

BAB I

PENDAHULUAN

1.1 LATAR BELAKANG

Depresi adalah sebuah *mood disorder* yang menyebabkan perasaan sedih yang berkelanjutan dan hilangnya motivasi [1][2]. Gejala umum pada depresi diantaranya adalah kesedihan, rasa kekosongan, dan perasaan marah, yang kemudian disertai dengan perubahan pada fitur kognitif dan somatik yang mempengaruhi kemampuan seorang individu untuk melakukan kegiatan sehari-hari secara signifikan [3]. Menurut estimasi WHO, sekitar 280 juta orang didunia memiliki depresi [4], dan lebih dari 700 ribu orang meninggal tiap tahunnya karena bunuh diri, yang juga merupakan penyebab kematian terbesar ke-4 didunia bagi rentang usia 15 hingga 29 tahun [5].

Sementara itu, aspek kehidupan yang erat kaitannya dengan kualitas kehidupan, dideskripsikan oleh WHO sebagai persepsi seseorang terhadap posisi mereka didalam kehidupan melalui konteks budaya dan sistem penilaian yang ditetapkan ditempat dimana mereka hidup dan kaitannya dengan tujuan, ekspektasi, standar, dan kekhawatiran yang dimiliki [7]. Kualitas kehidupan sendiri dipengaruhi oleh berbagai aspek kehidupan mulai dari kesehatan, hubungan interpersonal, status edukasi, lingkungan kerja, kekayaan, keamanan, autonomi dalam mengambil keputusan, hingga status sosial [6]. Dalam dunia kerja, aspek kehidupan berkaitan dengan standar pekerjaan dan dan ekspektasi

yang disematkan pada profesi seorang individu, dimana WHO mendeskripsikan bahwa berlebihnya jam kerja, hingga tidak cocoknya keahlian dengan profesi yang dijalani dapat mempengaruhi kesehatan mental [8]. Hal yang sama juga berpengaruh bagi mahasiswa, dimana kesulitan dalam kegiatan edukasional, hingga pola tidur dapat menjadi determinan dalam analisa kondisi kesehatan mental [49].

Dalam menganalisis dampak dari penyakit atau kondisi tertentu, AI dan dalam hal ini *Machine Learning*, dapat menjadi jawaban untuk membantu proses diagnosis dalam bidang medis [9]. *Machine Learning* adalah sebuah subset dari AI yang menggunakan data sebagai sumber input yang menghasilkan klasifikasi ataupun regresi yang pada umumnya sulit dilakukan oleh manusia [10]. Sejak meningkatnya penggunaan *Deep Learning* yang merupakan salah satu perwujudan *Machine Learning*, terbukti MLBDD (*Machine Learning Based Disease diagnose*) memiliki akurasi lebih dari 90% dalam proses diagnosa penyakit [9].

Salah satu algoritma *Machine Learning* yang telah terbukti efektif dalam analisa dan identifikasi penyakit atau kondisi tertentu yang dinilai kompleks serta akurat dalam klasifikasi data adalah SVM [11]. Kemampuan SVM untuk memproses data berdimensi tinggi dan menangani stratifikasi, yaitu proses memisahkan atau mengatur populasi data menjadi sejumlah subgrup, yang kompleks menjadikannya alternatif menjanjikan untuk memahami pola mendasar dan variabilitas dalam penyakit atau kondisi tertentu [12]. Dengan fitur data yang beragam pada penelitian ini, mulai dari tingkat usia responden, hingga tingkat keparahan dari kondisi depresi pasca melahirkan, SVM juga mempertimbangkan

kecepatan pemrosesan dengan memaksimalkan margin antar kelas dan memberikan hasil yang akurat dan efisien [13].

Selain SVM, *K-Nearest Neighbours* atau sering juga disebut K-NN, juga merupakan algoritma yang dikenal akan kemampuannya memberikan hasil yang *reliable* dengan akurasi yang tinggi mengingat sifatnya yang secara otomatis memasang kelas dengan sampel yang labelnya mirip sebagai karakterisasi dari sampel yang tidak memiliki label [25]. Kemampuan K-NN juga sudah terbukti dalam menyelesaikan masalah klasifikasi ataupun prediksi hingga regresi dengan membahas kepastian statistik *nonasymptotic* dasar serta mengumpulkan metode-metode yang sudah tenar dan terbukti esensinya dalam mengukur sistem prediksi yang dapat menganalisa dataset berukuran besar [26].

Metode lainnya, yakni *Random Forest*, adalah model pembelajaran yang digunakan pada klasifikasi dan prediksi dengan konsep pohon keputusan yang dikonstruksikan menggunakan *subset* acak yang ada pada data dimana setiap pohon keputusan dilatih pada sebuah porsi atau bagian dari keseluruhan dataset [27]. *Random Forest* juga memiliki berbagai keunggulan seperti akurasi yang tinggi, kalkulasi terhadap kesalahan pada proses generalisasi, menentukan variabel penting, dan merepresentasikan *value* yang hilang menggunakan algoritma yang berdasar pada *proximity matrix* [28].

Penelitian sebelumnya yang dilakukan oleh Vaishnavi et al. [29], yang bertujuan untuk memprediksi penyakit mental menggunakan *machine learning* menunjukkan bahwa K-nn dan *Random Forest* memiliki tingkat akurasi masing-

masing pada angka 80.69% dan 81.22%, serta ROC masing-masing 0.87 untuk K-nn, dan 0.90 untuk *Random Forest*.

Sementara itu, pada penelitian yang dilakukan oleh Jain et al. [30], yang juga bertujuan untuk memprediksi kesehatan mental menggunakan *Machine Learning*, ditemukan bahwa diantara beberapa algoritma yang digunakan, SVM memberikan hasil akurasi tertinggi pada angka 87.38%.

Pada penelitian lainnya yang dilakukan oleh Madububambachu et al. [31], yang bertujuan untuk memprediksi hasil diagnosa kesehatan mental menggunakan *Machine Learning*, ditemukan bahwa SVM meraih akurasi mencapai 90% untuk *f-measure* dan 95% untuk AUC.

Berdasarkan uraian latar belakang, maka diusulkan judul tugas akhir “Analisis Dampak Aspek Kehidupan Terhadap Tingkat Depresi Dan Kecemasan Pada Mahasiswa Dan Pekerja Menggunakan Machine Learning Dan Hyperparameter Tuning”.

1.2 RUMUSAN MASALAH

Berdasarkan uraian latar belakang diatas maka dapat diidentifikasi masalah pada judul yang diangkat adalah :

1. Bagaimana menggunakan algoritma SVM, K-NN, dan *Random Forest* dan *hyperparameter tuning* untuk menentukan dampak dari aspek kehidupan pada tingkat depresi dan kecemasan pada dataset *final depression dataset*.

2. Bagaimana perbandingan hasil antara algoritma SVM, K-NN, dan random forest dalam menentukan dampak dari aspek kehidupan pada tingkat depresi dan kecemasan pada dataset *final depression dataset*.
3. Apa yang bisa dikatakan sebagai aspek kehidupan paling berpengaruh dalam menentukan tingkat depresi dan kecemasan pada dataset *final depression dataset*.

1.3 BATASAN MASALAH

Adapun untuk batasan masalah pada tugas akhir ini antara lain :

- 1) Analisis algoritma SVM, K-NN, dan *Random Forest* serta *Hyperparameter Tuning* ini menghasilkan klasifikasi dampak dari aspek kehidupan pada tingkat depresi dan kecemasan pada dataset *final depression dataset*.
- 2) Studi kasus hanya dilakukan pada dataset *final depression dataset* dengan total 2556 data serta 19 atribut yang diambil melalui website *Kaggle*.
- 3) Proses input kode menggunakan *Google Colab*..

1.4 TUJUAN DAN MANFAAT PENELITIAN

1.4.1 Tujuan Penelitian

Adapun tujuan yang ingin di capai dari penelitian ini, yaitu :

1. Dapat menganalisis tingkat akurasi algoritma SVM, K-NN, dan *Random Forest* yang dikombinasikan dengan *hyperparameter tuning* dalam

mengklasifikasi dampak dari tingkat depresi dan kecemasan yang didasarkan dari dataset *final depression dataset*.

2. Menentukan kombinasi parameter terbaik berdasarkan *Hyperparameter Tuning* pada performa dari algoritma yang digunakan.
3. Mengembangkan dan melatih SVM, K-NN, dan *Random Forest*, untuk menentukan tingkat depresi dan kecemasan dari dataset *final depression dataset* berdasarkan nilai aspek kehidupan yang ada pada dataset.

1.4.2 Manfaat Penelitian

Adapun manfaat dari penelitian ini sebagai berikut :

1. Memberikan informasi yang mempermudah klasifikasi dampak tingkat depresi dan kecemasan berdasarkan pada dataset mahasiswa.
2. Memahami pola yang mendasari variabilitas dampak dari tingkat depresi dan kecemasan pada mahasiswa menggunakan metode SVM, K-NN, dan *Random Forest*.
3. Dapat menjadi referensi bagi peneliti lain yang berminat melakukan penelitian sejenis.

1.5 SISTEMATIKA PENULISAN

Adapun sistematika penulisan pada tugas akhir ini terdiri dari 5 (lima) bab sebagai berikut :

BAB I : PENDAHULUAN

Pada bab ini berisi tentang latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian dan manfaat penelitian, dan sistematika penulisan.

BAB II : LANDASAN TEORI

Dalam bab landasan teori ini terdiri dari konsep konsep teoritis yang digunakan sebagai kerangka atau landasan untuk mendukung pemahaman terhadap penelitian yang penulis lakukan berupa pengertian dasar mengenai analisis, SVM, K-nn, dan *random forest* serta *hyperparameter tuning*, serta alat bantu pembuatan program mulai dari *Google Colab*, *Visual Studio Code*, *Excel*.

BAB III : METODOLOGI PENELITIAN

Pada bab ini menjelaskan tentang tahapan proses yang dilakukan selama mengerjakan penelitian, Teknik pengumpulan data, serta alat bantu yang digunakan untuk analisis.

BAB IV : ANALISIS

Pada bab ini berisi analisis dari topik yang telah diambil, profil data, pemilihan data, preprocessing data, serta interpretasi data..

BAB V : PENUTUP

Pada bab penutup ini terdiri atas Kesimpulan dan saran yang berhubungan dengan hasil penelitian yang dilakukan.

