

BAB I

PENDAHULUAN

1.1 Latar Belakang Masalah

Klasifikasi merupakan cara pengelompokan atau penggolongan benda berdasarkan ciri-ciri yang dimiliki oleh objek klasifikasi (Prasetya dkk., 2018). Tujuan klasifikasi ini adalah meningkatkan kehandalan hasil yang diperoleh dari data (Nata dan Suparmadi, 2022). Proses Klasifikasi memiliki empat komponen yaitu kelas, *predictor*/variabel bebas, *training dataset* untuk menentukan kelas yang cocok dari *predictor*, *testing dataset*/set data baru yang akan diklasifikasikan pada model yang sudah dibuat (Indah, dkk., 2019). Klasifikasi juga metode yang mencoba menemukan hubungan/korelasi antara hubungan variabel masukan dan variabel target. Klasifikasi dalam prosesnya dapat dilakukan dengan banyak cara baik secara manual maupun dengan bantuan teknologi. Klasifikasi yang dilakukan dengan cara manual yaitu klasifikasi yang dilakukan oleh manusia secara langsung tanpa adanya algoritma komputer diantaranya *decision trees* dan regresi logistik. Sedangkan, Klasifikasi yang dibantu oleh teknologi mempunyai beberapa algoritma, diantaranya Naive Bayes, *Support Vector Machine* (SVM), *Artificial Neural Network* (ANN), *Decision Tree*, *Fuzzy*, dan Jaringan saraf tiruan (Prasetya dkk., 2018).

Pohon keputusan/ *decision trees* adalah salah satu teknik yang paling banyak digunakan dalam masalah klasifikasi (Yuliani, 2014). Pohon keputusan dibantu oleh data data pelatihan $(x_i, y_i), i = 1, \dots, n$ (Bertsimas dan Dunn, 2017). Sebuah *Decision trees* terdapat beberapa simpul yaitu *tree's root*/simpul akar untuk *node* awal, *internal node*/node percabangan yang hanya terdiri satu input dan minimal dua output, dan *leafs/terminal node*/simpul daun yaitu node akhir yang hanya terdapat satu input dan tidak ada output. Setiap simpul internal membagi ruang menjadi dua atau lebih sesuai pada fungsi diskrit pada input atribut nilai. Klasifikasi menggunakan *decision trees* dilakukan oleh *routing*/alur dari simpul akar sampai di simpul daun. Pohon keputusan memberikan fleksibilitas dalam berbagai tugas *data mining* seperti klasifikasi,

clustering, regresi, dan seleksi fitur. Algoritma *decision tree* antaranya ID3, C4.5, C5.0, dan CART (Prasetya dkk., 2018)

Classification and Regression Trees (CART) merupakan salah satu metode atau algoritma dari salah satu teknik eksplorasi data *decision tree* (Indah dkk., 2019). Klasifikasi dapat menganalisis struktur data yang berukuran besar dan kompleks (Lewis,2000). Proses *Classification and Regression Trees* (CART) menghasilkan sebuah pohon klasifikasi, di mana saat variabel responnya berupa data kategoris, dan CART juga dapat menghasilkan pohon regresi (*regression trees*) jika variabel responnya kontinu (Breiman, 1984). Metode Klasifikasi CART merupakan metode non parametrik yang berguna untuk mendapatkan suatu kelompok data yang akurat untuk penciri dari suatu pengklasifikasian (Indah dkk., 2019). Timofeev (dalam Indah dkk., 2019) mengemukakan bahwa kelebihan CART tidak membutuhkan spesifikasi dari bentuk model, tidak harus menentukan variabel di awal proses, dapat menangani *outliers*, tidak memiliki asumsi dan memiliki komputasi yang cepat, dan data fleksibel sesuai kebutuhan. Namun, CART ini mungkin akan menghasilkan pohon keputusan yang tidak stabil dan membagi data hanya pada satu variabel.

Pada kasus pengklasifikasian kompleks dengan sejumlah kumpulan data yang besar sering kali ditemukan dalam kondisi himpunan data yang tidak seimbang, di mana kelas data satu lebih banyak atau lebih sedikit dari kelas data lainnya (Syifauly, 2022). Permasalahan ketidakseimbangan data mengakibatkan *classifier machine* lebih mengutamakan untuk memprediksi kelas data yang lebih banyak dibandingkan yang lebih sedikit (minoritas) alhasil tingkat akurasi prediksinya pun kecil (Chawla dkk.,2002). Solusi dari permasalahan data tidak seimbang ini dapat diselesaikan dengan beberapa metode salah satunya yaitu pendekatan seleksi fitur (Pratama dkk., 2019). Seleksi fitur dilakukan supaya dapat mengurangi dimensi data yang tidak relevan (Amini dkk., 2022). Gunawan (dalam Amini dkk., 2022) mengemukakan jika algoritma genetika merupakan metode yang mampu menyeleksi fitur untuk meningkatkan hasil akurasi dari klasifikasi.

Genetic Algorithm (GA) mampu menemukan solusi global pada suatu perhitungan yang tidak diaplikasikan oleh teknik optimasi metode *deterministic* (Noor,2016). Salah satu aplikasi yang digunakan oleh *Genetic Algorithm* (GA) yaitu

permasalahan optimasi kombinasi yang menghasilkan suatu nilai solusi optimal terhadap suatu permasalahan yang mempunyai banyak kemungkinan Solusi (Carneiro dkk., 1998). *Genetic Algorithm* (GA) ditemukan oleh John H.Holland dari Universitas Michigan. *Genetic Algorithm* (GA) adalah algoritma yang memanfaatkan proses seleksi alamiah atau disebut juga dengan proses evolusi (Sofwan dkk., 2008). Evolusi Darwin mengemukakan jika yang kuat adalah yang menang dalam kelangsungan hidup suatu makhluk, lalu Darwin juga mengatakan proses reproduksi yang digunakan untuk mempertahankan kelangsungan hidup suatu makhluk hidup adalah *crossover* (perkawinan silang) dan mutasi (pengganti gen yang hilang) (Paranduk dkk., 2018).

Penggunaan *Genetic Algorithm* (GA) sebagai pemilihan fitur pada optimasi klasifikasi Naïve Bayes sudah menunjukkan peningkatan akurasi (Nugroho dan Religia, 2021). Pada setiap generasi, kromosom akan melewati proses evaluasi dengan menggunakan alat ukur yang disebut dengan fungsi *fitness* atau alat untuk melihat nilai dari kromosom (Noor, 2016). Solusi dari *Genetic Algorithm* (GA) disebut dengan kromosom. Kromosom tersebut yang dibangkitkan secara acak pada tahap awal proses GA, semakin tinggi nilai *fitness* kromosom tersebut maka peluang terpilih juga akan semakin tinggi (Ahmad, 2016). CART dengan bantuan *Particle Swarm Optimization* (PSO) menghasilkan akurasi 78,2%, kemudian CART+*Artificial Bee Colony* (ABC) menghasilkan 82% (Irawan dkk., 2021). Kelemahan PSO kadang-kadang mengalami konvergensi prematur ke solusi lokal optimal yang berdampak PSO tidak menemukan solusi optimal global (Kurniawan dan suciati, 2017). GA menggunakan mekanisme *crossover* dan mutasi untuk menghasilkan populasi baru yang terus berulang untuk mencegah konvergensi prematur (UNS, hlm.11). Sedangkan, kelemahan ABC kurang efisien dalam pemilihan fitur atau parameter optimasi karena dalam mengeksplorasi solusi lebih sederhana. GA terdapat mekanisme seleksi berdasarkan fungsi *fitness* yang memungkinkan pemilihan fitur atau parameter optimasi yang lebih efisien melalui evaluasi terus menerus dan pemilihan individu terbaik (Harafani dan Maulana, 2019, hlm.53).

Proses CART-GA merupakan pengembangan sebuah *Genetic Algorithm* (GA) untuk membangun pohon keputusan dengan akurasi tinggi. Algoritma pohon Keputusan CART digunakan dalam populasi awal untuk meningkatkan kinerja pohon

tersebut, sedangkan dalam GA digunakan untuk menerapkan operasi evolusioner, kemudian pohon keputusan harus diwakili sebagai struktur kromosom. Proses ini mengusulkan struktur untuk mengkodekan pohon keputusan ke dalam kromosom dan juga untuk menghasilkan populasi awal, kemudian membagi dataset asli menjadi subset dengan ukuran yang wajar dan menghasilkan pohon dengan subset ini pada CART. Dengan demikian, dalam GA data set ukuran penuh dan populasi awal digunakan untuk membuat pohon yang lebih akurat (Savitri dkk.2020). Jadi, peneliti ingin menggunakan kekuatan pohon CART dalam algoritma genetika dan mencoba meningkatkan kinerja solusi dari CART yang berlebihan.

WHO atau *World Health Organization* mendefinisikan kesehatan mental sebagai adaptasi yang berhasil dari lingkungan eksternal maupun internal (Dewi dkk., hlm.29). Hidalgo mengungkapkan terdapat banyak hal yang berpengaruh pada kesehatan mental seperti stres, kejadian traumatik, depresi, keadaan ekonomi sosial, kemampuan individu dalam mengontrol lingkungan di sekitarnya agar tetap baik untuk dirinya (Marchelina, 2021). Depresi merupakan salah satu indikator penting atas rendahnya kesehatan mental seseorang dan telah menjadi masalah kesehatan utama di seluruh dunia karena jumlah penderitanya yang meningkat setiap tahun (Marchelina, 2021). Depresi ini melibatkan gangguan emosional yang sering terjadi pada remaja yang menyebabkan perasaan sedih, lelah, marah, kehilangan minat untuk melakukan aktivitas sehari-hari dan cenderung bunuh diri (Auliya dkk., 2021). Negara yang mempunyai total kasus gangguan mental gejala depresi tertinggi di wilayah Asia Tenggara dan sekitarnya adalah India dengan jumlah 56.675.969 kasus, yang disusul dengan negara Indonesia yaitu peringkat kedua dengan jumlah 9.162.886 kasus (Dewi dkk., hlm.29).

Remaja merupakan salah satu kelompok yang berisiko tinggi mengalami gangguan kesehatan mental yang didukung oleh hasil Riset Kesehatan Dasar (Riskesmas) tahun 2018 bahwa penduduk berusia lebih dari sama dengan 15 tahun mengalami gangguan mental emosional dengan jumlah lebih dari 19 juta dan penduduk berusia 15 tahun mengalami depresi dengan jumlah 12 juta penduduk (Auliya dkk, 2021). I-NAMHS (*Indonesia-National Adolescent Mental Health Survey*) menemukan di mana satu dari tiga remaja (34,9%) memiliki masalah kesehatan mental menjadikan kasus kesehatan masyarakat yang cukup serius untuk kelompok usia remaja (Ellyza,

2022). Menurut PPDGJ III (Pedoman Penggolongan Dalam Gangguan Jiwa) faktor-faktor gejala depresi dapat diklasifikasikan pada tingkatan depresi yang berbeda, yaitu depresi berat, depresi sedang, dan depresi ringan (Departemen Kesehatan R.I, 1993). Namun, gejala utama atau gejala yang dirasakan pada seluruh tingkatan yang ada (ringan, sedang, maupun berat) adalah afek depresif (cenderung merasa sedih), kehilangan minat, dan berkurangnya energi yang menuju meningkatnya keadaan mudah lelah. Oleh karena itu, peneliti tertarik untuk mengklasifikasikan faktor depresi pada remaja sesuai tingkat depresi dengan menerapkan metode CART (*Classification and Regression trees*) dengan pendekatan *Genetic Algorithm* (GA) atau CART-GA dengan data simulasi.

1.2 Rumusan Masalah

Permasalahan yang ditinjau dalam penelitian ini setelah diuraikan latar belakang tersebut yakni bagaimanakah implementasi CART-GA pada klasifikasi depresi remaja?

1.3 Tujuan Penelitian

Tujuan yang ditargetkan dalam penelitian ini yakni mengklasifikasikan faktor depresi pada setiap tingkatan depresi (Depresi Berat, Depresi Sedang, Depresi Ringan).

1.4 Manfaat Penelitian

Terdapat beberapa manfaat atau keuntungan yang didapatkan dari penelitian ini yakni responden (remaja) yang mengalami depresi untuk mendapatkan pengobatan kepada ahlinya (psikolog atau psikiater)