

BAB I PENDAHULUAN

1.1 Latar Belakang

Pertumbuhan teknologi informasi, koordinasi *framework* serta pertumbuhan pemrograman, sistem sudah membentuk masa *framework* komputer yg imajinatif. Spesialis teknologi informasi sudah ditawarkan sedikit tantangan oleh *framework* ini. Permasalahan *framework* tersebut merupakan *framework* layanan kesehatan. Baru-baru ini, terdapat kepedulian yang lebih besar untuk menggunakan informasi mining dalam *framework* perawatan kesehatan. Dengan demikian, tujuan dari upaya ini yaitu untuk mengenali aspek pemakaian informasi pelayanan kesehatan untuk menunjang masyarakat melalui metode pembelajaran mesin serta prosedur data mining. Diagnosa sering kali dibuat dengan berfokus pada naluri serta pengalaman dokter, bukan bersumber pada data yang dicakup dalam *database*. Penerapan ini memunculkan bias yang tidak diinginkan, biaya pengobatan yang tidak perlu, dan kualitas layanan yang diberikan kepada pasien tidak maksimal [1].

Data mining merupakan teknik untuk mendapatkan pola serta pengetahuan yang bermakna dari beberapa informasi yang besar di dalam *database* ataupun informasi yang ditaruh di repositori data yang berbeda. *Data mining* sangat berarti dalam banyak bidang riset untuk mendapatkan data tersembunyi dari kumpulan informasi besar yang membantu pemangku kepentingan buat menguasai serta mengambil informasi mereka dalam waktu pendek. Pada *data mining* juga digunakan untuk mengklasifikasikan, memprediksi, serta mengelompokkan informasi buat membuat pengambilan keputusan yang benar ataupun akurat di banyak organisasi. Di pusat medis (rumah sakit, ataupun pusat klinis yang lain) teknik *data mining* membantu merawat pasien untuk mengenali apakah pasien tersebut mempunyai penyakit ataupun tidak, serta untuk diagnosa dini pada pasien yang diambil dalam waktu singkat[2].

Dalam *data mining* terdapat metode *ensemble learning* dimana metode ini merupakan kombinasi dari beberapa model menjadi satu yang memberikan kinerja lebih baik daripada satu model[3]. *Logistic Regression Ensemble (LORENS)* cukup banyak digunakan untuk klasifikasi dimana metode ini dikembangkan untuk mengatasi masalah dimensi data yang tinggi (jumlah variabel melebihi ukuran sampel) dan proporsi kategori di variabel respon tidak seimbang, meskipun

LORENS dikembangkan untuk data berdimensi tinggi namun *LORENS* masih dapat diterapkan untuk variabel terbatas dengan ukuran sampel yang besar[4][5]. Pada penelitian sebelumnya yang berjudul “*Logistic Regression Ensemble for Predicting Customer Defection with Very Large Sample Size*” [4], dalam penelitian ini menggunakan metode *Logistic Regression Ensemble (LORENS)* untuk memprediksi pembelotan pelanggan. Hasil dari penelitian ini menunjukkan bahwa prediksi menggunakan metode *LORENS* memiliki akurasi antara 66% hingga 77%. Kemudian pada penelitian berjudul “*Logistic Regression Ensemble to Classify Alzheimer Gene Expression*” [5], Penelitian ini menerapkan *Logistic Regression Ensemble (LORENS)* untuk mengklasifikasikan gen Alzheimer. Dari pengujian yang telah dilakukan mendapatkan hasil akurasi terbaik sebesar 75,28%.

Oleh karena itu, penulis tertarik untuk mengangkat topik yang berjudul “Klasifikasi penyakit jantung menggunakan metode *Logistic Regression Ensemble (LORENS)*” diharapkan mampu untuk menyelesaikan permasalahan klasifikasi penyakit jantung.

1.2 Perumusan Masalah

1.2.1 Permasalahan

Permasalahan yang terjadi adalah diagnosa sering kali dibuat dengan berfokus pada naluri dan pengalaman dokter, bukan bersumber pada data yang dicakup dalam *database*. Hal ini menimbulkan dampak yang tidak diinginkan seperti kesalahan dalam mendiagnosis penyakit dan biaya pengobatan yg tidak perlu pada pasien. Jika ini terjadi maka pasien akan mengalami keterlambatan dalam menangani penyakitnya dan akan berakibat fatal. Oleh karena itu perlu adanya identifikasi gejala penyakit jantung secara dini dengan melakukan klasifikasi penyakit jantung menggunakan metode *Logistic Regression Ensemble (LORENS)*.

Metode ini dipilih karena penulis ingin mengetahui apakah metode *LORENS* mampu diterapkan pada *dataset* penyakit jantung dimana *dataset* tersebut mempunyai dimensi data rendah (jumlah variabel terbatas dengan ukuran sampel yang besar), sedangkan metode *LORENS* dikembangkan untuk mengatasi masalah dimensi data yang tinggi (jumlah variabel melebihi ukuran sampel) dan proporsi kategori di variabel respon tidak seimbang, meskipun *LORENS* dikembangkan untuk data berdimensi tinggi namun *LORENS* masih dapat diterapkan untuk variabel terbatas dengan ukuran sampel yang besar.

1.2.2 Metode Usulan

Penelitian ini mengusulkan metode *Logistic Regression Ensemble (LORENS)* untuk klasifikasi penyakit jantung.

1.2.3 Pertanyaan Penelitian

1. Bagaimana cara melakukan klasifikasi penyakit jantung menggunakan metode metode *Logistic Regression Ensemble (LORENS)*?
2. Berapa akurasi dari metode *Logistic Regression Ensemble (LORENS)* untuk klasifikasi penyakit jantung?

1.3 Tujuan dan Manfaat

1.3.1 Tujuan Penelitian

1. Melakukan klasifikasi penyakit jantung menggunakan metode *Logistic Regression Ensemble (LORENS)*.
2. Mengetahui akurasi dari metode *Logistic Regression Ensemble (LORENS)* untuk klasifikasi penyakit jantung.

1.3.2 Manfaat Penelitian

Manfaat yang diperoleh dari penelitian ini yaitu membantu para ahli dalam mendiagnosa penyakit jantung dan mengurangi resiko kesalahan para ahli dalam melakukan diagnosa terhadap pasiennya, serta proses diagnosa secara dini pada pasien dapat diambil dalam waktu singkat. Sehingga tidak memunculkan bias yang tidak diinginkan, biaya pengobatan yang tidak perlu, dan kualitas layanan yang diberikan kepada pasien akan lebih maksimal.

1.4 Batasan-batasan

Batasan masalah pada penelitian ini adalah sebagai berikut :

1. Data yang akan digunakan pada penelitian ini adalah data *Heart Disease* dari situs Kaggle Repository
(<https://www.kaggle.com/johnsmith88/heart-disease-dataset>)
2. Data yang di gunakan tahun 2019 dimana data memiliki dua kelas yaitu mempunyai penyakit jantung (1) dan tidak mempunyai penyakit jantung (0) dengan jumlah sampel sebanyak 1025 serta memiliki 13 fitur. Kemudian fitur yg berjenis kategorikal akan ditransformasi ke numerik dengan metode *One Hot Encoding* sehingga fitur yang akan digunakan menjadi 26 fitur.
3. Menggunakan metode *Logistic Regression Ensemble (LORENS)*.

1.5 Sistematika Penulisan

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 dan penelitian sebelumnya.

BAB III METODE USULAN

Pada bab ini membahas tentang 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 didapatkan dari proses uji coba klasifikasi penyakit jantung 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