

BAB 1 PENDAHULUAN

1.1 Latar Belakang

Penyakit *liver* di Indonesia umumnya tergolong tinggi. Penyakit *liver* menempati urutan keempat belas dengan angka kematian mencapai 23.024 pertahun, dengan prosentase 0,01%. Pada umumnya penyakit *liver* dikenal sebagai penyakit yang berhubungan dengan organ *liver*. Penyakit *liver* dengan komplikasi merupakan masalah kesehatan yang masih sulit diatasi di Indonesia. Secara umum, angka kasus penyakit *liver* di rumah sakit seluruh Indonesia diperkirakan berkisar antara 0,6-14,5%.

Penyakit *liver* umumnya susah dikenali sejak dini dan penyakit ini cepat menyebar. Sementara itu, diagnosa penyakit *liver* yang ditangani lebih awal dapat meningkatkan kelangsungan hidup pasien. Dengan perkembangan teknologi saat ini diagnose penyakit dapat menggunakan metode *data mining*. Salah satu pengembangan dari *data mining* adalah klasifikasi. Metode klasifikasi dapat melakukan pembelajaran dengan memetakan suatu item dan ke dalam kelas berdasarkan kelas data yang telah didefinisikan sebelumnya [1].

Klasifikasi merupakan salah satu bagian dari *data mining*. *Data mining* merupakan proses yang mempekerjakan satu ataupun lebih metode pembelajaran komputer (*machine learning*) untuk menganalisis dan menghasilkan pengetahuan (*knowledge*) secara otomatis. Definisi lain dari *data mining* adalah pembelajaran berbasis induksi (*induction-based learning*) merupakan proses pembentukan definisi-definisi konsep umum yang dilakukan dengan cara mengobservasi contoh-contoh spesifik dari konsep-konsep yang akan dipelajari. *Knowledge Discovery in Database* (KDD) merupakan penerapan metode saintifik pada *data mining*. Dalam konteks ini, *data mining* merupakan satu langkah dari proses *Knowledge Discovery in Database* (KDD) [2].

Metode yang digunakan untuk membangun system klasifikasi dalam penyakit *liver* yaitu *Modified K- Nearest Neighbor* atau MKNN merupakan salah satu metode yang digunakan untuk klasifikasi dan memberikan hasil diagnosis, alasan

menggunakan algoritma ini karena memiliki keunggulan untuk mengklasifikasikan data gejala penyakit dengan data uji yang ada. MKNN memiliki penambahan prosedur dalam melakukan klasifikasi, dengan adanya perhitungan nilai validitas pada data latih dan perhitungan *Weight Voting*. Dengan adanya dua proses baru diharapkan dapat menyempurnakan kekurangan pada algoritma K-NN. Algoritma MKNN masih lebih unggul dibandingkan dengan algoritma KNN, banyaknya data latih mempengaruhi hasil kinerja dari algoritma tersebut, apabila data bertambah banyak, maka hasilnya akan menjadi lebih baik [3].

1.2 Perumusan Masalah

1.2.1 Permasalahan

Hasil diagnosa tidak sesuai dengan data penyakit *liver* yang telah diteliti sehingga timbul kesalahan dalam memprediksi penyakit tersebut. Untuk menentukan penyakit *liver* membutuhkan alat bantu *machine learning* dalam mengklasifikasi penyakit pasien.

1.2.2 Metode Usulan

Penelitian ini mengusulkan metode klasifikasi *Modified K-Nearest Neighbor* dan seleksi fitur *F-Score* untuk identifikasi penyakit *liver*.

1.2.3 Pertanyaan Penelitian

Berapa *performance measure* (akurasi, *precision*, *recall*) yang diperoleh setelah menggunakan metode *Modified K-Nearest Neighbor* dengan seleksi fitur *F-Score* maupun tanpa menggunakan seleksi fitur *F-Score*?

1.3 Tujuan dan Manfaat

1.3.1 Tujuan Penelitian

Mengetahui *performance measure* yang dihasilkan dari penerapan metode *Modified K-Nearest Neighbor* dan seleksi fitur *F-Score*

1.3.2 Manfaat Penelitian

Penelitian ini dapat mengimplementasikan IT pada kasus-kasus nyata dengan memberikan solusi yang lebih efektif dan efisien serta akurat, khususnya klasifikasi penyakit *liver*.

1.4 Batasan-batasan

Batasan masalah pada penelitian ini adalah sebagai berikut:

1. Data yang akan digunakan pada penelitian ini adalah data *Indian Liver Patient Records* dari situs *Kaggle Repository*.

(<https://www.kaggle.com/uciml/indian-liver-patient-records>)

2. Banyak fitur yang akan digunakan setelah di seleksi fitur *f-score* adalah 10
3. Jumlah data terdiri dari 832 data dibagi 80% data *training* yaitu 665 data dan 20% data *testing* yaitu 167 data.
4. Metode evaluasi *K-Fold Cross Validation* menggunakan 5 *kfold*

1.5 Sistematika Proposal

Sistematika penulisan Proposal ini terdiri dari beberapa bab, yaitu:

BAB I PENDAHULUAN

Pada bab ini membahas tentang latar belakang, rumusan masalah, tujuan, manfaat, batasan masalah, serta sistematika penulisan laporan.

BAB II KAJIAN PUSTAKA

Pada bab ini membahas tentang landasan teori dalam penyusunan penelitian ini. Landasan teori tersebut diantaranya adalah penelitian-penelitian terkait prediksi penyakit *liver*, normalisasi *Min-Max*, *Modified K-Nearest Neighbor*, *F-Score*.

BAB III METODE USULAN

Pada bab ini membahas mengenai metode yang diusulkan (algoritma), arsitektur (*Input*, *Proses*, *Output*), data penelitian, tahapan penelitian serta rencana pengujian.

BAB IV HASIL DAN PEMBAHASAN

Pada bab ini membahas tentang hasil implementasi sistem dan hasil yang telah diperoleh dari proses uji coba klasifikasi penyakit *liver* berdasarkan dengan skenario pengujian.

BAB V PENUTUP

Pada bab ini membahas tentang kesimpulan dan saran berdasarkan hasil uji coba yang telah dilakukan sebagai pengembangan penelitian terkait untuk berikutnya.

DAFTAR PUSTAKA

Berisi tentang daftar jurnal-jurnal atau referensi yang dijadikan acuan dalam penelitian ini.