

BAB I

PENDAHULUAN

1.1 Latar Belakang Masalah

Stroke merupakan masalah kesehatan yang utama bagi masyarakat modern saat ini. Pada dewasa ini, stroke semakin menjadi masalah serius yang dihadapi hampir diseluruh dunia. Hal tersebut dikarenakan serangan stroke yang mendadak dapat mengakibatkan kematian, kecacatan fisik dan mental baik pada usia produktif maupun usia lanjut [1].

[2], kematian akibat stroke sebesar 51% di seluruh dunia disebabkan oleh tekanan darah tinggi. Selain itu, diperkirakan sebesar 16% kematian stroke disebabkan tingginya kadar glukosa darah dalam tubuh. Tingginya kadar gula darah dalam tubuh secara patologis berperan dalam peningkatan konsentrasi glikoprotein, yang merupakan pencetus beberapa penyakit vaskuler. Kadar glukosa darah yang tinggi pada saat stroke akan memperbesar kemungkinan meluasnya area infark karena terbentuknya asam laktat akibat metabolisme glukosa secara anaerobik yang merusak jaringan otak

Berdasarkan hasil Riskesdas tahun 2013, prevalensi penyakit stroke di Indonesia meningkat seiring bertambahnya umur. Kasus stroke tertinggi yang terdiagnosis tenaga kesehatan adalah usia 75 tahun keatas (43,1%) dan terendah pada kelompok usia 15-24 tahun yaitu sebesar 0,2%. Prevalensi stroke berdasarkan

jenis kelamin lebih banyak laki-laki (7,1%) 2 dibandingkan dengan perempuan (6,8%). Berdasarkan tempat tinggal, prevalensi stroke di perkotaan lebih tinggi (8,2%) dibandingkan dengan daerah pedesaan (5,7%). Berdasarkan data 10 besar penyakit terbanyak di Indonesia tahun 2013, prevalensi kasus stroke di Indonesia berdasarkan diagnosis tenaga kesehatan sebesar 7,0 per mill dan 12,1 per mill untuk yang terdiagnosis memiliki gejala stroke. Prevalensi kasus stroke tertinggi terdapat di Provinsi Sulawesi Utara (10,8%) dan terendah di Provinsi Papua (2,3%), sedangkan Provinsi Jawa Tengah sebesar 7,7%. Prevalensi stroke antara laki-laki dengan perempuan hampir sama [3].

Dapat dilihat dari data diatas untuk mengelolan data yang besar di Indonesia ataupun dunia diperlukan cara pengolahan atau metode tertentu sehingg data tersebut dapat disajikan dan dilihat secara umum. Pengolahan data tersebut biasanya menggunakan data mining. Data mining sendiri merupakan proses penggalian informasi dan pola yang bermanfaat dari data yang sangat besar. Data mining mencakup pengumpulan data, ekstraksi data, analisis data, dan statistik data. Data mining juga dikenal sebagai *knowledge discovery*, *knowledge extraction*, *data/pattern analysis*, information harvesting, dan lain-lain [4].

Untuk bisa mendapatkan informasi dari data-data yang ada perlu dilakukan proses data mining seperti klasifikasi. Klasifikasi merupakan sebuah proses untuk menentukan model yang menjelaskan atau membedakan konsep atau kelas data, dengan tujuan untuk dapat memperkirakan kelas dari suatu objek yang kelasnya tidak diketahui, di dalam klasifikasi juga diberikan sejumlah record yang

dinamakan training set, yang terdiri dari beberapa atribut, atribut dapat berupa kontinyu ataupun kategoris, salah satu atribut menunjukkan kelas untuk record [5].

Algoritma C4.5 adalah salah satu metode klasifikasi dari data mining yang digunakan untuk mengkonstruksikan pohon keputusan (*decision tree*). Menurut [6], pohon keputusan atau *decision tree* adalah pohon yang digunakan sebagai prosedur penalaran untuk mendapatkan jawaban dari masalah yang dimasukkan. Menurut [7], Algoritma C4.5 merupakan pengembangan dari ID3 yang mampu mengatasi nilai yang hilang (*missing value*), mengatasi data bertipe kontinyu, dan melakukan pemangkasan pohon (*pruning tree*). Selain itu, dengan menggunakan Algoritma C4.5 dapat diketahui pulanilai akurasi klasifikasinya.

Pada permasalahan diatas mengenai penyakit stroke maka untuk dapat mengatasi masalah tersebut banyak dilakukan penelitian dalam bidang ilmu komputer diantaranya adalah Klasifikasi Penderita Penyakit Stroke Menggunakan Algoritma *Decision Tree* C4.5 untuk mengklasifikasikan faktor paling penting untuk penyakit ini. Pengujian menghasilkan akurasi yang cukup tinggi yaitu sebesar 92,02 %. Merujuk hasil akurasi dari penelitian tersebut dapat dilihat bahwa *Decision Tree* menghasilkan akurasi yang tinggi namun hasil akurasi tersebut masih dapat ditingkatkan lagi dengan melakukan penelitian lanjutan untuk menghasilkan akurasi lebih tinggi dengan menambahkan *optimization feature weighting* (PSO) . dan menggunakan dataset *kaggle*. Berdasarkan hasil penelitian yang dilakukan peneliti sebelumnya maka masih perlu dikembangkan lagi agar klasifikasi prediksi penyakit stroke mendapatkan tingkat akurasi yang lebih tinggi .

1.2 Identifikasi Masalah

Pada penelitian ini diperlukan adanya indentifikasi masalah, dimana masalah yang terdapat pada penelitian sebelumnya yaitu akurasi yang dihasilkan belum cukup tinggi serta penambahasn seleksi fitur kurang optimal.

1.3 Rumusan Masalah

Rumusan masalah dalam penelitian ini adalah bagaimana cara menghasilkan model algoritma yang memiliki akurasi yang sangat tinggi menggunakan *Particle Swam Optimization* (PSO) dan algoritma *Decision Tree* C4.5 pada data prediksi stroke yang di dapat dari situs kaggle.com.

1.4 Batasan Masalah

Batasan masalah pada penelitian ini yaitu membatasi ruang lingkup penelitian hanya pada peningkatan nilai akurasi pada prediksi stroke menggunakan *Particle Swam Optimization* (PSO) dan algoritma *Decision Tree* C4.5. Dataset yang digunakan pada penelitian ini berasal dari situs Kaggle dataset yaitu data prediksi stroke dengan atribut berjumlah 12 (<https://www.kaggle.com/fedesoriano/stroke-prediction-dataset>), serta tools yang digunakan dalam penelitian ini adalah Rapid Miner 9.

1.5 Tujuan Penelitian

Berdasarkan permasalahan yang ada tujuan dari penelitian ini adalah untuk menghasilkan nilai akurasi yang tinggi menggunakan metode seleksi fitur *Particle Swam Optimization* (PSO) dan algoritma *Decision Tree C4.5*.

1.6 Manfaat Penelitian

Manfaat yang dapat diperoleh dari penelitian ini adalah mempermudah dalam melakukan klasifikasi penyakit stroke agar mendapatkan nilai akurasi yang tinggi dengan menggunakan data mining dan menguji beberapa variabel yang diperoleh dari dataset publik sehingga dapat dianalisis tingkat akurasinya dengan menggunakan aplikasi rapidminer, serta dapat memberikan masukan dalam ilmu pengetahuan untuk penelitian selanjutnya dan referensi ilmiah dalam penelitian penerapan data mining.

1.7 Sistematika Penulisan

Proposal ini terdiri dari tiga bagian yang secara garis besar sistematika penulisannya adalah sebagai berikut.

- a. Bab I Pendahuluan, berisi tentang latar belakang masalah, identifikasi masalah, rumusan masalah, identifikasi masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan.

- b. Bab II Landasan Teori, berisi penelitian terkait dan landasan teori yang menjadi acuan dalam pelaksanaan penelitian ini.
- c. Bab III Metode Penelitian, berisi metode atau kerangka kerja yang digunakan dalam penelitian ini. Selain itu, pada bagian ini juga disajikan rencana penelitian berikut tahapan dan waktu pelaksanaannya.
- d. Bab IV Hasil Dan Pembahasan, berisi mengenai hasil, implementasi dan pembahasan penelitian. Hasil dari implementasi ini berupa gambar alat/program dan aplikasinya. Untuk penelitian lapangan hasil dapat berupa data (kualitatif maupun kuantitatif). Analisis dan pembahasan berupa hasil pengolahan data.
- e. Bab V Kesimpulan dan Saran, berisi simpulan dan saran dari isi penelitian yang sudah di buat.