

BAB I

PENDAHULUAN

1.1. Latar Belakang

Seiring berjalannya waktu informasi yang beredar melalui komputer semakin banyak. Dampak yang disebabkan oleh arus informasi yang cepat ini adalah semakin banyaknya data-data yang tersimpan di dalam jaringan. Data yang dihasilkan oleh teknologi ini tidak hanya berguna di satu bidang saja, tetapi hampir di semua bidang kehidupan.

Stroke adalah penyakit yang menyerang otak berupa serangan fungsi syaraf lokal dan/atau global, munculnya secara tiba-tiba, progresif, dan cepat. Gangguan fungsi syaraf pada stroke disebabkan oleh gangguan peredaran darah otak non traumatik. Gangguan syaraf tersebut menghasilkan gejala antara lain: lumpuh pada wajah atau sebagian anggota badan, bicara tidak lancar, bicara tidak jelas (pelo), mungkin perubahan kesadaran, gangguan penglihatan, dan lain-lain. Diartikan sebagai stroke jika pernah didiagnosa menderita penyakit stroke oleh tenaga kesehatan (dokter/perawat/bidan) atau belum pernah didiagnosis menderita penyakit stroke oleh nakes tetapi pernah mengalami secara mendadak gejala kelumpuhan pada satu bagian atau disertai kesemutan atau baal satu sisi tubuh atau mulut menjadi miring tanpa kelumpuhan otot mata atau sulit bicara/komunikasi dan atau tidak mengerti pembicaraan. Jumlah penderita penyakit stroke di Indonesia tahun 2018 berdasarkan diagnosis tenaga kesehatan Proporsi stroke di Indonesia pada penduduk berusia di atas 15 tahun berdasarkan diagnosis dokter yaitu 10,9‰ dan diprediksi ada 2.120.362 jiwa mengalami stroke. Provinsi dengan proporsi paling tinggi di Indonesia adalah (14,7‰) Kalimantan Utara, (14,6‰) DI Yogyakarta, dan (14,2‰) Sulawesi Utara. Sedangkan provinsi Maluku Utara dan Papua memiliki jumlah penderita paling sedikit yaitu sebanyak 4,6% dan 4,1%. [1]

Maka itu, dilakukan penelitian untuk menganalisa data terkait dengan penyebab stroke. Adapun atribut yang terlibat dalam penyebab terjadinya stroke yakni, usia, jenis kelamin, kadar glukosa, riwayat penyakit jantung, hipertensi, tipe pekerjaan, tipe tempat tinggal, status merokok, *index* masa tubuh dan status

pernikahan. Diperlukan suatu algoritma tertentu untuk mengklasifikasikan semua atribut tersebut. *C4.5* merupakan Algoritma yang paling banyak digunakan, dalam kasus ini akurasi dari algoritma *C4.5* sebesar 99.07%. Selanjutnya *Algoritma C4.5* dioptimasi dengan menggunakan *Particle Swarm Optimization* sehingga memperoleh akurasi sebesar 99.28% dan Algoritma *C4.5* juga dioptimasi dengan menggunakan *Genetic Algorithm* sehingga memperoleh akurasi sebesar 99.38%. [2]

Selanjutnya Optimasi *Particle Swarm Optimization (PSO)* dapat meningkatkan kinerja akurasi *algoritma C4.5* dan *Naive Bayes*. Pada akurasi *algoritma C4.5* mengalami peningkatan dari 90,19% menjadi 94,29%, akurasi *Naive Bayes* dengan nilai 97,65% menjadi 97,96%. Hasil kinerja terbaik yang diuji menggunakan *T-Test* pada algoritma *C4.5* dan *Naive Bayes* berbasis *Particle Swarm Optimization (PSO)* dapat dihasilkan bahwa algoritma *Naive Bayes (PSO)* memiliki nilai tertinggi sebesar 0,980 dilanjutkan algoritma *C4.5 (PSO)* sebesar 0,943. Dengan demikian algoritma *Naive Bayes Particle Swarm Optimization (PSO)* dapat memberikan solusi terbaik terhadap akurasi pendeteksian penyakit kanker payudara. [3]

Penelitian metode *Naive bayes* dan *Support Vector Machine (SVM)* terhadap keberhasilan pengobatan imunoterapi penyakit kutil menggunakan pemrograman *R* pada *R-Studio*, maka diambil kesimpulan hasil pengujian menggunakan metode itu dapat memprediksi class yang tidak tepat dengan tingkat akurasi 0.8. Hasil tersebut mengindikasikan metode *Naive bayes* merupakan metode yang lebih unggul dibandingkan dengan metode *Support Vector Machine (SVM)*. [4]

Dari metode klasifikasi data mining dengan *algoritma C4.5* dan pengaplikasian pohon keputusan yang membentuk aturan tersebut terdapat akurasi pada data training yang berjumlah 130 dari 156 data pasien sebesar 82,31% sedangkan akurasi pada data testing yang berjumlah 26 dari 156 data pasien sebesar 76,92%. Perhitungan keduanya menggunakan *confusion matrix*. [5]

Berdasarkan penelitian dan pengujian Analisis Algoritma *Support Vector Machine* (SVM) Dalam Klasifikasi Penyakit Stroke yang telah penulis lakukan dapat disimpulkan bahwa, dalam penelitian menggunakan dataset dari *Kaggle* yang diterbitkan pada tahun 2021, menggunakan *Machine Learning* menggunakan metode *Support Vector Machine* untuk mengklasifikasi data. Penulis melakukan pengujian data dengan dua cara, yaitu menggunakan teknik *Unbalanced Data* untuk melakukan penelitian dengan data yang tidak sebanding, dan cara klasifikasi pada umumnya dengan data sebanding (*Balanced*). Untuk metode *Support Vector Machine* penulis menggunakan dua jenis kernel yaitu kernel *Linear* dan *Polynomial*, untuk data *unbalanced* pada kernel linear mendapatkan akurasi terbesar 76% dan untuk *polynomial* mendapatkan hasil akurasi 80%. Untuk data yang *balanced* dengan kernel linear penulis mendapatkan hasil akurasi tertinggi 77%, sedangkan pada kernel *polynomial* penulis mendapatkan hasil tertinggi 76%. [6]

Perawatan yang paling umum digunakan untuk mengobati penderita kutil adalah *cryotherapy* dan *imunoterapi*. peneliti menggabungkan dua dataset untuk membuat metode peramalan. Kami menggabungkan metode algoritma *C4.5* dengan *Random Forest Feature Weighting* (*C4.5+RFFW*) digunakan untuk memilih fitur terkait untuk meningkatkan akurasi. Hasil percobaan menunjukkan bahwa metode yang diusulkan dapat meningkatkan kinerja, akurasi dan informasi masing-masing adalah 87,22% dan 71,24%. Hasil ini akan memungkinkan dokter untuk dengan mudah membuat keputusan perawatan untuk pasien mereka dengan metode prediksi tunggal dan prediksi kinerja yang lebih baik. [7]

Berdasarkan beberapa penelitian terdahulu, Pada penelitian ini peneliti akan melakukan penelitian untuk menganalisis dan membandingkan *Decision Tree C4.5* dan *Support Vector Machine (SVM)* Berbasis PSO untuk melihat perbandingan tingkat akurasi, presisi, dalam klasifikasi penyakit stroke. Selain itu, hasil pengujian juga harus dianalisis untuk melihat seberapa efektif kedua algoritma tersebut.

1.2. Rumusan Masalah

Rumusan masalah pada penelitian ini adalah menentukan Manakah yang merupakan metode klasifikasi terbaik dari dua metode yaitu *Decision Tree C4.5* dan *Support Vector Machine (SVM)* dengan menggunakan seleksi fitur *Particel Sward Option (PSO)* pada deteksi penyakit Stroke?

1.3. Batasan Masalah

Adapun batasan masalah pada penelitian ini:

1. Penelitian ini menggunakan metode *Algortima C4.5* dan *Support Vector Machine (SVM)* berbasis *PSO*
2. Data uji yang digunakan adalah Data Set Stroke dari *Kaggle.com*
3. Penenlitian ini hanya membandingkan mana hasil yang lebih baik antar kedua metode yang berbasis *PSO*

1.4. Tujuan Penelitian

Tujuan yang ingin dicapai pada penelitian ini adalah:

1. Membandingkan antar dua metode mana yang memiliki hasil klasifikasi yang lebih akurat untuk digunakan dalam deteksi Stroke.
2. Menentukan hasil klasifikasi dengan metode *Decision Tree C4.5* dan *Support Vector Machine (SVM)* menggunakan *dataset* Stroke.

1.5. Manfaat Penelitian

Manfaat yang ingin dicapai dalam penelitian ini adalah:

1. Diharapkan dapat memberikan kontribusi khususnya dalam bidang Kesehatan yaitu deteksi penyakit stroke dengan membandingkan dua metode terkait.
2. Dapat memberikan data bantu sebelum dilakukan pemeriksaan lebih lanjut terkait deteksi stroke.
3. Dapat menjadi penelitian terbaru selanjutnya yang dapat dikembangkan.

1.6. Sistematika Penulisan

Untuk mempermudah penyusunan laporan penelitian ini, maka sistematika penulisan terbagi atas lima bab sebagai berikut.

Bab I Pendahuluan adalah membahas mengenai latar belakang, masalah penelitian yang terdiri dari identifikasi masalah, pembatasan masalah dan perumusan masalah, tujuan dan manfaat penelitian, serta sistematika penulisan.

Bab II Tinjauan Pustaka pada bagian ini berisi penelitian yang pernah dilakukan terkait topik penelitian; dan teori-teori pendukung yang digunakan yang berhubungan erat dengan pokok-pokok landasan berpikir, yang relevan dengan topik penelitian.

Bab III Metode Penelitian pada bagian ini berisi metode penelitian menggunakan Metode *Algoritma C4.5* dan *Support Vector Machine (SVM)* serta tahapan yang dilakukan dalam penelitian antara lain: langkah- Langkah penelitian dan pengumpulan data.

Bab IV Hasil dan Pembahasan pada bagian ini ditampilkan tahapan pemrosesan dan penjabaran hasil penelitian.

Bab V Kesimpulan bagian ini menjabarkan kesimpulan dan saran menurut penelitian yang sudah dilakukan.