

PERBANDINGAN KINERJA MACHINE LEARNING DALAM KLASIFIKASI MINAT MBKM MAHASISWA UNRI

Gustriza Erda¹, Siti Farika Sari², Syaftiani Dwi Astuti³, Lia Thearas⁴, Anne Mudya Yolanda⁵

¹²³⁴⁵Program Studi Statistika, Fakultas Matematika dan Ilmu Pengetahuan, Universitas Riau

Kampus Bina Widya KM. 12,5, Simpang Baru, Kec. Tampan, Kota Pekanbaru, Riau 28293

¹gustrizaerda@lecturer.unri.ac.id, ²siti.farika1358@student.unri.ac.id,
³syaftiani.dwi2192@student.unri.ac.id, ⁴lia.thearas2194@student.unri.ac.id,
⁵annemudyayolanda@lecturer.unri.ac.id

Abstract

Merdeka Belajar Kampus Merdeka (MBKM) program is a program to increase college students' potency and creativity to face hurdles in work life. However, Universitas Riau (UNRI) students' interest in participating this program is still low, so further research is needed to increase UNRI students' interest participation in MBKM. This research aims to compare the performances of some machine learning algorithms, such as Random Forest, Logistic Regression, Decision Tree, and Gradient Boost Machine (GBM), yet identifying important factor that affects UNRI students' interest towards MBKM. Data were collected through a questionnaire distributed to 396 active undergraduate UNRI students from the 2019-2023 cohort that was chosen by stratified random sampling technique. With 12 independent variables used, it was found that GBM model had the best performance with the test accuracy of 91,25% and recall of 98,63%, upperhand other models. The most influential factors on students' interest are self confident and family support. These findings showed that those factors can be a focus in the evaluation and MBKM program improvement in the future.

Keywords : accuracy, decision tree, gradient boost machine, logistic regression, MBKM, random forest.

Abstrak

Program Merdeka Belajar Kampus Merdeka (MBKM) merupakan program yang bertujuan meningkatkan potensi dan keterampilan mahasiswa untuk menghadapi tantangan dunia kerja. Namun, minat mahasiswa Universitas Riau (UNRI) dalam mengikuti program ini masih rendah sehingga perlu penelitian lebih lanjut untuk meningkatkan minat mahasiswa UNRI dalam mengikuti MBKM. Penelitian ini bertujuan membandingkan kinerja beberapa algoritma machine learning seperti Random Forest, Logistic Regression, Decision Tree, dan Gradient Boost Machine (GBM), serta mengidentifikasi faktor penting yang memengaruhi minat mahasiswa UNRI terhadap MBKM. Data dikumpulkan melalui kuisioner yang disebarkan kepada 396 mahasiswa aktif S1-UNRI angkatan 2019-2023 yang dipilih menggunakan teknik stratified random sampling. Dengan 12 variabel independen yang digunakan, diperoleh bahwa model GBM memiliki kinerja terbaik dengan akurasi uji 91,25% dan recall 98,63%, mengungguli model lainnya. Faktor yang paling berpengaruh terhadap minat mahasiswa adalah keyakinan diri dan dukungan keluarga. Temuan ini menunjukkan bahwa faktor-faktor tersebut dapat menjadi fokus dalam evaluasi dan pengembangan program MBKM di masa depan.

Kata kunci : akurasi, gradient boost machine, MBKM, pohon keputusan, random forest, regresi logistik.

1. PENDAHULUAN

Dalam era globalisasi dan persaingan yang semakin intens, pengembangan diri melalui pendidikan dan peningkatan keterampilan menjadi sangat penting bagi individu. Untuk meningkatkan kualitas pendidikan tinggi dan mempersiapkan mahasiswa menghadapi tantangan dunia kerja, Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi (Kemendikbud Ristek) meluncurkan program Kampus Merdeka pada akhir Januari 2020. Dengan adanya program ini, diharapkan dapat memiliki manfaat dalam mendorong juga mengembangkan kreativitas dan inovasi para mahasiswa di bermacam-macam bidang [1]. Program Merdeka Belajar Kampus Merdeka (MBKM) dirancang untuk menyokong perguruan tinggi dalam mempersiapkan mahasiswa dalam menjadikannya lulusan kompeten dalam bidang ilmu dan teknologi, memiliki karakter yang kuat, dan siap dalam menghadapi berbagai tantangan pada dunia pekerjaan [2].

Program MBKM merupakan inisiatif dari Kementerian Pendidikan, Kebudayaan, Riset dan Teknologi (Kemendikbud Ristek) untuk menciptakan SDM berkualitas dalam aspek keterampilan teknis (*hard skill*) dan keterampilan lunak (*soft skill*), serta membentuk karakter yang kompetitif. Program ini diharapkan dapat menstimulasi para mahasiswa maupun dosen untuk mendapatkan pengalaman baru, yang dapat bermanfaat dalam memperbanyak wawasan dan jaringan, serta keunggulan karakter [2].

Program MBKM adalah revolusi pendidikan yang berlandaskan perkembangan industri 4.0. Kebijakan MBKM diatur dalam Peraturan Menteri Pendidikan dan Kebudayaan Nomor 3 Tahun 2020 mengenai Standar Nasional Perguruan Tinggi, terutama pada Pasal 15 hingga 18. MBKM memiliki tujuan untuk dapat membantu mahasiswa dalam mendapatkan pengalaman belajar dari bermacam-macam kompetensi, baik di lingkungan program studi maupun di luar lingkungan perkuliahan [3]. Kebijakan MBKM mencakup beberapa aspek utama, termasuk pembukaan program studi baru, pembangunan, perubahan dan pencabutan izin Perguruan Tinggi Negeri (PTN), serta pembangunan, perubahan, dan pencabutan izin Perguruan Tinggi Swasta (PTS). Beberapa komponen yang terdapat pada MBKM meliputi magang, penelitian, pertukaran pelajar, asistensi mengajar di satuan pendidikan, studi atau proyek independen, kegiatan wirausaha, proyek kemanusiaan, dan program membangun desa [4].

Minat mahasiswa dalam mengikuti program MBKM merupakan salah satu indikator penting

dalam mengevaluasi keberhasilan program-program oleh pendidikan tinggi. Mengklasifikasikan hasil minat mahasiswa dalam mengikuti MBKM, serta memahami faktor-faktor yang mempengaruhinya dapat membantu lembaga pendidikan dan pembuat kebijakan dalam meningkatkan partisipasi dan efektivitas program. Langkah ini bukan sekedar akan meningkatkan partisipasi mahasiswa, tetapi juga akan memastikan bahwa program MBKM memberikan manfaat yang maksimal bagi perkembangan akademik dan profesionalitas mahasiswa [18].

Penelitian terdahulu yang berjudul "Respon Mahasiswa dalam Implementasi Kebijakan Program Merdeka Belajar Kampus Merdeka pada Program Studi Pendidikan Pancasila dan Kewarganegaraan", menunjukkan bahwa mahasiswa dari program studi tersebut memberikan tanggapan positif terhadap program MBKM. Para mahasiswa paham bahwa inti dari program MBKM adalah untuk meningkatkan kemampuan akademis serta memberikan tantangan dan kesempatan berharga yang dapat mengembangkan kreativitas, potensi, dan kebutuhannya pribadi. Program ini juga memberikan kebebasan bagi mahasiswa untuk memperoleh informasi melalui berbagai faktor dan situasi nyata, termasuk tuntutan yang dinamis, tantangan nyata, interaksi sosial, keterampilan, pengelolaan diri, kolaborasi, persyaratan kinerja, tujuan, dan pencapaian [5].

Penelitian ini dilakukan dengan memprediksi minat mahasiswa dalam mengikuti program MBKM menggunakan proses *statistical machine learning* dengan metode klasifikasi. Klasifikasi adalah proses pengelompokan data ke dalam kategori-kategori yang terkait antara satu sama lain berdasarkan masalah yang ada, dengan tujuan utama yaitu untuk mendapatkan model yang nantinya dapat memberikan gambaran dan menunjukkan perbedaan kelompok data, hingga mampu memprediksi kelompok data tersebut [6]. Beberapa klasifikasi yang digunakan dengan memanfaatkan *machine learning*, yaitu Regresi Logistik, *Decision Tree*, *Random Forest*, dan *Gradient Boosting Machine* (GBM), yang memiliki kemampuan untuk mengelompokkan data dan menghasilkan prediksi berdasarkan pola-pola yang ada. Pendekatan *machine learning* menjadi penting karena memungkinkan peneliti untuk menganalisis data secara efisien dan menghasilkan prediksi yang akurat dalam menganalisis terutama terkait minat mahasiswa dalam mengikuti program MBKM. Model-model tersebut memanfaatkan algoritma-algoritma

kompleks yang mampu memproses data dalam skala yang besar dan mampu mengenali pola-pola kompleks yang mungkin tidak terlihat secara langsung oleh manusia.

Regresi logistik merupakan jenis analisis regresi yang bertujuan untuk menunjukkan hubungan antara variabel terikat (target) dengan variabel bebas (fitur). Rentang hasilnya berkisar dalam rentang 0 hingga 1, yang mampu memberikan gambaran kebenaran atau ketidakbenaran, dan ukurannya dapat beragam. Variabel bebas (fitur) dalam regresi logistik biasanya bersifat kategorikal [7]. Dalam penelitian mengenai metode regresi logistik biner dalam menilai ketepatan pengklasifikasian jenis musik K-Pop dan Western di aplikasi Spotify, didapatkan nilai akurasi sebesar 79,04% pada data *training* dan 79,69% pada data *testing* dengan proporsi perbandingan pembagian datanya sebesar 80:20 [8].

Decision Tree adalah metode klasifikasi diagram alir yang memiliki bentuk pohon di mana pada tiap cabang pohonnya menyatakan pengujian terhadap suatu atribut dan menyatakan *output* dari pengujian yang diteliti. *Decision Tree* memiliki tiga bagian, yaitu node teratas yang disebut *root node* (node akar), cabang-cabang dari atribut yang disebut *internal node* yang memiliki satu *input* dan minimal dua *output*, dan hasil akhir dari tiap label kelas atribut disebut *leaf node* [9]. Penelitian terkait pengembangan sistem klasifikasi tipe kepribadian siswa secara psikologis menggunakan algoritma *decision tree* C.45, menunjukkan bahwa algoritma tersebut mencapai nilai *precision* yaitu 90%, *recall* yang mencapai 85% dan *accuracy* dengan nilai 88%. Hal tersebut memberikan bukti bahwa algoritma *decision tree* C.45 mampu mengklasifikasikan tipe kepribadian dengan baik [10].

Metode *Random Forest* (RF) merupakan pengembangan dari metode *Classification and Regression Tree* (CART) yang menggunakan metode *bootstrap aggregating (bagging)* dan *random feature selection* (pemilihan fitur secara acak). *Random Forest* adalah model algoritma untuk mengklasifikasikan data dengan membentuk pohon keputusan yang banyak. Hasil dari setiap pohon kemudian digabungkan, dan hasil yang paling sering muncul dipilih sebagai hasil akhir. Seperti halnya dengan metode *decision tree*, metode ini juga terdiri dari *root node* (akar), *internal node* (cabang), dan *leaf node* (simpul akhir) [6]. Dalam penelitian mengenai implementasi algoritma klasifikasi *random forest* untuk penilaian kelayakan kredit, model algoritma tersebut memberikan performa dengan nilai akurasi uji sebesar 78,60% dan nilai *area under the curve* (AUC) mencapai 0,907. Dari nilai yang diperoleh

tersebut, model algoritma *random forest* dikategorikan sebagai klasifikasi yang sangat baik [11].

Gradient Boost Machine (GBM) adalah algoritma *ensemble* yang berfungsi dengan cara mengurangi kesalahan model secara bertahap-tahap. GBM merupakan model pembelajaran mesin untuk permasalahan klasifikasi yang menghasilkan prediksi dalam bentuk pohon keputusan. Algoritma ini dioptimalkan dengan cara menggunakan kriteria *differentiable loss function*. Dalam upaya untuk membangun model prediktif, GBM menggunakan teknik *back-fitting* dan regresi nonparametrik. Alih-alih membangun model tunggal, algoritma GBM dimulai dengan model awal dan secara terus-menerus menyesuaikan model baru melalui mengurangi fungsi kerugian untuk menciptakan model yang paling sesuai [12]. Dalam penelitian terkait *gradient boosting machine* pada klasifikasi kelulusan mahasiswa, diperoleh bahwa model *gradient boost machine* (GBM) dalam mengklasifikasikan kelulusan mahasiswa Program Studi Statistika Universitas Tanjungpura menghasilkan nilai akurasi sebesar 71,09%. Nilai akurasi model GBM ini didapat lebih tinggi dibandingkan dengan model CART yang memiliki akurasi sebesar 67,97%. Hal ini menunjukkan bahwa model GBM dapat meningkatkan kinerja dari model CART [12].

Dalam penelitian yang terdahulu, yaitu mengenai prediksi kemampuan pembayaran klien *home credit* dengan model *random forest*, *decision tree*, dan *logistic regression* menunjukkan bahwa *random forest* memiliki akurasi uji sebesar 0.9967 dan juga *recall* sebesar 1.00 dan skor ROC sebesar 0.9967. Pada *decision tree*, tingkat keakurasian bernilai 0.9087, *recall* yang mencapai 0.91, dan skor ROC sebesar 0.9087. Sedangkan itu, model *logistic regression* memperoleh tingkat keakurasian sebesar 0.6745, *recall* sebesar 0.67, dengan skor ROC yang berkisaran 0.6745. Berdasarkan hasil keakurasian uji ini, metode *random forest* terbukti merupakan model dengan nilai keakuratan yang paling tinggi daripada model-model lainnya [7].

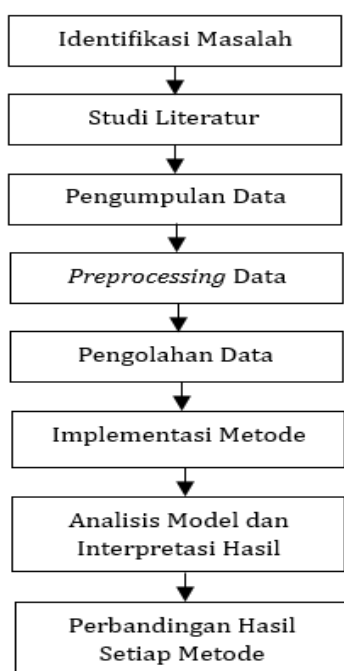
Penelitian ini berbeda dari penelitian terdahulu, di mana penelitian ini berfokus pada minat mahasiswa dalam mengikuti program MBKM di Universitas Riau yang menggunakan berbagai model pembelajaran mesin, yaitu dengan empat model algoritma yang berbeda. Selain pengklasifikasian, peneliti juga melakukan pengidentifikasian faktor-faktor yang memiliki pengaruh secara signifikan terhadap minat

mahasiswa, seperti dukungan keluarga dan persepsi manfaat. Berbeda dari penelitian sebelumnya yang sering berfokus pada satu model, penelitian ini menggabungkan algoritma *machine learning*, yaitu regresi logistik, *decision tree*, *random forest*, dan *gradient boosting machine* (GBM) dalam mengklasifikasikan, serta membandingkan akurasi dari setiap model, untuk memberikan pendekatan yang komprehensif dan terintegrasi.

Oleh sebab itu, penelitian ini memiliki tujuan yaitu untuk mengklasifikasikan minat mahasiswa dalam mengikuti program MBKM di Universitas Riau menggunakan berbagai algoritma model pembelajaran mesin, mengidentifikasi faktor-faktor yang mempengaruhi minat mahasiswa dalam mengikuti program MBKM, serta mengevaluasi ketepatan dari model-model algoritma tersebut. Manfaat penelitian ini meliputi peningkatan dan pemahaman lembaga pendidikan dan pembuat kebijakan tentang faktor-faktor yang dapat mendorong partisipasi mahasiswa dalam program MBKM, serta peningkatan efektivitas program melalui penerapan model prediksi yang akurat dan terintegrasi.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian



Gambar 1. Alur Penelitian

Penelitian ini diawali dengan mengidentifikasi masalah yang melatarbelakangi adanya penelitian ini, yakni identifikasi faktor pemengaruh minat

mahasiswa Universitas Riau mengikuti program MBKM. Selanjutnya, dilakukan studi literatur terhadap penelitian terdahulu yang pernah membahas topik serupa. Literatur yang dikaji dapat berupa jurnal, buku, dan sumber-sumber lainnya. Hal tersebut bertujuan untuk memperoleh perspektif baru terhadap masalah yang diteliti, mendapatkan argumen pendukung, serta menemukan kesenjangan penelitian.

Setelah mendapatkan referensi yang cukup, dilakukan pengumpulan data sekunder yang diperoleh dari civitas akademika Universitas Riau. Setelah mengumpulkan data yang diperlukan, kemudian dilakukan *preprocessing* terhadap data tersebut. *Preprocessing* dapat meliputi penanganan data hilang dan pengkodean untuk variabel kategorik.

Setelah data siap untuk digunakan, selanjutnya akan dilakukan pengolahan terhadap data. Pada tahap ini, data akan dibagi menjadi 80% data *training* dan 20% data *testing*. Selanjutnya, dengan algoritma model regresi logistik, *decision tree*, *random forest*, dan juga *gradient boosting machine* akan diimplementasikan untuk mengetahui variabel-variabel yang secara signifikan memiliki pengaruh terhadap minat mahasiswa dalam mengikuti program MBKM. Kemudian, model yang terbentuk akan selanjutnya dianalisis dan diinterpretasikan untuk kemudian akan saling dibandingkan terhadap metode lainnya.

2.2 Data dan Sumber Data

Pada penelitian ini, data yang digunakan merupakan data primer, yaitu data yang diperoleh secara langsung oleh peneliti dari sumber aslinya dengan tujuan spesifik untuk menjawab pertanyaan penelitian yang sedang dikerjakan. Data tersebut terdiri dari 11 variabel independen dan 1 variabel dependen. Variabel independennya adalah JK (jenis kelamin), angkatan, fakultas, pengetahuan MBKM, keyakinan diri, merasakan manfaat program, dukungan keluarga, dukungan teman, dukungan dosen, kemampuan diri, dan fasilitas. Dan variabel dependennya adalah Minat MBKM. Banyaknya amatan dalam data adalah 4752 amatan.

Data yang dianalisis dalam penelitian ini adalah data terkait program MBKM Universitas Riau pada tahun 2023. Teknik pengumpulan data dilakukan melalui studi pustaka, dengan mengumpulkan informasi dari literatur yang berkaitan dengan objek studi. Penelitian ini mencakup penelaahan dari berbagai sumber

seperti buku, jurnal ilmiah, artikel, laporan penelitian, dan dokumen relevan lainnya untuk mendapatkan pemahaman komprehensif terkait pembahasan yang diteliti.

2.3 Preprocessing

Preprocessing merupakan tahap pemilihan data sehingga dapat digunakan dengan cara yang lebih terstruktur dengan menghilangkan atribut yang tidak dibutuhkan, membuat data lebih sistematis, dan menghilangkan *noise*.

Proses *preprocessing* melibatkan beberapa langkah penting, seperti pembersihan data di mana mengidentifikasi data yang hilang, lalu mengimputasi data hilang (jika ada) dengan *mean* atau *median*, atau juga dapat dihapus. Selain itu, terdapat langkah *encoding* yaitu di mana mengubah variabel menjadi kategorik, dan langkah-langkah penting lain yang bertujuan untuk meningkatkan kualitas data [13].

Setelah melalui tahap *preprocessing*, data mentah telah diolah menjadi dataset baru yang terdiri dari 396 baris data dan sudah dalam bentuk *encoding* yang akan digunakan dalam proses pemodelan. Variabel prediktor atau variabel *independent* yang akan digunakan dalam proses pengolahan data meliputi JK, Angkatan, Fakultas, Pengetahuan MBKM, Keyakinan Diri, Merasakan Manfaat Program, Dukungan Keluarga, Dukungan Teman, Dukungan Dosen, Kemampuan Diri, Fasilitas, serta Minat MBKM sebagai variabel respon atau variabel *dependent*.

Dataset yang telah melalui tahap *preprocessing* akan menjadi landasan dalam pembangunan model klasifikasi. Memilih dan menerapkan langkah-langkah *preprocessing* yang sesuai sangat penting untuk menjamin bahwa model yang dibangun mampu beroperasi secara optimal, memberikan prediksi yang akurat sehingga dapat dipercaya, dan menyampaikan informasi-informasi berharga dari pembahasan yang diteliti.

2.4 Analisis Data

Analisis data dalam penelitian ini yaitu menggunakan empat model sebagai berikut: *logistic regression*, *decision tree*, *random forest* dan *gradient boost machine* (GBM) sebagai pendekatan untuk mengklasifikasikan minat mahasiswa terhadap program MBKM.

Metode *logistic regression* merupakan sebuah teknik analisis statistik yang digunakan untuk menjelaskan hubungan antara variabel terikat (*dependent variable*) yang terdiri dari dua kategori atau lebih dengan satu atau beberapa variabel bebas (*independent variable*) dengan skala kategori [14]. Metode ini melibatkan langkah-langkah analisis yang

berawal dari persiapan data (*import data* dan *preprocessing data*), membangun model dengan memilih variabel prediktor yang relevan, dan evaluasi model.

Metode *decision tree* atau juga yang sering disebut pohon keputusan, adalah sebuah representasi sederhana yang mudah dipahami dari teknik klasifikasi untuk kategori yang terbatas, di mana simpul-simpul internal dan simpul akar diberi label berdasarkan atribut, sementara cabang-cabangnya menunjukkan nilai-nilai atribut yang mungkin, dan simpul daun menandai berbagai kelas [15]. Metode *decision tree* melibatkan langkah-langkah analisis yang hampir sama dengan metode regresi logistik hanya saja pada pemodelan *decision tree* akan membangun pohon keputusan untuk klasifikasi.

Metode *random forest* merupakan perkembangan dari metode *decision tree* yang memanfaatkan beberapa pohon keputusan, di mana setiap pohon keputusan dilatih menggunakan sampel data individu dan setiap atribut dibagi di dalam pohon dengan memilih subset atribut secara acak [16]. Metode ini melibatkan langkah-langkah analisis yang hampir sama dengan dua metode sebelumnya, yaitu regresi logistik dan *decision tree*. Namun, pada metode *random forest* menggunakan *ensemble* dari pohon-pohon keputusan.

Metode *gradient boost machine* merupakan model yang termasuk dalam kategori algoritma *ensemble* yang mampu meningkatkan akurasi prediksi. Algoritma ini dinilai sebagai salah satu metode klasifikasi yang handal, yang mampu menangani berbagai jenis data, baik numerik maupun nominal [17]. Algoritma *gradient boost machine* ini bekerja dengan cara menyesuaikan model terhadap data dan kemudian secara berurutan membangun model berikutnya untuk mengurangi kesalahan (*residual*). Penelitian yang menggunakan metode ini mencatat ketiadaan kesalahan. Tujuan utama dari digunakannya metode ini yaitu untuk mengevaluasi nilai keakuratannya.

3. HASIL DAN PEMBAHASAN

Penelitian ini menggunakan empat algoritma model *machine learning* yang berbeda untuk mengklasifikasikan Minat MBKM. Berikut adalah hasil yang diperoleh dari masing-masing algoritma tersebut:

3.1 Regresi Logistik

Tujuan utama dari digunakannya metode ini ialah untuk mengevaluasi hasil tingkat akurasi modelnya. Hasilnya dapat dilihat yang terdapat pada Tabel 2. Namun nilai akurasi didapat dari tabel *confusion matrix*.

Tabel 1. merupakan tabel *confusion matrix* untuk algoritma model regresi logistik, dimana diperoleh bahwa 1 data diklasifikasikan “tidak berminat” yang dinilai sama dengan data yang sebenarnya, 6 data diperkirakan “berminat” akan tetapi sebenarnya “tidak berminat”. Kemudian, 71 data diklasifikasikan sebagai “berminat” diperkirakan sesuai dengan data riil, dan 2 data diperkirakan “tidak berminat” yang ternyata “berminat”.

TABEL I. *CONFUSION MATRIX* MODEL REGRESI LOGISTIK

	Pred. Tidak Berminat	Pred. Berminat	Presisi Kelas
<i>Actual.</i> Tidak Berminat	1	6	33%
<i>Actual.</i> Berminat	2	71	92%

Dari tabel di atas, dapat dihitung nilai akurasi dan *recall*, dengan rumusnya masing-masing.

Akurasi adalah rasio prediksi yang benar terhadap total prediksi. Dengan rumus:

$$\begin{aligned}
 \text{Akurasi} &= \frac{\text{Prediksi Benar}}{\text{Total Prediksi}} = \frac{TP + TN}{TP + TN + FP + FN} \\
 &= \frac{71+1}{71+1+6+2} \\
 &= \frac{72}{80} = 0,9
 \end{aligned}$$

Recall adalah ukuran yang mengukur seberapa baik model menangkap semua prediksi benar dari kelas Berminat. Rumusnya adalah:

$$\text{Recall} = \frac{TP}{TP + FN} = \frac{71}{71 + 2} = \frac{71}{73} \approx 0,9726$$

TABEL II. HASIL EVALUASI MODEL REGRESI LOGISTIK

Algoritma	Accuracy	Recall
Logistic Regression	90%	97,26%

Diperoleh bahwa akurasi uji dari model regresi logistik adalah 0.90 atau 90%, dengan *recall* mencapai 0.9726 atau 97,26%. Dengan tingkat

akurasi uji tersebut, menunjukkan bahwa model regresi logistik mampu mengklasifikasikan minat mahasiswa dalam mengikuti program MBKM dengan sangat baik.

Untuk menilai apakah variabel bebas secara bersama-sama (simultan) mampu mempengaruhi variabel terikat, dapat dilihat melalui nilai LLR p-value pada Tabel 3.

TABEL III. *LOGIT REGRESSION RESULTS*

Logit Regression Results			
Dep. Variable:	MinatMBKM	No. Observations:	316
Model:	Logit	Df Residuals:	304
Method:	MLE	Df Model:	11
Date:	Sun, 16 Jun 2024	Pseudo R-squ.:	0.4141
Time:	03:04:05	Log-Likelihood:	-52.636
converged:	True	LL-Null:	-89.839
Covariance Type:	nonrobust	LLR p-value:	1.763e-11

Nilai LLR P-value sebesar 1.763e-11 atau 1.763×10^{-11} . Jika nilai p-value (1.763×10^{-11}) < alpha (0.05) maka, dikatakan tolak hipotesis nol (H_0) yang menyatakan bahwa semua variabel *independent* sama dengan nol. Dapat disimpulkan pada taraf signifikansi 5% diperoleh cukup bukti untuk menyatakan bahwa setidaknya satu variabel bebas memiliki pengaruh signifikan secara bersama-sama terhadap variabel terikat (Minat MBKM).

Untuk melihat variabel-variabel yang memiliki pengaruh secara signifikan terhadap variabel *dependent*, digunakanlah Uji Wald yaitu uji signifikansi individual dari setiap koefisien dalam regresi logistik. Hasilnya dapat dilihat yang terdapat pada Tabel 4.

TABEL IV. VARIABEL BERPENGARUH

Variabel Independent	p> z
JK	0.834
Angkatan	0.267
Fakultas	0.065
Pengetahuan MBKM	0.712
Keyakinan Diri	0.004
MerasakanManfaatProgram	0.506
Dukungan Keluarga	0.067
Dukungan Teman	0.026
Dukungan Dosen	0.888
Kemampuan Diri	0.052
Fasilitas	0.785

Variabel dikatakan berpengaruh secara signifikan apabila nilai $p > |z| < \alpha$ (0.05). Dari analisis pada Tabel 4., dapat diambil kesimpulan bahwa di antara variabel-variabel yang diteliti, hanya keyakinan diri dan Dukungan teman yang secara signifikan mempengaruhi minat mahasiswa terhadap program MBKM. Variabel-variabel lain, meskipun mungkin memiliki beberapa dampak, tidak menunjukkan pengaruh yang signifikan secara statistik dalam konteks analisis ini. Dengan demikian, upaya untuk meningkatkan partisipasi mahasiswa dalam program MBKM sebaiknya difokuskan pada peningkatan keyakinan diri mereka dan memperkuat dukungan dari teman-teman sebaya.

3.2 Decision Tree

Tabel 5. merupakan tabel *confusion matrix* untuk algoritma model *decision tree*, dimana diperoleh bahwa 1 data diklasifikasikan “tidak berminat” yang dinilai sama dengan data yang sebenarnya, 6 data diperkirakan “berminat” tetapi sebenarnya “tidak berminat”. Kemudian, 72 data diklasifikasikan sebagai “berminat” diperkirakan sesuai dengan data riil, dan 2 data diperkirakan “tidak berminat” yang ternyata “berminat”.

TABEL V. *CONFUSION MATRIX* MODEL *DECISION TREE*

	Pred. Tidak Berminat	Pred. Berminat	Presisi Kelas
<i>Actual.</i> Tidak Berminat	1	6	33%
<i>Actual.</i> Berminat	2	71	92%

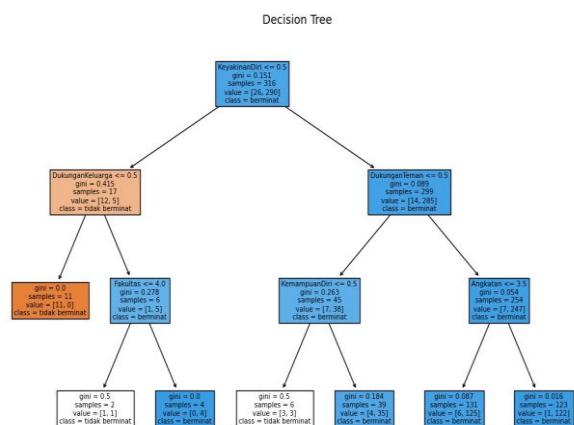
Nilai *accuracy* dan *recall* untuk model *decision tree* dapat dilihat pada Tabel 6.

TABEL VI. HASIL EVALUASI MODEL *DECISION TREE*

Algoritma	Accuracy	Recall
<i>Decision Tree</i>	90%	97,26%

Diperoleh bahwa akurasi uji dari model *decision tree* adalah 0.90 atau 90%, dengan *recall* mencapai 0.9726 atau 97,26%. Hal ini menunjukkan bahwa algoritma model *decision tree* mampu mengklasifikasikan minat mahasiswa dalam mengikuti program MBKM dengan sangat baik.

Untuk melihat variabel-variabel berpengaruh suatu model dengan metode *decision tree*, dapat dilihat pada variabel yang tercantum pada hasil pohon keputusan (Gambar 2).



Gambar 2. Hasil Pohon Keputusan

Dari hasil pohon keputusan, variabel yang berturut-turut dinilai paling berpengaruh terhadap minat mahasiswa Universitas Riau mengikuti MBKM adalah variabel KeyakinanDiri, DukunganKeluarga, KemampuanDiri, DukunganTeman, dan Angkatan. Oleh karenanya, dapat disimpulkan bahwa minat mengikuti MBKM mahasiswa Universitas Riau paling dipengaruhi oleh keyakinan diri mahasiswa tersebut terhadap dirinya sendiri.

3.3 Random Forest

Tabel 7. merupakan tabel *confusion matrix* untuk algoritma model *random forest*, dimana diperoleh bahwa 1 data diklasifikasikan “tidak berminat” yang dinilai sama dengan data yang sebenarnya, 6 data diperkirakan “berminat” tetapi sebenarnya “tidak berminat”. Kemudian, 71 data diklasifikasikan sebagai “berminat” diperkirakan sesuai dengan data riil, dan 2 data diperkirakan “tidak berminat” yang ternyata “berminat”.

TABEL VII. *CONFUSION MATRIX* MODEL *RANDOM FOREST*

	Pred. Tidak Berminat	Pred. Berminat	Presisi Kelas
<i>Actual.</i> Tidak Berminat	1	6	33%
<i>Actual.</i> Berminat	2	71	92%

Nilai *accuracy* dan *recall* untuk algoritma *random forest* dapat dilihat pada Tabel 8 sebagai berikut.

TABEL VIII. HASIL EVALUASI MODEL *RANDOM FOREST*

Algoritma	Accuracy	Recall
Random Forest	90%	97,26%

Diperoleh bahwa tingkat keakuratan dari model *random forest* adalah 0.90 atau 90%, dan *recall* sebesar 0.9726 atau 97,26%. Dari hasil analisis tersebut, diperoleh bahwa algoritma model *random forest* dapat mengklasifikasikan minat mahasiswa dalam mengikuti program MBKM dengan sangat baik.

TABEL IX. *FEATURE IMPORTANCE* MODEL *RANDOM FOREST*

Variabel fitur	Feature Importance
KeyakinanDiri	0.173350
Fakultas	0.170774
Angkatan	0.160228
KemampuanDiri	0.146260
DukunganKeluarga	0.121530
DukunganTeman	0.100268
JK	0.047031
DukunganDosen	0.042223
Fasilitas	0.017461
MerasakanManfaatProgram	0.012390
PengetahuanMBKM	0.008485

Berdasarkan analisis *feature importance* untuk model *random forest* di atas, variabel yang paling mempengaruhi minat mahasiswa dalam mengikuti program MBKM yang mana memiliki nilai *feature importance* tertinggi adalah KeyakinanDiri. Selanjutnya, diikuti dengan variabel Fakultas, Angkatan, KemampuanDiri DukunganKeluarga, DukunganTeman, JK, DukunganDosen, Fasilitas, dan MerasakanManfaatProgram. Variabel yang memiliki pengaruh paling rendah pada minat mahasiswa dalam mengikuti program MBKM adalah DukunganDosen, yang mana memiliki nilai *feature importance* terendah. Hal tersebut menggambarkan bahwa dukungan dari dosen tidak memiliki pengaruh secara signifikan dalam menentukan minat mahasiswa dalam mengikuti program MBKM.

3.4 Gradient Boost Machine

Tabel 10. merupakan tabel *confusion matrix* untuk algoritma model *gradient boost machine*, dimana diperoleh bahwa 1 data diklasifikasikan

“tidak berminat” yang dinilai sama dengan data yang sebenarnya, 6 data diperkirakan “berminat” tetapi sebenarnya “tidak berminat”. Kemudian, 71 data diklasifikasikan sebagai “berminat” diperkirakan sesuai dengan data riil, dan 1 data diperkirakan “tidak berminat” yang ternyata “berminat”.

TABEL X. *CONFUSION MATRIX* MODEL *GRADIENT BOOST MACHINE*

	Pred. Tidak Berminat	Pred. Berminat	Presisi Kelas
Actual. Tidak Berminat	1	6	50%
Actual. Berminat	1	72	92%

Nilai *accuracy* dan *recall* untuk algoritma *gradient boost machine* dapat dilihat pada Tabel 11 sebagai berikut.

TABEL XI. HASIL EVALUASI MODEL *GRADIENT BOOST MACHINE*

Algoritma	Accuracy	Recall
Gradient Boost Machine	91,25%	98,63%

Diperoleh bahwa akurasi uji dari model *gradient boost machine* (GBM) adalah sebesar 0.9125 atau 91,25%, dengan *recall* mencapai 0.9863 atau 98,63%. Hal tersebut menunjukkan bahwa algoritma model *gradient boost machine* (GBM) mampu mengklasifikasikan minat mahasiswa dalam mengikuti program MBKM dengan klasifikasi sangat baik.

TABEL XII. *FEATURE IMPORTANCE* MODEL *GRADIENT BOOST MACHINE*

Variabel fitur	Feature Importance
KeyakinanDiri	0.467172
DukunganKeluarga	0.225414
Fakultas	0.085594
Angkatan	0.066497
KemampuanDiri	0.060741
DukunganTeman	0.060687
DukunganDosen	0.019003
JK	0.010095
Fasilitas	0.002913

PengetahuanMBKM	0.001710
MerasakanManfaatProgram	0.000172

Berdasarkan analisis *feature importance* model *gradient boost machine* (GBM) di atas, diperoleh bahwa variabel yang paling mempengaruhi minat mahasiswa dalam mengikuti program MBKM yang mana memiliki nilai *feature importance* tertinggi (0.467172 atau 46,72%) adalah KeyakinanDiri. Dilanjutkan dengan variabel lainnya dengan urutan DukunganKeluarga, Fakultas, Angkatan, KemampuanDiri, DukunganTeman, DukunganDosen, JK, Fasilitas, kemudian PengetahuanMBKM. Variabel MerasakanManfaatProgram merupakan variabel fitur yang memiliki pengaruh paling rendah pada minat mahasiswa dalam mengikuti program MBKM, ditunjukkan dengan nilai *feature importance* paling rendah di antara variabel fitur lainnya (0.000172 atau 0.0172%).

3.5 Hasil Perbandingan

Setelah melakukan pengujian terpisah terhadap setiap algoritma yang digunakan, hasil tingkat akurasi kemudian dibandingkan untuk menentukan algoritma yang paling sesuai untuk mengklasifikasikan minat MBKM. Berikut adalah hasil perbandingan dari keempat metode algoritma tersebut berdasarkan penelitian yang telah dilakukan.

TABEL XIII. HASIL PERBANDINGAN

Algoritma	Accuracy	Recall
<i>Logistic Regression</i>	90%	97,26%
<i>Decision Tree</i>	90%	97,26%
<i>Random Forest</i>	90%	97,26%
<i>Gradient Boost Machine</i>	91,25%	98,63%

Berdasarkan Tabel 12., dapat dilihat bahwa dari keempat algoritma *machine learning* yang dianalisis, yaitu Regresi Logistik, *Decision Tree*, *Random Forest*, dan *Gradient Boost Machine* (GBM), model algoritma yang menunjukkan hasil terbaik adalah *Gradient Boost Machine* (GBM) yang dapat dilihat bahwa model tersebut memiliki tingkat akurasi uji sebesar 91,25%, dimana lebih tinggi dibandingkan dengan algoritma lainnya.

4. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa penggunaan berbagai model pembelajaran mesin, seperti regresi logistik, *decision tree*, *random forest*, dan *gradient boosting machine* (GBM) mampu mengklasifikasikan minat mahasiswa terhadap program MBKM dengan tingkat akurasi yang baik. Diperoleh bahwa akurasi uji dari model regresi logistik adalah 90% dengan *recall* 97,26%, lalu pada model *decision tree* dengan akurasi uji sebesar 90% dan *recall* mencapai 97,26%, model *random forest* dengan akurasi uji sebesar 90% dan *recall* 97,26%, serta model *gradient boost machine* (GBM) dengan tingkat akurasi uji sebesar 91,25% dan *recall* mencapai 98,63%. Terlihat hasil bahwa model *gradient boost machine* (GBM) menunjukkan kinerja yang paling unggul, dengan tingkat keakuratan yang lebih tinggi dibandingkan model-model lainnya, serta *recall* yang mencapai nilai yang tinggi. Hal ini mencerminkan bahwa model GBM memiliki kinerja yang baik dan kuat dalam memprediksi minat mahasiswa dalam mengikuti program MBKM.

Secara keseluruhan, penelitian ini dinilai dapat memberi kontribusi yang signifikan dalam mengidentifikasi minat mahasiswa dalam mengikuti program MBKM dan penggunaan berbagai model algoritma *machine learning* dalam menganalisisnya. Diharapkan hasil penelitian yang telah dilakukan ini mampu menjadi dasar terhadap pengembangan program-program pendidikan tinggi yang lebih efektif dan berdaya saing di masa depan.

Dari hasil penelitian ini, disarankan agar penelitian selanjutnya yang akan dilakukan dapat diperluas cakupannya dengan melibatkan lebih banyak perguruan tinggi dan responden, sehingga hasilnya dapat menjadi representatif secara lebih luas (mencakup keseluruhan). Selanjutnya, studi ini juga dapat melibatkan pertimbangan terhadap faktor-faktor tambahan yang memengaruhi minat mahasiswa, seperti variabel ekonomi, sosial, dan lingkungan, untuk memberikan pemahaman yang lebih lengkap. Selain itu, penelitian berikutnya dapat meningkatkan analisis terhadap faktor-faktor yang mempengaruhi minat mahasiswa dengan menggunakan pendekatan kualitatif, seperti wawancara atau studi kasus.

DAFTAR PUSTAKA

- [1] H. Triani, "Analisis Kesiapan Penerapan Program MBKM (Merdeka Belajar Kampus Merdeka) Do Fakultas Ekonomi Dan Bisnis Islam Iain Palopo", Institut Agama Islam Negeri Palopo, 2023. <http://repository.iainpalopo.ac.id/id/eprint/7787/1/HIKMAH%20TRIANI.PDF>
- [2] A. Oksari, L. Nurhayati, D. Susanty, dan G. A. P. K. Wardhani, "Analisis Implementasi Kurikulum Merdeka Belajar-Kampus Merdeka (MBKM) Program Studi Biologi Universitas Nusa Bangsa". *Jurnal Studi Guru Dan Pembelajaran*, vol. 5, no. 1, 78-85, 2022. <https://www.e-journal.my.id/jsgp/article/view/1556>
- [3] K. D. P. Meke, R. B. Astro, dan M. H. Daud, "Dampak Kebijakan Merdeka Belajar Kampus Merdeka (MBKM) pada Perguruan Tinggi Swasta di Indonesia", *Edukatif: Jurnal Ilmu Pendidikan*, vol. 4, no. 1, 675-685, 2022. <https://edukatif.org/index.php/edukatif/article/view/1940>
- [4] A. Irawan, dan H. Suharyati, "Analisis Dampak Kebijakan Kurikulum Merdeka Belajar Kampus Merdeka (MBKM) pada Perguruan Tinggi: Literatur Review", *Research and Development Journal Of Education*, vol. 9, no. 2, 1116-1123, 2023. <https://journal.lppmunindra.ac.id/index.php/RDJE/article/view/19419/6044>
- [5] E. Kusnadi, N. Nur'aeni, I. Asrofi, R. Ramdani, dan K. M. Lutfi, "Respon Mahasiswa dalam Implementasi Kebijakan Program Merdeka Belajar Kampus Merdeka pada Program Studi Pendidikan Pancasila dan Kewarganegaraan", *Edukatif: Jurnal Ilmu Pendidikan*, vol. 4, no. 3, 3595-3603, 2022. <https://edukatif.org/index.php/edukatif/article/view/2702>
- [6] M. Rahmayani, "PREDIKSI KUALITAS UDARA MENGGUNAKAN METODE ENSEMBLE PADA MODEL DECISION TREE ID.3, RANDOM FOREST DAN REGRESI LOGISTIK MULTINOMIAL", Universitas Sriwijaya, 2022. https://repository.unsri.ac.id/76549/56/RA_MA_44201_08011281823038_0019077302_0_020115903_01_front_ref.pdf
- [7] N. Chairunnisa, "Prediksi Kemampuan Pembayaran Klien Home Credit Menggunakan Model Random Forest, Decision Tree, Dan Logistic Regression", *Jurnal Elektronika dan Teknik Informatika Terapan*, vol. 1, no. 3, 140-147, 2023. <https://e-journal.poltek-kampar.ac.id/index.php/JENTIK/article/view/383>
- [8] M. Imron dan P. K. Intan, "Implementasi Metode Regresi Logistik Biner Pada Keakuratan Klasifikasi Jenis Musik (Studi Kasus Musik K-Pop dan Western Pada Spotify)", *Jurnal Matematika dan Statistika serta Aplikasinya*, vol. 11, no. 2, 30-36, 2023. <https://journal.uin-alauddin.ac.id/index.php/msa/article/view/36374/18063>
- [9] N. Giansyani, A. F. Hidayatullah, dan R. Rahmadi, "Komparasi Algoritma Machine Learning Dan Deep Learning Untuk Named Entity Recognition: Studi Kasus Data Kebencanaan", *JIRE (Jurnal Informatika & Rekayasa Elektronika)*, vol. 3, no. 1, 48-57, 2020. <https://e-journal.stmiklombok.ac.id/index.php/jire/article/view/222/109>
- [10] R. Nuraini, R. R. A. Hakim, T. Lisnawati, dan W. T. Fariati, "Pengembangan Sistem Klasifikasi Tipe Kepribadian Siswa Secara Psikologis dengan Algoritma Decision Tree C.45", *Building of Informatics, Technology and Science (BITS)*, vol. 3, no. 3, 220-227, 2021. <https://ejurnal.seminar-id.com/index.php/bits/article/download/1045/727/>
- [11] O. Pahlevi, Amrin, dan Y. Handrianto, "Implementasi Algoritma Klasifikasi Random Forest Untuk Penilaian Kelayakan Kredit", *Jurnal Infortech*, vol. 5, no. 1, 71-76, 2023. <https://ejournal.bsi.ac.id/ejournal/index.php/infortech/article/download/15829/6000>
- [12] V. Atlantic, E. Sulistianingsih, dan H. Perdana, "GRADIENT BOOSTING MACHINE PADA KLASIFIKASI KELULUSAN MAHASISWA", *Bimaster : Buletin Ilmiah Matematika, Statistika dan Terapannya*, vol. 13, no. 2, 165-174, 2024. <https://jurnal.untan.ac.id/index.php/jbmstr/article/viewFile/76955/75676601169>
- [13] A. Setiawan, R. F. Waleska, M. A. Purnama, Rahmaddeni, dan L. Efrizoni, "Komparasi Algoritma K-Nearest Neighbor (K-NN), Support Vector Machine (SVM), Dan Decision TREE Dalam Klasifikasi Penyakit Stroke", *JIRE (Jurnal Informatika & Rekayasa Elektronika)*, vol. 7, no. 1, 107-114, 2024. <https://e-journal.stmiklombok.ac.id/index.php/jire/article/view/1161/310>
- [14] D. W. Hosmer dan S. Lemeshow, *Applied Logistic Regression*, USA : John Wiley & Sons, 2000.
- [15] J. Eska, "Data Mining Untuk Prediksi Penjualan Wallpaper Menggunakan Algoritma C45",

- JURTEKSI (Jurnal Teknol. dan Sist. Informasi)*, vol. 2, 9-13, 2016.
- [16] R. Supriyadi, W. gata, N. Maulidah, dan A. Fauzi, "Penerapan Algoritma Random Forest Untuk Menentukan Kualitas Anggur Merah", *Jurnal Ilmiah Ekonomi dan Bisnis*, vol. 13, no. 2, 67-75, 2020.
<https://www.journal.stekom.ac.id/index.php/Bisnis/article/download/247/182>
- [17] L. Lumbaa, E. Mailoa, dan M. A. I. Pakereng, "Implementasi Metode SVM dan Gradient Boost Dalam Klasifikasi Bahasa Daerah (Halmahera, Kalimantan, Toraja)", *Jurnal Teknik Informatika dan Sistem Informasi*, Vol. 9, No. 2, 908-915, 2022.
<https://jurnal.mdp.ac.id/index.php/jatisi/article/view/1663/768>
- [18] Rochana, R. M. Darajatun, dan M. A. Ramdhany, "Pengaruh Implementasi Kebijakan Kampus Merdeka terhadap Minat dan Keterlibatan Mahasiswa", *Journal of Business Management Education*, Vo. 6, No. 3, 11-21, 2021.
<https://ejournal.upi.edu/index.php/JBME/article/viewFile/40165/17632>