

**ANALISIS KOMPARASI ALGORITMA *DECISION TREE*,
LOGISTIC REGRESSION, DAN *RANDOM FOREST* DALAM
DATA SCIENCE UNTUK KLASIFIKASI KESEHATAN
MENTAL MAHASISWA MENGGUNAKAN *PYTHON***

SKRIPSI

Skripsi diajukan untuk memenuhi persyaratan
memperoleh gelar sarjana



Disusun oleh:

Ahmad Syafei Nursuwanda

200111401003

**PROGRAM STUDI TEKNIK INFORMATIKA
UNIVERSITAS GLOBAL JAKARTA**

2024

PERNYATAAN ORISINALITAS SKRIPSI

Saya menyatakan dengan sebenar-benarnya bahwa berdasarkan hasil penelusuran berbagai karya ilmiah, gagasan dan masalah ilmiah yang diteliti dan diulas di dalam Naskah Skripsi ini adalah asli dari pemikiran saya. Tidak terdapat karya ilmiah yang pernah diajukan oleh orang lain untuk memperoleh gelar akademik disuatu Perguruan Tinggi, dan tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis dikutip dalam naskah ini dan disebutkan dalam sumber kutipan dan daftar pustaka.

Apabila ternyata di dalam naskah Skripsi ini dapat dibuktikan terdapat unsur-unsur jiplakan, saya bersedia Skripsi dibatalkan, serta diproses sesuai dengan peraturan perundang-undangan yang berlaku (UUNo. 20 Tahun 2003, pasal 25 ayat 2 dan pasal 70).

Depok, 15 Agustus 2024

Mahasiswa,



Ahmad Syafei Nursuwanda

NIM. 200111401003

HALAMAN PENGESAHAN PEMBIMBING

Skripsi ini diajukan oleh :
Nama : Ahmad Syafei Nursuwanda
NIM : 200111401003
Program Studi : Teknik Informatika
Judul Skripsi : Analisis Komparasi Algoritma *Decision tree*,
Logistic regression, Dan *Random forest* Dalam Data
Science Untuk Klasifikasi Kesehatan Mental
Mahasiswa Menggunakan Python

Telah berhasil dipertahankan di hadapan Dewan Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana Komputer pada Program Studi Teknik Informatika Universitas Global Jakarta.

DEWAN PEMBIMBING

Pembimbing 1 : Anindya Ananda Hapsari S.ST, M.IT ()
Pembimbing 2 : Halimatuz Zuhriyah S.KOM., M.IT ()
Ditetapkan di : Depok, Jawa Barat
Tanggal : 15 Agustus 2024

HALAMAN PENGESAHAN DEWAN PENGUJI

Skripsi ini diajukan oleh :
Nama : Ahmad Syafei Nursuwanda
NIM : 200111401003
Program Studi : Teknik Informatika
Judul Skripsi : Analisis Komparasi Algoritma Decision Tree, Logistic Regression, Dan Random Forest Dalam Data Science Untuk Klasifikasi Kesehatan Mental Mahasiswa Menggunakan Python

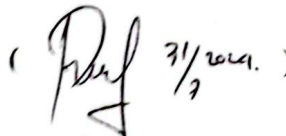
Telah berhasil dipertahankan di hadapan Dewan Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana Komputer pada Program Studi Teknik Informatika Universitas Global Jakarta.

DEWAN PENGUJI

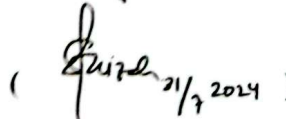
Penguji 1 : Dian Nugraha, S.S.T., MIT



Penguji 2 : Risna Oktaviati, S.S.T., MIT



Penguji 3 : Safira Faizah, S.Tr.Kom., MIT



Ditetapkan di : Depok, Jawa Barat

Tanggal : 15 Agustus 2024

KATA PENGANTAR/UCAPAN TERIMA KASIH

Puji syukur saya panjatkan kepada Tuhan Yang Maha Esa, karena atas berkat dan rahmat- Nya, saya dapat menyelesaikan skripsi ini. Penulisan skripsi ini dilakukan dalam rangka memenuhi salah satu syarat untuk mencapai gelar Sarjana Teknik Jurusan Teknik di Informatika Universitas Global Jakarta. Harap dimaklumi jika ada kekurangan dalam penelitian ini, karena penulis menyadari bahwa dunia data science dan teknologi terus berkembang. Penulis menyadari bahwa penulisan skripsi ini tidak sempurna. Dengan demikian, penulis memohon maaf sekiranya terdapat kesalahan dalam penulisan skripsi ini, dan dengan lapang dada penulis bersedia menerima kritik dan saran yang membangun. Selain itu, penulis mengucapkan terima kasih kepada semua orang yang telah membantu dan membantu dalam menyelesaikan skripsi ini dengan baik.

Saya menyadari bahwa, tanpa bantuan dan bimbingan dari berbagai pihak, dari masa perkuliahan sampai pada penyusunan skripsi ini, sangatlah sulit bagi saya untuk menyelesaikan skripsi ini. Oleh karena itu, saya mengucapkan terima kasih kepada:

- (1) Allah SWT. Tuhan semesta alam yang memberikan kesempatan, rahmat, dan rezeki sehingga peneliti dapat menyelesaikan penelitian ini dengan lancar;
- (2) Orang tua dan keluarga saya yang telah memberikan bantuan dukungan material dan moral;
- (3) Prof. Dr. apt., Eddy Yusuf, M.Pharm selaku rektor Universitas Global Jakarta.
- (4) Anindya Ananda Hapsari S.ST, M.IT dan Halimatuz Zuhriyah S.KOM, M.IT, selaku dosen pembimbing yang telah menyediakan waktu, tenaga, dan pikiran untuk mengarahkan serta membimbing saya dalam penyusunan skripsi ini;
- (5) Dosen penguji yang telah menguji dan memberikan arahan mengenai kekurangan dan kelemahan pada penelitian ini;
- (6) Ummy Gusti Salamah S.ST, M.IT selaku kepala prodi yang mengayomi dan membantu dalam pengumpulan data pada pen;
- (7) **Sriana Sihombing, S.Psi., M.M., M.Psi** selaku psikolog yang membantu dan mengarahkan peneliti secara mental dan tidak merasa terlalu tertekan selama melakukan penelitian; dan

- (8) Teman dan sahabat yang telah banyak membantu serta mendukung saya dalam menyelesaikan skripsi ini.

Akhir kata, saya berharap Tuhan Yang Maha Esa berkenan membalas segala kebaikan semua pihak yang telah membantu. Semoga skripsi ini membawa manfaat bagi pengembangan ilmu pengetahuan.

Depok,2024

Penulis

PERNYATAAN PERSETUJUAN PUBLIKASI AKADEMIS

Sebagai sivitas akademika Universitas Global Jakarta, saya yang bertanda tangan di bawah ini :

Nama : Ahmad Syafei Nursuwanda

NPM : 200111401003

Program Studi : Teknik Informatika

Jenis Karya Ilmiah : Skripsi

demikian pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Global Jakarta **Hak Bebas Royalti Non-eksklusif** (*None-exclusive Royalty Free Right*) atas karya ilmiah saya yang berjudul:

ANALISIS KOMPARASI ALGORITMA *DECISION TREE*, *LOGISTIC REGRESSION*, DAN *RANDOM FOREST* DALAM DATA SCIENCE UNTUK KLASIFIKASI KESEHATAN MENTAL MAHASISWA MENGGUNAKAN PYTHON

berserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti/Noneksklusif ini Universitas Global Jakarta berhak menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk pangkalan data (*database*), merawat dan mempublikasikan skripsi saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Jakarta, 26 Januari 2024

Yang menyatakan

A handwritten signature in black ink is written over a yellow revenue stamp. The stamp features the Garuda Pancasila emblem and the text '10000', 'METERAI TEMPEL', and the serial number '4CD09ALX239025238'.

Ahmad Syafei Nursuwanda
NIM. 200111401003

ABSTRAK

ANALISIS KOMPARASI ALGORITMA *DECISION TREE*, *LOGISTIC REGRESSION*, DAN *RANDOM FOREST* DALAM DATA SCIENCE UNTUK KLASIFIKASI KESEHATAN MENTAL MAHASISWA MENGUNAKAN PYTHON

Oleh

Ahmad Syafei Nursuwanda

Teknik Informatika

Di Indonesia, 6,1 persen penduduk berusia 15 tahun ke atas mengalami gangguan kesehatan mental, termasuk 12,69 persen mahasiswa yang rentan karena kondisi mental yang tidak stabil dan tekanan kehidupan. Kesehatan mental penting bagi mahasiswa karena berdampak langsung pada kesuksesan akademik mereka . Penelitian ini menggunakan algoritma *Decision tree*, *Random forest*, dan *Logistic regression* untuk menganalisis data kesehatan mental mahasiswa menggunakan Python. Hasil penelitian menunjukkan bahwa *Logistic regression* memiliki akurasi tertinggi 90%, diikuti *Decision tree* dan *Random forest* dengan akurasi 80%. Model *Logistic regression* juga memberikan prediksi kecemasan paling akurat berdasarkan TMAS, dengan akurasi 90%, dibandingkan 80% dari *Decision tree* dan *Random forest*.

Kata kunci : Kesehatan Mental; *Python*; Klasifikasi; Algoritma; *Data Science*.

ABSTRACT

COMPARATIVE ANALYSIS OF *DECISION TREE, LOGISTIC REGRESSION, AND RANDOM FOREST* ALGORITHMS IN DATA SCIENCE FOR STUDENT MENTAL HEALTH CLASSIFICATION USING PYTHON

By

Ahmad Syafei Nursuwanda

Informatics Engineering

In Indonesia, 6.1 percent of the population aged 15 years and above experience mental health disorders, including 12.69 percent of college students who are vulnerable due to unstable mental conditions and life pressures. Mental health is important for college students as it has a direct impact on their academic success. This research uses Decision tree, Random forest, and Logistic regression algorithms to analyze college students' mental health data using Python. The results showed that Logistic regression had the highest accuracy of 90%, followed by Decision tree and Random forest with 80% accuracy. The Logistic regression model also provides the most accurate prediction of anxiety based on TMAS, with 90% accuracy, compared to 80% from Decision tree and Random forest.

Keywords: *Mental Health; Python; Classification; Algorithm; Data Science.*

DAFTAR ISI

PERNYATAAN ORISINALITAS SKRIPSI	i
HALAMAN PENGESAHAN PEMBIMBING	ii
HALAMAN PENGESAHAN DEWAN PENGUJI.....	iii
KATA PENGANTAR/UCAPAN TERIMA KASIH	iv
PERNYATAAN PERSETUJUAN PUBLIKASI AKADEMIS	vi
ABSTRAK	vii
ABSTRACT	viii
DAFTAR ISI	ix
DAFTAR TABEL	xii
DAFTAR GAMBAR	xiii
DAFTAR RUMUS.....	xvi
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Tujuan Penelitian	3
1.4 Manfaat Penelitian	3
1.5 Batasan Masalah.....	4
BAB II KAJIAN PUSTAKA	5
2.1 Penelitian sebelumnya	5
2.2 Landasan Teori	13
2.2.1 <i>Data Science</i>	13
2.2.2 Klasifikasi.....	14
2.2.3 Algoritma <i>Decision tree</i>	14
2.2.4 <i>Gain</i>	15
2.2.5 <i>Entropy</i>	15
2.2.6 Algoritma <i>Logistic regression</i>	16
2.2.7 Algoritma <i>Random forest</i>	16
2.2.8 Dataset.....	17

2.2.9	Data Primer	18
2.2.10	Data Sekunder.....	19
2.2.11	<i>Confusion matrix</i>	19
2.2.12	<i>Cross Validation</i>	21
2.2.13	<i>Anxiety</i>	21
2.2.14	<i>Taylor Minnesota Anxiety Scale (TMAS)</i>	22
2.2.15	<i>Simple Random Sampling</i>	22
2.2.16	<i>Slovin</i>	23
2.2.17	Eksplanasi Berurutan (<i>Sequenti al E xplanatory Strategy</i>)	23
BAB III METODE PENELITIAN.....		24
3.1	Alur Penelitian	24
3.1.1	Identifikasi Masalah.....	24
3.1.2	Literature Review.....	25
3.1.3	Pengembangan Model Algoritma	25
3.1.4	Pengumpulan Data	27
3.1.5	Pre-Processing Data	34
3.1.6	Pengujian Model Algoritma.....	37
3.1.7	Analisis Evaluasi dan Validasi Kinerja Model Algoritma.....	40
3.1.8	Interpretasi Hasil	41
3.2	Lokasi & Objek Penelitian.....	42
3.2.1	Lokasi Penelitian.....	42
3.2.2	Objek Penelitian.....	42
BAB IV HASIL DAN PEMBAHASAN		43
4.1	Pengumpulan Data.....	43
4.2	<i>Pre-Processing</i> data	44
4.3	Pengujian Model Algoritma	48
4.3.1	<i>Decision tree</i>	52
4.3.2	<i>Random forest</i>	54

4.3.3	<i>Logistic regression</i>	55
4.4	Analisis Evaluasi dan Validasi Kinerja Model algoritma.....	57
4.4.1	<i>Confusion matrix</i>	57
4.4.2	<i>Cross Validation</i>	61
4.5	Interpretasi Hasil.....	62
BAB V KESIMPULAN DAN SARAN.....		67
5.1	Kesimpulan.....	67
5.2	Saran	68
DAFTAR PUSTAKA		70
LAMPIRAN		72

DAFTAR TABEL

Table 2. 1 Penelitian sebelumnya.....	5
Table 3. 1 table konversi pertanyaan.....	28
Table 3. 2 table translasi TMAS	30
Table 4. 1 Table Classification report DT <i>Train Test Size</i> 70-30	53
Table 4. 2 Table Classification report DT <i>Train Test Size</i> 80-20	54
Table 4. 3 Table Classification report RF <i>Train Test Size</i> 70-30.....	54
Table 4. 4 Table Classification report RF <i>Train Test Size</i> 80-20.....	55
Table 4. 5 Table Classification report LG <i>Train Test Size</i> 70-30	56
Table 4. 6 Table Classification report RF <i>Train Test Size</i> 80-20.....	57
Table 4. 7 table perbandingan <i>accuracy</i> model algoritma	62
Table 4. 8 table perbandingan <i>recall</i> model algoritma	63
Table 4. 9 table perbandingan <i>precision</i> model algoritma.....	64
Table 4. 10 table perbandingan <i>f1-score</i> model algoritma.....	64
Table 4. 11 table perbandingan performa keseluruhan model algoritma.....	65

DAFTAR GAMBAR

Gambar 2. 1 Algoritma <i>Decision tree</i> (Charbuty & Abdulazeez, 2021).	14
Gambar 2. 2 Algoritma <i>Random forest</i> (Ghosh, 2024).....	17
Gambar 2. 3 Proses Cross Validation(Nugroho & Amrullah, 2023)	21
Gambar 3. 1 bagan alur penelitian	24
Gambar 3. 2 google collab sebagai tools&media.....	26
Gambar 3. 3 library python yang digunakan.....	26
Gambar 3. 4 langkah-langkah pengolahan data sebelum masuk kedalam model.....	27
Gambar 3. 5 langkah-langkah evaluasi model	27
Gambar 3. 6 library python NumPy	34
Gambar 3. 7 library python pandas	35
Gambar 3. 8 proses cleaning data	35
Gambar 3. 9 proses <i>rename</i> fitur.....	35
Gambar 3. 10 proses encoder data fitur	36
Gambar 3. 11 proses normalisasi data fitur dengan MinMaxScaler	36
Gambar 3. 12 proses penambahan target	37
Gambar 3. 13 library <code>train_test_split</code> yang digunakan	38
Gambar 3. 14 proses split data sebelum masuk kedalam model	38
Gambar 3. 15 kodingan model <i>decision tree</i>	39
Gambar 3. 16 kodingan model <i>random forest</i>	39
Gambar 3. 17 kodingan model <i>logistic regression</i>	40
Gambar 3. 18 proses validasi cross validation	41
Gambar 4. 1 Data yang dikumpulkan	44
Gambar 4. 2 data sebelum di cleaning	45
Gambar 4. 3 data setelah di cleaning	45
Gambar 4. 4 data sebelum <i>rename</i> fitur	46
Gambar 4. 5 data setelah <i>rename</i> fitur	46
Gambar 4. 6 data sebelum di enncoder	46
Gambar 4. 7 data setelah di encoder	46

Gambar 4. 8 data sebelum di normalisasi	47
Gambar 4. 9 data setelah di normalisasi.....	47
Gambar 4. 10 data sebelum ditambahkan target	48
Gambar 4. 11 data setelah ditambahkan target	48
Gambar 4. 12 data setelah target diencoder	48
Gambar 4. 13 bentuk data rasio <i>Train Test Size</i> 70-30	49
Gambar 4. 14 contoh data <i>X_train</i> setelah di split rasio <i>Train Test Size</i> 70-30.....	49
Gambar 4. 15 contoh data <i>X_test</i> setelah di split rasio <i>Train Test Size</i> 70-30.....	50
Gambar 4. 16 contoh data <i>y_train</i> dan <i>y_test</i> rasio train test 70-30.....	50
Gambar 4. 17 bentuk data rasio <i>Train Test Size</i> 80-20	51
Gambar 4. 18 contoh data <i>X_train</i> setelah di split rasio <i>Train Test Size</i> 80-20.....	51
Gambar 4. 19 contoh data <i>X_test</i> setelah di split rasio <i>Train Test Size</i> 80-20.....	51
Gambar 4. 20 contoh data <i>y_train</i> dan <i>y_test</i> rasio train test 80-30.....	52
Gambar 4. 21 hasil uji model <i>decision tree</i> <i>Train Test Size</i> 70-30	53
Gambar 4. 22 hasil uji model <i>decision tree</i> rasio <i>Train Test Size</i> 80-20	53
Gambar 4. 23 hasil uji model <i>random forest</i> <i>Train Test Size</i> 70-30	54
Gambar 4. 24 hasil uji model <i>random forest</i> <i>Train Test Size</i> 80-20	55
Gambar 4. 25 hasil uji model <i>logistic regression</i> <i>Train Test Size</i> 70-30	56
Gambar 4. 26 hasil uji model <i>logistic regression</i> <i>Train Test Size</i> 80-20	56
Gambar 4. 27 hasil <i>confusion matrix decision tree</i> rasio <i>Train Test Size</i> 70-30	58
Gambar 4. 28 hasil <i>confusion matrix decision tree</i> rasio <i>Train Test Size</i> 80-20	58
Gambar 4. 29 hasil <i>confusion matrix random forest</i> rasio <i>Train Test Size</i> 70-30	59
Gambar 4. 30 hasil <i>confusion matrix random forest</i> rasio <i>Train Test Size</i> 80-20	59

Gambar 4. 31 hasil <i>confusion matrix logistic regression</i> rasio <i>Train Test Size</i> 70-30.....	60
Gambar 4. 32 hasil <i>confusion matrix logistic regression</i> rasio <i>Train Test Size</i> 80-20.....	60
Gambar 4. 33 hasil <i>confusion matrix</i> train size 70.....	61
Gambar 4. 34 hasil <i>confusion matrix</i> train size 80.....	61
Gambar 4. 35 bagan perbandingan akurasi model algoritma rasio <i>Train Test Size</i> 70-30	63
Gambar 4. 36 bagan perbandingan akurasi model algoritma rasio <i>Train Test Size</i> 80-20	63

DAFTAR RUMUS

(2.1) <i>Gain</i>	15
(2.2) <i>Shanon Entropy</i>	15
(2.3) <i>Entropy</i>	16
(2.4) <i>Logistic Regression</i>	16
(2.5) <i>Accuracy</i>	20
(2.6) <i>Precision</i>	20
(2.7) <i>Recall</i>	20
(2.8) <i>F1-Score</i>	20

BAB I

PENDAHULUAN

1.1 Latar Belakang

Di Indonesia banyak yang mengalami gangguan kesehatan mental mulai dari usia 15 tahun keatas yakni mencapai 6,1 persen, dan mahasiswa yang mempunyai rentang umur remaja akhir dan dewasa awal sangat rawan mengalami masalah dalam kesehatan mental mereka karena sedang berada dalam kondisi mental yang tidak stabil, dan diiringi dengan konflik yang terjadi dalam kehidupan pada masa tersebut, bahkan pada jurnal yang ditulis oleh (Salvia, 2021) ada sekitar 12,69 persen mahasiswa mengalami masalah kesehatan mental.

Kesehatan mental khususnya bagi mahasiswa sangat penting terutama dalam jangka pendek. Hal ini dikarenakan kesehatan mental berkontribusi dalam kesuksesan dan keberhasilan para civitas akademika, dikarenakan stress mental atau masalah pada kesehatan mental dapat berdampak negatif pada kemampuan berbicara dan mendengarkan orang lain, sehingga dapat sangat menghambat aktivitas mahasiswa (Dwi Putra dkk., 2023). Oleh karena itu, terdapat beberapa penelitian yang menerapkan metode data science untuk menganalisa permasalahan kesehatan mental (Salvia, 2021) agar dapat di deteksi sedari dini untuk mendapatkan pengobatan yang tepat dan mencegahnya sebelum menjadi lebih buruk (Thenata & Suryadi, 2022). Seperti pada beberapa penelitian sebelumnya yang menguji potensi algoritma pembelajaran mesin untuk mendiagnosis dan memprediksi berbagai penyakit mental, serta mengungkap perspektif baru tentang bagaimana algoritma tersebut dapat digunakan dalam pendidikan (Muzumdar dkk., 2022) dan kesehatan (Salvia, 2021).

Salah satu contoh penelitian lain yang telah dilakukan dengan menggunakan algoritma *data science* adalah penelitian yang dilakukan oleh (Dwi Putra dkk., 2023) pada jurnal "*Perbandingan Algoritma Naive Bayes Classifier dan Support Vector Machine untuk Klasifikasi Data Kesehatan Mental Mahasiswa*" untuk melihat dan menganalisis perbandingan algoritma *Naive Bayes Classifier* dan *Support Vector Machine* dalam klasifikasi data kesehatan mental mahasiswa, menunjukkan bagaimana berbagai metode dapat memberikan hasil yang berbeda

tergantung pada sifat data dan kasus penggunaan. Hasil dari perbandingan dari penelitian tersebut algoritma *Support Vector Machine* mengungguli *Naive Bayes* classifier dengan nilainya sebesar 94,37% (Dwi Putra dkk., 2023). Ada juga penelitian lain yang dilakukan oleh (Wibowo dkk., 2023) pada jurnal yang berjudul “*Perbandingan Metode Klasifikasi Untuk Deteksi Stress Pada Mahasiswa di Perguruan Tinggi*” yang melakukan klasifikasi menggunakan algoritma *Naive Bayes*, *Decision tree*, *SVM*, *Neural Network*, *Random Tree*, *Random forest*, dan *KNN data science* untuk mendeteksi stress pada mahasiswa, dan hasil yang didapat ialah algoritma *Decision tree* dan *Random Tree* mendapat nilai tertinggi dengan nilai 100% (Wibowo dkk., 2023).

Dari penelitian yang dilakukan (Dwi Putra dkk., 2023) dan (Wibowo dkk., 2023) kita bisa mengetahui bahwa penggunaan algoritma klasifikasi *data science* *Decision tree*, *Random forest*, dan *Logistic regression* dapat digunakan untuk mengolah data kesehatan mental mahasiswa. Sehingga saat penelitian akan dilakukan, algoritma *Decision tree*, *Logistic regression*, dan *Random forest* akan diterapkan pada dataset *Student Mental Health* dan akan dianalisis persebaran datanya guna memudahkan melihat dan mendapat hasil komparasi dari tiap nilai dari ketiga algoritma yang digunakan dengan menggunakan bahasa pemrograman python, dikarenakan python merupakan salah satu bahasa pemrograman teratas dalam pengolahan data termasuk *data science* (Syamsu, 2021). Nilai yang digunakan sebagai acuan komparasi adalah nilai *accuracy* dari tiap algoritma, dikarenakan nilai *accuracy* dinilai lebih populer digunakan dalam algoritma machine learning data science untuk menilai performa dan hasil di kebanyakan dataset (Muzumdar dkk., 2022). Selain dari nilai *accuracy*, terdapat juga nilai metrik lain yang bisa di ambil dan di analisis dalam algoritma tersebut seperti *precision*, *recall*, dan *F1 score* yang kelak nantinya akan divalidasi menggunakan *cross-validation*. Agar hasil dari performa dan prediksi dari model algoritma *decision tree*, *random forest*, dan *logistic regression* lebih akurat, penelitian ini akan membandingkan hasil performa prediksi model algoritma tersebut dengan metode perhitungan dari *psychology tools* menilai seberapa baik ketepatan prediksi model yang digunakan (Thenata & Suryadi, 2022). Berdasarkan latar belakang tersebut peneliti melakukan penelitian dengan judul “Analisis Komparasi

Algoritma *Decision tree*, *Logistic regression*, dan *Random forest* Dalam Data Science Untuk Klasifikasi Kesehatan Mental Mahasiswa Menggunakan Python”. Yang dimana peneliti akan membandingkan performa algoritma *Decision tree*, *Logistic regression*, dan *Random forest* dalam memprediksi kesehatan mental mahasiswa yang hasil dari algoritma tersebut akan dibandingkan dengan *psychology tools* untuk menilai tingkat ketepatannya.

1.2 Rumusan Masalah

Rumusan masalah dalam Penelitian ini terdapat beberapa masalah yang dapat kita selesaikan, antara lain:

1. Di antara algoritma *Decision tree*, *Logistic regression*, dan *Random forest*, algoritma manakah yang memberikan performa terbaik dalam klasifikasi kesehatan mental mahasiswa berdasarkan metrik *accuracy*, *precision*, *recall*, *F1-score* pada dataset yang sama dengan teknik validasi *cross-validation* dari model python yang digunakan?
2. Bagaimana hasil prediksi *anxiety* dari model algoritma *Decision tree*, *Logistic regression*, dan *Random forest* dibandingkan dengan *anxiety level* dari *TMAS*?

1.3 Tujuan Penelitian

Berikut adalah beberapa tujuan penelitian ini, antara lain:

1. Menentukan algoritma yang memberikan performa terbaik dalam klasifikasi kesehatan mental mahasiswa di antara *Decision tree*, *Logistic regression*, dan *Random forest*, dengan menggunakan metrik evaluasi *accuracy*, *precision*, *recall*, *F1-score* pada dataset yang sama dan dengan teknik validasi *cross-validation*.
2. Mengetahui hasil perbandingan antara prediksi algoritma *Decision tree*, *Logistic regression*, dan *Random forest* dalam mengukur *anxiety level* dengan *TMAS*.

1.4 Manfaat Penelitian

Adapun berikut beberapa manfaat yang didapatkan dari penelitian ini, antara lain:

1. Sebagai tambahan pondasi dalam penelitian data science dibidang kesehatan mental mahasiswa yang kemudian bisa dikembangkan lebih lanjut menjadi bentuk teknologi lainnya menggunakan algoritma python seperti AI, chatbot, dan lain-lain.
2. Sebagai tambahan referensi algoritma mana yang memiliki performa paling baik dalam memprediksi kesehatan mental mahasiswa

1.5 Batasan Masalah

Batasan masalah yang diidentifikasi dalam penelitian ini, adalah:

1. Berfokus untuk mencari algoritma mana yang lebih baik dalam mengolah dataset kesehatan mental mahasiswa dari algoritma *Decision tree*, *Logistic regression*, dan *Random forest* yang digunakan.
2. Pengolahan data berfokus pada dataset yang didapat melalui pengumpulan data yang dilakukan melalui survey Gform dan wawancara bersama pakar.
3. Merancang code python untuk mengolah dataset kesehatan mental mahasiswa mulai dari mencari persebaran data sampai muncul hasil dari komparasinya dan validasi menggunakan teknik validasi *cross validation* python.
4. Data yang diambil berfokus untuk meneliti hasil model algoritma yang dibuat pada kesehatan mental mahasiswa di bagian kecemasan atau *anxiety*.