

BAB I

PENDAHULUAN

Bab ini menyajikan tentang latar belakang masalah, rumusan masalah, tujuan dan manfaat dari dilakukannya penelitian, ruang lingkup, dan sistematika penulisan penelitian tugas akhir.

1.1 Latar Belakang

Makhluk hidup di dunia ini tidak akan pernah terlepas dari gangguan penyakit. Menurut Plianbangchang, penyakit dapat didefinisikan sebagai kegagalan mekanisme adaptif organisme untuk mengatasi rangsangan atau stres, yang mengakibatkan gangguan fungsi atau struktur tubuh (Plianbangchang, 2018). Pada tubuh manusia penyakit dapat terjadi di manapun. Salah satunya adalah penyakit yang terjadi pada punggung.

Penyakit nyeri punggung bawah merupakan gejala yang sangat umum dimiliki berbagai macam kalangan. Penyakit ini terjadi pada negara berpendapatan tinggi, sedang, maupun rendah dan pada berbagai macam usia dari anak-anak hingga lansia. *Years Lived with Disability* (Tahun hidup sehat yang hilang karena disabilitas) yang disebabkan oleh penyakit nyeri punggung bawah meningkat sebanyak 54% dari tahun 1990 sampai 2015, biasa disebabkan karena kenaikan populasi dan penuaan (Hartvigsen dkk., 2018).

Penyakit nyeri punggung bawah merupakan penyakit yang sangat umum dimiliki oleh orang-orang dari berbagai macam kalangan (Carr, 2022). Seiring berjalannya waktu, penderita penyakit nyeri punggung bawah akan terus bertambah. Penyakit nyeri punggung ataupun penyakit yang lainnya sebaiknya dilakukan pendekatan dari berbagai sudut pandang, sudut pandang medis maupun sudut pandang non-medis demi menambah pemahaman yang lebih tinggi tentang penyakit ini. Pendekatan medis dilakukan oleh orang yang memiliki latar belakang di bidang kesehatan sedangkan pendekatan non-medis dilakukan oleh orang yang memiliki bidang non-medis. Salah satu contoh dari pendekatan non-medis adalah sains data (Wertli, 2021).

Sains data merupakan suatu bidang interdisipliner yang berfokus pada metode ilmiah, proses, algoritma, dan sistem untuk mengekstraksi pengetahuan atau wawasan dari data dalam berbagai bentuk, baik terstruktur maupun tidak terstruktur (Cao, 2017). Sains data memiliki berbagai macam sub-bidang. Salah satu sub-bidang di dalam ilmu sains data adalah *data mining* atau penambangan data. Berdasarkan para ahli, *data mining* dapat didefinisikan sebagai berikut; Menurut Aggarwal, *Data Mining* adalah ilmu untuk mengumpulkan, membersihkan, memproses, menganalisis dan memperoleh pengetahuan yang berguna dari sebuah data (Aggarwal, 2015).

Berdasarkan para ahli, dapat disimpulkan bahwa *data mining* adalah proses pengolahan data dengan tujuan untuk menemukan suatu pola yang dapat membantu untuk pengambilan keputusan. Contoh penerapan *data mining* yang berhasil pada bidang medis antara lain; penelitian yang dilakukan oleh Pal dkk pada tahun 2020 berhasil melakukan klasifikasi penyakit jantung menggunakan algoritma *Random Forest* dengan nilai akurasi sebesar 86,9%. Penelitian yang dilakukan oleh Singh dkk pada tahun 2017 berhasil melakukan klasifikasi penyakit jantung menggunakan algoritma *Random Forest* dan *10 Fold Cross Validation* dengan nilai akurasi sebesar 85,81%. Berdasarkan penelitian yang dilakukan dengan pendekatan *data mining* dalam bidang medis memberikan hasil yang positif.

Proses *data mining*, terdiri dari beberapa tahap, salah satu tahapnya adalah *preprocessing*. *Preprocessing* diperlukan untuk meningkatkan akurasi dan efisiensi teknik *Data Mining* selanjutnya yang digunakan (García dkk., 2020). Fase *preprocessing* data terdiri dari 3 tahap antara lain, *Feature Extraction*, *Data Cleaning* dan *Feature Selection and Transformation*. (Aggarwal, 2015). Salah satu teknik yang digunakan secara umum dalam melakukan *Data Mining* adalah teknik *Feature Selection*. Penelitian lainnya yang dilakukan oleh Hassan dkk. (2016) berhasil meningkatkan akurasi klasifikasi pada dataset KDD'99 menggunakan metode *Feature Selection* dari 91,4% menjadi 91,9%. Penelitian yang dilakukan oleh Al-Taharwa dkk. (2022) berhasil meningkatkan akurasi model klasifikasi pada kelangsungan hidup pasien kanker payudara menggunakan *Feature Selection Kendall* meningkat dari 80,3% menjadi 81,8%. Penerapan *Feature Selection Kendall* mampu meningkatkan performa akurasi model, peningkatan ini terjadi karena *Feature Selection Kendall* berhasil mengidentifikasi variabel-variabel paling relevan dan mengurangi *noise*, sehingga model klasifikasi yang dihasilkan menjadi lebih akurat (Handayana, 2018).

Penelitian-penelitian yang telah dilakukan sebelumnya membuktikan bahwa, Algoritma *Random Forest* terbukti memberikan hasil klasifikasi yang lebih baik, dan teknik *Feature Selection* terbukti dapat menambah performa. Penelitian ini melakukan klasifikasi menggunakan algoritma. *Random Forest* dan *Feature Selection Kendall* pada dataset *Lower Back Pain* (Penyakit Punggung Bawah) untuk mencari seberapa besar perubahan akurasi dan performa dalam melakukan klasifikasi.

1.2 Rumusan Masalah

Rumusan masalah dalam penelitian ini adalah; mencari seberapa besar perubahan akurasi yang diperoleh sebelum dan sesudah diterapkan *Feature Selection Kendall* pada pemodelan?

1.3 Tujuan dan Manfaat

Tujuan dari penelitian ini adalah mendapatkan seberapa besar perubahan akurasi dan performa pemodelan setelah dilakukan *Feature Selection Kendall*.

Manfaat dari penelitian ini adalah menambah ragam pemodelan dengan *Feature Selection Kendall* diharapkan dapat berguna bagi peneliti-peneliti lain yang nantinya diterapkan di skenario yang berbeda dengan tujuan untuk meningkatkan performa pemodelan pada dataset *Lower Back Pain*.

1.4 Ruang Lingkup

Ruang lingkup dalam penelitian ini adalah sebagai berikut :

1. Data yang diambil adalah dataset *Lower back Pain* yang bersumber dari situs web Kaggle.com (<https://www.kaggle.com/datasets/sammy123/lower-back-pain-symptoms-dataset/data>). Data diambil dari pada tahun 2018. Dataset memiliki 310 baris data.
2. Proses data mining menggunakan perangkat lunak google.colab.com.
3. Fungsi *Random Forest* dari kode yang digunakan menggunakan parameter *default* karena membandingkan 2 model (Model 1 tanpa menggunakan *Feature Selection*, Model 2 menggunakan *Feature Selection*). Parameter harus sama agar pemodelan adil.

1.5 Sistematika Penulisan

Tujuan dari sistematika penulisan adalah untuk memberikan gambaran tentang pembahasan yang ada di dalam laporan penelitian secara urut dan jelas. Berikut adalah sistematika penulisan laporan penelitian ini:

BAB I PENDAHULUAN

Bab ini menyajikan tentang latar belakang masalah, rumusan masalah, tujuan dan manfaat dari dilakukannya penelitian, ruang lingkup, dan sistematika penulisan laporan penelitian.

BAB II LANDASAN TEORI

Bab ini menyajikan tinjauan pustaka dan dasar dari teori-teori yang akan digunakan dalam penelitian, antara lain penelitian terkait, penyakit nyeri punggung bawah (*Lower Back Pain*), *Data Mining*, Metode Klasifikasi, *Feature Selection Kendall*, *Random Forest*, *Decision tree*, *Confusion Matrix*, *K-fold Cross Validation*.

BAB III METODOLOGI PENELITIAN

Bab ini menyajikan tentang alur penelitian yang dilakukan untuk mencapai akhir penelitian, antara lain jenis dan sumber data, variabel, tahapan penelitian, diagram alir tahapan penelitian.

BAB IV HASIL DAN PEMBAHASAN

Bab ini menyajikan hasil dari penelitian dan pembahasan tentang hasil penelitian tersebut.

BAB V PENUTUP

Bab ini menyajikan tentang kesimpulan, saran, dan kritik.