program studi memperoleh informasi yang lebih akurat untuk memprediksi dan meningkatkan hasil belajar mahasiswa [1]. Kajian ulasan juga menunjukkan bahwa prediksi *dropout* merupakan salah satu tema penting dalam *educational data mining* karena berkaitan dengan keberlanjutan studi, mutu layanan akademik, dan efektivitas kebijakan retensi mahasiswa [2]. Oleh karena itu, perguruan tinggi

memerlukan sistem yang mampu mengidentifikasi mahasiswa berisiko secara proaktif agar intervensi dapat dilakukan lebih tepat waktu [3].

Permasalahan *dropout* mahasiswa menjadi isu penting dalam pengelolaan pendidikan tinggi karena berdampak pada keberhasilan studi mahasiswa sekaligus mutu institusi. Dengan tersedianya data akademik dan nonakademik dalam sistem informasi, berbagai pendekatan prediktif berbasis pembelajaran mesin mulai dikembangkan untuk mendukung identifikasi dini mahasiswa berisiko. de Oliveira dkk. menunjukkan bahwa *learning analytics* berperan dalam mengidentifikasi pola risiko dan mendukung intervensi yang lebih tepat sasaran [4], sedangkan Vaarma dkk. menegaskan bahwa data akademik, demografis, dan aktivitas *learning management system* dapat dimanfaatkan secara efektif untuk membangun model prediksi *dropout* [5]. Villar dan Andrade juga menunjukkan bahwa algoritma berbasis *boosting*, termasuk CatBoost, memiliki performa yang kompetitif dalam prediksi *dropout* mahasiswa [6].

Penelitian lain memperlihatkan bahwa persoalan *dropout* bersifat multidimensi dan memerlukan dukungan informasi yang dapat dimanfaatkan dalam pengambilan keputusan akademik. Gutierrez-Pachas dkk. menunjukkan bahwa risiko *dropout* dipengaruhi oleh faktor akademik, sosial ekonomi, dan kesetaraan [7], sedangkan Bañeres dkk. menunjukkan bahwa nilai praktis prediksi meningkat ketika model analitik diintegrasikan ke dalam sistem informasi untuk mendukung pemantauan dan intervensi [8]. Selain itu, Mosia melaporkan akurasi 92,81% pada pendekatan probabilistik untuk identifikasi dini mahasiswa berisiko [9], Daza Santacoloma mengembangkan sistem peringatan dini dengan sensitivitas 61,97% [10], dan Mary menunjukkan bahwa model ansambel mampu mendukung intervensi tepat waktu [11].

Meskipun demikian, penelitian sebelumnya masih lebih banyak berfokus pada performa model prediksi. Berbagai algoritma seperti *Decision Tree*, *Random Forest*, *Support Vector Machine*, dan *XGBoost* telah digunakan untuk memprediksi tingkat *dropout* dengan akurasi yang beragam [12], [6]. Namun, integrasi model prediktif ke dalam sistem informasi berbasis *web* yang mendukung pemantauan akademik secara langsung masih relatif terbatas. Selain itu, meskipun CatBoost telah menunjukkan performa yang kompetitif, penerapannya dalam sistem informasi prediksi risiko *dropout* berbasis *web* belum banyak dibahas secara spesifik [6].

Berdasarkan kondisi tersebut, penelitian ini mengembangkan sistem informasi prediksi risiko *dropout* mahasiswa berbasis *web* dengan mengintegrasikan algoritma CatBoost sebagai mesin prediksi. Berbeda dengan penelitian sebelumnya yang umumnya berfokus pada model prediksi dan belum mengarah pada pengembangan sistem yang mendukung pemantauan akademik secara langsung [13], penelitian ini penting karena *prototype* sistem yang dirancang menunjukkan potensi untuk mendukung deteksi dini risiko *dropout*, sehingga dapat menjadi dasar pengembangan sistem yang membantu perguruan tinggi menekan angka putus studi, meningkatkan layanan akademik, dan memperkuat pengelolaan pendidikan berbasis data.

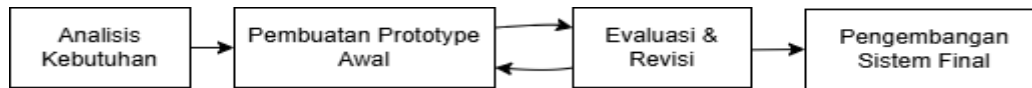
Oleh karena itu, *state of the art* penelitian ini terletak pada pengembangan sistem informasi prediksi risiko *dropout* mahasiswa berbasis *web* dengan integrasi algoritme CatBoost untuk mendukung pemantauan dan pengambilan keputusan akademik secara langsung. Kebaruan penelitian ini terletak pada penggabungan model prediksi berbasis CatBoost dengan sistem multi-pengguna berbasis *web* yang tidak hanya menghasilkan prediksi, tetapi juga menyediakan sarana pemantauan risiko bagi pihak akademik.

2. Metodologi

Penelitian ini berfokus pada pengembangan *software* sistem cerdas berupa sistem informasi prediksi risiko *dropout* mahasiswa berbasis *web* yang mengintegrasikan algoritme CatBoost sebagai model prediksi. Sistem ini dirancang untuk mendukung pemantauan dan pengambilan keputusan akademik melalui proses pengelolaan data dan prediksi risiko *dropout* mahasiswa.

2.1 Metode Penelitian dan Tahapan Pengembangan Sistem

Penelitian ini merupakan penelitian pengembangan yang bertujuan membangun sistem informasi prediksi risiko *dropout* mahasiswa berbasis *web*. Metode pengembangan sistem yang digunakan adalah *Prototyping*, karena proses pembangunan sistem dilakukan melalui pembuatan rancangan awal, evaluasi, dan penyempurnaan bertahap hingga diperoleh sistem yang sesuai dengan kebutuhan pengguna. Pendekatan ini digunakan karena selama proses pengembangan terdapat penyesuaian pada fitur, alur kerja, dan tampilan antarmuka.



Gambar 1. Tahapan Pengembangan Sistem dengan Metode Prototyping

Berdasarkan gambar 1, tahapan pengembangan sistem meliputi analisis kebutuhan, pembuatan *prototype* awal, evaluasi dan revisi, serta pengembangan sistem final. Pada tahap analisis kebutuhan dilakukan identifikasi masalah, pengumpulan data, dan penentuan kebutuhan sistem. Tahap pembuatan *prototype* awal dilakukan untuk menyusun rancangan awal sistem yang menggambarkan fungsi utama yang akan dikembangkan. *Prototype* yang dihasilkan kemudian dievaluasi dan direvisi sesuai kebutuhan pengguna. Proses evaluasi dan revisi dilakukan secara berulang hingga diperoleh *prototype* akhir yang mewakili kebutuhan sistem. Setelah itu, dilakukan tahap pengembangan sistem final dengan menyempurnakan rancangan *prototype* ke dalam bentuk sistem yang memuat fitur utama, seperti prediksi, monitoring, dan laporan.

2.2 Analisis Kebutuhan dan Perancangan Sistem

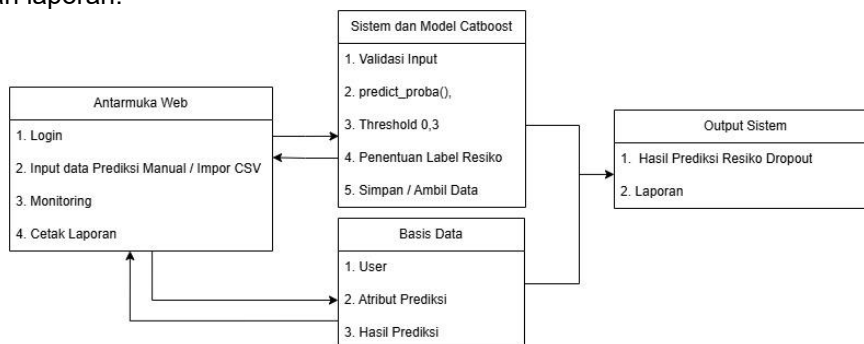
2.2.1 Kebutuhan Sistem

Analisis kebutuhan dilakukan untuk menentukan fungsi utama yang harus dimiliki sistem serta kebutuhan nonfungsional yang mendukung penggunaannya. Kebutuhan fungsional sistem meliputi autentikasi pengguna, input data mahasiswa, input atribut prediksi secara manual maupun melalui impor file *CSV*, proses prediksi risiko *dropout*, tampilan hasil prediksi, monitoring data hasil prediksi, pengeditan data hasil prediksi, serta penyajian laporan dalam bentuk tampilan monitoring, cetak, dan unduhan data. Sistem juga mendukung pembagian hak akses berdasarkan peran pengguna, yaitu admin dan kaprodi.

Kebutuhan nonfungsional mencakup kemudahan penggunaan antarmuka, keamanan data dan akses pengguna, keandalan sistem berbasis *web*, serta kemampuan sistem dalam memproses data prediksi secara efisien. Sistem dirancang agar dapat dijalankan secara lokal melalui peramban *web* dan memiliki tampilan yang mendukung proses kerja pengguna secara jelas.

2.2.2 Desain Arsitektur Sistem

Desain arsitektur sistem menggambarkan hubungan antara antarmuka web, sistem dan model *CatBoost*, basis data, serta output sistem. Rancangan ini digunakan untuk menunjukkan bagaimana data dari pengguna diproses oleh sistem hingga menghasilkan hasil prediksi risiko *dropout* dan laporan.



Gambar 2. Desain Arsitektur Sistem

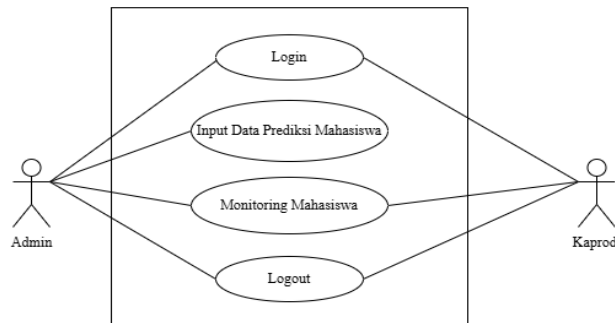
Berdasarkan Gambar 2, antarmuka web menjadi bagian awal yang digunakan pengguna untuk mengakses sistem. Pada bagian ini tersedia fitur login, input data prediksi secara manual atau melalui impor *CSV*, monitoring, dan cetak laporan. Data prediksi yang dimasukkan melalui antarmuka web kemudian diteruskan ke bagian sistem dan model *CatBoost*. Pada bagian tersebut, sistem melakukan validasi input, menjalankan fungsi *predict_proba()* untuk memperoleh probabilitas *dropout*, menerapkan nilai *threshold* 0,3, serta menentukan label risiko mahasiswa.

Hasil proses prediksi selanjutnya disimpan dan dapat diambil kembali melalui basis data. Basis data pada sistem terdiri atas tabel *users*, *atribut_prediksi*, dan *hasil_prediksi*. Tabel

users digunakan untuk menyimpan data pengguna, tabel atribut_prediksi menyimpan atribut masukan model, sedangkan tabel hasil_prediksi menyimpan hasil prediksi mahasiswa. Data yang tersimpan kemudian digunakan untuk menampilkan output sistem berupa hasil prediksi risiko *dropout* dan laporan. Dengan rancangan ini, proses input data, prediksi, penyimpanan hasil, monitoring, dan pelaporan dapat berjalan secara terintegrasi.

2.2.2.1 Use Case Diagram

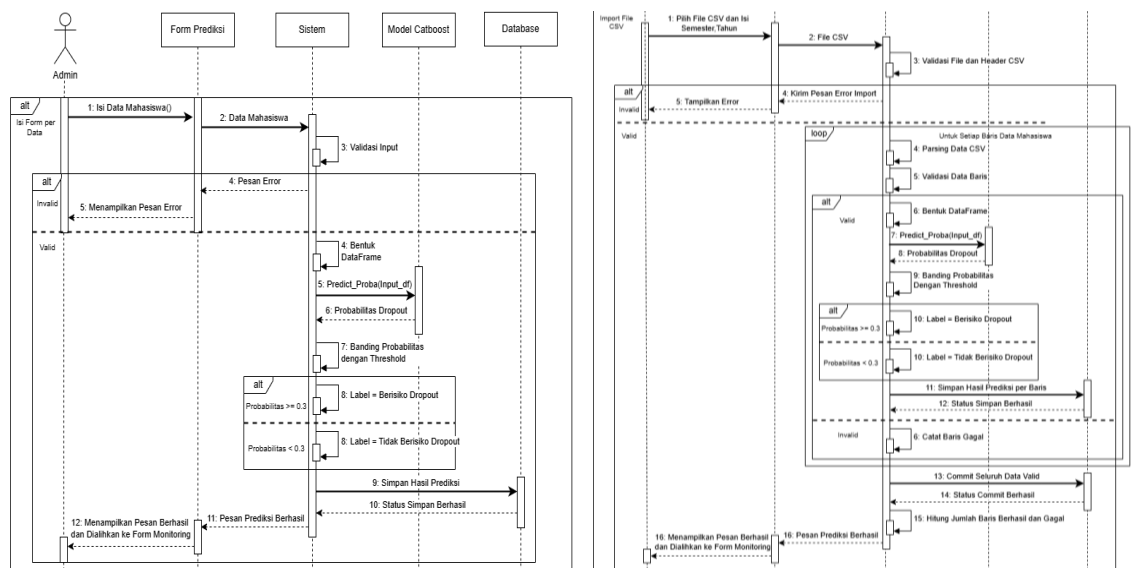
Gambar 3 menunjukkan *use case diagram* sistem yang menggambarkan hubungan antara aktor dengan fungsi utama yang tersedia. Pada diagram tersebut terdapat dua aktor, yaitu admin dan kaprodi. Admin memiliki hak akses untuk melakukan *login*, input data prediksi mahasiswa, monitoring mahasiswa, dan *logout*. Sementara itu, kaprodi memiliki hak akses untuk melakukan *login*, monitoring mahasiswa, dan *logout*.



Gambar 3. Use Case Diagram Sistem

Keberadaan *use case diagram* membantu memperjelas cakupan fungsi sistem dari sudut pandang pengguna. Melalui diagram ini dapat diketahui bahwa admin berperan lebih aktif dalam proses pengelolaan data prediksi, sedangkan kaprodi lebih berfokus pada pemantauan hasil prediksi yang telah dihasilkan sistem. Perancangan ini mendukung penerapan hak akses yang berbeda agar setiap pengguna hanya dapat menjalankan fungsi yang sesuai dengan kewenangannya.

2.2.2.2 Sequence Diagram

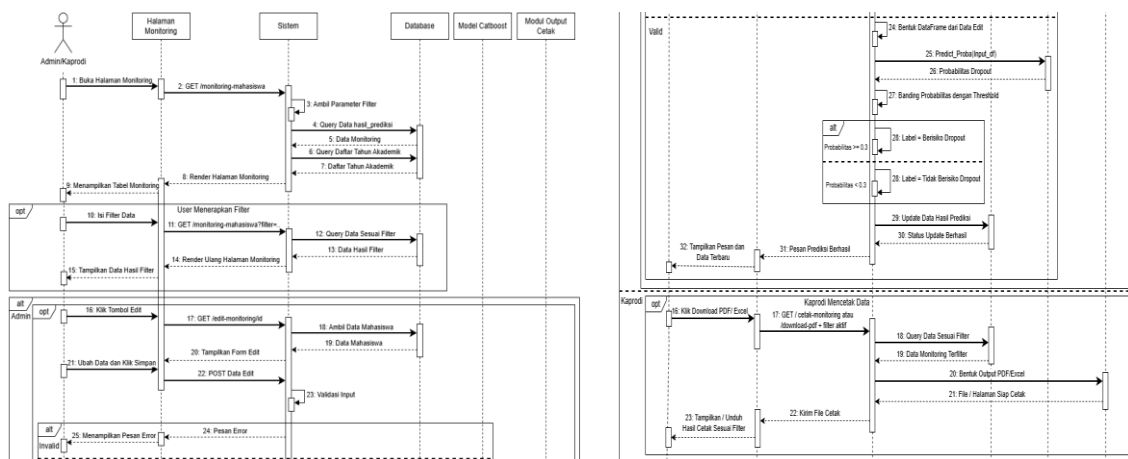


Gambar 4. Sequence Diagram Prediksi Mahasiswa

Sequence diagram digunakan untuk menggambarkan urutan interaksi antar komponen sistem pada proses-proses inti. Pada penelitian ini, *sequence diagram* difokuskan pada dua proses utama, yaitu proses prediksi mahasiswa dan proses monitoring mahasiswa.

Pada proses prediksi mahasiswa, alur dimulai ketika admin memasukkan data mahasiswa, baik secara manual maupun melalui fitur impor file CSV dapat dilihat pada gambar 4. Data yang masuk selanjutnya divalidasi oleh sistem. Apabila data valid, sistem membentuk data ke dalam format yang sesuai untuk diproses model, lalu meneruskannya ke model CatBoost untuk menghitung probabilitas risiko *dropout*. Hasil probabilitas tersebut kemudian dibandingkan dengan nilai *threshold* yang digunakan sistem untuk menentukan label prediksi, yaitu berisiko *dropout* atau tidak berisiko *dropout*. Setelah itu, hasil prediksi disimpan ke basis data dan ditampilkan kembali kepada pengguna melalui sistem.

Pada proses monitoring mahasiswa, sistem menampilkan data hasil prediksi yang telah tersimpan di basis data dan mendukung penggunaan filter agar data dapat ditampilkan sesuai kebutuhan pengguna, dapat dilihat pada gambar 5 diagram monitoring mahasiswa. Dalam proses ini, admin dan kaprodi memiliki alur lanjutan yang berbeda. Admin dapat melakukan monitoring sekaligus mengedit data mahasiswa, kemudian sistem akan memvalidasi data yang diperbarui, memproses ulang prediksi, dan memperbarui hasilnya pada basis data. Sementara itu, kaprodi menggunakan halaman monitoring untuk melihat data hasil prediksi dan secara opsional dapat mencetak atau mengunduh data sesuai filter yang dipilih. Perbedaan alur ini menunjukkan bahwa proses monitoring dirancang fleksibel sesuai hak akses dan kebutuhan masing-masing aktor.



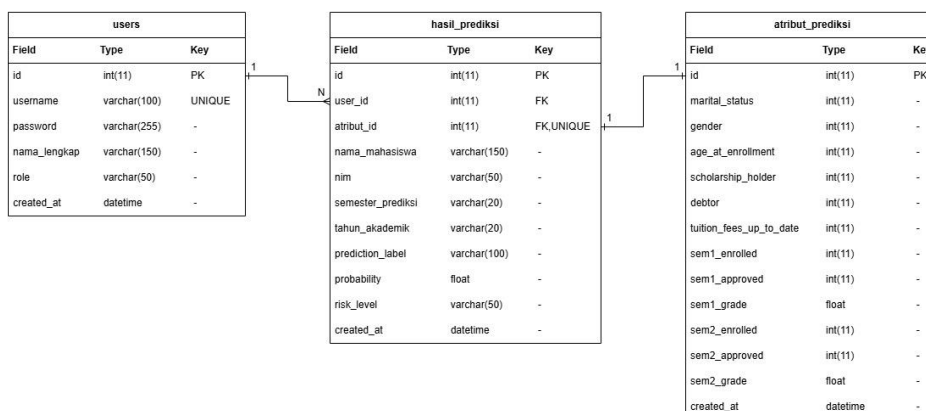
Gambar 5. Sequence Diagram Monitoring Mahasiswa

Dengan adanya *sequence diagram*, mekanisme kerja sistem dapat divisualisasikan secara lebih rinci dan sistematis, sehingga mempermudah pemahaman terhadap urutan proses, perpindahan data, serta interaksi antarkomponen dalam sistem.

2.2.2.3 Desain Basis Data

Desain basis data pada *prototype* sistem ditunjukkan pada Gambar 6. Basis data dirancang untuk mendukung penyimpanan data pengguna, atribut masukan prediksi, dan hasil prediksi risiko *dropout* mahasiswa. Dalam rancangan ini terdapat tiga tabel utama, yaitu *users*, *atribut_prediksi*, dan *hasil_prediksi*.

Tabel *users* digunakan untuk menyimpan data pengguna dan hak akses sistem. Tabel *atribut_prediksi* digunakan untuk menyimpan dua belas atribut masukan model CatBoost, sedangkan tabel *hasil_prediksi* digunakan untuk menyimpan identitas mahasiswa, periode prediksi, status prediksi, probabilitas *dropout*, tingkat risiko, dan waktu prediksi. Data monitoring dan laporan tidak dibuat sebagai tabel terpisah karena keduanya dihasilkan dari data yang telah tersimpan pada tabel *hasil_prediksi*. Dengan demikian, rancangan basis data ini mendukung proses prediksi, penyimpanan hasil, monitoring, cetak, dan unduhan data secara terintegrasi.



Gambar 6. Desain Basis Data

Tabel *users* digunakan untuk menyimpan data pengguna dan hak akses sistem. Tabel *atribut_prediksi* digunakan untuk menyimpan dua belas atribut masukan model CatBoost, sedangkan tabel *hasil_prediksi* digunakan untuk menyimpan identitas mahasiswa, periode prediksi, status prediksi, probabilitas *dropout*, tingkat risiko, dan waktu prediksi. Data monitoring dan laporan tidak dibuat sebagai tabel terpisah karena keduanya dihasilkan dari data yang telah tersimpan pada tabel *hasil_prediksi*. Dengan demikian, rancangan basis data ini mendukung proses prediksi, penyimpanan hasil, monitoring, cetak, dan unduhan data secara terintegrasi.

2.3 Pengembangan Prototype Akhir dan Evaluasi Model

Dalam penelitian ini, pengembangan sistem dilakukan hingga menghasilkan *prototype* akhir yang dapat dicobakan kepada pengguna. *Prototype* tersebut memuat fitur utama yang dibutuhkan, seperti input data mahasiswa, proses prediksi, monitoring hasil, dan penyajian laporan. Pengembangan dilakukan secara bertahap melalui evaluasi dan perbaikan hingga diperoleh rancangan sistem yang sesuai dengan kebutuhan pengguna. Penelitian ini tidak sampai pada tahap implementasi sistem secara operasional maupun pengujian fungsional sistem secara menyeluruh. Oleh karena itu, evaluasi dalam penelitian ini difokuskan pada model prediksi CatBoost sebagai komponen utama dalam sistem yang dirancang. Pengujian dilakukan untuk menilai kemampuan model dalam mengklasifikasikan mahasiswa ke dalam kategori risiko *dropout* berdasarkan metrik *accuracy*, *precision*, *recall*, dan *F1-score* yang dapat dilihat pada bagian hasil pengujian model.

2.3.1 Data Penelitian

Dataset yang digunakan dalam penelitian ini diperoleh dari Kaggle dengan judul *Predict Students Dropout and Academic Success*. *Dataset* tersebut memuat data akademik, karakteristik, dan kondisi ekonomi mahasiswa yang relevan terhadap keberlanjutan studi. Secara keseluruhan, *dataset* terdiri atas 35 atribut sebagai variabel fitur dan 1 variabel target yang mengklasifikasikan status mahasiswa menjadi *dropout*, *enrolled*, dan *graduate*.

Berdasarkan hasil seleksi fitur, penelitian ini menggunakan 12 atribut yang dipilih berdasarkan pada landasan teoritis dari penelitian-penelitian terdahulu yang menunjukkan bahwa faktor akademik, karakteristik mahasiswa, dan kondisi administratif atau finansial merupakan prediktor penting terhadap risiko *dropout*, dinilai paling relevan terhadap prediksi risiko *dropout*, yaitu *Curricular units 2nd sem (approved)*, *Curricular units 2nd sem (grade)*, *Curricular units 1st sem (approved)*, *Age at enrollment*, *Curricular units 1st sem (enrolled)*, *Curricular units 2nd sem (enrolled)*, *Tuition fees up to date*, *Curricular units 1st sem (grade)*, *Scholarship holder*, *Debtor*, *Gender*, dan *Marital status* [5],[14].

Studi sebelumnya menegaskan bahwa jumlah mata kuliah yang diselesaikan, nilai semester awal, usia saat pendaftaran, serta kondisi pembayaran biaya kuliah merupakan prediktor utama dalam model prediksi *dropout* pada pendidikan tinggi [15]. Seleksi atribut dilakukan untuk juga membantu model dalam menghindari penggunaan atribut yang tidak relevan atau redundan, sehingga hasil prediksi yang dihasilkan menjadi lebih optimal [16], [17].

Berdasarkan distribusi variabel target, *dataset* yang digunakan terdiri atas 2.153 data *Graduate* (63,34%) dan 1.246 data *Dropout* (36,66%), sehingga total data yang digunakan dalam pemodelan berjumlah 3.399 setelah kelas *Enrolled* tidak disertakan. Komposisi ini

menunjukkan adanya ketidakseimbangan kelas pada tingkat sedang (*moderately imbalanced dataset*), namun masih berada dalam rentang yang dapat ditangani dengan baik oleh algoritma CatBoost.

2.3.2 Evaluasi Model

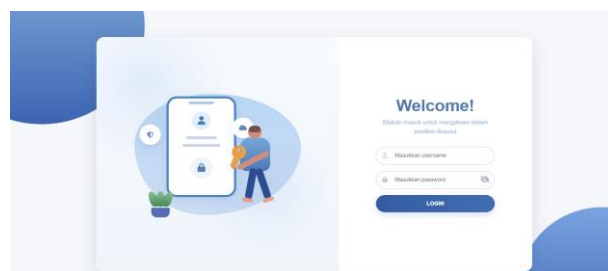
Evaluasi model dilakukan untuk menilai kemampuan algoritme CatBoost dalam mengklasifikasikan mahasiswa ke dalam kategori risiko *dropout*. Pada penelitian ini, data dibagi menjadi data latih dan data uji dengan metode *train-test split* menggunakan komposisi 80% data latih dan 20% data uji. Pembagian ini dilakukan agar model dapat mempelajari pola data dari data latih, kemudian diuji menggunakan data yang belum pernah dilihat sebelumnya.

Performa model dinilai menggunakan *confusion matrix* serta metrik *accuracy*, *precision*, *recall*, dan *F1-score*. *Accuracy* digunakan untuk melihat tingkat ketepatan model secara keseluruhan, *precision* digunakan untuk mengukur ketepatan model dalam memprediksi mahasiswa yang berisiko *dropout*, *recall* digunakan untuk menilai kemampuan model dalam mendeteksi mahasiswa yang benar-benar berisiko *dropout*, sedangkan *F1-score* digunakan untuk melihat keseimbangan antara *precision* dan *recall*. Penggunaan metrik tersebut bertujuan agar evaluasi model tidak hanya berfokus pada ketepatan umum, tetapi juga pada kemampuan model dalam mendukung deteksi dini mahasiswa berisiko.

3. Hasil dan Pembahasan

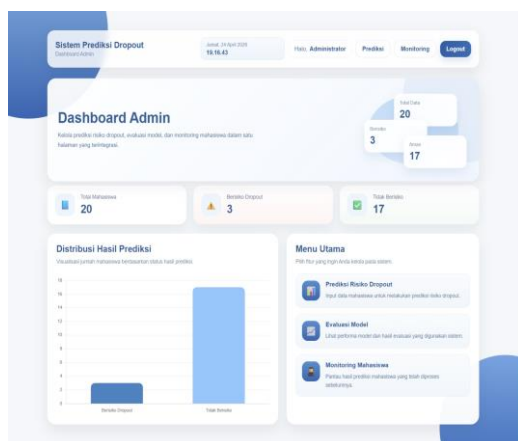
3.1 Antarmuka Prototype Sistem

Hasil pengembangan sistem pada penelitian ini ditampilkan dalam bentuk antarmuka *prototype* berbasis *web* yang dirancang untuk mendukung proses prediksi risiko *dropout* mahasiswa dan pemantauan hasil prediksi.

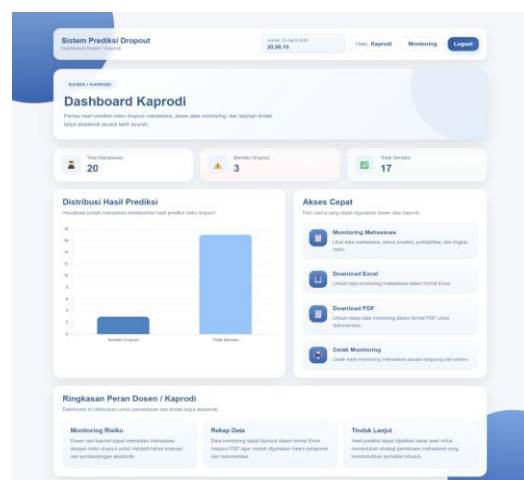


Gambar 7. Halaman Login

Gambar 7 menunjukkan halaman *login* yang digunakan sebagai proses autentikasi pengguna sebelum mengakses sistem. Pada halaman ini, pengguna memasukkan *username* dan *password* untuk proses autentikasi. Setelah data *login* dinyatakan *valid*, sistem akan mengarahkan pengguna ke halaman utama sesuai dengan peran yang dimiliki, yaitu admin atau kaprodi.

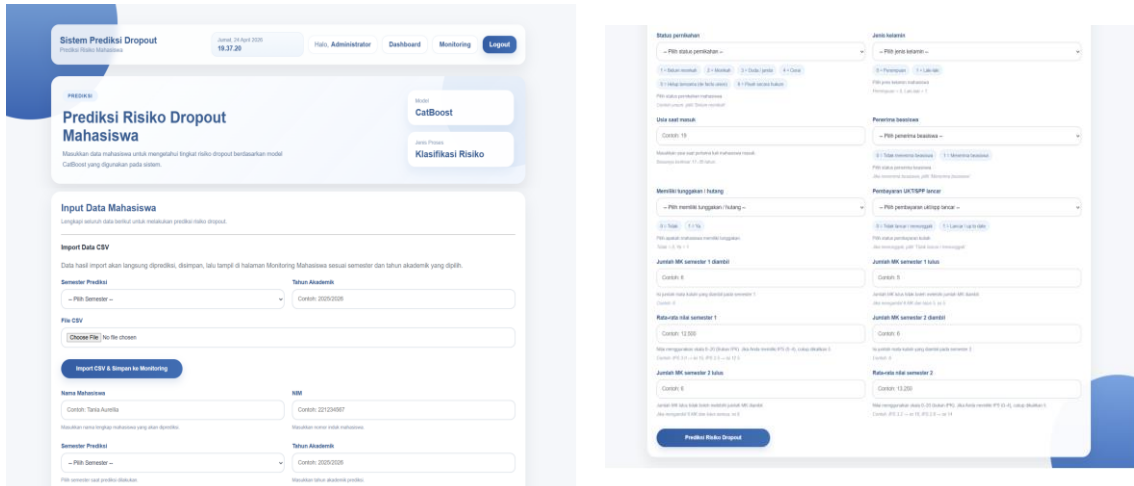


Gambar 8. Dashboard Admin



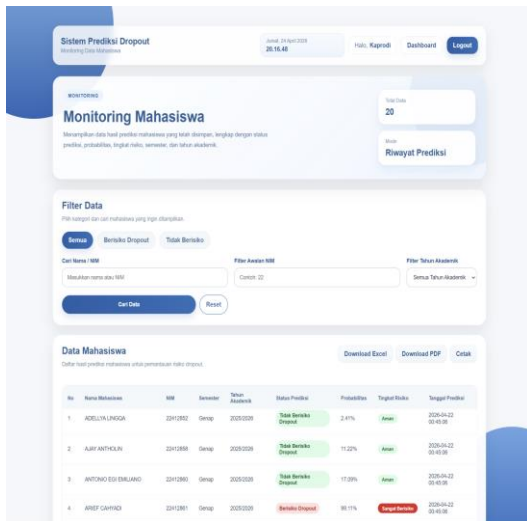
Gambar 9. Dashboard Kaprodi

Tampilan *dashboard* setelah proses *login* dapat dilihat pada gambar 8 dan 9. Keduanya sama-sama menyajikan ringkasan hasil prediksi, tetapi dengan cakupan akses yang berbeda. *Dashboard* admin pada gambar 8 digunakan untuk mendukung pengelolaan sistem secara lebih lengkap, seperti mengakses fitur prediksi mahasiswa, monitoring, dan evaluasi model. Sementara itu, *dashboard* kaprodi pada gambar 9 lebih difokuskan pada kebutuhan pemantauan dan evaluasi akademik, sehingga akses yang tersedia dibatasi pada monitoring dan laporan. Perbedaan ini menunjukkan bahwa antarmuka dirancang sesuai hak akses masing-masing pengguna.

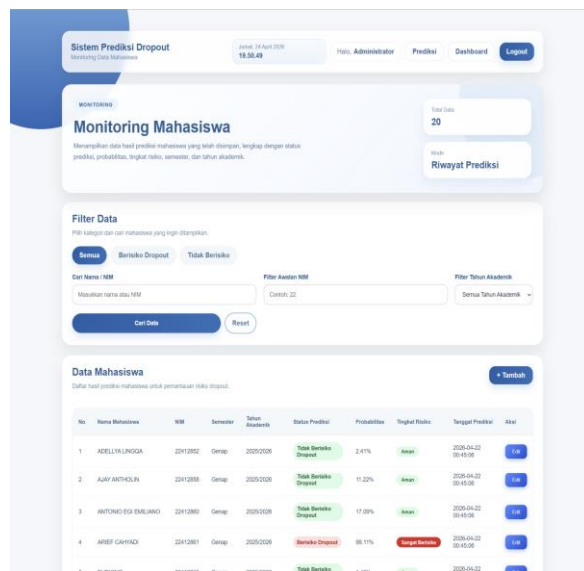


Gambar 10. Halaman Prediksi Mahasiswa dan Import CSV

Proses prediksi mahasiswa ditampilkan pada gambar 10. Halaman ini digunakan oleh admin untuk melakukan prediksi risiko *dropout* bisa melalui dua cara, yaitu *input* manual atau *impor* file CSV. Data yang dimasukkan akan divalidasi terlebih dahulu, kemudian diproses menggunakan model CatBoost untuk menghasilkan status prediksi, probabilitas *dropout*, dan tingkat risiko mahasiswa.



Gambar 11. Halaman Monitoring Mahasiswa Kaprodi



Gambar 12. Halaman Monitoring Mahasiswa Admin

Fitur monitoring hasil prediksi dapat dilihat pada gambar 11 dan gambar 12. Kedua halaman ini sama-sama menampilkan data hasil prediksi yang telah tersimpan di basis data dan dilengkapi dengan fitur pencarian serta filter berdasarkan status prediksi, semester, dan tahun akademik. Perbedaannya terletak pada fungsi lanjutan yang tersedia. Pada sisi kaprodi, seperti

terlihat pada gambar 11, halaman monitoring digunakan untuk melihat mahasiswa yang berisiko *dropout* sebagai dasar evaluasi akademik. Adapun pada sisi admin, seperti pada gambar 12, halaman monitoring juga mendukung proses pengeditan data mahasiswa sehingga data prediksi dapat diperbarui bila diperlukan.

No	Nama Mahasiswa	NIM	Status Prediksi	Probabilitas	Tingkat Risiko	Tanggal Prediksi
1	ADELLYA LUNIGA	22412852	Tidak Berisiko Dropout	2.41%	Aman	2026-04-22 00:45:00
2	AJAY ANTHOLIN	22412858	Tidak Berisiko Dropout	11.22%	Aman	2026-04-22 00:45:00
3	ANTONIO EGI EMILIANO	22412860	Tidak Berisiko Dropout	17.09%	Aman	2026-04-22 00:45:00
4	ARIEF CAHYADI	22412861	Berisiko Dropout	99.11%	Sangat Berisiko	2026-04-22 00:45:00
5	BUDIONO	22412865	Tidak Berisiko Dropout	1.49%	Aman	2026-04-22 00:45:00
6	DANIEL	22412866	Tidak Berisiko Dropout	2.74%	Aman	2026-04-22 00:45:00
7	ERRA JULYANTI	22412871	Tidak Berisiko Dropout	2.47%	Aman	2026-04-22 00:45:00
8	EVA GULTOM	22412872	Tidak Berisiko Dropout	12.64%	Aman	2026-04-22 00:45:00
9	FERNANDO LUIS VIDO	22412873	Berisiko Dropout	99.93%	Sangat Berisiko	2026-04-22 00:45:00
10	FLORENTIUS PEBRO	22412874	Tidak Berisiko Dropout	3.56%	Aman	2026-04-22 00:45:00
11	FRISLIA AGNES MARISA SINAGA	22412875	Tidak Berisiko Dropout	3.48%	Aman	2026-04-22 00:45:00
12	GEBBY PATRICIA PRADAMA	22412876	Tidak Berisiko Dropout	13.54%	Aman	2026-04-22 00:45:00
13	GERRY STEVEN MARTIN	22412878	Tidak Berisiko Dropout	9.37%	Aman	2026-04-22 00:45:00
14	GREGORIUS ADRIAN PADUA	22412881	Tidak Berisiko Dropout	1.62%	Aman	2026-04-22 00:45:00
15	JESSEN HERO PRATAMA	22412883	Berisiko Dropout	99.00%	Sangat Berisiko	2026-04-22 00:45:00
16	JONG WILLY PRAYOGA	22412884	Tidak Berisiko Dropout	3.0%	Aman	2026-04-22 00:45:00

Gambar 13. Halaman Laporan / Cetak Monitoring

Penyajian hasil monitoring dalam bentuk laporan ditampilkan pada gambar 13. Halaman ini digunakan untuk menampilkan hasil prediksi yang telah difilter sesuai kebutuhan pengguna dan disiapkan dalam format yang lebih ringkas untuk dicetak atau disimpan. Keberadaan fitur ini memudahkan pihak akademik dalam mendokumentasikan hasil monitoring dan memanfaatkannya sebagai bahan pendukung pengambilan keputusan.

3.2 Mekanisme Prediksi pada Prototype Sistem

Penelitian ini menghasilkan sistem dalam bentuk *prototype* akhir yang dapat dicobakan kepada pengguna dan belum sampai pada tahap implementasi operasional maupun pengujian fungsional sistem secara menyeluruh. Oleh karena itu, bagian ini menjelaskan mekanisme prediksi yang dirancang pada *prototype* sebagai alur kerja model CatBoost dalam menghasilkan klasifikasi risiko *dropout*.

Mekanisme prediksi pada *prototype* sistem dirancang menggunakan model CatBoost yang telah dilatih dengan data mahasiswa berstatus akhir *Dropout* dan *Graduate*. Data dengan status *Enrolled* tidak disertakan dalam proses pelatihan maupun pengujian karena belum merepresentasikan hasil akhir studi mahasiswa. Dengan demikian, model difokuskan untuk membedakan mahasiswa yang berisiko *dropout* dan mahasiswa yang berhasil menyelesaikan studi.

Pada *prototype* yang dikembangkan, data mahasiswa yang dimasukkan melalui form prediksi maupun impor file CSV terlebih dahulu divalidasi oleh sistem. Setelah data dinyatakan valid, data tersebut dibentuk ke dalam format *DataFrame* agar sesuai dengan struktur input yang dibutuhkan oleh model CatBoost. Selanjutnya, sistem menggunakan fungsi *predict_proba()* untuk memperoleh nilai probabilitas mahasiswa terhadap kelas *dropout*. Nilai probabilitas ini kemudian dibandingkan dengan nilai *threshold* sebesar 0,3. Apabila probabilitas *dropout* lebih besar atau sama dengan 0,3, maka mahasiswa diklasifikasikan sebagai "Berisiko *Dropout*". Sebaliknya, apabila probabilitas *dropout* berada di bawah 0,3, maka mahasiswa diklasifikasikan sebagai "Tidak Berisiko *Dropout*".

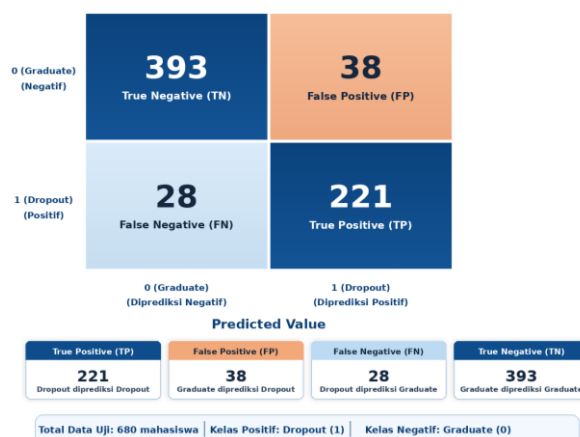
Pemilihan *threshold* 0,3 didasarkan pada hasil pengujian beberapa nilai *threshold*. Pada nilai tersebut dapat dilihat hasilnya pada Tabel 1 Hasil Evaluasi Model CatBoost. Nilai *recall* yang lebih tinggi menunjukkan bahwa model lebih sensitif dalam mendeteksi mahasiswa yang benar-benar berisiko *dropout*. Hal ini sesuai dengan tujuan sistem sebagai alat bantu deteksi dini, karena mahasiswa yang berpotensi *dropout* sebaiknya dapat teridentifikasi lebih awal agar pihak akademik dapat melakukan pemantauan atau intervensi. Dengan pendekatan ini, *prototype* sistem tidak hanya menghasilkan label klasifikasi, tetapi juga menyajikan probabilitas dan tingkat risiko sebagai informasi pendukung dalam proses monitoring akademik.

3.3 Hasil Pengujian Model

Bagian ini menyajikan hasil pengujian model CatBoost yang digunakan dalam sistem prediksi risiko *dropout* mahasiswa. Pengujian dilakukan untuk menilai performa model berdasarkan confusion matrix dan metrik evaluasi klasifikasi.

3.3.1 Confusion Matrix dan Evaluasi Metrik

Pengujian model dilakukan untuk mengevaluasi kemampuan algoritma CatBoost dalam memprediksi risiko *dropout* mahasiswa berdasarkan data uji yang dipisahkan dari data latih. Pada penelitian ini digunakan 3.399 data mahasiswa yang terdiri dari kategori *Dropout* dan *Graduate*, sedangkan data berstatus *Enrolled* tidak disertakan. *Dataset* dibagi menggunakan metode *train-test split* dengan komposisi data latih (2.719 data) dan 20% data uji (680 data). Pembagian ini dipilih karena 80% data dinilai cukup untuk melatih model agar mengenali pola data secara optimal, sementara 20% data digunakan untuk menguji kemampuan model dalam memprediksi data yang belum pernah dilihat sebelumnya. Evaluasi model dilakukan menggunakan confusion matrix serta metrik *accuracy*, *precision*, *recall*, dan *F1-score*.



Gambar 14. Confusion Matrix Model CatBoost

Hasil pengujian model melalui *confusion matrix* ditunjukkan pada gambar 14. Dari gambar tersebut terlihat bahwa model mampu mengklasifikasikan 393 mahasiswa non-*dropout* dengan benar sebagai *True Negative* (TN) dan 221 mahasiswa *dropout* dengan benar sebagai *True Positive* (TP). Sementara itu, terdapat 38 data yang termasuk *False Positive* (FP), yaitu mahasiswa yang sebenarnya tidak *dropout* tetapi diprediksi *dropout*, serta 28 data yang termasuk *False Negative* (FN), yaitu mahasiswa yang sebenarnya *dropout* namun diprediksi tidak berisiko.

Selain menunjukkan performa model secara umum, hasil *confusion matrix* juga memperlihatkan adanya data yang belum berhasil dikenali secara tepat oleh model. Kesalahan klasifikasi pada kategori *False Positive* terjadi ketika mahasiswa yang sebenarnya berstatus *Graduate* diprediksi sebagai *Dropout*. Kondisi ini dapat terjadi karena sebagian mahasiswa *Graduate* memiliki pola atribut yang menyerupai mahasiswa *Dropout*, terutama pada atribut akademik semester awal. Sebaliknya, kesalahan pada kategori *False Negative* terjadi ketika mahasiswa yang sebenarnya berstatus *Dropout* diprediksi sebagai *Graduate* atau tidak berisiko *dropout*. Kesalahan ini dapat muncul karena sebagian mahasiswa *Dropout* memiliki karakteristik akademik dan administratif yang mendekati mahasiswa *Graduate*, sehingga probabilitas *dropout* yang dihasilkan model berada di bawah nilai *threshold* 0,3.

Berdasarkan hasil tersebut, evaluasi model dilanjutkan menggunakan metrik *accuracy*, *precision*, *recall*, dan *F1-score*. Nilai *accuracy* digunakan untuk melihat tingkat ketepatan model secara keseluruhan, *precision* digunakan untuk menilai ketepatan prediksi terhadap kelas *dropout*, *recall* digunakan untuk melihat kemampuan model dalam mendeteksi mahasiswa yang benar-benar berisiko, sedangkan *F1-score* digunakan untuk melihat keseimbangan antara *precision* dan *recall*.

Rumus *accuracy* digunakan untuk mengukur tingkat ketepatan prediksi model secara keseluruhan.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \quad (1)$$

Rumus *precision* digunakan untuk mengukur tingkat ketepatan model dalam memprediksi mahasiswa yang berisiko *dropout*.

$$Precision = \frac{TP}{TP + FP} \times 100\% \tag{2}$$

Rumus *recall* digunakan untuk mengukur kemampuan model dalam mendeteksi mahasiswa yang benar-benar berisiko *dropout*.

$$Recall = \frac{TP}{TP + FN} \times 100\% \tag{3}$$

Sementara itu, *F1-score* digunakan untuk melihat keseimbangan antara *precision* dan *recall*.

$$F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \tag{4}$$

Nilai-nilai pada tabel 1 menunjukkan bahwa model memiliki performa yang baik dalam memprediksi risiko *dropout* mahasiswa. Ketepatan model secara umum tergolong tinggi, dan kemampuan model dalam mendeteksi mahasiswa yang benar-benar berisiko juga cukup kuat. Hasil evaluasi ini disajikan kembali secara ringkas pada Tabel 1.

Tabel 1. Hasil Evaluasi Model CatBoost

Metrik	Nilai
Accuracy	90,29%
Precision	85,33%
Recall	88,76%
F1-Score	87,01%

Secara umum, hasil pada Tabel 1 memperlihatkan bahwa model CatBoost sudah cukup andal untuk digunakan dalam sistem pendukung pemantauan risiko *dropout*. Nilai *recall* yang tinggi menjadi poin penting karena dalam konteks deteksi dini, kemampuan sistem mengenali mahasiswa yang benar-benar berisiko lebih dibutuhkan agar intervensi akademik dapat dilakukan lebih cepat.

No	Nama Mahasiswa	NIM	Semester	Tahun Akademik	Status Prediksi	Probabilitas	Tingkat Risiko	Tanggal Prediksi
1	RACHEL RATTU VEBYOLA	22412902	Ganjil	2026/2027	Tidak Berisiko Dropout	15,01%	Aman	2026-06-21 15:15:21
2	RONSEN NIAGA	22412904	Ganjil	2026/2027	Berisiko Dropout	35,01%	Peringatan	2026-06-21 15:15:21
3	TANIA AURELLIA	22412910	Ganjil	2026/2027	Berisiko Dropout	45,12%	Tidak Aman	2026-06-21 15:15:21
4	THESSA THERESIA MEDIANA	22412911	Ganjil	2026/2027	Berisiko Dropout	55,07%	Sangat Tidak Aman	2026-06-21 15:15:21
5	TRISSIANTO	22412912	Ganjil	2026/2027	Berisiko Dropout	64,99%	Risiko Tinggi	2026-06-21 15:15:21

Gambar 15. Contoh Hasil Prediksi Mahasiswa pada Prototipe Sistem

Berdasarkan Gambar 15, sistem menghasilkan keluaran berupa status prediksi, probabilitas *dropout*, dan tingkat risiko mahasiswa. Status prediksi ditentukan berdasarkan *threshold* 0,3, yaitu mahasiswa dengan probabilitas *dropout* $\geq 0,3$ diklasifikasikan sebagai berisiko *dropout*, sedangkan probabilitas di bawah 0,3 diklasifikasikan tidak berisiko *dropout*. Tingkat risiko ditampilkan untuk membantu pengguna membaca hasil prediksi secara lebih mudah.

Jenis Kesalahan	Status Aktual	Prediksi Model	Probabilitas Dropout	Karakteristik Data	Interpretasi
False Positive	Graduate	Dropout	30,48%	Semester 1: 5 dari 6 mata kuliah disetujui, nilai 11,50. Semester 2: 4 dari 6 mata kuliah disetujui, nilai 14,60. Tidak memiliki tunggakan biaya dan bukan debitur.	Model memprediksi data sebagai <i>Dropout</i> karena probabilitas <i>dropout</i> berada sedikit di atas <i>threshold</i> 0,3. Namun, status aktual data uji menunjukkan bahwa mahasiswa tersebut termasuk kelas <i>Graduate</i> .
False Negative	Dropout	Graduate	29,69%	Semester 1: 3 dari 5 mata kuliah disetujui, nilai 11,00. Semester 2: 4 dari 5 mata kuliah disetujui, nilai 12,75. Tidak memiliki tunggakan biaya dan bukan debitur.	Model memprediksi data sebagai <i>Graduate</i> karena probabilitas <i>dropout</i> berada sedikit di bawah <i>threshold</i> 0,3. Namun, status aktual data uji menunjukkan bahwa mahasiswa tersebut termasuk kelas <i>Dropout</i> .

Gambar 16. Contoh Kesalahan Klasifikasi pada Data Uji Model

Berdasarkan Gambar 16, contoh kesalahan klasifikasi diperoleh dari data uji model, bukan dari data prediksi baru pada prototype sistem. Kesalahan diketahui dengan membandingkan status aktual pada data uji dengan hasil prediksi model. Pada kasus *False Positive*, data dengan status aktual *Graduate* diprediksi sebagai *Dropout* karena probabilitasnya berada sedikit di atas *threshold* 0,3. Sebaliknya, pada kasus *False Negative*, data dengan status aktual *Dropout* diprediksi sebagai *Graduate* karena probabilitasnya berada sedikit di bawah *threshold* 0,3. Hal ini menunjukkan bahwa data yang berada di sekitar batas keputusan model perlu ditafsirkan secara hati-hati.

3.4 Pembahasan

Berdasarkan hasil pengembangan *prototype* dan pengujian model, penelitian ini menunjukkan bahwa sistem yang dirancang telah memuat fungsi utama untuk mendukung deteksi dini mahasiswa berisiko *dropout*. Permasalahan yang diangkat dalam penelitian ini tidak hanya berkaitan dengan performa model prediksi, tetapi juga bagaimana hasil prediksi dapat disajikan melalui sistem informasi agar dapat digunakan dalam proses pemantauan akademik. Oleh karena itu, fitur prediksi, monitoring, dan laporan pada *prototype* menjadi bagian penting karena menghubungkan proses pengolahan data dengan kebutuhan pengguna dalam membaca dan menindaklanjuti hasil prediksi.

Dari sisi fungsional, fitur input manual dan impor CSV mendukung proses pengolahan data mahasiswa secara lebih fleksibel. Admin dapat memasukkan data secara individual maupun dalam jumlah banyak, sehingga proses prediksi tidak terbatas pada pengisian data satu per satu. Fitur monitoring yang dilengkapi filter semester dan tahun akademik membantu kaprodi atau dosen menelusuri hasil prediksi sesuai periode tertentu, sedangkan fitur laporan mendukung dokumentasi hasil monitoring sebagai bahan evaluasi akademik. Dengan demikian, rancangan sistem tidak hanya berfungsi sebagai alat prediksi, tetapi juga sebagai sarana penyajian informasi risiko *dropout* secara lebih terstruktur.

Dari sisi model, CatBoost memperoleh *accuracy* sebesar 90,29%, *precision* 85,33%, *recall* 88,76%, dan *F1-score* 87,01%. Nilai tersebut menunjukkan bahwa model memiliki performa yang baik dalam mengklasifikasikan mahasiswa berisiko *dropout*. Dalam konteks deteksi dini, nilai *recall* menjadi penting karena sistem diharapkan mampu mengenali sebanyak mungkin mahasiswa yang benar-benar berisiko agar dapat dipantau lebih awal. Hasil ini sejalan dengan penelitian Cannistrà dkk. [18] yang menekankan pentingnya prediksi dini *dropout* agar institusi dapat melakukan tindakan pencegahan sebelum mahasiswa benar-benar keluar dari proses studi.

Hasil penelitian ini juga relevan dengan penelitian Putra dkk. [19] yang menunjukkan bahwa algoritma berbasis pohon keputusan dan *ensemble learning* dapat digunakan untuk memprediksi *dropout* mahasiswa dengan performa yang baik. Perbedaannya, penelitian ini tidak hanya menempatkan model sebagai alat klasifikasi, tetapi mengintegrasikannya ke dalam *prototype* sistem berbasis web dengan pembagian peran pengguna. Selain itu, penelitian Carballo-Méndivil dkk. [20] menunjukkan bahwa pendekatan *early warning system* berbasis pembelajaran mesin dapat membantu institusi mengidentifikasi mahasiswa berisiko sejak tahap awal. Hal ini mendukung arah penelitian ini, yaitu menjadikan hasil prediksi sebagai dasar awal pemantauan akademik, bukan sebagai keputusan akhir yang berdiri sendiri.

Kontribusi penelitian ini terletak pada pengembangan *prototype* sistem informasi prediksi risiko *dropout* mahasiswa yang menggabungkan model CatBoost dengan fitur prediksi, monitoring, dan pelaporan dalam satu rancangan sistem. Admin berperan dalam pengelolaan data dan proses prediksi, sedangkan kaprodi memanfaatkan hasil prediksi untuk pemantauan dan evaluasi akademik. Secara praktis, *prototype* ini dapat menjadi dasar pengembangan sistem operasional yang lebih lengkap. Secara akademik, penelitian ini memperkuat gagasan bahwa model prediksi akan memiliki nilai guna lebih tinggi apabila diintegrasikan ke dalam sistem informasi akademik yang mendukung proses pemantauan dan pengambilan keputusan akademik.

4. Simpulan

Penelitian ini menghasilkan *prototype* sistem informasi prediksi risiko *dropout* mahasiswa berbasis *web* dengan mengintegrasikan algoritme CatBoost sebagai model prediksi. *Prototype* yang dikembangkan telah memuat fitur utama, yaitu input data mahasiswa secara manual maupun melalui impor file CSV, prediksi risiko *dropout*, monitoring hasil prediksi, serta

laporan yang dapat dimanfaatkan untuk mendukung evaluasi akademik. Fitur-fitur tersebut menunjukkan bahwa rancangan sistem telah diarahkan untuk membantu identifikasi dini mahasiswa berisiko dan penyajian hasil prediksi secara lebih terstruktur.

Hasil pengujian model menunjukkan bahwa CatBoost memperoleh *accuracy* sebesar 90,29%, *precision* 85,33%, *recall* 88,76%, dan *F1-score* 87,01%. Nilai tersebut menunjukkan bahwa model memiliki performa yang baik, khususnya dalam mendeteksi mahasiswa yang benar-benar berisiko *dropout*. Dengan demikian, penelitian ini menunjukkan bahwa integrasi model prediksi CatBoost ke dalam *prototype* sistem informasi berpotensi mendukung proses pemantauan akademik dan menjadi dasar pengembangan sistem operasional yang lebih lengkap di lingkungan perguruan tinggi.

Sebagai pengembangan lanjutan, penelitian berikutnya dapat diarahkan pada implementasi sistem secara operasional, penambahan variabel nonakademik yang lebih beragam, penggunaan data internal perguruan tinggi, serta pengembangan fitur notifikasi otomatis agar proses deteksi dini mahasiswa berisiko *dropout* dapat dilakukan secara lebih efektif.

Daftar Referensi

- [1] M. N. Waruwu, Y. Zega, R. N. Mendrofa, and Y. N. Telaumbanua, "Implementasi Algoritma Machine Learning Untuk Deteksi Performa Akademik Mahasiswa," *teknimedia*, vol. 5, no. 2, pp. 181–186, Dec. 2024, doi: doi.org/10.46764/teknimedia.v5i2.214.
- [2] M. P. Colpo, T. Thompsen Primo, M. S. de Aguiar, and C. Cechinel, "Educational data mining for dropout prediction: Trends, opportunities, and challenges," *Rev. Bras. Inform. Na Educ.*, vol. 32, pp. 220–256, May 2024, doi: 10.5753/rbie.2024.3559.
- [3] D. I. Saputra and D. A. Haris, "Perancangan Business Intelligence untuk Prediksi Risiko Drop Out Mahasiswa Berdasarkan Analisis Data Historis," *J. Ekon. Manaj. Sist. Inf.*, vol. 7, no. 2, pp. 1572–1579, Dec. 2025, doi: doi.org/10.38035/jemsi.v7i2.6991.
- [4] C. F. de Oliveira, S. R. Sobral, M. J. Ferreira, and F. Moreira, "How does learning analytics contribute to prevent students' dropout in higher education: A systematic literature review," *Big Data Cogn. Comput.*, vol. 5, no. 4, p. 64, Nov. 2021, doi: doi.org/10.3390/bdcc5040064.
- [5] M. Vaarma and H. Li, "Predicting student dropouts with machine learning: An empirical study in Finnish higher education," *Technol. Soc.*, vol. 76, no. 102474, p. 102474, Mar. 2024, doi: doi.org/10.1016/j.techsoc.2024.102474.
- [6] A. Villar, C. Robledo, and V. De Andrade, "Supervised machine learning algorithms for predicting student dropout and academic success: a comparative study," *Discov. Artif. Intell.*, vol. 4, no. 2, Jan. 2024, doi: 10.1007/s44163-023-00079-z.
- [7] D. A. Gutierrez-Pachas, G. Garcia-Zanabria, E. Cuadros-Vargas, G. Camara-Chavez, and E. Gomez-Nieto, "Supporting decision-making process on higher education dropout by analyzing academic, socioeconomic, and equity factors through machine learning and survival analysis methods in the Latin American context," *Educ. Sci.*, vol. 13, no. 2, p. 154, Feb. 2023, doi: doi.org/10.3390/educsci13020154.
- [8] D. Bañeres, M. E. Rodríguez-González, A.-E. Guerrero-Roldán, and P. Cortadas, "An early warning system to identify and intervene online dropout learners," *Int. J. Educ. Technol. High. Educ.*, vol. 20, no. 1, p. 3, Jan. 2023, doi: 10.1186/s41239-022-00371-5.
- [9] E. Nimy and M. Mosia, "applied sciences Identifying At-Risk Students for Early Intervention — A Probabilistic Machine Learning Approach," *Appl. Sci.*, vol. 13, no. 6, p. 3869, Mar. 2023, doi: doi.org/10.3390/app13063869.
- [10] J. Keider and G. Santacoloma, "Predictive Model to Identify College Students with High Dropout Rates Modelo predictivo para identificar estudiantes universitarios con alto grado de deserción Modelo predictivo para identificar estudantes universitários com alto risco de evasão," *SciELO*, vol. 25, pp. 1–10, May 2023, doi: 10.24320/redie.2023.25.e13.5398.
- [11] A. C. M. T and A. L. R. P. J, "Ensemble Machine Learning Model for University Students' Risk Prediction and Assessment of Cognitive Learning Outcomes," *IJIET*, vol. 13, no. 6, pp. 948–958, Jun. 2023, doi: 10.18178/ijiet.2023.13.6.1891.
- [12] S. Wang and J. He, "Evaluating and Forecasting Undergraduate Dropouts Using Machine Learning for Domestic and International Students," *technologies*, vol. 13, no. 11, p. 480, Oct. 2025, doi: doi.org/10.3390/technologies13110480.

-
- [13] N. Kurbanazarova, P. Sethuraman, D. P, N. A. Khamitdkhanovich, Y. Mahmood Younus, and A. D. Ghate, "Student Dropout Risk Classification Using CatBoost Algorithm in Higher Education Retention Systems," in *2025 International Conference on AI-Driven STEM Education and Learning Technologies (AISTEMEDU)*, Dec. 2025, pp. 1–7. doi: 10.1109/AISTEMEDU67077.2025.11403945.
- [14] E. Seo, J. Yang, J. Lee, and G. So, "Predictive modelling of student dropout risk: Practical insights from a South Korean distance university," *Heliyon*, vol. 10, no. 11, p. e30960, Jun. 2024, doi: 10.1016/j.heliyon.2024.e30960.
- [15] S. Kim, E. Yoo, S. Kim, L. Palma, and O. Academy, "Associated Factor Analysis Using Machine Learning Techniques", arXiv preprint arXiv:2310.10987, Oct. 2023, doi: 10.48550/arXiv.2310.10987.
- [16] D. Andrade-girón *et al.*, "Predicting Student Dropout based on Machine Learning and Deep Learning: A Systematic Review," *EAI Endorsed Trans.*, vol. 10, no. 5, pp. 1-11, Jul. 2023, doi: 10.4108/eetsis.3586.
- [17] M. A. Hassan and A. H. Muse, "Insights from the 2022 National Education Accessibility Survey in Somaliland," *Appl. Sci. Learn.*, vol. 14, no. 17, p. 7593, Aug. 2024, doi: doi.org/10.3390/app14177593.
- [18] M. Cannistrà, C. Masci, F. Ieva, T. Agasisti, and A. M. Paganoni, "Early-predicting dropout of university students: an application of innovative multilevel machine learning and statistical techniques," *Stud. High. Educ.*, vol. 47, no. 9, pp. 1935–1956, Sep. 2022, doi: doi.org/10.1080/03075079.2021.2018415.
- [19] L. G. R. Putra, D. D. Prasetya, and M. Mayadi, "Student dropout prediction using Random Forest and XGBoost method," *INTENSIF*, vol. 9, no. 1, pp. 147–157, Feb. 2025, doi: doi.org/10.29407/intensif.v9i1.21191.
- [20] B. Carballo-Méndivil, A. Arellano-González, N. J. Ríos-Vázquez, and M. del P. Lizardi-Duarte, "Predicting student dropout from day one: {XGBoost-based} early warning system using pre-enrollment data," *Appl. Sci.*, vol. 15, no. 16, p. 9202, Aug. 2025, doi: doi.org/10.3390/app15169202.