Building of Informatics, Technology and Science (BITS)
Volume 6, No 1, Juni 2024 Page: 559-566
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i1.5332

Analisis Tingkat Ketertarikan Mahasiswa Terhadap Bidang Artifcial Intelligence dalam Penulisan Skripsi dengan Random Forest

Abdul Halim Anshor*, Tri Ngudi Wiyatno

Program Studi Teknik Informatika, Fakultas Teknik, Universitas Pelita Bangsa, Bekasi, Indonesia
Email: [1,*]abdulhalimanshor@pelitabangsa.ac.id , [2]tringudi@pelitabangsa.ac.id
Email Penulis Korespondensi: abdulhalimanshor@pelitabangsa.ac.id
Submitted: **12/06/2024**; Accepted: **30/06/2024**; Published**: 30/06/2024**

Abstrak−Perkembangan teknologi informasi maju dengan sangat pesat sesuai dengan kebutuhan manusia yang terus meningkat. Kecerdasan Buatan atau Artificial Intelligence (AI) adalah fenomena yang sangat umum pada saat ini, dimana hampir semua pekerjaan yang dilakukan oleh manusia dikerjakan oleh komputer. Komputer diberikan pembelajaran sedemikian rupa sehingga dengan sistemnya dapat melakukan atau menggantikan beberapa pekerjaan manusia. Perkembangan AI yang terus menerus meningkat membutuhkan tenaga ahli yang dapat memahami bidang ini. Universitas dengan jurusan Teknik informatika merupakan salah satu lembaga pendidikan yang mempunya kontribusi terbesar dalam mencetak tenaga ahli dibidang AI. Penelitian ini dilakukan dengan tujuan untuk mengukur dan mengetahui seberapa besar minat ketertarikan mahasiswa dalam bidang AI, yang nantinya diharapkan dapat melahirkan tanaga ahli dibidang AI dan dapat mengimplementasikannya dalam kehidupan sehari hari. Untuk itu peneliti melakukan analisis ketertarikan mahasiswa pada bidang AI yang diiddentifikasi sebagai bidang yang diminati untuk penulisan skripsi. Dalam penelitian ini peneliti menggunakan metode machine learning Random Forest dengan data sampling kuisioner yang berikan kepada 200 mahasiswa Universitas Pelita Bangsa jurusan Teknik Informatika pada semester gassal tahun ajaran 2023 - 2024. Data yang akan dianalisis adalah data mahasiswa tertarik dan tidak tertarik pada bidang AI. Dari penelitian ini didapatkan nilai akurasi sebesar 80.05% ketertarikan mahasiswa pada bidang AI. Diharapkan dari penelitian ini dapat menjadi informasi yang bermanfaat bagi Universitas Pelita Bangsa dalam mengembangkan kurikulum di bidang AI kedepannya sehingga dapat memenuhi kebutuhan akan tenaga ahli dibidang AI.

Kata Kunci: Artificial Intelligence; Kuisioner; Machine Learning; Random Forest; Skripsi

Abstract−Artificial Intelligence (AI) is an advanced phenomenon of information technology that is currently very fast. With AI human work can be replaced by computers. Universities are superior providers in providing experts in the field of AI. One indicator that can describe how the curriculum in the field of AI is implemented on campus is by assessing how interested students are in taking the field of AI for writing their thesis. In this research, researchers used the Random Forest machine learning method with questionnaire sampling data from 200 students interested and not interested in the field of AI. The results of this research will provide accuracy values for classifying students' interest in the field of AI. Questionnaire data will be classified into two classes, namely interested and not interested classes. Results from classification trials with the WEKA application. It is known that the classification results have an accuracy value of 80.5%, this shows that the random forest algorithm has worked effectively in the process of classifying Pelita Bangsa University student interest data in the field of AI in writing these**s.**

Keywords: Artificial Intelligence; Questionnaires; Machine Learning; Random Forest; Thesis

1. PENDAHULUAN

Kebutuhan akan tenaga ahli dibidang Artificial Intelligence sangat diperlukan pada era saat ini, hal ini menuntut setiap univeritas terutama jurusan Teknik informatika untuk berperan aktif dalam mempersiapkan tenaga ahli dibidang Artificial Intelligence . Universitas Pelita Bangsa adalah salah satu lembaga pendidikan yang dapat memberikan kontribusi dalam mencetak tenaga ahli dibidang Artificial Intelligence dengan kurikulum yang diberikan kepada mahasiwanya. Untuk mengetahui mahasiswa yang dihasilkan apakah dapat memenuhi kebutuhan tenaga ahli di bidang Artificial Intelligence salah satu caranya adalah dengan dilakukan sebuah analisis terhadap data ketertarikan mahasiswa terhadap bidang Artificial Intelligence dalam penulisan skripsinya. kemampuan dan peminatan mahasiswa pada bidang Artificial Intelligence sehingga lulusan yang dihasilkan dapat menjadi tenaga ahli dibidang Artificial Intelligence yang handal dan dapat mengimplementasikannya dalam kehidupan sehari-hari. Kurangnya data mengenai tingkat minat mahasiswa terhadap bidang kecerdasan buatan (AI) saat menulis skripsi. Sulit untuk memahami faktor-faktor yang mempengaruhi minat siswa terhadap AI. Kurangnya model prediksi yang akurat untuk memperkirakan minat siswa terhadap AI. Penelitian ini bertujuan untuk menganalisis tingkat minat mahasiswa terhadap bidang AI saat menulis skripsi. Identifikasi faktor-faktor yang mempengaruhi minat siswa terhadap AI. Kembangkan model prediksi yang akurat untuk memperkirakan minat siswa terhadap AI. Penelitian ini memiliki kepentingan untuk memahami tingkat minat mahasiswa terhadap AI, yang penting untuk mengembangkan kebijakan dan program penelitian yang sesuai dengan minat dan kebutuhannya. Selain itu, penelitian ini dilakukan untuk mengetahui faktor-faktor yang mempengaruhi minat siswa terhadap AI, yang dapat membantu meningkatkan minat dan partisipasi siswa dalam bidang AI. Terakhir, model prediksi yang akurat dapat digunakan untuk membantu mahasiswa memilih topik skripsi yang sesuai dengan minat dan potensinya. Penelitian ini akan fokus pada tingkat minat mahasiswa terhadap bidang AI saat menulis skripsi. Hal yang akan diteliti meliputi latar belakang pengetahuan, minat dan motivasi mahasiswa terhadap penulisan skripsi dibidang AI. Model prediksi yang akan dikembangkan akan menggunakan algoritma Random Forest untuk mengklasifikasikan mahasiswa berdasarkan tingkat minatnya terhadap AI. Random forest adalah algoritma machine learning yang menggabungkan hasil dari banyak decision tree untuk mencapai satu hasil. Random forest merupakan teknik penting untuk identifikasi dan prediksi dalam pembelajaran

Building of Informatics, Technology and Science (BITS)
Volume 6, No 1, Juni 2024 Page: 559-566
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i1.5332

mesin [1] [2] [3] [4]. Untuk membangun pembelajaran mandiri dan sistem pengambilan keputusan yang tidak memerlukan pemrograman ulang manusia adalah tujuan pembelajaran mesin. Selain dapat memutuskan apa yang terbaik untuk dilakukan ketika membuat penilaian, mesin juga dapat beradaptasi dengan perubahan di lingkungan mereka [5].

Metode machine learning dengan menggunakan semua pohon optimal digabungkan menjadi satu model yang dikenal sebagai random forest (hutan acak) Random Forest bergantung pada nilai vektor acak yang memiliki distribusi yang sama di semua pohon keputusan, yang masing-masing memiliki kedalaman maksimum. Pengklasifikasi yang terdiri dari pengklasifikasi berbentuk pohon disebut hutan acak [6] [7] [8] [9] [10]. Penelitian ini mengacu pada penelitian sebelumnya yang menggunalkan metode random forest. Penelitian yang dilakukan oleh Sena Wijayanto, penelitian ini bertujuan untuk melakukan analisis sentimen terhadap ulasan yang diberikan oleh pengunjung hotel. Data ulasan tersebut diambil dari Traveloka menggunakan web scrapping. Metode yang digunakan untuk ekstraksi fitur adalah word2vec. Untuk klasifikasi sentimen, metode yang digunakan adalah random forest. Penelitian dimulai mengumpulkan ulasan hotel dari platform online seperti Traveloka menggunakan web scraping. Selanjutnya memilah data ulasan berdasarkan aspek layanan hotel. Data yang ada selanjutnya dibersihkan dari noise seperti karakter khusus, HTML tag, dan tanda baca yang tidak perlu. Tahap selanjutnya melakukan tokenisasi untuk memecah teks menjadi kata-kata individu. Ekstraksi fitur dilakukan dengan menggunakan metode Word2Vec untuk mengubah kata-kata menjadi representasi vektor numerik. Mereka melakukan klasifikasi data ulasan menjadi sentiment positif dan negative. Hasil percobaan terbaik didapatkan dari hasil percobaan dengan menggunakan jumlah tree 100, 200, dan 300 dengan hasil akurasi sebesar 82%-83% [1]. Perbedaan penelitian ini dengan penelitian yang peneliti lakukan adalah adanya pengujian sampel dengan pembagian jumlah tree sebanyak tiga bagian, atribut yang dikelompokkan berdasarkan rating untuk menentukan pelabelan kelas posisitf dan negatif, sedangkan peneliti membuat kelas dengan membuat kelas positif dan kelas negatif dari data survei yang bersifat optional tertarik atau tidak tertarik.

Selanjutnya penelitian yang dilakukan oleh Muhamad Azhar dkk, penelitian ini menerapkan teknologi oversampling berbasis Random Forest untuk pengenalan dialek. Untuk optimasi hyper-parameter dari algoritma random forest, kami menerapkan Grid Search. Percobaan pada data Speech Accent Archive menggunak ulasan an fitur MFCC menghasilkan akurasi 0.91 dan AUC 0.95 [2]. Penelitian tersebut menunjukkan nilai AUC rendah 0.57 sehingga perlu dilakukan resampling data dengan pendekatan SMOTE dan ROS, sedangkan pada peneian ini nilai AUC sudah tinggi 0.801 sehingga tidak perlu dilakukan metode resampel.

Penelitian selanjutnya dilakukan oleh Robi Aziz Zuama dkk, penelitian ini adalah membuat sebuah model prediksi penyakit stroke yang efektif, sistem menggunakan parameter dari faktor gaya hidup, faktor yang dapat dikendalikan seperti faktor risiko medis dan faktor-faktor yang tidak dapat dikendalikan. Empat algoritma pengklasifikasi di usulkan yaitu multi layer perceptron, KNN, Decisson Tree dan Random Forest. Hasil menunjukkan bahwa algoritma pengklasfikasi dapat bekerja efektif dengan hasilkan nilai akurasi mencapar sempurna 99,99% pada tingkat validasi 10K-Fold Validation [11]. Penelitian tersebut berisi perbandingan kinerja algoritma random forest dengan algoritma lainnya, sedangkan pada penelitian ini hanya membahas kinerja algoritma random forest.

Penelitian yang dilakukan oleh . Chandra Ayunda Apta Soemedhy dkk, dalam penelitian tersebut metode yang dipakai yaitu algoritme pembelajaran mesin seperti algoritme SVM, Naive Bayes, dan Random Forest. Ketiga algoritme tersebut kemudian dibandingkan untuk mengetahui algoritme mana yang paling baik dalam menganalisis sentiment masyarakat terhadap kasus pelecehan seksual. Hasil dari perbandingan ketiga algoritme Pembelajaran mesin, Random Forest memperoleh tingkat akurasi terbaik sebesar 78% dibandingkan kedua algoritme lainnya dalam melakukan analisis sentimen terhadap komentar YouTube tentang bahasan pelecehan seksual [12]. Pada penelitian tersebut membandingkan ketiga algoritma, naive bayes, SVM dan random forest.

Kemudian penelitian yang dilakukan oleh Kurniabudi, penelitian ini menghasilkan sistem deteksi anomali yang memiliki performa yang tinggi. Untuk data eksperimen digunakan dataset CICIDS-2017, yang memiliki dimensionalitas data yang tinggi dan mengandung data tidak seimbang. Hasil pengujian memperlihatkan Random Tree memiliki akurasi 99,983% dan Random Forest 99,984% [13].

Penelitian yang dilakukan Syakirah Fachid menyebutkan algoritma Random Forest menghasilkan nilai RMSE sebesar 1886.555, MAPE 14.85 dan tingkat akurasi sebesar 97.7%.[14]. Penelitian yang dilakukan oleh Ajeng Citra Mawani dkk dengan judul penelitian " Deteksi Dini Gejala Awal Penyakit Diabetes Menggunakan Algoritma Random Forest" penelitian ini menggunakan data didapatkan melalui website www.kaggle.com dengan total 520 record dataset yang terdiri dari 17 attribut, terdapat 320 dataset dengan positif diabetes dan 200 dataset dengan negative diabetes. Klasifikasi dilakukan dengan dengan komposisi data training dan data testing 90:10 menggunakan teknik stratified random sampling dengan number of trees 5, maximal depth 5, dan dilakukannya apply pruning. Diperoleh akurasi 90.38%, precision 100%, recall 84.38% dan niai AUC 1.00 [15].

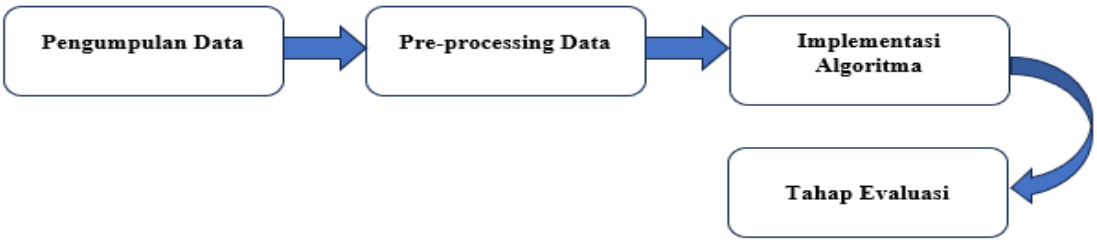
Penelitian ini memiliki perbedaan metodologi yang signifikan dengan penelitian lain dalam hal pengumpulan data, pra-prosesing data, ekstraksi fitur, klasifikasi, dan evaluasi. Hal ini menunjukkan bahwa penelitian ini menggunakan pendekatan yang unik dan inovatif untuk meneliti topiknya Dari hasil beberapa penelitian yang telah dilakukan di atas ,peneliti menilai bahwa metode random forest sangat baik dalam memberikan memprediksi dan mengklasifikasikan dengan nilai akurasi lebih baik disbanding metode yang lain. Algoritma ini juga dapat mengatasi permasalahan untuk data tidak seimbang, untuk itu peneliti memutuskan untuk menggunakan metode random forest pada penelitian kali ini. Tujuan dari penelitian ini adalah untuk memahami tingkat ketertarikan mahasiswa

Building of Informatics, Technology and Science (BITS)
Volume 6, No 1, Juni 2024 Page: 559-566
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i1.5332

terhadap bidang Artificial Intelligence (AI) dalam penulisan skripsi dan mengembangkan model prediksi untuk memperkirakan ketertarikan mahasiswa terhadap AI.

2. METODOLOGI PENELITIAN

Dalam melakukan penelitian ini peneliti menggunakan metode eksperimen dan penelitian ini bersifat kuantitatif dan kualitatif untuk mendapatkan hasil yang maksimum. Dalam penelitian ini peneliti menggunaka software WEKA 3.8.6 [4]. Penelitian ini menggunakan Algoritma C.45. Kinerja kedua metode diuji dengan cross-validation, yang ditunjukkan oleh Confusion matrix dan Kurva ROC. Tahapan penelitian yang dilakukan ditunjukan oleh gambar 1.



Gambar 1. Tahapan penelitian

2.1 Pengumpulan Data

Perjalanan penelitian ini dimulai dengan pertanyaan mendasar , Seberapa besar ketertarikan mahasiswa terhadap bidang Artificial Intelligence (AI) dalam penulisan skripsi. Pertanyaan ini mengantarkan peneliti pada tujuan penelitian, yaitu untuk mengukur tingkat ketertarikan mahasiswa terhadap AI dan mengidentifikasi faktor-faktor yang memengaruhinya. Langkah selanjutnya dengan merancang Kuisioner yang komprehensif dan terstruktur. Kuisioner ini terbagi menjadi beberapa item, yaitu NIM , Nama, Kelas, dan Respon tingkat ketertarikan terhadap AI melalui pertanyaan tentang minat terhadap aplikasi AI dalam menulis skripsi. Langkah selanjutnya adalah memilih metode pengumpulan data. Pada penelitian ini peneliti memilih untuk menggunakan platform google forms. Kuisioner selanjutnya disebar pada mahasiswa Teknik Informatika Universitas Pelita Bangsa sebanyak 200 orang dengan cara menyebarkan tautan melalui platform media sosial WhatsApp.

2.2 Pre-processing Data

Pada tahap ini data survei dibersihkan dari data yang hilang, tidak lengkap, atau tidak konsisten.Data dikodekan ulang atau diubah formatnya untuk analisis selanjutnya. Normalisasi data numerik untuk memastikan semua fitur memiliki skala yang sama. Identifikasi variabel dependen (tingkat ketertarikan) dan variabel independen (pertanyaan kuisioner). Hasil dari kuioner diubah format data fiturnya agar sesuai dengan algoritma machine learning. Tahap selanjutnya proses Encoding, yaitu mengubah data kategorikal menjadi data numerik agar dapat diproses oleh algoritma random forest. Peneliti memilih label encoding karena data kategori yang ada sedikit jumlahnya.

2.3 Implementasi Algoritma

Pada penelitian ini akan dipilih algoritma Random Forest untuk menyelesaikan masalah yang ada. Random Forest adalah algoritma machine learning yang menggabungkan keluaran dari beberapa decision tree untuk mencapai satu hasil [4] [5] [6]. Random forest forest atau 'hutan' dibentuk dari banyak tree (pohon) yang diperoleh melalui proses bagging atau bootstrap aggregating [7]. Random Forest bekerja dalam dua fase. Fase pertama yaitu menggabungkan sejumlah N decision tree untuk membuat Random Forest. Kemudian fase kedua adalah membuat prediksi untuk setiap tree yang dibuat pada fase pertama. Random Forest merupakan salah satu teknik pembelajaran mesin yang cerdas dan serbaguna adalah. Karena Random Forest dapat melakukan tugas klasifikasi dan regresi secara bersamaan, maka Random Forest dikenal sebagai multifungsi. Dibandingkan dengan teknik klasifikasi lainnya, kinerja Random Forest menghasilkan kesalahan klasifikasi yang lebih rendah [16] [17]. Rata-rata dari semua pohon keputusan di hutan acak adalah hasil dari algoritma yang menggabungkan ratusan hingga ribuan pohon keputusan [8] [9] [10]. Random Forest dapat bekerja dengan membentuk cluster dari decision tree pada saat training dan dapat menghasilkan class yang prediksi rata-rata dari individual trees. Random forest memiliki relasi proporsional langsung antara jumlah trees yang tercipta dan akurasi yang dihasilkan [15]. Persamaan (1) dan (2) merupakan pehitungan algoritma Random Forest [3].

$$Entropy(Y) = -\sum pi\ (xY) \log 2\text{p(c V Y)} \quad (1)$$

Entropy menyatakan ukuran ketidakpastian secara probabilistik. Dari persamaan nomor 2 dapat ditentukan nilai entropy dari variabel kategorikal Y, dimana Pi(xY) merupakan probabilitas kemunculan kategori Y.

Information Gain, Y merupakan kumpulan kasus dan dan p(c|Y) meupakan rasio nilai Y terhadap kelas c.

$$InformationGain(y,a) = Entropy(Y) - \sum |Yv|\,|Ya|\ Entropy(Yv) veValues \quad (2)$$

Building of Informatics, Technology and Science (BITS)
Volume 6, No 1, Juni 2024 Page: 559-566
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i1.5332

Sedangkan Values (a) yaitu seluruh nilai yang memungkinan pada suatu kumpulan kasus: aYv adalah subkelas dari Y dengan kelas v yang berhubungan dengan kelas a. Ya adalah semua nilai yang sesuai dengan a.

Pemilihan atribut sebagai simpul, baik akar (root) maupun simpul internal akan disesuaikan dengan nilai information gain tertinggi pada atribut – atribut yang tersedia. Untuk menenetukannya akan digunakan rumus gain ratio, sedangkan gain ratio diperoleh melalui penentuan nilai split information yang dijelaskan pada persamaan (3) dan (4).

$$Split\ Information(S,A) = \sum(|Si||S|c\ i = 1 \log 2(|Si||S|) \quad (3)$$

Split Information (S, A) merupakan nilai untuk estimasi entropy pada variable input S dengan kelas c dan $|Si|$ $|S|$ yaitu probabilitas kelas i pada atribut.

$$GainRatio(S,A) = \frac{InformationGain(S,A)}{SplitInformation(S,A)} \quad (4)$$

Information Gain (Gain(S, A)) mengukur penurunan rata-rata ketidakpastian dalam dataset setelah dibagi berdasarkan nilai atribut S. Semakin tinggi nilai Gain, semakin informatif atribut S. Split Information (Split_Info(S, A)) mengukur entropi rata-rata dari subset data yang dihasilkan setelah dibagi berdasarkan nilai atribut S. Semakin tinggi nilai Split Info, semakin beragam subset data, dan semakin kecil efektivitas atribut Sdalam membagi data. Gain Ratio dihitung dengan membagi Gain dengan Split Info. Tujuannya adalah untuk menormalisasi Gain dengan mempertimbangkan tingkat keragaman yang dihasilkan oleh atribut S. Atribut dengan Gain Ratio yang lebih tinggi umumnya lebih disukai karena menghasilkan pohon keputusan yang lebih kecil dan lebih akurat.

2.5 Evaluasi

Evaluasi merupakan langkah penting dalam penelitian untuk menilai kinerja model yang telah dibangun. Dalam penelitian ini tujuan utama evaluasi adalah untuk memastikan akurasi prediksi dengan cara mengukur seberapa tepat model dalam memprediksi tingkat ketertarikan mahasiswa terhadap AI dalam penulisan skripsi. Langkah-langkah yang dilakukan dalam evaluasi adalah pembagian data, data dibagi menjadi dua set data training dan data testing. Data training digunakan untuk melatih model, sedangkan data testing digunakan untuk mengevaluasi kinerja model. Selanjutnya memilih metrik yang sesuai untuk mengukur kinerja model. Metrik yang digunakan dalam klasifikasi penelitian ini adalah nilai akurasi yaitu persentase prediksi yang benar, presisi yaitu Persentase prediksi positif yang benar. Recall merupakan persentase sampel positif yang teridentifikasi dengan benar dan F1-Score merupakan abungan presisi dan recall. Evaluasi dari hasil pengujian di atas akan diukur tingkat akurasinya menggunakan 2 (dua) model yaitu menggunakan confusion matrix dan ROC. Confusion matrix adalah metode yang digunakan untuk mengukur kinerja pemroses klasifikasi yang digunakan untuk mengevaluasi model klasifikasi untuk memperkirakan apakah suatu objek benar atau salah . Hasil yang dihasilkan dari metode ini berupa 2 pengklasifikasi atau lebih. Evaluasi kinerja pengklasifikasi menghitung True Positives (TP), True Negatives (TN), False Positives (FP) dan False Negatives (FN) [18]. Keempat item tersebut dibentuk dalam tabel Confusion Matrix dengan pengklasifikasi 2 kelas, dijelaskan pada Tabel 1.

Tabel 1 . Confusion Matrix

	Prediction Class	
Actual Class	**Positive (+)**	**Negative (-)**
Positive (+)	TP (True Positive)	TN (True Negative)
Negative (-)	FP (False Positive)	FN (False Negative)

Pengukuran kinerja dapat dilakukan untuk mengetahui nilai akurasi, presisi, recall, dan F1 Score [19] [20]. Akurasi merupakan penentuan seberapa akurat model klasifikasi dalam memprediksi data secara akurat dan akan dinyatakan pada persamaan (5). Penghitungan presisi melibatkan penentuan seberapa akurat hasil data yang diminta dibandingkan dengan prediksi yang diberikan model dan akan disajikan pada persamaan (6). Perhitungan recall untuk mengetahui keberhasilan model dalam mengambil informasi disajikan pada persamaan (7). F1 – Score adalah rata-rata harmonik dari presisi dan perolehan dan dinyatakan dalam persamaan (8).

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (5)$$

$$Presisi = \frac{TP}{TP+FP} \quad (6)$$

$$Recall = \frac{TP}{TP+FN} \quad (7)$$

$$F1 - Score = \frac{2x(Presisi\ x\ Recall)}{(Presisi+Recall)} \quad (8)$$

3. HASIL DAN PEMBAHASAN

Building of Informatics, Technology and Science (BITS)
Volume 6, No 1, Juni 2024 Page: 559-566
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i1.5332

3.1 Pengumpulan Data

Data yang digunakan adalah data kuisioner ketertarikan mahasiswa akan penulisan skripsi dibidang Artificial Intelligence yang disebarkan kepada 200 responden yang merupakan mahasiswa Universitas Pelita bangsa fakultas Teknik Industri. Dari data kuisioner tersebut penulis membuat data set dengan atribut berupa NIM, Nama, Jenis kelamin, dan Class (label yang diberikan tertarik dan tidak tertarik pada bidang AI seperti yang ditunjukaoleh tabel 2.

Tabel 2. Variabel Data Kuisioner Mahasiswa

Nama Variabel	Type	Keterangan
NIM	Numeric	Nomor induk mahasiswa
Nama	Character	Nama Mahasiswa
Jenis Kelamin	Character	Jenis Kelamin
Respon	Character	Ketertarikan mahasiswa

3.2 Pre-processing Data

Tahap ini dilakuka dengan pemeriksaan apakah ada data yang hilang atau tidak lengkap untuk setiap variabel dalam kuesioner dan memastikan data yang dikumpulkan konsisten dan tidak ada duplikasi. Jika ada duplikasi data akan dihapus.Selanjutnya data akan diperiksa apakah nilai data yang dimasukan valid dan sesuai dengan format yang diharapkan. Data "Respon ketertarikan mahasiswa dalam menulis Skripsi AI" diubah menjadi variabel kategorikal tertarik dan tidak tertarik, hasil data kuisioner sebelum dan sedudah tahap pre-processing data ditunjukan oleh tabel 3 dan tabel 4.

Tabel 3. Hasil Sebelum Pre-Processing Data Kuisioner

Timestamp	Email Address	Nama	Jenis Kelamin	NIM	Kelas	Label
11/11/2023 5:16:18	kingawpbekasi@gmail.com	Muhamad David Ali	Laki-laki	312210291	TI 22 A1	Sangat Tidak Tertarik
11/11/2023 10:24:43	nukelaili@gmail.com	Nuke Fitri Laili	Perempuan	312210519	TI22CB1	Tertarik
11/11/2023 10:47:23	adenurlaela2576@gmail.com	ADE NURLAELA	Perempuan	312210652	TI CB 1	Sedikit tertarik
12/11/2023 12:57:31	herumansyah76@gmail.com	Muhammad Choeruman Syah	Laki-laki	312010031	TI.20.RPL.1	Tidak Tertarik
12/11/2023 12:57:33	kinantianing@mhs.pelitabangsa.ac.id	aning kinanti	Perempuan	312010364	TI.20.A.RPL.2	Tertarik
12/11/2023 13:00:03	bismaputra1986@gmail.com	Muhammad Bisma Putra Haryana	Laki-laki	312010443	TI.20.A.RPL-1	Tertarik

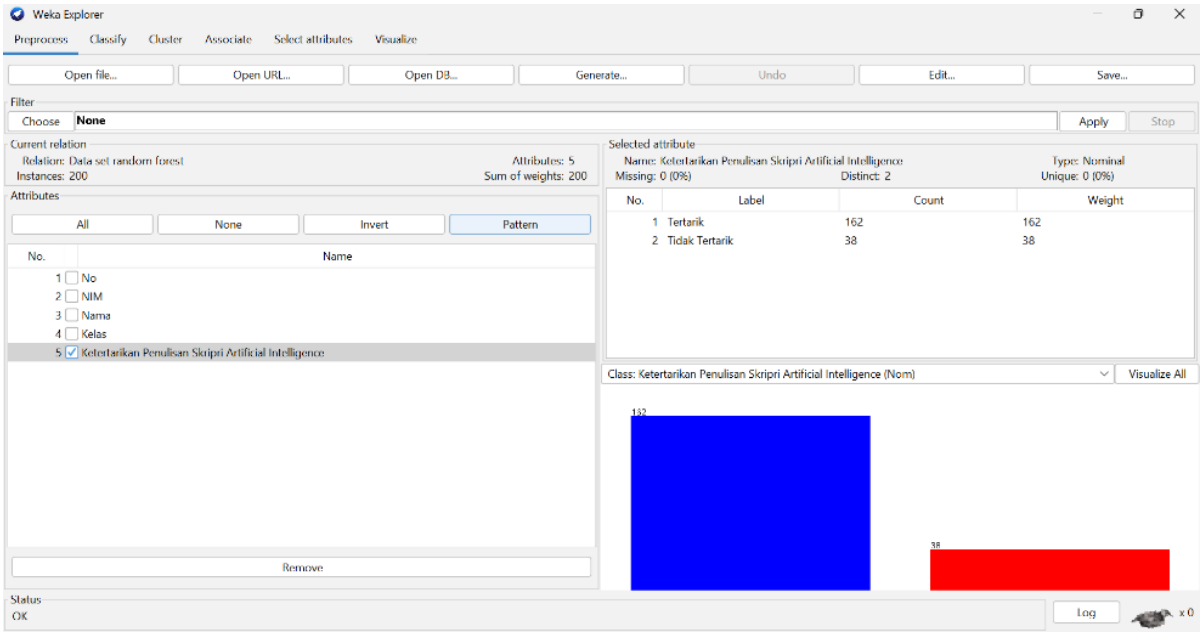
Tabel 4. Hasil Setelah Pre-Processing Data Kuisioner

Nama	Jenis Kelamin	NIM	Kelas	Label
Muhamad David Ali	Laki-laki	312210291	TI 22 A1	Tidak Tertarik
Nuke Fitri Laili	Perempuan	312210519	TI22CB1	Tertarik
Ade Nurlaela	Perempuan	312210652	TI CB 1	Tertarik
Muhammad Choeruman Syah	Laki-laki	312010031	TI.20.RPL.1	Tidak Tertarik
aning kinanti	Perempuan	312010364	TI.20.A.RPL.2	Tertarik
Muhammad Bisma Putra Haryana	Laki-laki	312010443	TI.20.A.RPL-1	Tertarik

3.3 Implementasi Algoritma

Data yang sudah melalui tahap preprocessing data selanjunya akan digunakan untuk implementasi algoritma random forest.Tujuan dari penerapan teknik machine learning ini adalah untuk melakukan klasifikasi ketertarikan mahasiswa untuk mengambil polis kendaraan menggunakan algoritma *random forest classifier*. Data yang digunakan adalah data kuisioner dalam bentuk file excel, kemudian data tersebut diubah dalam bentuk csv untuk selanjutnya diproses pada tools WEKA. Bentuk pemodelan dari data yang uji dapat dilihat pada gambar 2.

Building of Informatics, Technology and Science (BITS)
Volume 6, No 1, Juni 2024 Page: 559-566
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i1.5332



Gambar 2. Tahap Pemodelan

Dari gambar 2 merupakan bentuk visual dari data yang sudah di preprocessing kemudian dilakukan dimuat ke dalam aplikasi pengolah machine learning WEKA, pada penelitian ini peneliti menggunakan tools WEKA 3.8.6. sebelum data dimuat, data kuisioner dalam bentuk Ms.Excel diubah formatnya kedalam bentuk csv. agar dihasilkan kinerja dari algoritma yang *Random Forest*, penelitian ini melakukan pembagian persentasi data testing dan data training. Atribut yang dipilih menjadi Class adalah atribut respon mahasiswa, kelas ini dikelompokan menjadi kelas tertarik dan kelas tidak tertarik.

Dari hasil pengujian dengan aplikasi WEKA 3.8.6 akan di lakukan evaluasi dari data *stratified cross-validation* yang ditunjukkan oleh gambar 3, selanjutnya avaluasi akan dilajutkan dengan analisa dari dua (dua) model, confusion matrix dan ROC (*Receiver Operating Characteristic*).

```
=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         161               80.5    %
Incorrectly Classified Instances        39               19.5    %
Kappa statistic                         -0.0098
Mean absolute error                      0.2956
Root mean squared error                  0.3802
Relative absolute error                 95.3646 %
Root relative squared error             96.8869 %
Total Number of Instances              200
```

Gambar 3. Startified Cross-Validation

Hasil uji validasi yang ditunjukkan oleh gambar 3 dengan jelas menunjukkan bahwa persentase yang sampel diklasifikasikan dengan benar bernilai 80.5% dari total 100% sedangkan sampel yang diklasifikasikan bernilai salah bernilai 19.5%, hal ini membuktikan bahwa proses klasifikasi bernilai valid. Selanjutnya evaluasi dilanjutkan dengan melakukan analisa pada *confusion matrix*. *Confusion matrix* merupakan alat yang digunakan untuk mengevaluasi kinerja model klasifikasi dalam *machine learning*. Ini adalah tabel yang menunjukkan bagaimana data aktual dibandingkan dengan prediksi model. Pada penelitian ini *confusion matrix* yang dihasilkan seperti ditunjukkan pada gambar 4 di bawah ini.

```
  a   b   <-- classified as
161   1 |   a = Tertarik
 38   0 |   b = Tidak Tertarik
```

Gambar 4. Confusion matrix

Building of Informatics, Technology and Science (BITS)
Volume 6, No 1, Juni 2024 Page: 559-566
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i1.5332

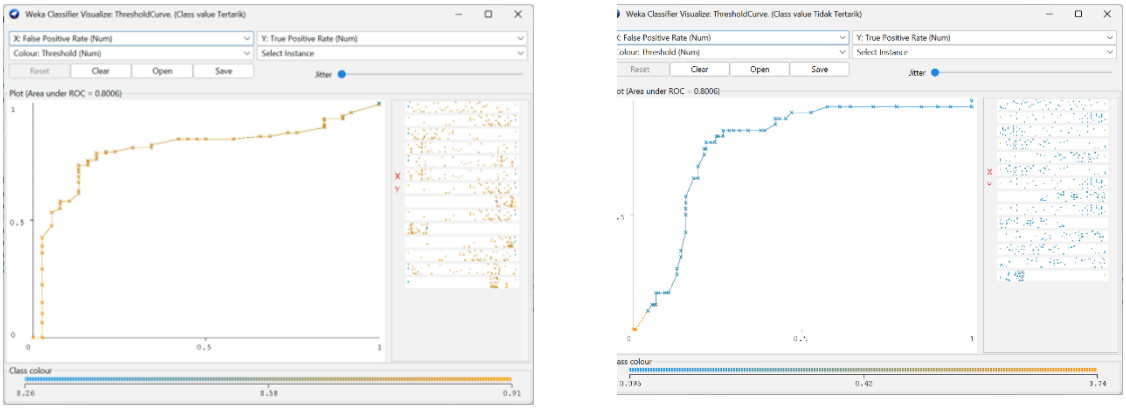
Dari hasil uji coba yang ditunjukan oleh gambar 4 dapat diketahui bahwa jumlah mahasiswa tertarik untuk menulis skripsi dibidang *AI* yang memiliki nilai benar sebanyak 161 orang mahasiswa dan mahasiswa tertarik untuk menulis skripsi dibidang *AI* yang memiliki nilai salah sebanyak 1 orang. Sedangkan untuk mahasiswa tidak tertarik untuk menulis skripsi di bidang *AI* yang memiliki nilai benar sebanyak 38 orang mahasiswa dan mahasiswa tidak tertarik untuk menulis skripsi di bidang *AI* yang memiliki nilai salah sebanyak 0 orang mahasiswa atau tidak ada.

Tahap evaluasi selanjutnya yaitu dengan menganalisa hasil ROC *(Receiver Operating Characteristic curve)*. ROC merupakan kurva yang digunakan untuk mengevaluasi kinerja model klasifikasi. Fungsinya adalah untuk menganalisis kemampuan model dalam membedakan antara kelas positif dan kelas negatif. Nilai ROC bernilai baik jika nilainya lebih besar dari 0.7. Berikut dibawah ini gambar 5 yang menunjukan nilai ROC yang dihasilkan dari hasil uji coba :

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.994	1.000	0.809	0.994	0.892	-0.034	0.801	0.938	Tertarik
	0.000	0.006	0.000	0.000	0.000	-0.034	0.801	0.383	Tidak Tertarik
Weighted Avg.	0.805	0.811	0.655	0.805	0.722	-0.034	0.801	0.833	

Gambar 5. Nilai ROC

Dari gambar 5 dapat diketuahui bahwa nilai ROC dari kedua data set tertarik dan tidak tertarik bernilai 0.801 . Hal ini menunjukkan bahwa model klasifikasi secara keseluruhan mampu membedakan antara kelas positif dan kelas negatif dengan baik. Evaluasi yang terakhir dilakukan dengan menganalisa grafik ROC yang dihasilkan dari proses klasifikasi.



Gambar 6. Grafi ROC Area untuk Kelas Tertarik dan Kelas Tidak Tertarik

Kurva ROC pada gambar 6 menunjukan bahwa model Random Forest memiliki performa yang baik. Kurva mendekati sudut kiri atas grafik, menunjukkan bahwa model dapat mengidentifikasi mahasiswa yang tertarik dengan AI dengan baik sambil meminimalkan kesalahan klasifikasi mahasiswa yang tidak tertarik. Nilai AUC untuk model Random Forest tinggi, menunjukkan bahwa model memiliki kemampuan diskriminatif yang baik dalam membedakan antara mahasiswa yang tertarik dan tidak tertarik dengan AI.Dengan hasil ini, maka kemungkinan terjadinya penurunan minat mahasiswa dalam mengambil bidang *Artificial Intelligence* (*AI*) dalam penulisan skripsi dapat dikurangi dan menekan jumlah mahasiswa yang mengalami kendala dalam penulisan skripsi dibidang AI. Dengan demikian, Algoritma Random Forest dapat memberikan solusi untuk permasalahan analisa ketertarikan mahasiswa dalam mengambil bidang *AI* dalam penulisan skripsi.

4. KESIMPULAN

Algoritma Random Forest dapat melakukan klasifikasi dengan akurat dalam hal menganalisis ketertarikan mahasiswa dalam mengambil bidang *Artificial Intelligence* dalam penulisan skripsi. Algoritma Random Forest mampu menganalisis mahasiswa yang tertarik dan mahasiswa yang tidak tertarik dengan menghasilkan nilai akurasi sebesar 80,5%. Hasil penelitian ini dapat bermanfaat bagi program studi Teknik Informatika Universitas Pelita Bangsa dalam merancang kurikulum dan program pembinaan minat mahasiswa terhadap Artificial Intelligence, Sehingga Universitas Pelita bangsa dapat menjadi kampus yang unggul yang dapat bersaing dalam penyediaan tenaga-tenaga alhi di bidang AI. Universitas Pelita Bangsa perlu memberikan edukasi lebih lanjut tentang AI kepada mahasiswa. Hal ini dapat dilakukan melalui seminar, workshop, atau mata kuliah khusus tentang AI. Dosen perlu mendorong mahasiswa untuk menggunakan AI dalam penelitian skripsi. Hal ini dapat dilakukan dengan memberikan contoh penelitian skripsi yang menggunakan AI atau dengan memberikan bimbingan khusus kepada mahasiswa yang ingin menggunakan AI dalam penelitian mereka. Disamping itu mahasiswa perlu meningkatkan pengetahuan dan

Building of Informatics, Technology and Science (BITS)
Volume 6, No 1, Juni 2024 Page: 559-566
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i1.5332

keterampilan mereka tentang AI. Hal ini dapat dilakukan dengan mengikuti pelatihan, membaca buku dan artikel tentang AI, atau dengan bergabung dengan komunitas AI. Penelitian ini menunjukkan bahwa mahasiswa Universitas Pelita Bangsa Prodi Teknik memiliki tingkat ketertarikan yang tinggi terhadap AI dan memiliki minat yang tinggi untuk menggunakan AI dalam penulisan skripsi. Diperlukan kerjasama dari berbagai pihak untuk meningkatkan pengetahuan dan keterampilan mahasiswa tentang AI agar mereka dapat memanfaatkan AI untuk penelitian skripsi mereka dan untuk meningkatkan peluang kerja mereka di masa depan

REFERENCES

[1] S. Wijayanto, D. A. Prabowo, D. Y. Kristiyanto, and M. Y. Fathoni, "Analisis Sentimen Berbasis Aspek pada Layanan Hotel di Wilayah Kabupaten Banyumas dengan Word2Vec dan Random Forest," *J. Inform. J. Pengemb. IT*, vol. 8, no. 1, pp. 1–3, Dec. 2022, doi: 10.30591/jpit.v8i1.4186.

[2] M. Azhar and H. F. Pardede, "Klasifikasi Dialek Pengujar Bahasa Inggris Menggunakan Random Forest," *J. MEDIA Inform. BUDIDARMA*, vol. 5, no. 2, p. 439, Apr. 2021, doi: 10.30865/mib.v5i2.2754.

[3] M. Azhari, Z. Situmorang, and R. Rosnelly, "Perbandingan Akurasi, Recall, dan Presisi Klasifikasi pada Algoritma C4.5, Random Forest, SVM dan Naive Bayes," *J. MEDIA Inform. BUDIDARMA*, vol. 5, no. 2, p. 640, Apr. 2021, doi: 10.30865/mib.v5i2.2937.

[4] F. Baharuddin and A. Tjahyanto, "Peningkatan Performa Klasifikasi Machine Learning Melalui Perbandingan Metode Machine Learning dan Peningkatan Dataset," *J. Sisfokom Sist. Inf. Dan Komput.*, vol. 11, no. 1, pp. 25–31, Mar. 2022, doi: 10.32736/sisfokom.v11i1.1337.

[5] A. Pratama, A. A. Wicaksana, and A. Razi, "Analisa Kesesuaian Lahan Tanah Untuk Tanaman Padi (Oryza Sativa L.) Dengan Metode Decision Tree Berbasis Web (Studi Kasus Kabupaten Aceh Utara)," *J. Inform. Kaputama JIK*, vol. 6, no. 1, pp. 1–23, Jan. 2022, doi: 10.59697/jik.v6i1.128.

[6] U. Erdiansyah, A. Irmansyah Lubis, and K. Erwansyah, "Komparasi Metode K-Nearest Neighbor dan Random Forest Dalam Prediksi Akurasi Klasifikasi Pengobatan Penyakit Kutil," *J. MEDIA Inform. BUDIDARMA*, vol. 6, no. 1, p. 208, Jan. 2022, doi: 10.30865/mib.v6i1.3373.

[7] S. Abro, S. Shaikh, Z. Hussain, Z. Ali, S. Khan, and G. Mujtaba, "Automatic Hate Speech Detection using Machine Learning: A Comparative Study," *Int. J. Adv. Comput. Sci. Appl.*, vol. 11, no. 8, 2020, doi: 10.14569/IJACSA.2020.0110861.

[8] A. Sanjurjo-de-No, A. M. Pérez-Zuriaga, and A. García, "Analysis and prediction of injury severity in single micromobility crashes with Random Forest," *Heliyon*, vol. 9, no. 12, p. e23062, Dec. 2023, doi: 10.1016/j.heliyon.2023.e23062.

[9] A. U. Zailani and N. L. Hanun, "PENERAPAN ALGORITMA KLASIFIKASI RANDOM FOREST UNTUK PENENTUAN KELAYAKAN PEMBERIAN KREDIT DI KOPERASI MITRA SEJAHTERA," *Infotech J. Technol. Inf.*, vol. 6, no. 1, pp. 7–14, Jun. 2020, doi: 10.37365/jti.v6i1.61.

[10] C. Cai, C. Yang, S. Lu, G. Gao, and J. Na, "Human motion pattern recognition based on the fused random forest algorithm," *Measurement*, vol. 222, p. 113540, Nov. 2023, doi: 10.1016/j.measurement.2023.113540.

[11] R. A. Zuama, S. Rahmatullah, and Y. Yuliani, "Analisis Performa Algoritma Machine Learning pada Prediksi Penyakit Cerebrovascular Accidents," *J. MEDIA Inform. BUDIDARMA*, vol. 6, no. 1, p. 531, Jan. 2022, doi: 10.30865/mib.v6i1.3488.

[12] C. A. A. Soemedhy, N. Trivetisia, N. A. Winanti, D. P. Martiyaningsih, T. W. Utami, and S. Sudianto, "Analisis Komparasi Algoritma Machine Learning untuk Sentiment Analysis (Studi Kasus: Komentar YouTube 'Kekerasan Seksual')," *J. Inform. J. Pengemb. IT*, vol. 7, no. 2, pp. 80–84, May 2022, doi: 10.30591/jpit.v7i2.3547.

[13] K. Kurniabudi, A. Harris, and V. Veronica, "Komparasi Performa Tree-Based Classifier Untuk Deteksi Anomali Pada Data Berdimensi Tinggi dan Tidak Seimbang," *J. MEDIA Inform. BUDIDARMA*, vol. 6, no. 1, p. 370, Jan. 2022, doi: 10.30865/mib.v6i1.3473.

[14] S. Fachid and A. Triayudi, "Perbandingan Algoritma Regresi Linier dan Regresi Random Forest Dalam Memprediksi Kasus Positif Covid-19," *J. MEDIA Inform. BUDIDARMA*, vol. 6, no. 1, p. 68, Jan. 2022, doi: 10.30865/mib.v6i1.3492.

[15] A. C. Mawarni, R. Rusdah, L. L. Hin, and D. Anubhakti, "DETEKSI DINI GEJALA AWAL PENYAKIT DIABETES MENGGUNAKAN ALGORITMA RANDOM FOREST," *IDEALIS Indones. J. Inf. Syst.*, vol. 6, no. 2, pp. 165–171, Jul. 2023, doi: 10.36080/idealis.v6i2.3018.

[16] I. P. A. M. Utama, S. S. Prasetyowati, and Y. Sibaroni, "Multi-Aspect Sentiment Analysis Hotel Review Using RF, SVM, and Naïve Bayes based Hybrid Classifier," *J. MEDIA Inform. BUDIDARMA*, vol. 5, no. 2, p. 630, Apr. 2021, doi: 10.30865/mib.v5i2.2959.

[17] H. A. Salka and K. M. Lhaksmana, "Work Readiness Prediction of Telkom University Students Using Multinomial Logistic Regression and Random Forest Method," *J. MEDIA Inform. BUDIDARMA*, vol. 6, no. 4, p. 1903, Oct. 2022, doi: 10.30865/mib.v6i4.4546.

[18] M. P. K. Dewi and E. B. Setiawan, "Feature Expansion Using Word2vec for Hate Speech Detection on Indonesian Twitter with Classification Using SVM and Random Forest," *J. MEDIA Inform. BUDIDARMA*, vol. 6, no. 2, p. 979, Apr. 2022, doi: 10.30865/mib.v6i2.3855.

[19] D. A. Rachmawati, N. A. Ibadurrahman, J. Zeniarja, and N. Hendriyanto, "IMPLEMENTATION OF THE RANDOM FOREST ALGORITHM IN CLASSIFYING THE ACCURACY OF GRADUATION TIME FOR COMPUTER ENGINEERING STUDENTS AT DIAN NUSWANTORO UNIVERSITY," *J. Tek. Inform. Jutif*, vol. 4, no. 3, pp. 565–572, Jun. 2023, doi: 10.52436/1.jutif.2023.4.3.920.

[20] N. Widjiyati, "Implementasi Algoritme Random Forest Pada Klasifikasi Dataset Credit Approval," *J. Janitra Inform. Dan Sist. Inf.*, vol. 1, no. 1, pp. 1–7, Apr. 2021, doi: 10.25008/janitra.v1i1.118.