

BAB I

PENDAHULUAN

1.1 LATAR BELAKANG

Saat ini, kemajuan teknologi telah berkembang pesat dan mengubah cara pengolahan informasi dari manual menjadi terkomputerisasi. Penerapan teknologi informasi ini sudah meluas ke berbagai bidang, terutama dalam sektor kesehatan [1]. Di rumah sakit, misalnya, sistem informasi terkomputerisasi memudahkan penyimpanan dan pengolahan data administrasi, mulai dari pendaftaran pasien hingga penjadwalan tindakan medis. Selain itu, teknologi ini juga mendukung penelitian kedokteran, membantu proses diagnosis penyakit, menentukan obat yang paling tepat, bahkan menganalisis area dalam organ tubuh yang sulit diakses secara konvensional[2]. Tumor otak adalah pertumbuhan jaringan akibat sel abnormal di dalam atau sekitar otak. Penyebabnya belum pasti, namun diduga dipengaruhi faktor keturunan dan paparan radiasi kimia. Tumor ini juga bisa timbul dari *metastasis* sel kanker organ lain[3]. Jika sel-sel tersebut tetap terbatas di area asal tanpa menyusup ke jaringan lain, disebut tumor jinak (*benigna*). Sebaliknya, jika terus berkembang, menyerang jaringan sekitar, dan berpotensi menyebar, kondisi ini disebut tumor ganas (*maligna*).[4]. Di Indonesia setiap tahun terdiagnosis sekitar 300 kasus tumor otak, termasuk anak-anak. Penanganannya membutuhkan data lokasi spasial tumor untuk merencanakan operasi dan memilih metode pengobatan efektif. Selain itu, tingkat keganasan tumor penting untuk menyusun rencana pengobatan komprehensif dan

memprediksi perkembangan penyakit.[5].

Dalam mengatasi masalah tersebut, peneliti mengusulkan klasifikasi menggunakan *Machine Learning* (ML), cabang dari kecerdasan buatan (*artificial intelligence*) yang memungkinkan mesin belajar tanpa instruksi eksplisit. ML memanfaatkan algoritma berbasis data untuk meniru cara manusia belajar. Akurasi model sangat bergantung pada algoritma yang digunakan serta pola dari data historis. Oleh karena itu, dibutuhkan data yang cukup dan berkualitas tinggi agar model prediktif yang dihasilkan efektif[6]. Metode ini dapat diterapkan menggunakan berbagai algoritma, seperti *Support Vector Machine* (SVM), *Random Forest*, *K-Nearest Neighbor* (KNN), *Naïve Bayes*, dan algoritma lainnya[7]. *Support Vector Machine* (SVM) adalah metode pembelajaran mesin yang menggunakan fungsi linier dalam ruang fitur berdimensi tinggi untuk memisahkan kelas dengan margin terbesar, berdasarkan teori optimasi dan statistik. Sementara itu, *K-Nearest Neighbor* (KNN) adalah metode klasifikasi yang menentukan kelas suatu data berdasarkan mayoritas kategori dari k tetangga terdekat, dengan mengukur jarak dari data pelatihan[8].

Adapun beberapa penelitian yang telah dilakukan dalam upaya mendeteksi berbagai penyakit : Seperti penelitian yang dilakukan Simeon dan Ghinaa [9] bertujuan mengembangkan alat diagnostik tumor otak berbasis citra MRI dengan ekstraksi ciri bentuk menggunakan *Discrete Cosine Transform* (DCT), dengan menguji beberapa model *machine learning* yaitu *Naive Bayes*, *Decision Tree*, *K-Nearest Neighbor*, dan *Support Vector Machine*. Hasil akhir yang didapatkan bahwa algoritma *Support Vector Machine* menghasilkan akurasi sebesar 93%.

Kemudian penelitian yang dilakukan oleh Monikka dkk [10] mengembangkan alat diagnostik tumor otak berbasis citra MRI, dengan mengimplementasikan model *machine learning* yaitu *Linear Regression*, *K-Nearest Neighbor*, *Decision Tree*, *Random Forest*, *Naïve Bayes* dan *Support Vector Machine*. Hasil akhir yang didapatkan bahwa algoritma *K-Nearest Neighbor* menghasilkan akurasi sebesar 81%. Selanjutnya penelitian yang dilakukan Dewi Nasien dkk [11] mengimplementasikan perbandingan *machine learning* pada penyakit diabetes, berikut algoritma yang digunakan yaitu *K-Nearest Neighbor*, *Naive Bayes*, dan *Logistic Regression*. Hasil akhir yang didapat mengatakan bahwa algoritma *K-Nearest Neighbor* memiliki akurasi sebesar 71%. Kemudian penelitian yang dilakukan oleh Yufis Azhar dkk [12] bertujuan untuk memanfaatkan teknik data mining untuk memprediksi pasien yang terkena penyakit stroke, dimana mereka membandingkan algoritma *Decision Tree C4.5*, *Logistic Regression*, *Random Forest*, *Naïve Bayes*, *Support Vector Machine*, dan *K-Nearest Neighbours*. Hasil dari penelitian itu didapatkan bahwa algoritma *Support Vector Machine* memiliki akurasi sebesar 76,52%.

Berdasarkan hasil kinerja algoritma pada penelitian sebelumnya yang membandingkan beberapa algoritma klasifikasi dalam mendeteksi penyakit tumor otak, penelitian ini juga akan melakukan hal serupa, yaitu membandingkan kinerja algoritma klasifikasi *Support Vector Machine* dan *K-Nearest Neighbor*[13]. Pemilihan algoritma *Support Vector Machine* dan *K-Nearest Neighbor* dalam penelitian ini didasarkan pada hasil penelitian sebelumnya yang menunjukkan keduanya memiliki performa baik dalam klasifikasi, terutama untuk tumor otak.

Selain itu, kedua algoritma tersebut menggunakan pendekatan yang berbeda. SVM memaksimalkan margin pemisah, sedangkan KNN mengklasifikasikan berdasarkan kedekatan data. Dengan membandingkan kedua metode ini, diharapkan dapat diperoleh algoritma yang paling efektif untuk membantu diagnosa tumor otak. Selanjutnya, performa kedua algoritma tersebut akan dievaluasi dengan menganalisis nilai *confusion matrix*, yang mencakup perhitungan akurasi, *recall*, *precision*, dan *f1-score*.

Implementasi kedua algoritma dilakukan menggunakan bahasa *Python* dengan bantuan *library Scikit-Learn*. *Scikit-Learn* mempermudah penerapan algoritma *machine learning* melalui API yang efisien, dan banyak digunakan di industri maupun akademik. *Library* ini dibangun di atas *NumPy*, sehingga perhitungan menjadi lebih cepat dan optimal[14]. Berdasarkan latar belakang tersebut, peneliti tertarik untuk melakukan penelitian yang berjudul **“PERBANDINGAN PERFORMA ALGORITMA SUPPORT VECTOR MACHINE DAN K-NEAREST NEIGHBOR DALAM KLASIFIKASI TUMOR OTAK.”**

1.2 RUMUSAN MASALAH

Berdasarkan latar belakang yang telah diuraikan, maka rumusan masalah dalam penelitian ini adalah:

- 1) Bagaimana kinerja metode *Support Vector Machine* (SVM) dalam mendeteksi penyakit tumor otak?
- 2) Bagaimana kinerja metode *K-Nearest Neighbor* (KNN) dalam mendeteksi penyakit tumor otak?

- 3) Bagaimana kinerja kedua metode tersebut dalam hal *accuracy*, *recall*, *precision*, dan *F1-score* dalam mendeteksi penyakit tumor otak?

1.3 BATASAN MASALAH

Agar penelitian lebih terfokus dan terarah, maka batasan-batasan dalam penelitian ini adalah:

- 1) Dataset yang digunakan dalam penelitian ini adalah *Brain Tumor Dataset* yang diperoleh dari situs *Kaggle*
- 2) Penilaian performa algoritma difokuskan pada empat metrik evaluasi utama, yaitu *accuracy*, *recall*, *precision*, dan *F1-score*.
- 3) Evaluasi performa menggunakan *confusion matrix*.
- 4) Penelitian hanya dilakukan pada data tabular, bukan pada citra medis atau data visual lainnya.
- 5) Alat analisis yang digunakan meliputi *Python*, *Scikit-learn*, *Pandas*, *NumPy*.

1.4 TUJUAN PENELITIAN

Penelitian ini memiliki tujuan sebagai berikut:

- 1) Membandingkan algoritma *Support Vector Machine* (SVM) dan *K-Nearest Neighbor* (KNN) dalam mendeteksi penyakit tumor otak.
- 2) Mengevaluasi kinerja kedua algoritma berdasarkan metrik *accuracy*, *recall*, *precision*, dan *F1-score*.
- 3) Mengetahui hasil performa terbaik diantara kedua algoritma *Support Vector Machine* (SVM) dan *K-Nearest Neighbor* (KNN).

1.5 MANFAAT PENELITIAN

Manfaat yang diharapkan dari penelitian ini adalah:

- 1) Memberikan informasi dan wawasan kepada peneliti lain mengenai kinerja dan perbandingan algoritma SVM dan KNN dalam klasifikasi penyakit, khususnya tumor otak.
- 2) Menjadi referensi bagi pengembangan sistem pendukung keputusan berbasis *machine learning* dalam bidang kesehatan, terutama dalam deteksi dan klasifikasi tumor otak.
- 3) Memberikan kontribusi dalam pemanfaatan teknologi *machine learning* untuk membantu proses diagnosis awal terhadap tumor otak secara lebih cepat dan efisien.

1.6 SISTEMATIKA PENULISAN

Sistematika penulisan memberikan gambaran umum mengenai isi yang akan dibahas oleh penulis dalam setiap bab laporan tugas akhir, yang terdiri dari lima bab. Struktur penulisan disusun sebagai berikut:

BAB I : PENDAHULUAN

Bab ini membahas latar belakang penelitian, rumusan masalah, batasan masalah, tujuan dan manfaat penelitian, serta sistematika penulisan yang digunakan dalam laporan tugas akhir.

BAB II : LANDASAN TEORI

Bab ini berisi konsep-konsep teori yang relevan sebagai dasar pemahaman dan kerangka kerja yang digunakan untuk mendukung penelitian. Teori-teori yang digunakan antara lain mengenai Tumor Otak, *Machine Learning*, *Data Preprocessing*, *Support Vector Machine*, *K-Nearest Neighbor*, *Confusion Matrix*, *Python*.

BAB III : METODOLOGI PENELITIAN

Bagian ini menjelaskan kerangka kerja penelitian, metode pengumpulan data, prosedur penelitian, proses yang dilakukan, serta alat dan bahan yang digunakan dalam menyelesaikan permasalahan yang diteliti.

BAB IV : ANALISIS DAN HASIL

Bab ini menyajikan gambaran umum mengenai data yang dianalisis, hasil pengujian yang telah dilakukan, serta pencapaian dari pengolahan data berdasarkan metode yang telah ditentukan.

BAB V : PENUTUP

Pada bagian BAB V berisi kesimpulan dari hasil penelitian dan saran yang ditujukan pada semua pihak yang bersangkutan.