p-ISSN: 2746-9204
e-ISSN: 2746-9190
Volume 4, No. 1, 2024

OPTIMALISASI PREDIKSI KELULUSAN MAHASISWA TEPAT WAKTU MENGGUNAKAN *BINNING* DAN *SYNTHETIC MINORITY OVERSAMPLING TECHNIQUE* (SMOTE)

Faidhul Rahman[1], Mustikasari[2]

[1,2,]Teknik Informatika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Alauddin Makassar, Jl. H.M. Yasin Limpo No. 36, Gowa, Sulawesi Selatan, Indonesia, 92118.
Email: [1]faidhulrahman@gmail.com, [2]mustikasari@uin-alauddin.ac.id

***Koresponden Author:** Faidhul Rahman, faidhulrahman@gmail.com
Accepted: 22 02, 2024 ; Revised: 27 02, 2024; Published: 28 02, 2024

Abstrak

Kelulusan mahasiswa tepat waktu adalah situasi dimana seorang mahasiswa lulus dari program pendidikan mereka pada waktu yang direncanakan atau yang ditentukan oleh institusi pendidikan terkait. Penelitian ini bertujuan untuk mengoptimalkan prediksi kelulusan mahasiswa tepat waktu menggunakan metode Binning untuk mengelompokkan variabel ke dalam kategori diskrit dan Synthetic Minority Oversampling Technique (SMOTE) untuk mengatasi ketidakseimbangan kelas pada dataset. Data berisi beberapa variabel dianalisis menggunakan algoritma Machine learning Naïve Bayes, Decision Tree, dan Random Forest. Evaluasi model dilakukan dengan menggunakan metrik seperti precision, Recall, accuracy, dan F1-score. Hasilnya menegaskan bahwa kombinasi Binning dan SMOTE memberikan dampak yang signifikan terhadap peningkatan akurasi prediksi. Hasil penelitian ini diharapkan dapat memberikan kontribusi dalam meningkatkan ketepatan prediksi kelulusan mahasiswa tepat waktu. Dengan mengoptimalkan penggunaan Binning dan SMOTE, diharapkan model prediksi dapat mengatasi kendala ketidakseimbangan data dan memberikan informasi yang lebih akurat kepada institusi pendidikan tinggi untuk mengambil tindakan preventif yang diperlukan guna meningkatkan tingkat kelulusan mahasiswa serta menjadi referensi bagi penelitian serupa di masa depan.
Kata kunci*: prediksi, kelulusan, binning, SMOTE*

Abstract

[Optimization of On-Student Graduation Predictions on Time Using Binning And Synthetic Minority Oversampling Technique (SMOTE)] *On-time student graduation is a situation where a student graduates from their educational program at the time planned or determined by the relevant educational institution. This research aims to optimize predictions of student graduation on time using the Binning method to group variables into discrete categories and Synthetic Minority Oversampling Technique (SMOTE) to overcome class imbalances in the dataset. Data containing several variables was analyzed using the Naïve Bayes, Decision Tree and Random Forest machine learning algorithms. Model evaluation is carried out using metrics such as precision, Recall, accuracy, and F1-score. The results confirm that the combination of Binning and SMOTE has a significant impact on increasing prediction accuracy. It is hoped that the results of this research can contribute to increasing the accuracy of predicting student graduation on time. By optimizing the use of Binning and SMOTE, it is hoped that the prediction model can overcome the problem of data imbalance and provide more accurate information to higher education institutions to take the necessary preventive actions to increase student graduation rates and become a reference for similar research in the future.*

Keywords*: prediction, graduation, binning, SMOTE*

p-ISSN: 2746-9204
e-ISSN: 2746-9190
Volume 4, No. 1, 2024

1. PENDAHULUAN

Kelulusan mahasiswa tepat waktu adalah situasi dimana seorang mahasiswa lulus dari program pendidikan mereka pada waktu yang direncanakan atau yang ditentukan oleh institusi pendidikan terkait. Ini berarti bahwa mahasiswa tersebut lulus dalam jangka waktu yang sesuai dengan jadwal pendidikan mereka, tanpa terlambat atau mengalami keterlambatan. Konsep ini sangat penting bagi mahasiswa dan institusi pendidikan. Bagi mahasiswa, kelulusan tepat waktu memungkinkan mereka untuk berlanjut ke tahap berikutnya dalam hidup dan karir mereka secepat mungkin. Bagi institusi pendidikan, kelulusan tepat waktu membantu dalam perencanaan dan pengelolaan sumber daya yang tersedia.

Mahasiswa dituntut untuk melakukan evaluasi terhadap usaha dan tindakan yang telah dilakukan dalam perjalanan pendidikan mereka. Evaluasi ini bermanfaat untuk menyempurnakan tindakan yang sudah baik dan memperbaiki hal-hal yang masih kurang sehingga pada saat kelulusan tiba, mahasiswa telah mencapai prestasi yang maksimal dan meraih kelulusan dengan penuh kualitas. Namun Tingkat kelulusan mahasiswa bisa sangat sulit untuk dideteksi secara dini, sehingga mengakibatkan keterlambatan lulusan. Untuk mengatasi hal tersebut perlu ada teknik untuk dapat melakukan prediksi terhadap kelulusan mahasiswa.

Sebagai contoh [1] telah meneliti ketepatan waktu lulus mahasiswa yang menghasilkan lebih banyak kesalahan dalam memprediksi kelulusan mahasiswa tepat waktu ketimbang mahasiswa yang tidak lulus tepat waktu. Sama halnya dengan penelitian [2] yang memprediksi tingkat kelulusan mahasiswa tepat waktu dan menghasilkan akurasi prediksi mahasiswa tidak lulus tepat waktu lebih besar dari mahasiswa lulus tepat waktu. Hal ini terjadi karena masalah data tidak seimbang belum teratasi.

Prediksi kelulusan mahasiswa adalah proses memperkirakan atau menentukan kemungkinan waktu seorang mahasiswa akan lulus dari program studinya. Ini biasanya didasarkan pada analisis data seperti indeks prestasi di tiap semester, jumlah kredit semester, dan faktor-faktor lain yang mempengaruhi kelulusan mahasiswa.

Data yang tidak seimbang memiliki pengaruh yang signifikan terhadap hasil prediksi data kelulusan mahasiswa. Biasanya, algoritma *machine learning* cenderung memprediksi kelas yang lebih sering muncul pada dataset sebagai kelas dominan. Misalnya, jika data mahasiswa yang tidak lulus tepat waktu lebih banyak dibandingkan dengan data mahasiswa yang lulus tepat waktu, maka model prediksi akan cenderung memprediksikan bahwa setiap mahasiswa akan tidak lulus tepat waktu. Dalam hal ini, data kelulusan mahasiswa yang tepat waktu yang merupakan kelas minoritas dalam dataset dapat terprediksi dengan kurang baik. Oleh karena itu, perlu dilakukan tindakan seperti *oversampling* atau *Binning* untuk mengatasi masalah ini dan meningkatkan akurasi prediksi.

Binning adalah sebuah proses untuk mengelompokkan data ke dalam bagian-bagian yang lebih kecil yang disebut bin berdasarkan kriteria tertentu. *Binning* data merupakan salah satu teknik praproses data yang digunakan untuk meminimalisasi kesalahan dalam pengamatan serta terkadang dapat meningkatkan akurasi dari model prediktif [3]. *Binning* dapat membantu mengoptimalkan prediksi data kelulusan mahasiswa dari segi pengelompokan data numerik menjadi beberapa kelas. Dalam prediksi data kelulusan mahasiswa, banyak variabel yang memiliki nilai numerik yang besar dan memiliki rentang yang luas. Oleh karena itu, dalam membuat model prediksi, data tersebut perlu dikelompokkan menjadi beberapa kelas. Dengan menggunakan

p-ISSN: 2746-9204
e-ISSN: 2746-9190
Volume 4, No. 1, 2024

Binning, data numerik dapat dibagi menjadi beberapa kelas, seperti kelas rendah, sedang, dan tinggi. Hal ini akan membantu memperbaiki akurasi prediksi dengan memperkecil perbedaan antar kelas yang besar dan memastikan bahwa data tidak terlalu banyak terdistribusi pada beberapa kelas saja. Selain itu, Binning juga dapat membantu meminimalisir perpengaruhan outliers dalam data.

Synthetc Minority Oversampling Technique (SMOTE). Pada dasarnya SMOTE adalah salah satu penerapan dari metode *oversampling*. Sehingga salah satu kelebihan metode ini adalah tidak akan menyebabkan adanya informasi yang hilang [4]. SMOTE dapat meningkatkan akurasi prediksi data karena SMOTE membuat representasi dari kelas minoritas dengan menambahkan instance baru yang akan membantu model untuk lebih baik memahami dan memprediksi kelas minoritas.

SMOTE memecahkan masalah data tidak seimbang dengan membuat sintesis dari data minoritas. Dalam hal ini, data minoritas adalah data untuk kelas yang sangat sedikit. SMOTE menghitung jarak antar titik data dan membuat titik-titik baru yang berada di antara titik-titik yang ada, memperbanyak jumlah data untuk kelas minoritas. Dengan memperbanyak jumlah data untuk kelas minoritas, model prediksi akan memiliki lebih banyak data untuk melatih dan membuat prediksi yang lebih akurat. Ini akan memastikan bahwa model tidak memiliki bias terhadap data mayoritas dan dapat membuat prediksi yang lebih akurat bagi kelas minoritas.

Penelitian pada kali ini mengkombinasikan Binning dan SMOTE dalam melakukan prediksi data kelulusan mahasiswa tepat waktu dengan optimal. Perbedaan kondisi akurasi prediksi sebelum dan setelah menggunakan kombinasi teknik Binning dan SMOTE dapat sangat signifikan. Sebelum menggunakan kombinasi ini, prediksi kelulusan mahasiswa seringkali tidak akurat karena data yang digunakan tidak seimbang dan cenderung memiliki kelas mayoritas yang lebih besar daripada kelas minoritas. Kombinasi ini dapat membantu memperbaiki masalah imbalance pada data yang digunakan untuk memprediksi dan meningkatkan akurasi hasil prediksi menjadi lebih optimal.

2. METODE

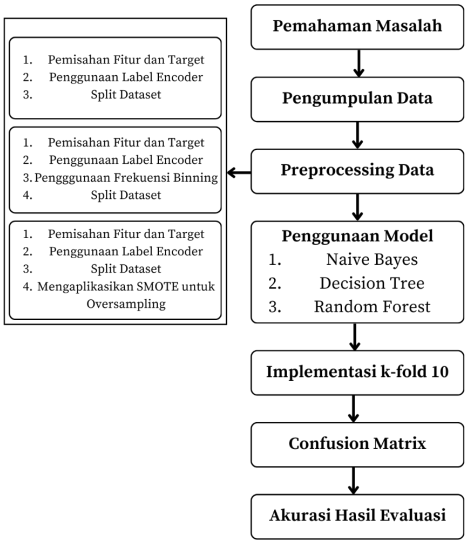
Metode optimalisasi dalam prediksi kelulusan mahasiswa tepat waktu dengan menggabungkan teknik *Binning* dan *Synthetic Minority Oversampling Technique* (SMOTE). Hasil analisis yang diharapkan bahwa penggunaan metode ini dapat mengatasi permasalahan umum dalam pemodelan prediksi kelas ketika terdapat ketidakseimbangan yang signifikan antara kelas mayoritas dan kelas minoritas dalam dataset. Penggunaan akan Binning membantu menghasilkan prediksi yang lebih seimbang antara kedua kelas, sementara SMOTE digunakan untuk menciptakan data sintetis yang serupa dengan kelas minoritas.

Dalam konteks prediksi, optimasi model klasifikasi sangat penting untuk mencapai keseimbangan yang optimal antara *True Positive*s (TP), *True Negatives* (TN), *False Positive*s (FP), dan *False Negatives* (FN). Menemukan keseimbangan yang tepat antara komponen-komponen ini adalah kunci untuk menghasilkan model yang efektif dan akurat. Melalui optimasi, kita berusaha untuk meningkatkan TP dan TN, yang mengindikasikan ketepatan prediksi yang tinggi, sambil sebisa mungkin mengurangi FP dan FN, yang mengukir kemungkinan kesalahan dalam prediksi

Jika terdapat lebih banyak *False Positive*s (FP) dalam prediksi kelulusan mahasiswa, hal ini dapat berdampak pada mahasiswa yang sebenarnya gagal, membuat mereka merasa terkejut dan mungkin menghasilkan biaya tambahan serta menurunkan kualitas

p-ISSN: 2746-9204
e-ISSN: 2746-9190
Volume 4, No. 1, 2024

pendidikan universitas. Di sisi lain, jika ada lebih banyak *False Negatives* (FN), mahasiswa yang berpotensi berhasil tetapi dianggap gagal dapat mengalami dampak negatif pada perkembangan karier mereka, dengan risiko kehilangan bakat berharga dan penurunan kepuasan mahasiswa. Oleh karena itu, menjaga keseimbangan yang baik antara FP dan FN penting untuk menghasilkan keputusan kelulusan yang adil dan akurat.



Gambar 2.1. Pengolahan dan Analisis Data

1. Pemahaman Data

Langkah pertama adalah memahami data dengan baik. Ini mencakup pemahaman tentang atribut-atribut yang ada, jenis data yang dihadapi, serta konteks dan tujuan analisis.

Dalam penelitian ini dirumuskan definisi operasional variabel utama yang menjadi fokus penelitian. Variabel pertama adalah "Prediksi Kelulusan Mahasiswa Tepat Waktu," yang diukur dalam bentuk biner, yakni "1" untuk prediksi kelulusan tepat waktu dan "0" untuk prediksi kelulusan tidak tepat waktu. Variabel kedua adalah "Teknik *Binning*," yang merujuk pada penggunaan metode *Binning* dalam pemrosesan data, dengan metrik pengukuran yang mencakup jenis *Binning* yang digunakan dan jumlah interval atau bin yang terbentuk. Variabel ketiga adalah Synthetic Minority Oversampling Technique (SMOTE). *Synthetic Minority Oversampling Technique* (SMOTE) adalah satu diantara turunan dari metode *oversampling*. Metode SMOTE merupakan metode untuk mengatasi data tidak seimbang dengan replikasi data buatan atau data sintetis untuk kelas data minoritas[5].

Definisi operasional yang tegas untuk ketiga variabel ini memungkinkan penelitian untuk dilakukan secara terstruktur, memfasilitasi pengumpulan dan analisis data yang konsisten, serta memungkinkan interpretasi hasil penelitian yang lebih akurat dan bermakna.

2. Pengumpulan Data

Pengumpulan data merupakan langkah awal dalam proses. Data harus relevan dengan tujuan analisis, terstruktur dengan baik, dan mencakup atribut-atribut yang berpotensi mempengaruhi prediksi kelulusan mahasiswa tepat waktu.

3. Preprocessing

Preprocessing mencakup berbagai langkah seperti pembersihan data dari nilai yang hilang atau noise, transformasi data (*Binning*), dan penerapan teknik *oversampling* (SMOTE) untuk menangani ketidakseimbangan kelas. Ini membantu mempersiapkan data untuk tahap pemodelan.

4. Penggunaan Model (*Naive Bayes*, *Random Forest*, *Decision Tree*)

Menggunakan berbagai model seperti *Naive Bayes*, *Random Forest*, dan *Decision Tree* memungkinkan untuk menguji pendekatan yang berbeda dalam memodelkan data. Setiap model memiliki karakteristiknya sendiri dan dapat memberikan wawasan tentang hubungan antara atribut dan hasil prediksi.

5. Implementasi *K-fold Cross Validation* (*K-fold* 10)

Dengan menerapkan *K-fold* Cross Validation dengan k=10, hal ini membagi data menjadi 10 bagian (*fold*) dan melakukan pelatihan dan pengujian sebanyak 10 kali. Ini

p-ISSN: 2746-9204
e-ISSN: 2746-9190
Volume 4, No. 1, 2024

membantu menghindari bias dalam evaluasi model dan memberikan perkiraan yang lebih baik tentang kinerja rata-rata model.

6. *Confusion matrix*

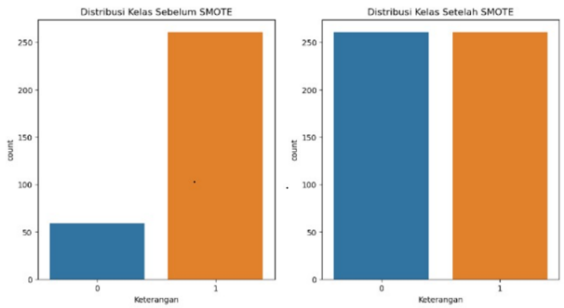
Confusion matrix adalah alat untuk mengukur performa model dengan membandingkan prediksi model dengan kenyataan. Ini membantu menghitung metrik seperti akurasi, presisi, *Recall*, dan F1-*score*, yang semuanya memberikan wawasan yang lebih rinci tentang performa model.

7. Akurasi Hasil Evaluasi

Akurasi merupakan metrik penting untuk mengukur seberapa baik model dalam melakukan prediksi secara keseluruhan. Namun, juga penting untuk melihat metrik lain seperti presisi dan *Recall*, terutama mengingat kelas yang tidak seimbang (kelulusan mahasiswa tidak tepat waktu sebagai kelas minoritas).

3. HASIL DAN PEMBAHASAN

Berikut merupakan distribusi kelas sebelum dan setelah menggunakan SMOTE



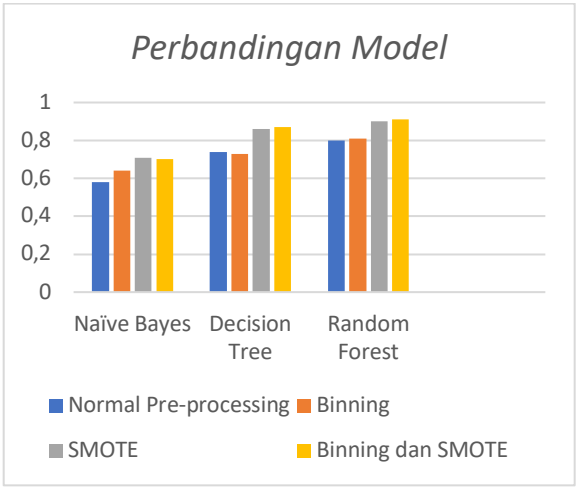
Gambar 3.1 Distribusi SMOTE

SMOTE diterapkan pada dataset, dengan metrik pengukuran mencakup apakah SMOTE diterapkan, besarnya *oversampling* yang dilakukan, dan dampaknya terhadap kinerja model.

Berikut disajikan tabel perbandingan tiap model:

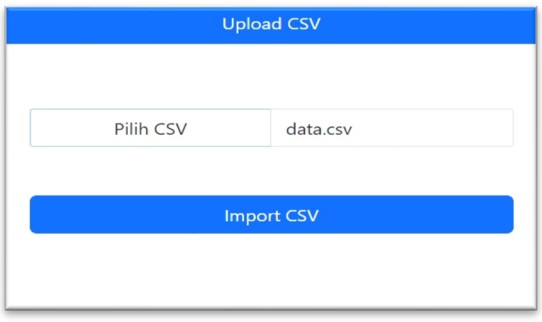
Tabel 3.1 Perbandingan Model

No	Model	Normal *Preprocessing*	*Binning*	SMOTE	*Binning* dan SMOTE
1	Naïve Bayes	0,58	0,64	0,71	0,70
2	*Decision Tree*	0,74	0,73	0,86	0,87
3	*Random Forest*	0,80	0,81	0,90	**0,91**



Gambar 3.2 Perbandingan Model

Penting untuk memahami apa yang diprediksi oleh model dan mengapa. Ini membantu dalam mengambil tindakan yang sesuai berdasarkan prediksi, terutama dalam konteks pengambilan keputusan. Dapat dilihat dari tabel yang telah diberikan, terdapat beberapa kesimpulan yang dapat ditarik terkait dengan kinerja berbagai model machine learning dalam memprediksi kelulusan mahasiswa tepat waktu dengan variasi dalam pemrosesan data.



Gambar 3.3 Tampilan Halaman Utama

Pertama, hasil menunjukkan bahwa penggunaan teknik *Binning* dan *Synthetic Minority Oversampling Technique* (SMOTE)

p-ISSN: 2746-9204
e-ISSN: 2746-9190
Volume 4, No. 1, 2024

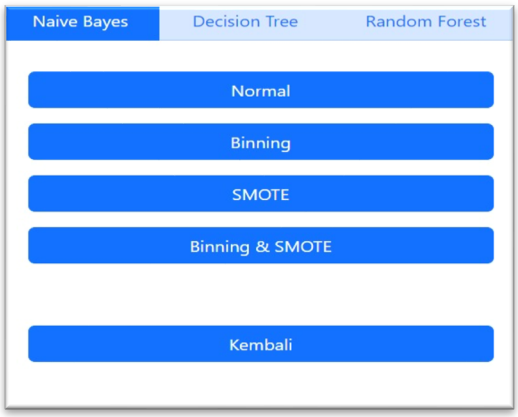
secara bersamaan (*Binning* dan SMOTE) cenderung menghasilkan tingkat akurasi yang lebih tinggi dibandingkan dengan penggunaan teknik tersebut secara terpisah atau bahkan dengan pemrosesan data yang biasa (Normal *Preprocessing*). Ini menunjukkan bahwa strategi gabungan ini efektif dalam mengatasi permasalahan ketidakseimbangan kelas dalam dataset dan meningkatkan kemampuan model dalam memprediksi kelulusan mahasiswa tepat waktu.

Kedua, dari tiga model yang dievaluasi, yaitu Naïve Bayes, *Decision Tree*, dan *Random Forest*, *Decision Tree* dan *Random Forest* menunjukkan performa yang lebih baik dalam semua kondisi pemrosesan data. Bahkan, ketika teknik *Binning* dan SMOTE diterapkan, *Decision Tree* dan *Random Forest* mencapai tingkat akurasi yang sangat tinggi. Ini mengindikasikan bahwa model berbasis pohon keputusan cenderung lebih sesuai untuk tugas prediksi ini, terutama ketika diterapkan dengan pemrosesan data yang tepat.

Terakhir, hasil ini juga menekankan pentingnya pemrosesan data yang optimal dalam memengaruhi kinerja model. *Binning* dan SMOTE adalah teknik yang efektif dalam mengatasi masalah ketidakseimbangan kelas dan memproses data kategori, yang berperan penting dalam meningkatkan akurasi prediksi kelulusan mahasiswa tepat waktu.

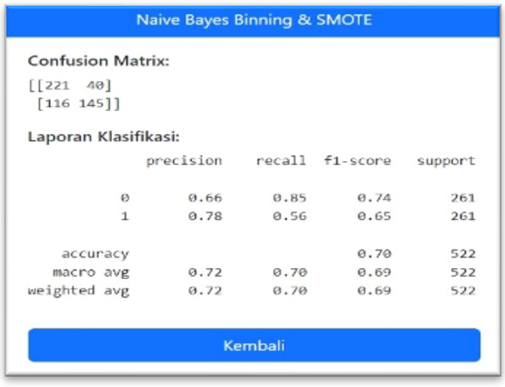
Pembuatan model dan pengujian model telah dilaksanakan, untuk penggunaan agar lebih bermanfaat untuk pengguna yang lebih luas selanjutnya dilakukan implementasi model tersebut dalam deployment *interface* website sederhana, Adapun tampilan website tersebut antara lain:

Tampilan awal merupakan tampilan pengi*nput*an data yang akan diprediksi adalah antarmuka yang memungkinkan pengguna untuk memasukkan informasi atau data yang diperlukan untuk melakukan prediksi dengan model *machine learning*. Tampilan ini dapat beragam tergantung pada jenis aplikasi atau model.



Gambar 3.4. Tampilan Halaman Utama

Tampilan utama untuk menguji tiap model prediksi adalah antarmuka yang memungkinkan pengguna menguji model *machine learning* yang telah dikembangkan dengan berbagai data uji. Ini adalah bagian penting dari aplikasi machine learning karena memungkinkan pengguna mengukur kinerja model pada data yang belum pernah dilihat sebelumnya. Model prediksi yang tersedia antara lain Naïve Bayes, *Decision Tree*, dan *Random Forest*. Setiap model tersebut memiliki metode preprosesing yang berbeda beda bisa diterapkan antara lain secara normal, menggunakan *Binning*, SMOTE, *Binning* dan SMOTE



Gambar 3.5. Tampilan Halaman Hasil

Tampilan hasil prediksi adalah antarmuka yang menampilkan hasil dari prediksi yang

p-ISSN: 2746-9204
e-ISSN: 2746-9190
Volume 4, No. 1, 2024

dilakukan oleh model *machine learning*. Tampilan ini dapat berisi informasi seperti *Confusion matrix* serta metrik evaluasi seperti accuracy, *precision*, *Recall*, F1-score, dan *support*.

4. KESIMPULAN

Kesimpulan dari analisis yang telah dilakukan dapat dirangkum dalam tiga poin utama, pertama yaitu peningkatan performa dengan Teknik *Binning* dan SMOTE, implementasi teknik *Binning* dan *Synthetic Minority Oversampling Technique* (SMOTE) pada model *Random Forest* dan *Decision Tree* secara signifikan meningkatkan kinerja prediksi kelulusan mahasiswa tepat waktu. Penggunaan Binning membantu dalam menghasilkan prediksi yang lebih seimbang antara kelas mayoritas dan minoritas, sementara SMOTE berhasil mengatasi ketidakseimbangan kelas dengan menghasilkan data sintetis untuk kelas minoritas; kedua yaitu efektivitas model *Random Forest* dan *Decision Tree* di mana hasil menunjukkan bahwa model *Random Forest* memiliki performa yang baik dalam mengklasifikasikan kelas mayoritas. Namun, ketika digabungkan dengan teknik SMOTE, model ini mampu meningkatkan kemampuan dalam mengklasifikasikan kelas minoritas dengan tingkat *Recall* yang lebih baik. Model *Decision Tree* juga mengalami peningkatan yang signifikan dalam kinerja prediksi setelah penerapan teknik SMOTE; ketiga yaitu kombinasi optimal untuk prediksi kelulusan di mana hasil yang paling mengesankan terjadi ketika teknik *Binning* dan SMOTE digabungkan dengan model *Random Forest* dan *Decision Tree*. Kombinasi ini menghasilkan hasil yang sangat baik dalam hal *precision*, *Recall*, F1-*score*, dan akurasi keseluruhan. Penggunaan kedua teknik ini bersama-sama mengatasi masalah ketidakseimbangan kelas dan menghasilkan prediksi yang lebih kuat dan handal untuk kedua kelas.

5. DAFTAR PUSTAKA

[1] W. Agwil, H. Fransiska, and N. Hidayati, "Analisis Ketepatan Waktu Lulus Mahasiswa Dengan Menggunakan Bagging Cart," *Fibonacci : Jurnal Pendidikan Matematika dan Matematika*, vol. 6, no. 2, pp. 155–166, 2020.

[2] S. Salmu and A. Solichin, "Prediksi Tingkat Kelulusan Mahasiswa Tepat Waktu Menggunakan Naïve Bayes: Studi Kasus UIN Syarif Hidayatullah Jakarta," *Prosiding Seminar Nasional Multidisiplin Ilmu*, 2017.

[3] L. Afifah, "Cara Binning Data di Python dengan Pandas," IlmudataPy. Accessed: Feb. 05, 2023. [Online]. Available: https://ilmudatapy.com/binning-data-di-python/

[4] N. P. Y. T. Wijayanti, E. N. Kencana, and W. Sumarjaya, "SMOTE : POTENSI DAN KEKURANGANNYA PADA SURVEI," *E-Jurnal Matematika*, vol. 10, no. 4, pp. 235–240, 2021.

[5] R. D. Permatasari, S. W. Rizki, and N. N. Debataraja, "Penerapan Synthetic Minority Oversampling Technique Dalam Mengatasi Data Tidak Seimbang Pada Metode Classification And Regression Tree," 2020.