

Penerapan Teknik *Random Oversampling* Untuk Memprediksi Ketepatan Waktu Lulus Menggunakan Algoritma *Random Forest*

Sri Diantika^{1*}, Hiya Nalatissifa², Nurlaelatul Maulidah³, Riki Supriyadi⁴, Ahmad Fauzi⁵

^{1,2,3,5}Universitas Bina Sarana Informatika
Jl. Kramat Raya No. 98, Jakarta Pusat, DKI Jakarta-10450, Indonesia

⁴Universitas Nusa Mandiri
Jl. Jatiwaringin No.2, Cipinang Melayu, Makasar, Jakarta Timur-13620, Indonesia

e-mail: ¹sri.szd@bsi.ac.id, ²hiya.hys@bsi.ac.id, ³nurlaelatul.nlt@bsi.ac.id, ⁴riki.rsd@nusamandiri.ac.id,
⁵ahmad.fzx@bsi.ac.id,

(*) Corresponding Author

Artikel Info : Diterima : 29-05-2023 | Direvisi : 13-11-2023 | Disetujui : 03-01-2024

Abstrak - Ketepatan waktu lulus merupakan suatu hal yang didambakan mahasiswa, Selain penting bagi mahasiswa, ketepatan waktu lulus ini juga sangat penting bagi perguruan tinggi, hal ini dikarenakan aspek kelulusan mahasiswa menjadi salah satu aspek penilaian dalam sebuah proses akreditasi institusi suatu perguruan tinggi untuk menunjukkan kualitasnya. Salah satu kendala yang dihadapi untuk mengetahui apakah seorang mahasiswa dapat lulus tepat waktu atau tidak adalah masa studi yang belum dapat dideteksi secara dini, hal ini akan berdampak pada terlambatnya kelulusan mahasiswa. Untuk menanggulangi tantangan tersebut, penelitian ini bertujuan untuk mengidentifikasi ketepatan waktu kelulusan mahasiswa sejak dini dengan memanfaatkan nilai Indeks Prestasi Kumulatif (IPK) yang dicapai mahasiswa selama masa *study* mereka. Penelitian perihal prediksi ketepatan waktu lulus mahasiswa ini menggunakan model *random forest*. Algoritma *random forest* menggabungkan beberapa algoritma dengan tipe yang sama, seperti *multiple decision tree*. Data yang digunakan sebagai objek riset ini, memiliki jumlah kelas data yang tidak seimbang, untuk mengatasi hal ini maka diterapkan teknik *resampling random oversampling* (ROS) dan juga menetapkan *Split validation* dimana akan dilakukan pembagian antara data pembelajaran sebesar 50% untuk data pengujian 50%. Untuk mengkaji model yang dibentuk, Penulis memakai *metrics evaluation* semacam akurasi, *precision*, serta *recall*. Hasil dari riset menampilkan jikalau model yang diusulkan bisa dengan baik melaksanakan prediksi dibanding dengan model lainnya, yaitu dengan hasil nilai *precision* sebesar 87,05%, uji akurasi sebesar 90,04%, *recall* 90,04%. Dari hasil ini dapat diartikan bahwa algoritma *random forest* dinilai baik dalam memprediksi ketepatan waktu lulus seorang mahasiswa

Kata Kunci : *Imbalanced*, Ketepatan Waktu Lulus, Prediksi, *Random Forest*, *Random Oversampling*

Abstracts - *Punctuality of graduation is something that students yearn for, besides being important for students, punctuality of graduation is also very important for universities, this is because the aspect of student graduation is one aspect of assessment in an institutional accreditation process of a university to show its quality. One of the obstacles faced to find out whether a student can graduate on time or not is because the study period cannot be detected early, this will have an impact on late student graduation. To analyze this, a lot of research was conducted on the accuracy of student graduation, through the cumulative grade point average (GPA) obtained by students during their studies. This research on the prediction of student graduation timeliness uses a random forest algorithm model. The data used in this research object has an unbalanced number of data classes, to overcome this, a random oversampling (ROS) resampling technique is applied and also applies Split validation or division between learning data by 50% for test data and 50%. To evaluate the model built, the author uses evaluation metrics such as accuracy, recall, and precision. The results of the study showed that the proposed model can well predict compared to other models, namely with the results precision of 87.05%, accuracy test values of 90.04%, recall of 90.04%. From these results, it can be interpreted that the random forest algorithm is considered good in predicting the timeliness of a student's graduation*

Keywords : *Imbalanced*, *On Time Graduation*, *Prediction*, *Random Forest*, *Random Oversampling*



PENDAHULUAN

Pendidikan saat ini menjadi aspek yang dibutuhkan dalam suatu kehidupan. Dalam dunia pendidikan akan banyak ditemui ilmu-ilmu pengetahuan yang nantinya akan bermanfaat untuk masa depan kita. Dalam proses menempuh pendidikan akan banyak ditemukan kendala, dari kendala-kendala ini diharapkan pelajar atau mahasiswa dapat menerapkan ilmu yang didapatkannya dan menemukan solusi untuk menyelesaikannya. Selain ilmu pengetahuan, hal lain yang didambakan pelajar atau mahasiswa adalah dapat lulus dengan tepat waktu (Heryana et al., 2020)

Ketepatan waktu lulus ini merupakan impian dari mahasiswa, sebagai wujud atau bentuk dari perjuangan mereka selama beberapa tahun belajar di bangku kuliah. Selain ketepatan waktu kuliah ini penting bagi mahasiswa, penting juga bagi perguruan tinggi, hal ini disebabkan karena aspek kelulusan mahasiswa merupakan salah satu aspek penilaian dalam sebuah proses akreditasi institusi suatu perguruan tinggi untuk menunjukkan kualitasnya (S.A Amosea, 2021) Kualitas dari perguruan tinggi ini dapat dilihat dari banyak aspek, selain dilihat dari rata-rata lama waktu alumninya mendapatkan pekerjaan, aspek penilaian lainnya yang tak kalah penting adalah presentase ketepatan waktu lulus dari mahasiswanya (Jananto, 2021)

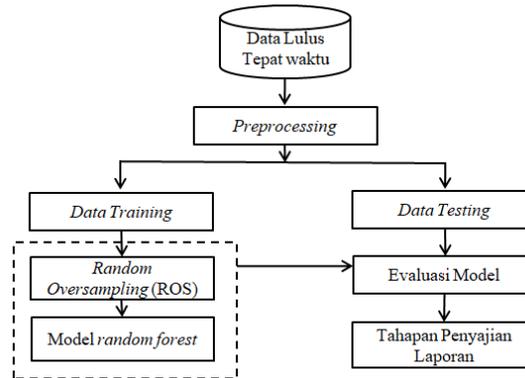
Jumlah besar persentase mahasiswa yang bisa menuntaskan masa *study* tepat waktu akan membuat tingkat evaluasi akreditasi naik. Pada lain sisi, mahasiswa yang bisa menuntaskan masa *study* tepat waktu tidak akan terbebani dengan Uang Kuliah Tunggal ataupun UKT yang wajib senantiasa dibayarkan sepanjang mahasiswa belum dinyatakan lulus tes skripsi. Tetapi masa *study* mahasiswa belum bisa dideteksi secara dini, sehingga bisa berakibat pada terlambatnya kelulusan mahasiswa (Satria et al., 2020) Untuk menganalisa hal tersebut, maka banyak dilakukan penelitian mengenai ketepatan kelulusan mahasiswa, salah satunya pada penelitian Klasifikasi Ketepatan Waktu Kelulusan Mahasiswa yang berstudi kasus di suatu Universitas, membandingkan antara 2 metode yaitu metode *Naïve Bayes* dan *K-Nearest Neighbor*. Hasilnya menunjukkan bahwa menggunakan Algoritma *K-Nearest Neighbor* dapat lebih baik mengklasifikasikan ketepatan waktu lulus mahasiswa dibuktikan dengan perolehan akurasi sebesar 97,68% dengan tingkat *recall* sebesar 86,11% dan presisi 100% (Anwarudin et al., 2022) adapula penelitian dengan fokus yang sama yaitu mengenai ketepatan waktu lulus mahasiswa pada program studi Teknik Elektro dan Teknik Sipil. Penelitian ini memiliki 3 *cluster* atau kelas data yaitu lulus tidak tepat waktu, tepat waktu dan melebihi masa *study*. Hasil penelitian menunjukkan bahwa algoritma K-Means dan Fuzzy C-Means dapat mengidentifikasi ketepatan lulusan dengan baik (Setiaji et al., 2019)

Fiqih Satria sebelumnya melaksanakan riset serupa, di mana ia memproyeksikan ketepatan kelulusan mahasiswa di Fakultas Dakwah dan Ilmu Komunikasi UIN Raden Intan Lampung dengan memanfaatkan algoritma C4.5. Temuan dari penelitian tersebut menunjukkan bahwa algoritma C4.5 dapat digunakan untuk memprediksi waktu kelulusan mahasiswa, dengan tingkat *recall* sebesar 60.4%, presisi 70.70%, dan akurasi 58.2% (Satria et al., 2020) kemudian pada penelitian lain yang juga ingin memprediksi ketepatan waktu lulus dengan membandingkan 2 algoritma, yaitu algoritma C4.5 dan ID3, Temuan dari riset ini menunjukkan bahwa informasi yang diperoleh menunjukkan bahwa nilai akurasi yang dicapai oleh algoritma C4.5 cenderung lebih tinggi secara keseluruhan jika dibandingkan dengan nilai akurasi algoritma ID3 (Faizah & Jananto, 2021) selain itu ada pada penelitian lain menggunakan Algoritma *Naïve Bayes* menyatakan bahwa ia mampu menciptakan model prediksi untuk ketepatan waktu lulus mahasiswa, dengan tingkat akurasi mencapai 93,75%, presisi sebesar 93,75%, *recall* 100%, dan nilai AUC sebesar 1,000 (*excellent classification*) sehingga dapat dikatakan algoritma ini mampu memprediksi secara akurat (Mulyadi & Sugiarto, 2021) dilain penelitian dengan tema yang sama yaitu untuk memprediksi ketepatan waktu lulus menggunakan metode *Decision Tree* C4.5. Penelitian ini menggunakan SMOTE (*Synthetic Minority Oversampling Technique*) pada aplikasi WEKA untuk menyeimbangkan data di kelas minor, Hasil penelitian menunjukkan bahwa atribut IPK, merupakan akar tertinggi yang memberikan pengaruh tertinggi. Akurasi pada validasi silang 10 Lipatan dengan SMOTE dan tanpa SMOTE hasilnya menunjukkan bahwa: dengan SMOTE 83,055% dan tanpa SMOTE 82,644% (Samuel et al., 2019)

Dari beberapa studi kasus yang sudah ada, maka akan dilakukan penelitian dengan tujuan yang sama yaitu untuk memprediksi ketepatan waktu lulus mahasiswa, akan tetapi, dalam penelitian ini akan dicoba beberapa hal baru guna menghasilkan suatu prediksi yang lebih akurat, yaitu dengan menerapkan teknik menyeimbangkan kelas data, hal ini dilakukan karena *dataset* yang digunakan dalam penelitian ini memiliki kelas data yang tidak seimbang. Penelitian ini menggunakan algoritma *random forest*.

METODE PENELITIAN

Pada bagian tata cara riset akan dijabarkan secara mendetail mengenai proses yang dicoba dalam riset ini, dari titik awal akan dilakukan pengumpulan data yang digunakan pada riset sampai bagaimana alur ataupun tahapan dari riset prediksi ketepatan waktu lulus mahasiswa ini berjalan. Buat digram alur riset bisa dilihat pada gambar 1 berikut ini:



Sumber: Penelitian (2023)

Gambar 1. Diagram Alur Penelitian

1. Pengumpulan *Dataset*

Kumpulan data yang dijadikan sebagai objek atau bahan dari riset ini adalah data ketepatan waktu lulus pada salah satu universitas, akan tetapi untuk nama universitas tersebut dirahasiakan. Data ini diperoleh dari data terbuka atau umum dengan url <https://www.kaggle.com/yvirgantara/on-time-graduation-classification/>. *Dataset* tersebut berisi kumpulan data Indeks Prestasi Kumulatif (IPK) mahasiswa. Jumlah data yang digunakan sebanyak 1687 data. Setiap sampel memiliki 4 *parameter* dan label kelas yang mengidentifikasinya ada sebanyak 2 kelas data, yaitu tepat dan tidak tepat.

2. *Preprocessing*

Salah satu proses yang tidak boleh dilewatkan sebelum masuk ke tahap prediksi adalah *preprocessing*. Tahap ini dianggap penting karena tidak semua data yang digunakan sebagai objek dalam penelitian sudah siap (Hermawan & Bellanar Ismiati, 2020), terkadang terdapat data yang kosong hingga data yang *double*. Proses ini pula dicoba supaya informasi yang hendak digunakan cocok dengan kebutuhan riset. Beberapa langkah *preprocessing* juga diuji, seperti mengisi informasi yang kosong, menghapus data ganda atau *duplicate*, memeriksa konsistensi informasi, membersihkan data, dan memperbaiki kesalahan yang mungkin terdapat dalam *dataset* (Nasir & Saharuna, 2020)

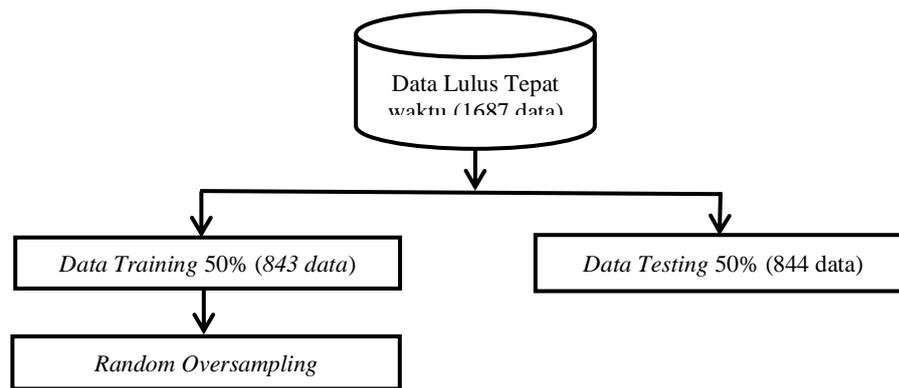
Proses *preprocessing* awal yang dilakukan pada penelitian ini adalah dengan melakukan proses pembersihan data, proses ini dilakukan guna menghilangkan data yang memiliki *missing value* dan *noise*. kemudian juga dilakukan pembersihan data ganda atau *double*.

Alat yang dioperasikan dalam penelitian ini untuk penulisan kode program melibatkan penggunaan Google *Colaboratory* dengan penerapan bahasa pemrograman *Python* dan model *Random Forest*.

3. *Split Validation*

Setelah menyelesaikan persiapan data melalui proses *preprocessing*, langkah selanjutnya adalah memisahkan data menjadi dua bagian. Bagian pertama berfungsi sebagai data pembelajaran, yang akan digunakan untuk melatih model, sementara bagian kedua akan dijadikan sebagai data pengujian. Data pembelajaran ini digunakan untuk menginstruksikan model, sedangkan data pengujian berperan dalam mengevaluasi sejauh mana model yang diusulkan mampu melakukan prediksi dengan akurat (Diantika et al., 2023)

Setelah selesai pada tahap *preprocessing* ditemukan data sebanyak 1687 data siap digunakan dan 5 parameter dengan 2 kelas data, yaitu kelas tepat dan kelas tidak tepat. Data yang sudah siap digunakan ini kemudian di bagi menjadi 2 bagian, yaitu 50% data digunakan untuk mengajarkan model penelitian untuk mengenali data dan 50% lagi data untuk menguji model penelitian, agar dapat menilai apakah suatu model penelitian sudah dengan akurat mengenali suatu data, dalam penelitian ini didapatkan data pembelajaran sebanyak 1349 data dan data pengujian sebanyak 338 data.

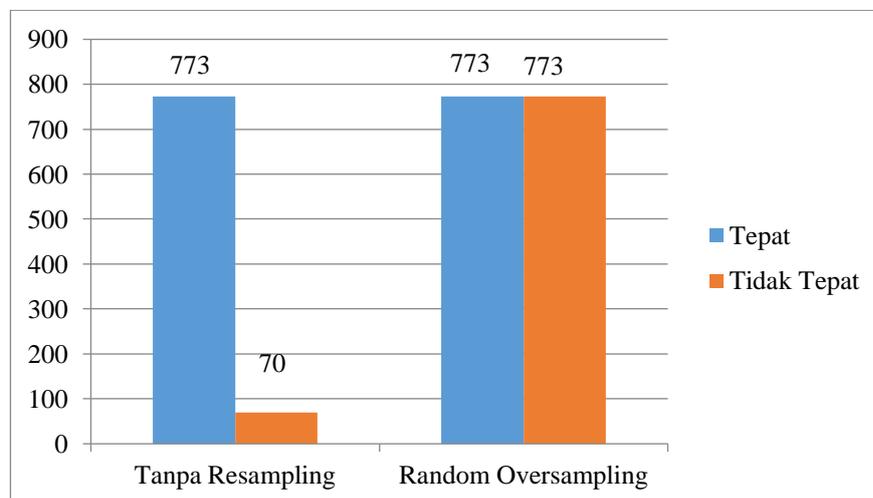


Sumber: Penelitian (2023)
Gambar 2. Split Validation

4. Random Oversampling

Setelah pembagian data selesai didapatkan data pembelajaran sebanyak 843 data dan data pengujian sebanyak 844 data. Pada data pembelajaran terdapat kelas data tepat sebanyak 773 data dan kelas data tidak tepat sebanyak 70 data. Dapat dilihat bahwa data ini berjumlah tidak seimbang antara kelas data tepat dan tidak tepat. Untuk mengatasi hal ini maka perlu dilakukannya *resampling* data. Teknik *resampling* dapat dikatakan bagian dari proses *preprocessing* untuk membuat distribusi data seimbang dalam proses pembelajaran (Bagui & Li, 2021) Ketidakseimbangan kelas data akan menghasilkan efek yang kurang baik bagi hasil prediksi seperti, kelas data minoritas sering disalah prediksi sebagai kelas mayoritas (Chen & Li, 2022) Oleh karena itu perlu diterapkannya teknik *resampling* yang diterapkan adalah teknik *random oversampling* (ROS).

Proses menyeimbangkan data dilakukan dengan metode menaikkan jumlah data dari kelas minoritas ke dalam data pembelajaran secara acak. Proses pemberian data ini dilakukan secara berulang sampai didapatkan jumlah data kelas minoritas *balance* dengan jumlah kelas mayoritas. Penggambaran dari proses *random oversampling* (ROS) bisa dilihat pada gambar 3



Sumber: Penelitian (2023)
Gambar 3. Random Oversampling

Setelah didapatkan kelas data yang seimbang maka, langkah selanjutnya adalah proses pengujian model *random forest* menggunakan data yang telah melalui tahap *resampling* sebelumnya dengan jumlah data masing-masing kelas sebanyak 773 data.

5. Modelling

Setelah kelas data seimbang, langkah selanjutnya adalah menerapkan model *random forest*. Random forest adalah metode data mining dan merupakan hasil pengembangan dari metode *decision tree* (Mao et al., 2020) Cara kerjanya, *random forest* memilih atribut secara acak untuk membuat jumlah pohon K dengan atribut yang berbeda tanpa pemangkasan. Pada *random forest* data pengujian akan diuji pada semua pohon yang dibangun dan kemudian output yang paling sering akan ditugaskan berulang hingga model yang dihasilkan memenuhi kriteria

tertentu, contohnya nilai *loss function* yang cukup kecil. Penelitian ini menggunakan *Parameter random forest* seperti dapat dilihat pada Tabel 1

Tabel 1. *Parameter Random Forest*

<i>Parameter</i>	<i>Nilai</i>
<i>N Estimator</i>	10
<i>Min sample split</i>	2
<i>Num trees</i>	10
<i>Boostrape</i>	<i>True</i>
<i>Random state</i>	10
<i>Seed</i>	10
<i>Criterion</i>	Gini
<i>Min sample leaf</i>	1
<i>Max Features</i>	<i>None</i>

Sumber: Penelitian (2023)

6. *Evaluation*

Langkah terakhir dalam serangkaian tahapan penelitian ini adalah evaluasi. Pada tahap evaluasi ini, hasil pengujian dibandingkan untuk menentukan kinerja model yang lebih baik. Evaluasi kinerja model yang dibangun melibatkan pemeriksaan *confusion matrix*, yang memberikan informasi tentang kebenaran prediksi data terhadap data sebenarnya. Dengan menggunakan *confusion matrix*, kita dapat memperoleh nilai *F1-Score*, Akurasi, dan *Recall*.

HASIL DAN PEMBAHASAN

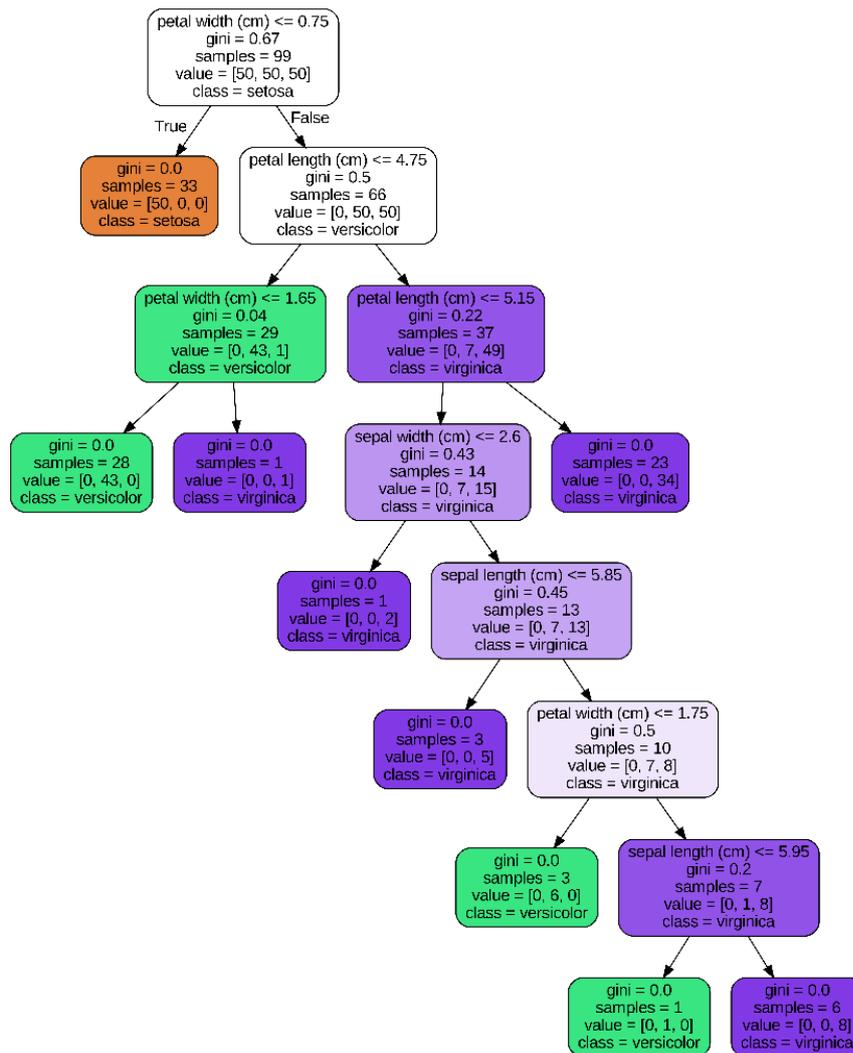
Dataset yang digunakan sebagai objek atau materi dalam penelitian ini adalah data mengenai ketepatan waktu lulus di salah satu universitas yang tidak akan diungkapkan namanya. Pada Tabel 2, dijabarkan deskripsi mengenai *dataset* yang digunakan dalam penelitian ini.

Tabel 2. Data Penelitian

No	Nama	Keterangan
1	Nama Variabel	IP1, IP2, IP3 dan IP4 yaitu Nilai Indeks Prestasi Kumulatif (IPK) mahasiswa
2	Kelas Data	Lulus dan Tidak Lulus
3	Jumlah Data	1687

Sumber: Penelitian (2023)

Model yang diterapkan dalam penelitian ini adalah model algoritma *random forest*, dan pengujian dilaksanakan dengan menggunakan Python dengan desain model sebagai berikut:



Sumber: Penelitian (2023)

Gambar 3. Model Penelitian *Random Forest*

Dengan menerapkan teknik *resampling* yaitu *random Oversampling* (ROS) untuk mengatasi ketidakseimbangan kelas data pada *dataset* yang digunakan dan juga model *random forest* untuk melakukan prediksi ketepatan waktu lulus, peneliti mendapatkan hasil yang diuraikan melalui *matris evaluation*, yang menampilkan bahwa model *random forest* lebih baik dalam melakukan prediksi ketepatan waktu lulus jika dibandingkan dengan model lain.

1. Akurasi *random forest*

Untuk mengetahui bahwa model melakukan prediksi dengan benar, yaitu memprediksi ketepatan waktu lulus sesuai dengan kelas datanya, maka dapat dilihat dari besaran akurasi yang dihasilkan oleh model yang dipakai. Pada Tabel 3 diuraikan hasil pengujian model terhadap data ketepatan waktu lulus menggunakan teknik *resampling random oversampling* (ROS) dan mencoba beberapa teknik *split data*. Nilai akurasi tertinggi didapatkan dari model algoritma *random forest* dengan teknik *split data* sebesar 50% untuk *data pengujian* dan 50% untuk *data pembelajaran* yaitu sebesar 90,04%.

Tabel 3. Nilai Akurasi Hasil Pengujian Model

		Akurasi		
		Algoritma		
Data training	Data Testing	Random Forest	Gradient Boosting	LGBM
50	50	90.04%	73.70%	89.80%
60	40	88.30%	70.80%	88.90%
70	30	89.90%	69.60%	88.95%
80	20	89.05%	68.9%	88.8%

Sumber: Penelitian (2023)

2. Recall Random Forest

Untuk mengetahui bahwa model melakukan prediksi dengan *actual*, yaitu memprediksi ketepatan waktu lulus sesuai dengan kelas datanya dibandingkan dengan kelas ketepatan waktu lulusnya tersebut, maka dapat dilihat dari besaran *recall* yang dihasilkan oleh model yang dipakai. Pada Tabel 4 diuraikan hasil pengujian model terhadap data ketepatan waktu lulus menggunakan teknik *resampling random oversampling* (ROS) dan mencoba beberapa teknik *split data*. Nilai akurasi tertinggi didapatkan dari model algoritma *random forest* dengan teknik *split data* sebesar 50% untuk *data pengujian* dan 50% untuk data pembelajaran yaitu sebesar 90,04%.

Tabel 4. Nilai Recall Hasil Pengujian Model

		Recall		
Data training	Data Testing	Algoritma		
		Random Forest	Gradient Boosting	LGBM
50	50	90.04%	73.70%	89.80%
60	40	88.30%	70.80%	88.90%
70	30	89.90%	69.60%	88.95%
80	20	89.05%	68.9%	88.8%

Sumber: Penelitian (2023)

3. Precision Random Forest

Untuk memahami perbandingan antara prediksi *true positif* dan jumlah data yang diprediksi positif, kita dapat mengobservasi besaran nilai presisi yang dihasilkan oleh model yang digunakan. Pada Tabel 4, hasil pengujian model terhadap data ketepatan waktu lulus menggunakan teknik *resampling random oversampling* (ROS) dan beberapa teknik *split data*. Model algoritma *random forest* dengan teknik *split data* sebesar 50% untuk data pengujian dan 50% untuk data pembelajaran menghasilkan nilai akurasi tertinggi, yakni sebesar 90,04%.

Tabel 5. Nilai Presisi Hasil Pengujian Model

		Presisi		
Data training	Data Testing	Algoritma		
		Random Forest	Gradient Boosting	LGBM
50	50	87.05%	87.90%	86.70%
60	40	86.40%	86.20%	85.10%
70	30	85.10%	88.80%	85.10%
80	20	89.05%	68.9%	88.8%

Sumber: Penelitian (2023)

KESIMPULAN

Penelitian yang dilakukan untuk memprediksi ketepatan waktu lulus menggunakan model *random forest* memberikan hasil bahwa prediksi terbaik didapatkan dengan menerapkan parameter *criterion gini* dengan *n_estimators* sebanyak 10, *bootstrap true*, *num trees* 10, *seed* sebanyak 10, *min sample leaf* sebanyak 2 dan *max features* adalah *none*. Serta dengan menerapkan *Split validation* atau pembagian antara data pembelajaran sebesar 50% untuk data *pengujian* dan 50%. Selain parameter dan pembagian data, hal yang perlu dilakukan lagi adalah menyeimbangkan data yaitu menggunakan teknik *random oversampling* (ROS). Untuk mengevaluasi model yang disusun, Penulis menggunakan *metrics evaluation* seperti akurasi, *recall*, dan *precision*.

Hasil dari penelitian menunjukkan bahwa model yang diusulkan dapat dengan baik melakukan prediksi dibandingkan dengan model lainnya, yaitu dengan hasil nilai uji akurasi sebesar 90,04%, *recall* 90,04%, dan *precision* sebesar 87,05%. Dengan nilai ini maka tingkat keberhasilan untuk memprediksi ketepatan waktu lulus mahasiswa akan semakin akurat. Untuk revisi hasil yang lebih akurat, diharapkan pada riset selanjutnya terdapat hal-hal yang perlu dicoba khususnya yang belum ada pada riset ini, seperti akumulasi *dataset* yang dijadikan objek riset sehingga sebaran data bisa *balance* serta *representative*

REFERENSI

- Anwarudin, A., Andriyani, W., DP, B. P., & Kristomo, D. (2022). The Prediction on the Students' Graduation Timeliness Using Naive Bayes Classification and K-Nearest Neighbor. *Journal of Intelligent Software Systems*, 1(1), 75. <https://doi.org/10.26798/jiss.v1i1.597>
- Bagui, S., & Li, K. (2021). Resampling imbalanced data for network intrusion detection datasets. *Journal of Big Data*, 8(1). <https://doi.org/10.1186/s40537-020-00390-x>
- Chen, W., & Li, K. (2022). Self-supervised Learning for Medical Image Classification Using Imbalanced Training Data. *Communications in Computer and Information Science*, 1590 CCIS, 242–252. https://doi.org/10.1007/978-981-19-4109-2_23
- Diantika, S., Nalatissifa, H., Supriyadi, R., Maulidah, N., & Fauzi, A. (2023). Implementasi Multi-Class Gradient Boosting Untuk Mengklasifikasikan Jenis Hewan Pada Kebun Binatang. *17(1)*, 32–40.
- Faizah, T., & Jananto, A. (2021). Perbandingan Algoritma C4.5 Dan Id3 Untuk Prediksi. *Jatiti*, 8(2), 980–990.
- Hermawan, L., & Bellanar Ismiati, M. (2020). Pembelajaran Text Preprocessing berbasis Simulator Untuk Mata Kuliah Information Retrieval. *Jurnal Transformatika*, 17(2), 188. <https://doi.org/10.26623/transformatika.v17i2.1705>
- Heryana, N., Ghassani, F. Z., Usrahmawan, M. A., Juardi, D., & Solehudin, A. (2020). Kajian Model Prediksi Ketepatan Kelulusan Mahasiswa: Sebuah Literatur Review. *Syntax: Jurnal Informatika*, 9(1), 54. <https://doi.org/10.35706/syji.v9i1.3388>
- Jananto, A. (2021). Algoritma Naive Bayes Untuk Mencari Perkiraan Waktu Studi Mahasiswa. *INTECOMS: Journal of Information Technology and Computer Science*, 4(1), 20–29. <https://doi.org/10.31539/intecom.v4i1.2219>
- Mao, Y., He, Y., Liu, L., & Chen, X. (2020). Disease Classification Based on Eye Movement Features With Decision Tree and Random Forest. *Frontiers in Neuroscience*, 14(August), 1–11. <https://doi.org/10.3389/fnins.2020.00798>
- Mulyadi, C., & Sugiarto, L. (2021). Penggunaan algoritma naïve bayes untuk prediksi ketepatan waktu lulus mahasiswa diploma 3 STMIK Cipta Darma Surakarta. *Teknomatika*, 11(01), 21–30. <http://ojs.palcomtech.ac.id/index.php/teknomatika/article/view/512>
- Nasir, H. D., & Saharuna, Z. (2020). Prediksi Tingkat Polusi Udara Dengan Data Mining. *Prosiding Seminar Nasional Teknik Elektro Dan Informatika (SNTEI)*, 90–95.
- S.A Amosea. (2021). Prediksi Ketepatan Waktu Lulus Mahasiswa Teknik Informatika Menggunakan Algoritma Naive Bayes. *Doctoral Dissertation, Institut Teknologi Sepuluh Nopember*.
- Samuel, Y. T., Hutapea, J. J., & Jonathan, B. (2019). Predicting the Timeliness of Student Graduation Using Decision Tree C4.5 Algorithm in Universitas Advent Indonesia. *International Conference on Information, Communication Technology and System (ICTS)*. <https://doi.org/10.1109/ICTS.2019.8850948>
- Satria, F., Zamhariri, Z., & Syaripudin, M. A. (2020). Prediksi Ketepatan Waktu Lulus Mahasiswa Menggunakan Algoritma C4.5 Pada Fakultas Dakwah Dan Ilmu Komunikasi UIN Raden Intan Lampung. *Jurnal Ilmiah Matrik*, 22(1), 28–35. <https://doi.org/10.33557/jurnalatrik.v22i1.836>
- Setiaji, G. G., Khoirudin, K., & Vydia, V. (2019). Komparasi Metode Clustering K-Means Dan Fuzzy C-Means Untuk Mempredesi Ketepatan Waktu Lulus. *Jurnal Pengembangan Rekayasa Dan Teknologi*, 15(1), 38. <https://doi.org/10.26623/jprt.v15i1.1488>