

BAB I

PENDAHULUAN

1.1 Latar Belakang

Gangguan jiwa menjadi isu kesehatan masyarakat global yang terus meningkat dalam dua dekade terakhir karena prevalensinya yang tinggi serta memiliki dampak terhadap individu, keluarga, dan sistem kesehatan. Menurut DSM-5 (American Psychiatric Association, 2013), gangguan jiwa merupakan kondisi yang ditandai oleh adanya gangguan pada proses kemampuan berpikir, pengendalian emosi, atau perilaku seseorang yang menunjukkan adanya disfungsi dalam proses psikologis, biologis, atau perkembangan yang mendasari fungsi mental individu (Stein *et al.*, 2021). Pada tingkat global, lebih dari 1 miliar penduduk dunia hidup dengan gangguan kesehatan jiwa, terutama kecemasan dan depresi (WHO, 2025). Selain menyebabkan penurunan kualitas hidup dan disabilitas, gangguan jiwa juga secara signifikan memengaruhi beban ekonomi. Kehilangan produktivitas akibat ketidakhadiran kerja dan penurunan kinerja saat bekerja pada individu dengan kecemasan dan depresi diperkirakan menimbulkan kerugian ekonomi hingga IDR 463,8 triliun per tahun (Arulsamy *et al.*, 2025). Riset Kesehatan Dasar (Riskesdas) 2018 menunjukkan bahwa lebih dari 19 juta penduduk berusia lebih dari 15 tahun mengalami gangguan mental emosional. Indonesia memiliki prevalensi orang dengan gangguan jiwa sekitar 1 dari 5 penduduk, artinya sekitar 20% populasi di Indonesia mempunyai potensi masalah gangguan jiwa (Kemenkes, 2021). Angka ini menunjukkan bahwa gangguan jiwa

merupakan masalah kesehatan masyarakat yang perlu mendapatkan perhatian komprehensif.

Gangguan jiwa sering kali disertai dengan komorbiditas yaitu kondisi di mana dua atau lebih gangguan jiwa secara bersamaan terjadi pada satu individu. Sekitar 78% individu dengan *Clinical High Risk for Psychosis* (CHR-P) memiliki setidaknya satu gangguan jiwa komorbid seperti gangguan mood, kecemasan, dan trauma (Solmi *et al.*, 2023). Kondisi tersebut menimbulkan tantangan bagi sistem klasifikasi tradisional yang cenderung menggunakan satu label diagnosis padahal satu pasien dapat mengalami beberapa gangguan sekaligus. Dalam *machine learning*, klasifikasi adalah proses membangun model yang menghubungkan input (X) ke label output (Y), sehingga model mampu menentukan kelas suatu data baru berdasarkan hubungan yang telah dipelajari dari data sebelumnya (Alpaydin, 2014). Namun, pendekatan klasifikasi tradisional hanya memungkinkan suatu data diklasifikasikan hanya ke dalam satu label, sehingga tidak sesuai ketika pasien memiliki komorbiditas.

Komorbiditas pada pasien gangguan jiwa memerlukan metode klasifikasi *multi-label* yang mampu memodelkan banyak label secara bersamaan karena pendekatan klasifikasi konvensional yang hanya memprediksi satu kategori untuk setiap individu tidak lagi efektif. Klasifikasi *multi-label* adalah metode pembelajaran mesin yang digunakan ketika satu data dapat dikaitkan dengan lebih dari satu label pada saat yang bersamaan sehingga proses prediksi dilakukan dengan mempertimbangkan kemungkinan kemunculan beberapa label secara bersamaan (Zhang & Zhou, 2014). Klasifikasi *multi-label* sudah dimanfaatkan oleh beberapa penelitian sebelumnya untuk mengklasifikasikan data medis karena banyak data

medis melibatkan lebih dari satu diagnosis yang perlu diprediksi secara bersamaan. Klasifikasi *multi-label* telah digunakan untuk menganalisis data medis yang memuat lebih dari satu diagnosis guna meningkatkan ketepatan identifikasi komorbiditas pada pasien (Wosiak *et al.*, 2018). Selain itu, klasifikasi *multi-label* juga telah digunakan untuk kasus gangguan jiwa di Nigeria, yang menunjukkan bahwa pendekatan ini telah digunakan dalam penelitian terkait kesehatan jiwa (Folorunso *et al.*, 2020). *Multi-label K-Nearest Neighbor* (ML-KNN) merupakan salah satu algoritma klasifikasi yang dikembangkan secara khusus untuk menangani masalah data dengan lebih dari satu label pada setiap objek. Algoritma ini pertama kali diperkenalkan dengan menunjukkan kinerja yang lebih baik dibanding beberapa algoritma *multi-label* lain pada berbagai dataset (Zhang & Zhou, 2007). Penelitian tersebut juga menjabarkan bahwa ML-KNN memperoleh nilai evaluasi yang tinggi pada beberapa indikator seperti *hamming loss*.

Hingga saat ini, ML-KNN terus mengalami perkembangan dan digunakan dalam berbagai penelitian modern yang menunjukkan relevansi dan daya adaptasinya terhadap permasalahan *multi-label* yang semakin kompleks. Misalnya, ML-KNN diterapkan untuk klasifikasi *multi-label* pengguna media sosial dan terbukti mampu memberikan akurasi sebesar 91% yang lebih tinggi dibandingkan metode *multi-label* lain yaitu *Binary Relevance* dengan akurasi sebesar 80% dan *Label Powerset* dengan akurasi sebesar 70% (Huang *et al.*, 2023). Penelitian lain juga mengembangkan penentuan nilai k secara lokal untuk meningkatkan performa ML-KNN pada ruang fitur yang memiliki distribusi label berbeda-beda, dan hasilnya secara signifikan meningkatkan performa evaluasi dengan *Hamming Loss* sebesar 13% yang lebih rendah dibandingkan ML-KNN standar dengan nilai

Hamming Loss sebesar 15% (Romero-del-Castillo *et al.*, 2022). Selain itu, terdapat penelitian yang memperkenalkan varian *generalized weighted distance KNN* yang mampu memperbaiki performa pada data *multi-label* dengan nilai *F1 Score macro* sebesar 70,14% pada dataset emotions dimana lebih tinggi dibandingkan nilai pada ML-KNN yang sebesar 67,77% (Rastin *et al.*, 2021). Selain penelitian yang berfokus pada pengembangan ML-KNN, evaluasi komprehensif terhadap berbagai algoritma *multi-label* terus dilakukan. Algoritma *multi-label*, termasuk ML-KNN, tetap menjadi perhatian utama dalam kajian perbandingan performa klasifikasi modern (García-pedrajas *et al.*, 2024). Berdasarkan perkembangan penelitian tersebut, ML-KNN tidak hanya terbukti relevan, tetapi juga terus dikaji dan disempurnakan sehingga memiliki landasan teoritis dan empiris yang kuat untuk digunakan dalam penelitian ini.

Dalam ML-KNN, penentuan label dilakukan menggunakan pendekatan *Maximum A Posteriori* (MAP) yang menghitung probabilitas kemunculan sebuah label berdasarkan informasi statistik dari k tetangga terdekat pada data latih. Permasalahan utama pada data *multi-label* adalah ketidakseimbangan distribusi label (Jiang *et al.*, 2019). Pada ML-KNN, proses inferensi dilakukan melalui perhitungan jumlah tetangga yang memiliki label tertentu, sehingga ketidakseimbangan distribusi label secara langsung mempengaruhi nilai prediksi di mana label dengan frekuensi kemunculan lebih besar memiliki peluang terpilih lebih tinggi sementara label minoritas memiliki kemungkinan kecil untuk diprediksi. Oleh karena itu, perbaikan terhadap ML-KNN diperlukan agar proses prediksi tidak hanya bergantung pada jumlah kemunculan label, tetapi juga mempertimbangkan proporsi label yang tidak seimbang dalam data.

Selain permasalahan ketidakseimbangan label, ML-KNN juga tidak mempertimbangkan struktur distribusi data latih dalam ruang fitur. ML-KNN dapat menghasilkan prediksi yang kurang tepat ketika data pelatihan dengan suatu label berkumpul lebih padat dalam area tertentu (Jiang *et al.*, 2019). Hal ini terjadi karena ML-KNN memperlakukan semua tetangga secara sama tanpa menilai apakah tetangga tersebut benar-benar mewakili pola label yang relevan atau hanya berada dalam wilayah yang padat. Akibatnya, sebuah objek bisa diprediksi memiliki label mayoritas hanya karena jumlah tetangga berlabel tersebut lebih banyak, meskipun secara karakteristik data lebih sesuai diprediksi memiliki label lain.

Untuk mengatasi permasalahan dalam ML-KNN, yaitu ketidakseimbangan jumlah label yang terdapat dalam data latih serta distribusi spasial data latih yang tidak merata di ruang fitur, dikembangkan metode *Weighted ML-KNN* dengan menambahkan dua mekanisme pembobotan pada proses klasifikasi (Jiang *et al.*, 2019). Pembobotan pertama diberikan berdasarkan distribusi jumlah label dalam data pelatihan. Melalui pendekatan ini, label dengan frekuensi kemunculan yang rendah memperoleh bobot lebih besar agar pengaruhnya dalam proses penentuan label tidak terabaikan oleh dominasi label mayoritas, sehingga perhitungan probabilitas posterior tidak bias terhadap distribusi frekuensi. Pembobotan kedua diterapkan berdasarkan kondisi distribusi spasial data latih dalam ruang fitur. Dalam pendekatan ini, kepadatan data di sekitar data uji serta tingkat keterkaitan antara data uji dan tetangga terdekatnya dihitung untuk menentukan bobot setiap tetangga. Tetangga yang dinilai lebih relevan terhadap pola data yang sedang diprediksi akan memperoleh bobot yang lebih besar dibandingkan tetangga lain yang hanya berada dalam jumlah lebih banyak tetapi tidak memiliki kedekatan

informasi yang kuat. Dengan dua jenis pembobotan tersebut, proses inferensi menjadi lebih seimbang karena mampu mempertimbangkan perbedaan frekuensi label sekaligus pola penyebaran data dalam ruang fitur.

Hasil penelitian menggunakan *Weighted* ML-KNN pada beberapa dataset menunjukkan bahwa *Weighted* ML-KNN memberikan peningkatan performa dibandingkan ML-KNN, terutama dalam menurunkan kesalahan klasifikasi serta meningkatkan metrik evaluasi *multi-label* (Jiang *et al.*, 2019). Hal tersebut menunjukkan bahwa penambahan bobot merupakan pendekatan yang tepat secara teoritis dan empiris untuk mengatasi keterbatasan ML-KNN, khususnya pada data dengan distribusi label tidak seimbang dan distribusi spasial yang tidak merata dalam ruang fitur. Berdasarkan pertimbangan tersebut, penelitian ini mengimplementasikan pembobotan yang dikembangkan oleh Jiang *et al.* (2019) agar proses klasifikasi *multi-label* pada data penelitian dapat menghasilkan hasil prediksi yang lebih akurat dan sesuai dengan karakteristik data yang digunakan.

Penelitian ini menggunakan empat penyakit dalam gangguan jiwa yaitu skizofrenia, gangguan skizoafektif, gangguan bipolar, dan episode depresi karena keempat label tersebut merupakan gangguan yang sering muncul bersamaan pada satu pasien. Pemilihan keempat label tersebut beserta variabel prediktor (X) disesuaikan dengan ketersediaan data rekam medis di RSUD dr. Loekmono Hadi Kudus, sehingga variabel penelitian merepresentasikan kondisi pasien yang sebenarnya dan relevan untuk dianalisis menggunakan metode klasifikasi *multi-label*.

Seluruh variabel prediktor yang tersedia dalam data rekam medis di RSUD dr. Loekmono Hadi Kudus merupakan variabel kategorik, sehingga diperlukan

ukuran jarak yang sesuai dengan karakteristik tersebut. Metode jarak seperti *Euclidian* atau *Manhattan* tidak tepat digunakan karena keduanya hanya bisa digunakan pada data numerik. Jarak Gower merupakan metode perhitungan jarak untuk menghitung kedekatan antar objek berdasarkan kesamaan nilai tanpa mensyaratkan bentuk skala tertentu (Gower, 1971). Koefisien kemiripan Gower merupakan salah satu ukuran kedekatan yang banyak digunakan untuk data bertipe kategorik maupun campuran karena mampu menghitung kemiripan antara objek tanpa mensyaratkan skala data tertentu (Hamid *et al.* (2017). Oleh karena itu, penelitian ini menggunakan jarak Gower sebagai dasar perhitungan kedekatan antar data agar proses penentuan tetangga terdekat pada ML-KNN tetap sesuai dengan karakteristik variabel yang seluruhnya bersifat kategorik.

Berdasarkan uraian latar belakang di atas, penelitian ini dilakukan untuk mengkaji penerapan algoritma *Weighted* ML-KNN dengan pendekatan jarak Gower dalam mengklasifikasikan data pasien dengan diagnosis gangguan jiwa di RSUD dr. Loekmono Hadi Kudus dan mengevaluasi performa hasil klasifikasinya. Penelitian ini juga membandingkan kinerja model ML-KNN dengan *Weighted* ML-KNN, sehingga dapat diketahui sejauh mana penambahan bobot mampu meningkatkan performa hasil klasifikasi.

1.2 Rumusan Masalah

Berdasarkan uraian latar belakang di atas, rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Bagaimana kinerja metode *Weighted* ML-KNN dalam melakukan klasifikasi *multi-label* pada data gangguan jiwa?

2. Bagaimana evaluasi performa klasifikasi menggunakan metode *Weighted* ML-KNN dalam mengklasifikasikan penyakit gangguan jiwa?
3. Bagaimana perbandingan evaluasi performa klasifikasi menggunakan metode *Weighted* ML-KNN dengan metode ML-KNN?

1.3 Batasan Masalah

Batasan masalah yang digunakan dalam penelitian ini adalah:

1. Analisis klasifikasi pada penelitian ini menggunakan metode *Weighted Multi-label K-Nearest Neighbor (Weighted ML-KNN)* sebagai metode utama klasifikasi *multi-label*.
2. Data yang digunakan dalam penelitian ini merupakan data rekam medis pasien pada poli jiwa di RSUD dr. Loekmono Hadi Kudus periode Januari 2023- Oktober 2025 yang memiliki diagnosis *multi-label*.
3. Variabel target yang dianalisis dibatasi pada empat jenis gangguan jiwa yaitu skizofrenia, gangguan skizoafektif, gangguan bipolar, dan episode depresi.
4. Variabel prediktor pada penelitian ini merupakan gejala klinis yang tercatat selama masa perawatan pasien yang disesuaikan dengan ketersediaan data rekam medis di RSUD dr. Loekmono Hadi Kudus.
5. Penelitian ini tidak mencakup proses seleksi variabel karena fokus utama penelitian difokuskan pada penerapan serta evaluasi kinerja metode tanpa modifikasi struktur fitur.
6. Nilai k yang digunakan dalam penerapan klasifikasi *Weighted Multi-label K-Nearest Neighbor* adalah 3 hingga 13.

7. Nilai faktor koreksi bobot (σ) yang digunakan untuk dipilih melalui *hyperparameter tuning* adalah 0,0001, 0,001, 0,01, 0,1, dan 1.

1.4 Tujuan Penelitian

Berdasarkan latar belakang dan rumusan masalah yang telah diuraikan, penelitian ini memiliki tujuan sebagai berikut:

1. Menerapkan metode *Weighted* ML-KNN untuk mengklasifikasikan pasien gangguan jiwa di RSUD dr. Loekmono Hadi Kudus.
2. Mengevaluasi performa hasil klasifikasi menggunakan metode *Weighted* ML-KNN dalam mengklasifikasikan pasien gangguan jiwa di RSUD dr. Loekmono Hadi Kudus.
3. Membandingkan hasil performa metode *Weighted* ML-KNN dengan ML-KNN untuk mengetahui sejauh mana pendekatan pembobotan mampu meningkatkan performa klasifikasi pasien gangguan jiwa di RSUD dr. Loekmono Hadi Kudus.