

BAB I

PENDAHULUAN

1.1 Latar Belakang

Asam mefenamat (*mefenamic acid*), merupakan salah satu obat-obat inflamasi non-steroid yang sulit untuk larut di dalam air, namun mudah untuk larut dalam pelarut organik hidroliflik seperti etanol (Putranti et al., 2019). Sifat fisiokimianya inilah yang menyebabkan tantangan besar bagi perusahaan farmasi dalam pengembangan dan pembuatan obat berbentuk sediaan yang menggunakan bahan baku *mefenamic acid* (Srivastava et al., 2019). Dengan dosis yang diperbolehkan antara 250 mg sampai 500 mg dalam kurun waktu pemakaian 7 hari, penentuan kandungan *mefenamic acid* di dalam obat bentuk sediaan, khususnya obat untuk menghilangkan rasa nyeri, harus diperhatikan dengan baik agar dapat dihasilkan obat bentuk sediaan yang efektif namun tetap aman untuk dikonsumsi bagi manusia (Mohammed & Kareem, 2020). Dalam memproduksi obat seperti Mefenamic Acid kaplet 500mg, Mefenamic Acid kaptab 500mg, Omestan 500, Omestan Mefenamic 500mg, dan Omestan Sirup, PT. Mutifa Industri Farmasi mengalami masalah dalam memperkirakan berapa banyak bahan baku *mefenamic acid* yang dibutuhkan untuk produksi obat-obatan tersebut, berdasarkan tingkat penjualan obat-obatan tersebut pada periode produksi tahun-tahun sebelumnya.

Data mining merupakan sebuah proses untuk menemukan pola-pola menarik dari sekumpulan data berjumlah besar yang tersimpan di dalam *database*, *data warehouse*, atau penyimpanan informasi lainnya, dan berhubungan erat dengan bidang *machine learning*, khususnya *supervised learning* (Prasetya & Ridwan, 2019). Pada *supervised learning*, dataset akan di latih (*training*) berdasarkan informasi-informasi yang diberikan pada setiap data inputan dan target keluaran yang diinginkan, berdasarkan hasil pelatihan inilah kemudian sistem menghasilkan sebuah pola dari data yang dapat digunakan sebagai acuan untuk data-data berikutnya (Abijono et al., 2021). Sebagai salah satu teknik *supervised learning*, klasifikasi dapat dimanfaatkan untuk membentuk sebuah model yang mampu mengklasifikasikan data ke dalam target-target objek (Al-Khowarizmi et al., 2021). Penelitian ini memanfaatkan teknik klasifikasi dalam membangun sebuah model yang dapat mengklasifikasikan tingkat pemakaian bahan baku *mefenamic acid* dalam produksi obat-obatan pada perusahaan farmasi PT. Mutifa Industri Farmasi.

Beberapa algoritma klasifikasi diantaranya adalah *Support Vector Machine* (SVM) (Demidova et al., 2017), *Neural Network* (Kaya et al., 2019), *Genetic Algorithm* (Jude Hemanth & Anitha, 2019), C4.5 (Meng et al., 2020), dan *Random Forest* (Sathish kumar et al., 2022). Pendekatan dari algoritma yang dikelompokkan adalah pendekatan berbasis *classifier*. Diantara pendekatan tersebut maka C4.5 dan *Random Forest* menggunakan pendekatan *ensemble classifier*, dimana *Random Forest* melakukan proses *ensemble classifier* berdasarkan pada *multiple weak classifier* dan C4.5 melakukan proses *ensemble*

classifier dengan berdasarkan pada *node selection algorithm*. Kedua algoritma ini memiliki keunggulan dibandingkan yang lain karena tidak membutuhkan proses *data preprocessing* yang kompleks (Zhou et al., 2022).

Algoritma C4.5 sangat baik digunakan untuk menghasilkan sebuah model yang mampu menghasilkan keputusan, seperti pada penelitian klasifikasi ras anjing *bulldog* (Jinan et al., 2023), klasifikasi sentimen di media sosial Twitter untuk masalah pemblokiran PayPal oleh Kominfo (M. Z. Lubis et al., 2023), klasifikasi data jasa pengiriman pada perusahaan pelayaran (Khasanah et al., 2022), analisis minat siswa SMA untuk melanjutkan pendidikan ke perguruan tinggi (Lintang et al., 2022), dan klasifikasi potensi tsunami di Indonesia berdasarkan gempa bumi (Irfiani et al., 2022). Algoritma *random forest* juga memiliki performa yang cukup baik untuk digunakan sebagai model klasifikasi, seperti yang ditunjukkan pada hasil penelitian mengenai klasifikasi data fungsional (Maturo & Verde, 2022), pemetaan hutan mangrove di Taman Nasional Sembilang (Purwanto et al., 2023), klasifikasi anjungan minyak berbasis citra SAR C-Band (da Silva et al., 2022), prediksi sentimen pengguna jasa GoFood pada media sosial Twitter (Setiawan et al., 2022), dan klasifikasi sentimen masyarakat terhadap penyelenggaraan event MotoGP Mandalika 2022 (Margolang & Zarlis, 2023).

Penulis tertarik untuk membandingkan kinerja kedua algoritma ini untuk menghasilkan sebuah model klasifikasi tingkat pemakaian bahan baku *mefenamic acid* dalam produksi obat. Metode *k-fold cross validation* digunakan untuk memperoleh nilai *accuracy*, *precision*, dan *recall* dari masing-masing model

klasifikasi, dan dianalisa model mana yang memiliki kinerja terbaik untuk diterapkan dalam solusi pemecahan permasalahan penelitian ini. Rancangan penelitian ini kemudian penulis tuangkan dalam bentuk proposal tesis yang berjudul **“Perbandingan Algoritma C4.5 Dan Random Forest Dalam Klasifikasi Tingkat Pemakaian Bahan Baku Mefenamic Acid Dalam Produksi Obat”**.

1.2 Rumusan Masalah

Adapun rumusan masalah pada penelitian ini, berdasarkan latar belakang masalah di atas, adalah bagaimana perbandingan kinerja antara algoritma C4.5 dan *random forest* dalam mengklasifikasikan tingkat pemakaian bahan baku *mefenamic acid* dalam produksi obat.

1.3 Tujuan

Adapun tujuan dari penelitian ingin dicapai diantaranya sebagai berikut :

1. Memperoleh perbandingan kinerja antara algoritma C4.5 dan *random forest* dalam mengklasifikasikan tingkat pemakaian bahan baku *mefenamic acid* dalam produksi obat.
2. Evaluasi algoritma C4.5 dan *random forest* dengan berbagai nilai K yang digunakan pada *k-fold validation*.
3. Menganalisa teknik *5-fold cross validation*, *10-fold cross validation* dan *20-fold cross validation* untuk memperoleh nilai *accuracy*, *precision*, *recall* dan *f-measure* masing-masing.

1.4 Manfaat

Adapun manfaat yang diharapkan dari penelitian ini adalah menghasilkan analisis berupa algoritma yang paling cocok untuk digunakan pada model klasifikasi tingkat pemakaian bahan baku *mefenamic acid* dalam produksi obat, antara algoritma C4.5 dan *random forest*.

1.5 Batasan Masalah

Adapun batasan masalah dalam penelitian ini adalah sebagai berikut:

1. Dataset yang digunakan dalam penelitian dibatasi pada data produksi obat yang menggunakan bahan baku *mefenamic acid* pada periode lima tahun sebelumnya.
2. Output yang digunakan dalam memprediksi produksi obat adalah kategori besaran ketersediaan *mefenamic acid* yang terbagi ke dalam golongan rendah, sedang, dan tinggi.
3. Analisa yang dilakukan dibatasi pada perbandingan nilai *accuracy*, *precision*, dan *recall* dari hasil prediksi masing-masing algoritma.
4. Nilai K yang digunakan dalam *k-fold cross validation* dibatasi pada angka 5, 10 dan 20.

1.6 Sistematika Pembahasan

Adapun sistematika pembahasan yang digunakan di dalam penelitian ini adalah sebagai berikut:

BAB I : PENDAHULUAN

Bab ini menjelaskan tentang Latar Belakang, Rumusan Masalah, Tujuan, Manfaat serta Batasan Masalah penelitian.

BAB II : TINJAUAN PUSTAKA

Bab ini berisikan tentang teori-teori yang berkaitan langsung dengan penelitian seperti teori mengenai penelitian terkait, *mefenamic acid*, *machine learning*, klasifikasi, algoritma C4.5, dan algoritma *random forest* dari literatur yang diperoleh melalui buku-buku, jurnal dan penelitian-penelitian sebelumnya sebagai rujukan dalam penelitian ini.

BAB III : METODOLOGI PENELITIAN

Bab ini membahas tentang bentuk penerapan algoritma C4.5 dan *random forest* dalam klasifikasi tingkat pemakaian bahan baku *mefenamic acid* dalam produksi obat serta bentuk analisa yang digunakan untuk membandingkan hasil klasifikasi kedua algoritma tersebut.

BAB IV : HASIL

Bab ini membahas tentang hasil penelitian berupa data sampling, confusion matrix, serta nilai performa masing-masing model dalam bentuk accuracy, precision, dan recall.

BAB V : PEMBAHASAN

Bab ini membahas tentang analisa terhadap hasil klasifikasi masing-masing model untuk melihat dan menentukan model terbaik yang

diperoleh dari hasil penelitian klasifikasi tingkat pemakaian bahan baku Mefenamic Acid dalam produksi obat.

BAB VI : KESIMPULAN DAN SARAN

Bab ini berisi kesimpulan yang penulis peroleh dari hasil analisa penelitian serta saran-saran yang membangun untuk pengembangan penelitian berikutnya.