Jutif_Syafitri,_Farradinna,_Jayanti,_Arta_2023.pdf

by

Submission date: 28-Apr-2023 09:15AM (UTC+0700)
Submission ID: 2077790941
File name: Jutif_Syafitri,_Farradinna,_Jayanti,_Arta_2023.pdf (791.51K)
Word count: 5376
Character count: 30714

Jurnal Teknik Informatika (JUTIF)
Vol. 4, No. 2, April 2023, hlm. 381-390

DOI: https://doi.org/10.52436/1.jutif.2023.4.2.590
p-ISSN: 2723-3863
e-ISSN: 2723-3871

MACHINE LEARNING TO CREATE DECISION TREE MODEL TO PREDICT OUTCOME OF ENTERPRENEURSHIP PSYCHOLOGICAL READINESS (EPR)

Nesi Syafitri*[1], Syarifah Farradinna[2], Wella Jayanti[3], Yudhi Arta[4]

[1,4]Teknik Informatika, Fakultas Teknik, Universitas Islam Riau, Indonesia
[2,3]Psikologi, Fakultas Psikologi, Universitas Islam Riau, Indonesia
Email: [1]nesisyafitri@eng.uir.ac.id, [2]syarifah.farradinna@psy.uir.ac.id, [3]wellajayanti17@gmail.com, [4]yudhiarta@eng.uir.ac.id

(Naskah masuk: 31 Agustus 2022, Revisi: 11 Oktober 2022, Diterbitkan: 23 Maret 2023)

Abstract

This study aims to create a decision tree model using machine learning to predict psychological readiness for entrepreneurship in college graduates. This research was conducted through several stages of research. In the early stages, a survey was conducted on 700 students from several universities in Riau aged between 17-25 years. The survey was conducted using the Entrepreneur Psychology Readiness (EPR) instrument. Furthermore, the survey data was validated and obtained 604 valid data to be used in forming machine learning models The urgency of this research is to find a number of decision rules from the best decision tree model to be used in building AI-based counseling applications in measuring entrepreneurial psychology readiness for college graduates. In this research, the decision tree model that is formed is divided into 2 models, namely: decision tree with pruning model and decision tree with unpruning. The pruning decision tree model produces 180 decision rules, while the unpruning model produces 121 decision rules. Good accuracy results are obtained in the pruned decision tree, which is above 99% in the use training set mode, and 82.87% in the percentage split mode. Meanwhile, the accuracy results on the unpruned decision tree are 90.18% with the use training set mode test, and 80.38% in the percentage split mode. The decision tree model with pruning technique has better performance than the unpruning decision tree model.

Keywords: *Accuracy*, *Decision Tree, Enterprenuer Readiness, Machine Learning, Psychological assessment*

MACHINE LEARNING UNTUK MEMBUAT MODEL DECISION TREE GUNA MEMPREDIKSI KESIAPAN PSIKOLOGI BERWIRAUSAHA

Abstrak

Penelitian ini bertujuan untuk membuat sebuah model pohon keputusan (Decision Tree Model) dengan menggunakan machine learning untuk memprediksi kesiapan psikologi untuk berwirausaha pada lulusan perguruan tinggi. Penelitian ini dilakukan melalui beberapa tahap penelitian. Pada tahapan awal dilakukan survey terhadap 700 orang mahasiswa dari beberapa perguruan tinggi di Riau yang berusia antara 17-25 tahun. Survey dilakukan dengan menggunakan instrument *Enterprenuer Psikology Readiness* (EPR). Selanjutnya data survey divalidasi dan diperoleh 604 data valid untuk digunakan dalam membentuk model *machine learning*. Urgensi dari penelitian ini adalah menemukan sejumlah aturan keputusan (*decision rules*) dari model pohon keputusan (*decision tree*) terbaik untuk digunakan dalam membangun aplikasi konseling berbasis AI dalam mengukur kesiapan psikologi kewirausahaan bagi lulusan perguruan tinggi. Pada penelitian ini model *decision tree* yang dibentuk, dibedakan dalam 2 model yaitu: decision tree dengan model *pruning* dan decision tree dengan *unpruning*. *Decision tree* model *pruning* menghasilkan 180 aturan keputusan (*decision rules*), sedangkan model unpruning menghasilkan 121 decision rules. Hasil akurasi yang baik diperoleh pada decision tree dengan pruned yaitu diatas 99% pada *mode use training set*, dan 82.87% pada *mode percentage split*. Sementara hasil akurasi pada *decision tree unpruned* yaitu 90.18% dengan pengujian *mode use training set*, dan 80.38% pada *mode percentage split*. Model *decision tree* dengan teknik *pruning* memiliki kinerja lebih baik dibandingkan dengan model *decision tree unpruning*.

Kata kunci: Akurasi, *decision tree, machine learning, kesiapan berwirausaha, pengukuran psikologi*.

1. PENDAHULUAN

Kewirausahaan dipandang sebagai kekuatan pendorong terpenting dalam pertumbuhan ekonomi,

karena kemampuannya untuk menciptakan pekerjaan, meningkatkan pendapatan individu, mempercepat pengentasan kemiskinan dan mendukung inovasi kreatif[1].

Pendidikan kewirausahaan di Indonesia lebih berfokus pada pengembangan wawasan kognitif, pengetahuan manajemen, rencana bisnis, pemasaran, keuangan, pengembangan inovasi dan kreatifitas. Sementara, faktor psikologis juga memegang peran yang sama pentingnya dalam pendidikan kewirausahaan, hal ini terbukti bahwa ketahanan kepribadian (personality hardiness) menawarkan pengaruh yang tinggi terhadap intensi kewirausahaan[2]. Faktor psikologis seperti kebutuhan untuk berprestasi, kecenderungan untuk mengambil risiko dan kemampuan memegang kendali serta faktor sosial seperti latar belakang keluarga, sistem pendidikan dan status social, merupakan indicator yang cukup menonjol dan signifikan untuk menjadi wirausaha[3].

Berbagai instrumen kesiapan psikologi kewirausahaan terus dikembangkan para ahli dan mulai berinovasi dengan teknologi terkini, salah satunya dengan penerapan kecerdasan buatan / artificial intelligence (AI). Sistem berbasis AI diketahui sebagai system yang memberikan layanan evaluasi yang akurat dan cerdas, mengurangi pekerjaan berulang dalam konseling kesiapan psikologis dan memberikan informasi perbandingan dengan data yang tersedia[4].

AI ini telah banyak digunakan diberbagai bidang seperti diagnosa di bidang medis, simulasi dalam matematika dan fisika, klasifikasi gambar dalam biologi dan kimia, dan banyak lagi. Beberapa tahun terakhir bidang psikologi pun telah mulai eksperimen dengan menggunakan AI untuk memprediksi dan mengklasifikasikan di banyak bidang studi: mulai dari mengukur tingkat rasa sakit dari pemindaian otak[5], menerapkan teknik pembelajaran mesin untuk lebih memahami kepribadian [6], dan mendeteksi kebutuhan manusia dalam peristiwa kritis[7], untuk memprediksi penggunaan media sosial yang bermasalah, dan penyalahgunaan alkohol di masa depan[8]–[11], untuk membantu diagnosis dan prognosis penyakit dan gangguan mental, mendeteksi tingkat depresi, dan memprediksi risiko perilaku bunuh diri dan melukai diri sendiri[12]. Para peneliti bahkan telah menganalisis bagaimana membuat model AI ini lebih baik khusus untuk bidang psikologi [13]–[15].

Salah satu disiplin ilmu dalam AI adalah machine learning. Machine learning merupakan teknik untuk meningkatkan kinerja sistem dengan mempelajari pengalaman melalui metode komputasi [16]. Tujuan utama dari machine learning adalah untuk belajar dari sejumlah data yang besar[17]. Algoritma machine learning yang dapat digunakan dalam menemukan pola data, salah satunya adalah decision tree. Decision tree, jenis machine learning tertentu, didasarkan pada kovariat untuk membuat model dalam memprediksi hasil[18].

Tujuan dari penelitian ini untuk menemukan sejumlah aturan keputusan (*decision rules*) dari model pohon keputusan (*decision tree*) terbaik untuk memprediksi kesiapan psikologi kewirausahaan menggunakan 8 (delapan) parameter pengukuran psikologi. Sehingga model pohon keputusan ini nanti dapat digunakan dalam membangun aplikasi konseling psikologi berbasis cerdas.

2. METODE PENELITIAN

2.1. Data Penelitian

Data primer yang digunakan dalam penelitian ini bersumber dari data hasil kuisioner yang disebarkan kepada sejumlah mahasiswa dari perguruan tinggi yang ada di wilayah Riau menggunakan alat ukur Kesiapan Psikologi Kewirausahaan / Enterprenuership Psycological Readiness (EPR) instrument. Alat ukur ini menilai kesiapan individu secara psikologis dan sosiologis, untuk menentukan apakah seorang individu menunjukkan kesiapan menjadi seorang wirausaha atau tidak. Instrumen EPR mengukur pada 8 parameter dengan 59 item pertanyaan yang dijawab menggunakan skala Likert lima poin, yaitu: Sangat Tidak Setuju (1), Tidak Setuju (2), Ragu-ragu (3), Setuju (4), dan Sangat Setuju (5). Uraian 8(delapan) parameter denagn jumlah item pertanyaan setiap parameter pada EPR instrument terlihat pada Tabel 1.

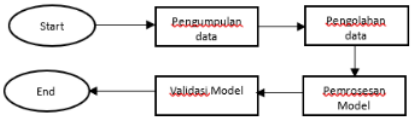
Tabel 1. Parameter pada Instrumen EPR

Parameter	Jumlah item pertanyaan
Personal Knowledge	10
Personal Adversity	13
Commited Certain Action	8
Willingness to learn	8
Personal Relationship to Others	8
Personal growth	6
Passion Achieved	3
Related Person Support	3
Total	59

Output prediksi kesiapan psikologis berwirausaha ini menghasilkan 4 (empat) kategori yaitu: tidak siap, diberi pelatihan, direkomendasikan, dan siap berwirausaha.

2.2. Tahap Penelitian

Penelitian ini terdiri dari beberapa tahapan, yaitu: pengumpulan data, pengolahan data, pemrosesan model, evaluasi model. Gambar 1 menunjukkan bagan alir dari tahap penelitian.



Gambar 1. Bagan Alir Tahap Penelitian

a. Pengumpulan Data

Pada tahap pengumpulan data dilakukan dengan menyebarkan kuisioner kepada responden mahasiswa yang berasal dari beberapa perguruan tinggi di wilayah Riau. Karakteristik responden berusia antara 17 sampai 25 tahun yang terdiri dari laki-laki dan perempuan. Jumlah data yang diperoleh sebanyak 604 data. Kuisioner digunakan untuk menilai tingkat kesiapan psikologis dalam berwirausaha setelah mahasiswa lulus dari perguruan tinggi. Kuisioner berisikan 59 item pertanyaan untuk mengukur setiap responden pada 8 parameter psikologi. Setiap pertanyaan dijawab menggunakan skala Likert lima poin, yaitu: Sangat Tidak Setuju (1), Tidak Setuju (2), Ragu-ragu (3), Setuju (4), dan Sangat Setuju (5).

b. Pengolahan data

Langkah awal yang dilakukan dalam pengolahan data adalah memverifikasi data untuk memastikan tidak ada jawaban yang kosong atau terlewati. selanjutnya menentukan skoring dari setiap parameter berdasarkan nilai skor dari jawaban responden. Setelah skor diperoleh dilakukan penentuan kategorisasi setiap parameter. Norma kategorisasi parameter yang digunakan berpedoman pada norma kategorisasi yang disusun oleh Azwar[16], dalam 5 (lima) kategori diagnosis, yaitu: sangat rendah, rendah, sedang, tinggi dan sangat tinggi. Skala norma kategorisasi ini mengacu pada Tabel 2.

Tabel 2. Skala Norma Kategorisasi

Kategorisasi	Norma Kategorisasi
Sangat Rendah	$X \leq \mu - 1.5\sigma$
Rendah	$\mu - 1.5\sigma < X \leq \mu - 0.5\sigma$
Sedang	$\mu - 0.5\sigma < X \leq \mu + 0.5\sigma$
Tinggi	$\mu + 0.5\sigma < X \leq \mu + 1.5\sigma$
Sangat Tinggi	$X > \mu + 1.5\sigma$

dengan X adalah skor parameter, μ(means) adalah nilai rata-rata teoritis dari skor maksimum dan minimum dan σ(standar deviasi) adalah luas jarak rentang yang dibagi dalam 6 satuan deviasi sebaran.

Berdasarkan skala likert lima poin yang digunakan untuk menjawab item pertanyaan, maka skor minimum dan maksimum yang diperoleh dari setiap parameter instrument EPR digunakan untuk memperoleh nilai μ(means) dan nilai σ(standar deviasi) seperti terlihat pada Tabel 3.

Tabel 3. Deskripsi Data Penelitian

Parameter	X min	X max	μ (Mean)	σ (Sd)
Personal Knowledge	10	50	30	6.7
Personal Adversity	13	65	26	8.67
Commited Certain Action	8	40	24	5.33
Willingness to learn	8	40	24	5.33
Personal Relationship to Others	8	40	24	5.33
Personal growth	6	30	18	4
Passion Achieved	3	15	9	2
Related Person Support	3	15	9	2

Pada parameter Personal knowledge terdapat 10 item pertanyaan, maka skor minimum (x min): 1 x 10 = 10; skor maksimum (x max): 5 x 10 = 50; nilai mean teoritis (nilai μ) : (50+10)/2 = 30; dan standar deviasi (σ / sd) : (50-10)/6 = 6.7. Perhitungan yang sama dilakukan pada parameter yang lain.

Dengan mengacu pada Tabel 2 dan Tabel 3, maka rentang skor untuk kategorisasi pada parameter Personal Knowledge terlihat pada Tabel 4.

Tabel 4. Skor Kategorisasi pada Parameter Personal Knowledge

Kategorisasi	Norma Kategorisasi	Rentang Skor
Sangat Rendah	$X \leq \mu - 1.5\sigma$	$X \leq 20$
Rendah	$\mu - 1.5\sigma < X \leq \mu - 0.5\sigma$	$20 < X \leq 26.67$
Sedang	$\mu - 0.5\sigma < X \leq \mu + 0.5\sigma$	$26.67 < X \leq 33.33$
Tinggi	$\mu + 0.5\sigma < X \leq \mu + 1.5\sigma$	$33.33 < X \leq 40$
Sangat Tinggi	$X > \mu + 1.5\sigma$	$X > 40$

Tingkat kategori sangat rendah diperoleh jika skor kecil sama 20, kategori "rendah" dengan skor diantara 20 dan 26.67, kategori sedang berada antara skor 26.67 dan 33.33, kategori tinggi dengan skor antara 33.33 dan 40, dan kategori sangat tinggi diatas skor 40.

Rentang skor untuk kategorisasi pada parameter Personal Adversity dengan acuan nilai mean = 26 dan standar deviasi =8.67 adalah seperti yang terlihat pada Tabel 5.

Tabel 5. Skor Kategorisasi pada Parameter Personal Adversity

Kategorisasi	Norma Kategorisasi	Rentang Skor
Sangat Rendah	$X \leq \mu - 1.5\sigma$	$X \leq 26$
Rendah	$\mu - 1.5\sigma < X \leq \mu - 0.5\sigma$	$26 < X \leq 34.67$
Sedang	$\mu - 0.5\sigma < X \leq \mu + 0.5\sigma$	$34.67 < X \leq 43.33$
Tinggi	$\mu + 0.5\sigma < X \leq \mu + 1.5\sigma$	$43.33 < X \leq 52$
Sangat Tinggi	$X > \mu + 1.5\sigma$	$X > 52$

Tingkat kategori sangat rendah diperoleh jika skor kecil sama 26, kategori "rendah" dengan skor diantara 26 dan 34.67, kategori sedang berada antara skor 34.67 dan 43.33, kategori tinggi dengan skor antara 43.33 dan 52, dan kategori sangat tinggi diatas skor 52.

Pada parameter Commited Certain rentang skor kategorisasi yang diperoleh dengan acuan nilai mean = 24 dan standar deviasi = 5.33 terlihat pada Tabel 6.

Tabel 6. Skor Kategorisasi pada Parameter Commited Certain Action

Kategorisasi	Norma Kategorisasi	Rentang Skor
Sangat Rendah	$X \leq \mu - 1.5\sigma$	$X \leq 16$
Rendah	$\mu - 1.5\sigma < X \leq \mu - 0.5\sigma$	$16 < X \leq 21.33$
Sedang	$\mu - 0.5\sigma < X \leq \mu + 0.5\sigma$	$21.33 < X \leq 26.67$
Tinggi	$\mu + 0.5\sigma < X \leq \mu + 1.5\sigma$	$26.67 < X \leq 32$
Sangat Tinggi	$X > \mu + 1.5\sigma$	$X > 32$

Tingkat kategori sangat rendah diperoleh jika skor kecil sama 16, kategori "rendah" dengan skor diantara 16 dan 21.33, kategori sedang berada antara skor 21.33 dan 26.67, kategori tinggi dengan skor

antara 26.67 dan 32, dan kategori sangat tinggi diatas skor 32.

Selanjutnya pada parameter Willingness to Learn rentang skor kategorisasi yang diperoleh dengan acuan nilai mean = 24 dan standar deviasi = 5.33 terlihat pada Tabel 7.

Tabel 7. Skor Kategorisasi pada Parameter Willingness to Learn

Kategorisasi	Norma Kategorisasi	Rentang Skor
Sangat Rendah	$X \leq \mu - 1.5\sigma$	$X \leq 16$
Rendah	$\mu - 1.5\sigma < X \leq \mu - 0.5\sigma$	$16 < X \leq 21.33$
Sedang	$\mu - 0.5\sigma < X \leq \mu + 0.5\sigma$	$21.33 < X \leq 26.67$
Tinggi	$\mu + 0.5\sigma < X \leq \mu + 1.5\sigma$	$26.67 < X \leq 32$
Sangat Tinggi	$X > \mu + 1.5\sigma$	$X > 32$

Tingkat kategori sangat rendah diperoleh jika skor kecil sama 16, kategori "rendah" dengan skor diantara 16 dan 21.33, kategori sedang berada antara skor 21.33 dan 26.67, kategori tinggi dengan skor antara 26.67 dan 32, dan kategori sangat tinggi diatas skor 32.

Berikut pada parameter Personal Relationship to Others rentang skor kategorisasi yang diperoleh dengan acuan nilai mean = 24 dan standar deviasi = 5.33 terlihat pada Tabel 8.

Tabel 8. Skor Kategorisasi pada Parameter Personal Relationship to Others

Kategorisasi	Norma Kategorisasi	Rentang Skor
Sangat Rendah	$X \leq \mu - 1.5\sigma$	$X \leq 16$
Rendah	$\mu - 1.5\sigma < X \leq \mu - 0.5\sigma$	$16 < X \leq 21.33$
Sedang	$\mu - 0.5\sigma < X \leq \mu + 0.5\sigma$	$21.33 < X \leq 26.67$
Tinggi	$\mu + 0.5\sigma < X \leq \mu + 1.5\sigma$	$26.67 < X \leq 32$
Sangat Tinggi	$X > \mu + 1.5\sigma$	$X > 32$

Skor dibawah 16 berada pada tingkat kategori sangat rendah, skor antara 16 dan 21.33 ditingkat rendah, skor 21.33 sampai 26.67 berada ditingkat sedang, antara 26.67 dan 32 berada ditingkat tinggi dan diatas skor 32 berada ditingkat sangat tinggi.

Parameter Personal Grow dengan 6 (enam) item pertanyaan maka diperoleh nilai mean = 18 dan standar deviasi = 4. Rentang skor kategorisasi yang diperoleh terlihat pada Tabel 9

Tabel 9. Skor Kategorisasi pada Parameter Personal Growth

Kategorisasi	Norma Kategorisasi	Rentang Skor
Sangat Rendah	$X \leq \mu - 1.5\sigma$	$X \leq 12$
Rendah	$\mu - 1.5\sigma < X \leq \mu - 0.5\sigma$	$12 < X \leq 16$
Sedang	$\mu - 0.5\sigma < X \leq \mu + 0.5\sigma$	$16 < X \leq 20$
Tinggi	$\mu + 0.5\sigma < X \leq \mu + 1.5\sigma$	$20 < X \leq 24$
Sangat Tinggi	$X > \mu + 1.5\sigma$	$X > 24$

Skor dibawah 12 berada pada tingkat kategori sangat rendah, skor antara 12 dan 16 ditingkat rendah, skor 16 sampai 20 berada ditingkat sedang, antara 20 dan 24 berada ditingkat tinggi dan diatas skor 24 berada ditingkat sangat tinggi.

Pada parameter Personal Achieve dan parameter Related Person Support masing-masing terdapat 3 (tiga) item pertanyaan dengan nilai mean = 9 dan standar deviasi = 2. Rentang skor kategorisasi yang diperoleh terlihat pada Tabel 10 dan Tabel 11

Tabel 10. Skor Kategorisasi pada Parameter Personal Achieve

Kategorisasi	Norma Kategorisasi	Rentang Skor
Sangat Rendah	$X \leq \mu - 1.5\sigma$	$X \leq 6$
Rendah	$\mu - 1.5\sigma < X \leq \mu - 0.5\sigma$	$6 < X \leq 8$
Sedang	$\mu - 0.5\sigma < X \leq \mu + 0.5\sigma$	$8 < X \leq 10$
Tinggi	$\mu + 0.5\sigma < X \leq \mu + 1.5\sigma$	$10 < X \leq 12$
Sangat Tinggi	$X > \mu + 1.5\sigma$	$X > 12$

Tabel 11. Skor Kategorisasi pada Parameter Related Person Support

Kategorisasi	Norma Kategorisasi	Rentang Skor
Sangat Rendah	$X \leq \mu - 1.5\sigma$	$X \leq 6$
Rendah	$\mu - 1.5\sigma < X \leq \mu - 0.5\sigma$	$6 < X \leq 8$
Sedang	$\mu - 0.5\sigma < X \leq \mu + 0.5\sigma$	$8 < X \leq 10$
Tinggi	$\mu + 0.5\sigma < X \leq \mu + 1.5\sigma$	$10 < X \leq 12$
Sangat Tinggi	$X > \mu + 1.5\sigma$	$X > 12$

Skor dibawah 6 berada pada tingkat kategori sangat rendah, skor antara 6 dan 8 ditingkat rendah, skor 8 sampai 10 berada ditingkat sedang, antara 10 dan 12 berada ditingkat tinggi dan diatas skor 12 berada ditingkat sangat tinggi

Setelah skor dan kategorisasi setiap parameter diperoleh, kemudian dilanjutkan untuk menghitung total skor output prediksi dan menentukan kategorisasi output menggunakan norma kategori 4 tingkat, terlihat pada Tabel 12.

Tabel 12. Skor Kategorisasi pada Output Prediksi

Kategorisasi	Norma Kategorisasi	Rentang Skor
Tidak Siap	$X \leq \mu - 1\sigma$	$X \leq 138$
Diberi Pelatihan	$\mu - 1\sigma < X \leq \mu$	$138 < X \leq 177$
Direkomendasikan	$\mu < X \leq \mu + 1\sigma$	$177 < X \leq 216$
Siap berwirausaha	$X > \mu + 1\sigma$	$X > 216$

Data yang diperoleh sebanyak 604 data ditentukan kategorisasi setiap parameter dan output prediksinya sesuai rentang skor yang diperoleh.

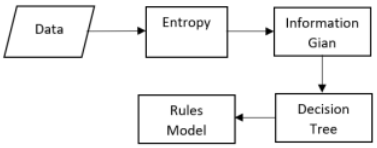
c. Pemrosesan Model

Model machine learning yang dapat digunakan dalam menemukan pola data, salah satunya adalah decision tree. Decision tree adalah algoritma yang umum digunakan untuk mengambil sebuah keputusan [19]. Decision tree merupakan algoritma yang baik digunakan untuk klasifikasi atau prediksi [20]. Decision tree adalah metode klasifikasi yang melibatkan konstruksi pohon keputusan yang terdiri dari node keputusan yang dihubungkan dengan cabang-cabang dari simpul akar (root node) sampai ke node daun (leaf node)[21]. Algoritma decision tree yang digunakan pada penelitian ini adalah algoritma ID3.

Konsep entropi digunakan untuk menentukan atribut mana yang menjadi root node dan branch node pada pohon keputusan, Atribut dengan nilai information gian tertinggi akan terpilih sebagai node pada decision tree yang dibentuk. Kemudian menentukan atribut yang akan menjadi internal node untuk setiap cabang dari parent node, dan membuat leaf node (simpul keputusan) ketika pemilihan atribut tidak dapat digunakan lagi.

Dari model decision tree yang terbentuk akan dilakukan proses pembentukan decision rules model.

Pembentukan rules dilakukan dengan menelusuri root node melaui branch node hingga mencapai leaf node. Gambar 4 menunjukkan alur kalkulasi pada algoritma ID3.



Gambar 4. Alur Kalkulasi ID3

Persamaan yang digunakan dalam algoritma ID3 ini meliputi:

1) Perhitungan Nilai Entropy

Entropi digunakan sebagai parameter untuk mengukur heterogenitas (keragaman) suatu sampel data. Semakin kecil nilai Entropy, semakin baik digunakan dalam mengekstraksi kelas. Nilai entropi ditentukan dengan persamaan:

$$Entropy(S) = \sum_{i=1}^{n} -(P_i log_2 P_{i)} \quad (1)$$

dengan Pi menyatakan proposi kelas ke-i pada output.

2) Perhitungan Nilai Information Gian

Nilai yang diperoleh dari perhitungan entropi masih belum asli tetapi pengukuran efektivitas atribut dalam mengklasifikasikan data latih dapat ditentukan oleh informasi yang telah diperoleh, Information gian diperoleh dengan persamaan:

$$Gian(S,A) = Entropy(S) - \sum_{i=1}^{n} \frac{|S_i|}{|S|} * Entropy(S)$$

(2)

dimana S adalah kumpulan data pelatihan, A adalah atribut, n adalah jumlah partisi dalam atribut A, dan Si adalah jumlah partisi ke-i.

d. Validasi Model

Validitas model untuk mengetahui kinerja dari model machine learning yang dihasilkan. Untuk evaluasi model dari decision tree, digunakan 4(empat) jenis evaluasi, yaitu: precision, recall, dan accuracy. Pengukuran menggunakan Confusion Matrix dengan tabel yang ditampilkan pada Tabel 13.

abel 13. Confusion Matrix

Kelas	Terklasifikasi Positif	Terklasifikasi Negatif
Positif	TP	FN
Negatif	FP	TN

Untuk menghitung precision, recall, dan accuracy menggunakan persamaan (3), (4), (5).

$$Precision = \frac{TP}{TP+FP} \quad (3)$$

$$Recall = \frac{TP}{TP+FN} \quad (4)$$

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

Confusion matrix sangat berguna untuk menganalisis kualitas model klasifikasi dalam mengenali tuple-tuple dari kelas yang ada. TP (True Positive) dan TN (True Negative) menyatakan model klasifikasi mengenali tuple dengan benar, dimana TP adalah data positif yang diprediksi benar, TN adalah data negative yang diprediksi benar. Sebaliknya FP (False Positive) dan FN (False Negative) menyatakan model klasifikasi salah dalam mengenali tuple. FP adalah data negative namun diprediksi sebagai data positif, FN adalah data positif namun diprediksi sebagai data negative.

Accuracy atau tingkat pengenalan menyatakan persentasse dari jumlah tuple dalam data uji yang diklasifikasikan dengan benar oleh model klasifikasi. Precision adalah ukuran kepastian, yaitu berapa persentase tuple dilabeli sebagai positif adalah benar pada kenyataanya. Recall adalah ukuran kelengkapan, yaitu berapa persentase tuple positif yang dilabeli sebagai positif.

3. HASIL DAN PEMBAHASAN

Penelitian ini menerapkan pendekatan machine learning dengan menggunakan algoritma decision tree ID3, dimana data yang digunakan adalah data yang didapatkan dari hasil kuisioner yang disebarkan kepada mahasiswa yang berasal dari perguruan tinggi di wilayah Riau dan mendapatkan 604 data valid. Alat bantu yang dipilih penulis dalam membentuk decision tree menggunakan RapidMiner.

3.1. Proses Pengolahan Data

Data kuisioner yang sudah terkumpul sebanyak 604 data, selanjutnya diolah untuk diskoring berdasarkan nilai skala dari jawaban responden. Data skoring kuisioner terlihat pada Gambar 2:

B	C	D	E	F	G	H	I	J	K
Responden	Faktor 1	Faktor 2	Faktor 3	Faktor 4	Faktor 5	Faktor 6	Faktor 7	Faktor 8	TOTAL
Responden 1	44	44	13	25	22	148	195	365	856
Responden 2	50	52	19	23	27	171	221	419	982
Responden 3	43	51	11	26	25	156	207	388	907
Responden 4	35	47	21	25	26	154	205	385	898
Responden 5	40	49	25	25	27	166	218	411	961
Responden 6	30	38	19	23	17	127	167	311	732
Responden 7	38	55	34	34	29	190	253	472	1105
Responden 8	35	39	28	28	24	154	206	384	898
Responden 9	19	52	9	22	28	122	164	306	714
Responden 10	41	46	18	33	32	170	235	437	1012
Responden 599	35	51	27	32	28	165	217	402	949
Responden 600	40	52	34	34	32	192	258	482	1124
Responden 601	30	39	20	24	26	139	189	354	821
Responden 602	39	53	15	27	26	160	213	399	932
Responden 603	50	55	20	31	35	191	257	483	1122
Responden 604	49	59	34	37	36	215	288	539	1257

Gambar 2. Data Skoring Kuisioner

Data skoring kuisioner kemudian akan dilakukan kategorisasi setiap parameter dengan mengacu pada skala norma kategorisasi di Tabel 2. Hasil kategorisasi setiap parameter dapat dilihat pada Gambar 3. Data inilah selanjutnya digunakan pada tahap pembentukan model machine learning.

A	B	C	D	E	F	G	H	I
Personal Knowledge	Personal Adversity	Committed Certain Action	Willingness to Learn	Personal Relationship to other	Personal Growth	Passion Achieved	Related Person Support	Kategori
P1	P2	P3	P4	P5	P6	P7	P8	
Sedang	Rendah	Sedang	Rendah	Sedang	Rendah	Rendah	Rendah	Diberi Pelatihan
Tinggi	Sedang	Sedang	Sedang	Sedang	Sedang	Sedang	Sedang	Direkomendasikan
Tinggi	Sangat Tinggi	Sangat Rendah	Tinggi	Rendah	Sangat Tinggi	Sedang	Tinggi	Direkomendasikan
Sangat Tinggi	Sangat Tinggi	Rendah	Sangat Tinggi	Tinggi	Sangat Tinggi	Rendah	Tinggi	Siap
Sangat Tinggi	Sangat Tinggi	Sedang	Tinggi	Tinggi	Sangat Tinggi	Sangat Tinggi	Rendah	Siap
Sangat Tinggi	Sangat Tinggi	Sangat Rendah	Tinggi	Sedang	Tinggi	Sedang	Rendah	Direkomendasikan
Sangat Tinggi	Tinggi	Tinggi	Tinggi	Sangat Tinggi	Sangat Tinggi	Sedang	Sangat Tinggi	Siap
Sangat Tinggi	Sangat Tinggi	Sangat Tinggi	Sangat Tinggi	Sangat Tinggi	Sangat Tinggi	Sangat Tinggi	Sangat Tinggi	Siap
Tinggi	Tinggi	Sedang	Tinggi	Sedang	Tinggi	Tinggi	Tinggi	Direkomendasikan
Sedang	Rendah	Sedang	Sedang	Sedang	Sedang	Sedang	Sangat Tinggi	Direkomendasikan
Sangat Tinggi	Sangat Tinggi	Tinggi	Tinggi	Sangat Tinggi	Sangat Tinggi	Tinggi	Rendah	Siap
Tinggi	Sangat Tinggi	Rendah	Sedang	Tinggi	Sedang	Sangat Tinggi	Sedang	Direkomendasikan
Sangat Tinggi	Sangat Tinggi	Rendah	Tinggi	Sangat Tinggi	Sangat Tinggi	Sedang	Tinggi	Siap
Sangat Tinggi	Sangat Tinggi	Sangat Tinggi	Tinggi	Sangat Tinggi	Sangat Tinggi	Tinggi	Sangat Tinggi	Siap
Sedang	Sedang	Sedang	Tinggi	Sedang	Sedang	Sedang	Sedang	Direkomendasikan
Rendah	Sedang	Rendah	Rendah	Sedang	Sangat Tinggi	Sangat Rendah	Tinggi	Diberi Pelatihan
Tinggi	Tinggi	Sangat Rendah	Sangat Rendah	Tinggi	Tinggi	Rendah	Sedang	Direkomendasikan
Sangat Tinggi	Tinggi	Sangat Rendah	Sedang	Rendah	Sedang	Sedang	Sangat Rendah	Direkomendasikan
Tinggi	Sangat Tinggi	Tinggi	Tinggi	Sangat Tinggi	Tinggi	Tinggi	Sangat Tinggi	Siap

Gambar 3. Data Hasil Kategorisasi

3.2. Proses Pemodelan Decision Tree

Pembentukan *decision tree* menggunakan algoritma ID3 dengan melibatkan 2 teknik yang berbeda. Pada penelitian ini diuji 2 teknik sebagai perbandingan, khususnya dari sisi akurasi. Model pertama dibuat sebuah decision tree dengan teknik *unpruning*, model kedua menggunakan teknik *pruning*.

Dengan menggunakan data training sebanyak 604, *decision tree model pruning* yang terbentuk menghasilkan 180 *decision rules* dengan node akar yang terpilih pada parameter "*Willingness to Learn*". Susunan decision rules yang dihasilkan dari decision tree pruning dapat dilihat pada Gambar 5.

```
Willingness to Learn = Rendah
| Personal Adversity = Rendah: Diberi Pelatihan {Direkomendasikan=0, Siap=0, Diberi Pelatihan=3, Tidak siap=0}
| Personal Adversity = Sangat Tinggi
| | Personal Growth = Rendah: Diberi Pelatihan {Direkomendasikan=0, Siap=0, Diberi Pelatihan=1, Tidak siap=0}
| | Personal Growth = Sangat Tinggi: Direkomendasikan {Direkomendasikan=5, Siap=0, Diberi Pelatihan=0, Tidak siap=0}
| | Personal Growth = Sedang: Direkomendasikan {Direkomendasikan=2, Siap=0, Diberi Pelatihan=0, Tidak siap=0}
| | Personal Growth = Tinggi: Direkomendasikan {Direkomendasikan=1, Siap=0, Diberi Pelatihan=0, Tidak siap=0}
| Personal Adversity = Sedang
| | Personal Relationship to other = Rendah: Diberi Pelatihan {Direkomendasikan=0, Siap=0, Diberi Pelatihan=7, Tidak siap=0}
| | Personal Relationship to other = Sangat Rendah
| | | Personal Growth = Rendah: Tidak siap {Direkomendasikan=0, Siap=0, Diberi Pelatihan=0, Tidak siap=1}
| | | Personal Growth = Sedang: Diberi Pelatihan {Direkomendasikan=0, Siap=0, Diberi Pelatihan=1, Tidak siap=0}
| | | Personal Growth = Tinggi: Diberi Pelatihan {Direkomendasikan=0, Siap=0, Diberi Pelatihan=1, Tidak siap=0}
| | Personal Relationship to other = Sedang: Diberi Pelatihan {Direkomendasikan=0, Siap=0, Diberi Pelatihan=8, Tidak siap=0}
| | Personal Relationship to other = Tinggi: Direkomendasikan {Direkomendasikan=1, Siap=0, Diberi Pelatihan=0, Tidak siap=0}
```

Gambar 5. Decision Rule dari Decision Tree Prunning

```
Willingness to Learn = Rendah
| Personal Adversity = Rendah: Diberi Pelatihan {Direkomendasikan=0, Siap=0, Diberi Pelatihan=3, Tidak siap=0}
| Personal Adversity = Sangat Tinggi: Direkomendasikan {Direkomendasikan=8, Siap=0, Diberi Pelatihan=1, Tidak siap=0}
| Personal Adversity = Sedang
| | Personal Knowledge = Rendah: Diberi Pelatihan {Direkomendasikan=0, Siap=0, Diberi Pelatihan=3, Tidak siap=0}
| | Personal Knowledge = Sangat Rendah: Diberi Pelatihan {Direkomendasikan=0, Siap=0, Diberi Pelatihan=1, Tidak siap=1}
| | Personal Knowledge = Sedang: Diberi Pelatihan {Direkomendasikan=1, Siap=0, Diberi Pelatihan=7, Tidak siap=0}
| | Personal Knowledge = Tinggi: Diberi Pelatihan {Direkomendasikan=0, Siap=0, Diberi Pelatihan=6, Tidak siap=0}
| Personal Adversity = Tinggi
| | Committed Certain Action = Rendah: Diberi Pelatihan {Direkomendasikan=0, Siap=0, Diberi Pelatihan=3, Tidak siap=0}
| | Committed Certain Action = Sangat Rendah
| | | Personal Relationship to other = Sedang: Direkomendasikan {Direkomendasikan=3, Siap=0, Diberi Pelatihan=0, Tidak siap=0}
| | | Personal Relationship to other = Tinggi: Direkomendasikan {Direkomendasikan=1, Siap=0, Diberi Pelatihan=1, Tidak siap=0}
| | Committed Certain Action = Sedang: Direkomendasikan {Direkomendasikan=3, Siap=0, Diberi Pelatihan=0, Tidak siap=0}
| | Committed Certain Action = Tinggi: Direkomendasikan {Direkomendasikan=2, Siap=0, Diberi Pelatihan=0, Tidak siap=0}
Willingness to Learn = Sangat Rendah
| Personal Adversity = Rendah: Tidak siap {Direkomendasikan=0, Siap=0, Diberi Pelatihan=1, Tidak siap=2}
```

Gambar 6. Decision Rule dari Decision Tree Unprunning

Berdasarkan Gambar 6, decision tree model unpruning menghasilkan 121 rules dengan node akar yang terpilih juga sama pada parameter "Willingness to Learn".

3.3. Proses Validasi Model Decision Tree

Bagian ini menjelaskan pengujian decision tree dengan teknik *unpruning* dan teknik *pruning*. Setiap model dilakukan pengujian dengan skenario: 1) *Use Training Test* yaitu mengevaluasi seberapa baik algoritma mampu memprediksi kelas dari instance setelah dilakukan pelatihan, dimana data uji yang digunakan sama dengan data pelatihan, 2) *Percentage Split* yaitu mengevaluasi seberapa baik algoritma mampu memprediksi data, dimana jumlah data pelatihan dan data uji dibagi sesuai persentase yang digunakan.

1. Perbandingan Decision Tree Pruned dan Unpruned pada Mode Use Training Test

Pada mode ini dilakukan pengujian dengan data set sejumlah 404, 504 dan 604 data akan dijadikan data training dan data testing. Hasil dari mode pengujian diharapkan akan mencapai nilai tingkat akurasi yang besar hingga mencapai 100%, dikarenakan data pelatihan dan data uji yang digunakan adalah data yang sama. Hasil dari mode pengujian ini dapat dilihat pada Tabel 14.

Tabel 14. Hasil Perbandingan Pruned dan Unpruned Tree pada Mode Use Training Test

Tree	Jml Sampel	Akurasi (%)	Recall (%)	Precisio (%)
Pruned	604	99.50	99.69	99.72
	504	99.50	99.69	99.72
	404	99.26	99.56	99.55
Unpruned	604	89.11	81.56	88.57
	504	91.09	85.41	88.35
	404	90.35	85.51	87.28

Pada Tabel 14 memperlihatkan bahwa model decision tree yang telah mengalami pemangkasan (*pruning*) menghasilkan nilai akurasi, recall dan precision yang lebih baik. Dimana hampir semua data terprediksi secara tepat dengan nilai akurasi diatas 99%. Begitu juga halnya dengan nilai *recall* dan *precision*.

2. Perbandingan Decision Tree Pruned dan Unpruned pada pengujian Mode Percentage Split

Pada mode ini dilakukan pengujian dengan data set sebanyak 404, 504 dan 604 dengan split data training sebesar 90% dan 80% dari data set. Hasil dari pengujian mode ini seperti terlihat pada Tabel 15.

Dari Tabel 15 menunjukkan pada model decision tree yang mengalami *pruned*, tingkat akurasi, recall dan precision lebih tinggi pada kondisi split data 90% dibandingkan menggunakan split data 80%. Dari pengujian diperoleh rata-rata tingkat akurasi pada decision tree pruned adalah 82.87% sedangkan pada decision tree unpruned adalah 80.38%.

Tabel 15. Hasil Perbandingan Pruned dan Unpruned Tree pada Mode Percentage Split

Tree	Jml Sample	Split Data Training (%)	Akurasi (%)	Recall (%)	Precision (%)
Pruned	604	90	83.61	77.37	71.61
		80	79.17	53.11	54.92
	504	90	88.24	87.22	86.02
		80	87.00	63.51	64.73
	404	90	80.49	82.99	82.99
		80	78.75	59.26	59.00
Unpruned	604	90	80.33	75.76	78.27
		80	81.67	55.08	59.38
	504	90	82.35	84.06	81.60
		80	80.00	59.54	58.37
	404	90	82.93	87.15	84.55
		80	75.00	58.49	54.22

4. DISKUSI

Penelitian ini dibuat berdasarkan pengalaman peneliti sebelumnya yang sudah pernah melakukan penelitian yang menghasilkan beberapa sumber teori hasil atau analisis.

Pada penelitian terdahulu oleh Mohammad Naufala Bairat, Roswan Latuconsina, Casi Setianingsih dengan judul "Perancangan dan Implementasi Tes Psikologi Myers-Briggs type Indicator (MBTI) Berbasis Komputer Dengan Metode Decision Tree". Tes psikologi Myers-Briggs Type Indicator (MBTI) yang berfungsi untuk mengetahui kecerdasan individu dan tipe kepribadian seseorang melalui beberapa aspek pada Myers-Briggs Type Indicator (MBTI). Penelitian ini membahas tentang perancangan dan implementasi tes psikologi berbasis komputer yang berdasarkan pada teori kepribadian Myers-Briggs Type Indicator (MBTI) dengan menggunakan metode Decision Tree sebagai pengambilan keputusan penilaian tes. Hasil pengujian akhir yang dilakukan nilai akurasi yang didapat dari metode decision tree sesuai dataset adalah 50%. Aplikasi tersebut mendapatkan 100% dari total 6 pengujian alpha dan 58,6% dari pengujian usability[22].

Suryadi Syamsu, Muhajirin dan Nyoman Suta Wijaya dengan judul " Rules Generation untuk Klasifikasi Data Bakat dan Minat Berdasarkan Rumpun Ilmu Dengan Decision Tree", Penelitian ini bertujuan untuk merancang aplikasi dan mengimplementasikan Rules Generation untuk klasifikasi data bakat dan minat berdasarkan rumpun ilmu dengan decision tree. Data ini diperoleh melalui penelitian dalam bentuk kuesioner pada objek penelitian dan pustaka. Data dianalisa dengan data bakat dan minat yang berasal dari ilmu psikologi ialah Tes IST (Intelligenz Struktur Test) merupakan salah satu tes psikologi untuk mengukur tingkat intelegensi seseorang. Hasil penelitian ini menunjukkan bahwa Implementasi Rules Generation terhadap Decision Tree (RGFDT) penelitian tentang penentuan jurusan seseorang yang menggunakan algoritma C4.5 dapat mengatasi cabang-cabang pohon keputusan yang tidak relevan menjadi relevan, sehingga dalam mengimplementasi ke dalam bahasa pemrograman dapat lebih mudah diterjemahkan., dan aplikasi tes bakat dan minat dapat berjalan dengan baik dan memudahkan responden dalam melakukan tes bakat dan minat dengan mudah[23].

M A Abdillah, Arief Setyanto, Sudarmawan dengan judul " Implementasi Decision Tree Algoritma C4.5 Untuk Memprediksi Kesuksesan Pendidikan Karakter". Perancanangan kurikulum Pendidikan Karakter dalam sistem pendidikan di Indonesia, adalah sesuatu hal yang baru dan belum banyak dikaji dalam penelitian Educational Data Mining (EDM). Sebagian besar penelitian dalam ranah EDM masih menggunakan faktor kognitif dalam penilaiannya, berbeda dengan pendidikan karakter yang lebih berorientasi kepada pengajaran nilai-nilai karakter, serta mempertimbangkan latar belakang peserta didik. Oleh karena itu, diperlukan cara atau metode untuk mengidentifikasi calon peserta didik, serta memprediksi kesuksesannya dalam sistem pendidikan karakter. Algoritma C4.5 dapat digunakan untuk melakukan prediksi dan klasifikasi terhadap calon siswa dengan cara membuat pohon keputusan berdasarkan data-data yang sudah ada dan melakukan prediksi terhadap calon siswa baru, dalam penelitian ini peneliti menggunakan data mahasiswa Unires Yogyakarta sebagai objek penelitian. Dengan penelitian ini, juga diharapkan dapat diketahui tingkat akurasi Decision Tree Algoritma C.45 dalam mengukur pengaruh atribut-atribut latar belakang siswa tersebut terhadap kesuksesan pendidikan karakter, sehingga akan diketahui apakah Decision Tree Algoritma C.45 memenuhi aspek reliabilitas dan validitas sebagai alat ukur kesuksesan pendidikan karakter. Dari hasil pengukuran, diketahui bahwa kombinasi atribut Bidang Bahasa dan Sosial, Latar belakang pendidikan agama dan orang tua yang menjadi seorang pendidik/guru, serta kemampuan untuk membaca Al-Qur'an berkorelasi positif terhadap kesuksesan pendidikan karakter. Nilai accuracy sebesar 60,91%, menunjukkan bahwa algoritma decision tree C4.5 layak digunakan untuk melakukan prediksi tingkat kesuksesan pada pendidikan karakter[24].

5. KESIMPULAN

Tujuan dilakukan penelitian ini untuk menemukan decision rule terbaik dari decision tree model yang memiliki tingkat akurasi lebih baik. Hasil yang didapatkan dari penelitian ini menunjukan *decision tree model prunning* memiliki tingkat akurasi 99.5%, nilai recall 99.69% dan nilai precision 99.72%. Sedangkan *decision tree model unprunning* memiliki tingkat akurasi 89.11%, nilai recall 81.56% dan nilai precision 88.57%. Dari hasil penelitian dapat disimpulkan bahwa *decision tree model prunning* menghasilkan nilai akurasi yang lebih tinggi dari pada *decision tree model*

unprunning. Dari *decision tree model prunning* diperoleh 180 decision rules, sedangkan *decision tree model unprunning* menghasilkan 121 rules. Dari hasil kesimpulan yang didapat, maka decision rule model yang terbentuk dapat digunakan untuk pengembangan Alat/aplikasi konseling berbasis AI nantinya.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih banyak kepada Universitas Islam Riau yang telah mendanai penelitian dan publikasi ini, serta terima kasih atas Kerjasama dan bantuan tim peneliti yang terlibat dalam penyelesaian penelitian ini.

DAFTAR PUSTAKA

[1] S. Farradinna and T. N. Fadhlia, "Entrepreneurial personality in predicting self-regulation on small and medium business entrepreneurs in pekanbaru, riau, indonesia," *J. Mgt. Mkt. Rev.*, vol. 3, no. 1, pp. 34–39, 2018.

[2] S. Farradinna, T. N. Fadhlia, and D. Azmansyah, "Psychological resilience predicted by personality traits, locus of control and self-regulation of young entrepreneurs in Pekanbaru," *Glob. J. Bus. Soc. Sci. Rev.*, vol. 7, no. 1, p. 1, 2019.

[3] W. Rokhman and F. Ahamed, "The role of social and psychological factors on entrepreneurial intention among Islamic college students in Indonesia," *Entrep. Bus. Econ. Rev.*, vol. 3, no. 1, p. 30, 2015.

[4] C. Xu and Z. Zhang, "The Effect of Law Students in Entrepreneurial Psychology Under the Artificial Intelligence Technology," *Front. Psychol.*, vol. 12, p. 731713, 2021.

[5] T. Goto, C. A. Camargo, M. K. Faridi, R. J. Freishtat, and K. Hasegawa, "Machine learning–based prediction of clinical outcomes for children during emergency department triage," *JAMA Netw. open*, vol. 2, no. 1, pp. e186937–e186937, 2019.

[6] W. Bleidorn and C. J. Hopwood, "Using machine learning to advance personality assessment and theory," *Personal. Soc. Psychol. Rev.*, vol. 23, no. 2, pp. 190–203, 2019.

[7] R. Alharthi, B. Guthier, and A. El Saddik, "Recognizing human needs during critical events using machine learning powered psychology-based framework," *IEEE Access*, vol. 6, pp. 58737–58753, 2018.

[8] M. Savci, A. Tekin, and J. D. Elhai, "Prediction of problematic social media use (PSU) using machine learning approaches," *Curr. Psychol.*, pp. 1–10, 2020.

[9] J. D. Elhai and C. Montag, "The compatibility of theoretical frameworks with machine learning analyses in psychological research," *Curr. Opin. Psychol.*, vol. 36, pp. 83–88, 2020.

[10] J. D. Elhai, H. Yang, D. Rozgonjuk, and C. Montag, "Using machine learning to model problematic smartphone use severity: The significant role of fear of missing out," *Addict. Behav.*, vol. 103, p. 106261, 2020.

[11] M. H. Afzali *et al.*, "Machine-learning prediction of adolescent alcohol use: A cross-study, cross-cultural validation," *Addiction*, vol. 114, no. 4, pp. 662–671, 2019.

[12] R. Dave, K. Sargeant, M. Vanamala, and N. Seliya, "Review on Psychology Research Based on Artificial Intelligence Methodologies," *J. Comput. Commun.*, vol. 10, no. 5, pp. 113–130, 2022.

[13] R. Jacobucci, A. K. Littlefield, A. J. Millner, E. Kleiman, and D. Steinley, "Pairing machine learning and clinical psychology: how you evaluate predictive performance matters," 2020.

[14] D. B. Dwyer, P. Falkai, and N. Koutsouleris, "Machine learning approaches for clinical psychology and psychiatry," *Annu. Rev. Clin. Psychol.*, vol. 14, pp. 91–118, 2018.

[15] A. Lavecchia, "Machine-learning approaches in drug discovery: methods and applications," *Drug Discov. Today*, vol. 20, no. 3, pp. 318–331, 2015.

[16] Z.-H. Zhou, *Machine learning*. Springer Nature, 2021.

[17] B. Mahesh, "Machine learning algorithms-a review," *Int. J. Sci. Res. (IJSR).[Internet]*, vol. 9, pp. 381–386, 2020.

[18] A. Venkatasubramaniam, J. Wolfson, N. Mitchell, T. Barnes, M. JaKa, and S. French, "Decision trees in epidemiological research," *Emerg. Themes Epidemiol.*, vol. 14, no. 1, pp. 1–12, 2017.

[19] F. Y. Pamuji and V. P. Ramadhan, "Komparasi Algoritma Random Forest dan Decision Tree untuk Memprediksi Keberhasilan Immunotheraphy," *J. Teknol. dan Manaj. Inform.*, vol. 7, no. 1, pp. 46–50, 2021.

[20] U. I. Lestari, A. Y. Nadhiroh, and C. Novia, "Penerapan Metode K-Nearest Neighbor Untuk Sistem Pendukung Keputusan Identifikasi Penyakit Diabetes Melitus," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 8, no. 4, pp. 2071–2082, 2021.

[21] E. Muningsih, "Kombinasi Metode K-Means Dan Decision Tree Dengan Perbandingan Kriteria Dan Split Data," *J. Teknoinfo*, vol. 16, no. 1, pp. 113–118, 2022.

[22] A. R. Rabbani, M. Nasrun, and C. Setianingsih, "Perancangan Dan Implementasi Tes Psikologi Myers-briggs Type Indicator Berbasis Komputer Dengan Metode NaÃ¯ ve Bayes Sebagai Pengambilan Keputusan," *eProceedings Eng*., vol. 7, no. 1, 2020.

[23] S. Syamsu, M. Muhajirin, and N. S. Wijaya, "Rules Generation Untuk Klasifikasi Data Bakat dan Minat Berdasarkan Rumpun Ilmu Dengan Decision Tree," *Inspir. J. Teknol. Inf. dan Komun*., vol. 9, no. 1, pp. 40–51, 2019.

[24] M. A. Abdillah, A. Setyanto, and S. Sudarmawan, "Implementasi Decision Tree Algoritma C4. 5 Untuk Memprediksi Kesuksesan Pendidikan Karakter," *Respati*, vol. 15, no. 2, pp. 59–69, 2020.

Jutif_Syafitri,_Farradinna,_Jayanti,_Arta_2023.pdf

ORIGINALITY REPORT

19%	17%	10%	8%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

MATCH ALL SOURCES (ONLY SELECTED SOURCE PRINTED)

5%

★ Submitted to Universitas Muhammadiyah Sidoarjo

Student Paper

Exclude quotes	On	Exclude matches	< 1%
Exclude bibliography	Off		