BAB I

PENDAHULUAN

1.1 LATAR BELAKANG

Pembelajaran mesin (*machine learning*) menjadi semakin penting dikarenakan kemampuannya dalam menganalisis serta memproses data dalam jumlah yang besar secara cepat dan efisien [1]. Dalam proses pembelajaran mesin (*machine learning*), digunakan sejumlah data untuk melatih model yang kemudian akan dilakukan prediksi atau pengambilan keputusan berdasarkan pola yang ditemukan dalam data tersebut. Terdapat beberapa kategori dalam algoritma *machine learning* diantaranya pembelajaran dengan pengawasan (*supervised learning*), pembelajaran tanpa pengawasan (*unsupervised*), serta pembelajaran penguatan (*reinforcement learning*). Dalam pembelajaran mesin (*machine learning*), data dibagi ke dalam beberapa kategori tertentu dalam klasifikasi yang merupakan salah satu aplikasi utama. Salah satu metode sederhana namun efektif dalam bidang klasifikasi yaitu algoritma *K-Nearest Neighbor* (KNN). Menghitung jarak antara data baru dengan data yang sudah ada serta mengelompokkannya berdasarkan "tetangga" terdekat merupakan cara kerja *K-Nearest Neighbor* (KNN) [2]. Pembelajaran mesin (*machine learning*) digunakan dalam berbagai bidang, salah satunya dalam bidang kesehatan.

Salah satu tantangan kesehatan terbesar saat ini adalah obesitas, yang merupakan kondisi kelebihan berat badan akibat penumpukan lemak tubuh berlebihan [3]. Obesitas telah menjadi masalah kesehatan global dengan angka

kejadian yang terus meningkat. Berdasarkan data *World Health Organization* (WHO), jumlah kasus obesitas telah meningkat lebih dari tiga kali lipat sejak tahun 1975. Obesitas dihubungkan dengan berbagai penyakit kronis seperti diabetes tipe 2, hipertensi, penyakit jantung, serta stroke, yang tidak hanya mempengaruhi kualitas hidup individu tetapi juga meningkatkan beban ekonomi dalam sistem perawatan kesehatan.

Terdapat beberapa penelitian sebelumnya mengenai klasifikasi menggunakan algoritma KNN. Pada penelitian pertama [4] menunjukkan bahwa model KNN dapat diperoleh dengan nilai k=1, serta atribut berat badan dan umur menjadi faktor yang paling berpengaruh. Berdasarkan percobaan yang dilakukan, akurasi maksimum yang diperoleh ialah 79,96%. Namun, pada hasil akurasi tersebut menunjukkan adanya peluang dalam melakukan pengembangan serta perbaikan lebih lanjut agar sistem dapat memberikan hasil yang lebih akurat. Pada penelitian kedua [5] hasil evaluasi kinerja terbaik diperoleh dengan k=21 dan k=11 dengan akurasi 98%. Hal tersebut menunjukkan bahwa algoritma KNN efektif dalam mengklasifikasikan kanker payudara. Namun diharapkan pada penelitian selanjutnya agar dapat menggabungkan metode ini dengan algoritma lain untuk meningkatkan akurasi dan presisi hasil prediksi. Pada penelitian ketiga [6] tahapan penelitian CRISP-DM dan penerapan algoritma *K-Nearest Neighbor* (KNN) yang menunjukkan bahwa tingkat akurasi yang dihasilkan sebesar 93,88% dengan jumlah data sebanyak 2.500 data. Dengan nilai *recall* tertinggi yang dicapai oleh klasifikasi harga terbesar 96,91% serta nilai *precision* tertinggi dicapai oleh kualifikasi pembayaran sebesar 98,67%. Pada penelitian keempat [7] proses

klasifikasinya dilakukan dengan melakukan pengisian data yang kosong setelah itu dilakukan pembobotan atribut serta normalisasi pada data yang terpilih. Pada penelitian ini, penguunaan normalisasi *min-max* dapat memperoleh nilai akurasi paling optimal. Pada seleksi fitur *Information Gain* dan *Gain Ratio* digunakan nilai rata-rata untuk mengisi kekosongan data dengan menggunakan nilai k = 10. Sedangkan pada seleksi fitur *Symmetrical Uncertainty* dapat menggunakan nilai 0 untuk kekosongan data dan nilai k = 5. Pada penelitian kelima [8] prediksi penyakit stroke dilakukan berdasarkan jenis kelamin, umur, hipertensi, riwayat penyakit jantung, status menikah, tipe pekerjaan, tipe tempat tinggal, rerata kadar glukosa , BMI dan status merokok dengan akurasi yang didapatkan sebesar 95% dengan nilai k=9 melewati hasil perhitungan klasifikasi-prediksi pada data penyakit stroke dengan data latih sebanyak 80 data serta data uji sebanyak 20 data.

Pada penelitian sebelumnya, algoritma *K-Nearest Neighbor* (KNN) telah banyak digunakan dalam berbagai klasifikasi data. Namun, pada penelitian sebelumnya [6] hanya menggunakan data sebanyak 2.500 sampel dalam proses klasifikasi. Jumlah tersebut dirasa belum cukup mewakili dalam penggambaran populasi yang lebih luas sehingga dapat memengaruhi akurasi hasil klasifikasi. Oleh karena itu, penelitian ini akan menggunakan dataset yang lebih besar dengan jumlah data sebanyak 20.758 sampel yang bertujuan untuk meningkatkan keakuratan model dalam proses klasifikasi tingkat obesitas menggunakan algoritma *K-Nearest Neighbor* (KNN).

Berdasarkan analisis sebelumnya [5] menunjukkan bahwa algoritma KNN

memiliki tingkat akurasi yang tinggi. Hal tersebut dapat membantu para professional kesehatan dalam mengklasifikasikan tingkat obesitas dalam upaya pencegahan serta penanganan yang tepat berdasarkan tingkat keparahan obesitas. Atas beberapa penelitian tersebut, penulis memutuskan untuk menggunakan algoritma *K-Nearest Neighbor* (KNN) dalam proses pengklasifikasian tingkat obesitas.

Oleh karena itu, penulis tertarik untuk melakukan klasifikasi serta melakukan penelitian mengenai obesitas yang akan dijelaskan dengan judul: **"Klasifikasi Tingkat Obesitas Menggunakan Algoritma *K-Nearest Neighbor* (KNN)"**.

1.2 RUMUSAN MASALAH

Berdasarkan uraian latar belakang diatas, dapat dirumuskan beberapa permasalahan yaitu " Bagaimana cara menganalisis serta mengklasifikasikan tingkat obesitas pada penelitian sebelumnya ? "

1.3 BATASAN MASALAH

Dalam proses pengklasifikasian tingkat obesitas menggunakan algoritma KNN, penulis membatasi ruang lingkup masalah supaya tetap sejalan dengan maksud serta tujuan. Beberapa batasan tersebut diantaranya :

1. Penelitian ini hanya menggunakan bahasa pemrograman *python*.
2. Penelitian ini hanya dilakukan menggunakan algoritma *K-Nearest Neighbour* (KNN) dalam mengevaluasi akurasi hasil klasifikasi tingkat obesitas.
3. Data yang digunakan dalam penelitian ini mencakup beberapa variabel seperti, jenis kelamin (*gender*), usia (*age*), tinggi badan (*height*), berat badan (*weight*), kebiasaan merokok (*SMOKE*), dan jenis transportasi (*MTRANS*).

4. Dataset yang digunakan berjumlah 20.758 baris.
5. Data yang digunakan pada klasifikasi tingkat obesitas ini hanya bersumber dari *kaggle*.

1.4 TUJUAN DAN MANFAAT PENELITIAN

1.4.1 Tujuan Penelitian

Tujuan dari klasifikasi tingkat obesitas menggunakan algoritma *K-Nearest Neighbor* (KNN), diantaranya:

1. Menganalisis akurasi serta efektivitas algoritma KNN dalam klasifikasi tingkat obesitas.
2. Mengidentifikasi beberapa variabel utama yang paling mempengaruhi hasil klasifikasi obesitas.
3. Mengimplementasikan algoritma KNN untuk mengimplementasikan tingkat obesitas berdasarkan data pasien.

1.4.2 Manfaat Penelitian

Beberapa manfaat yang dapat dihasilkan dari klasifikasi tingkat obesitas menggunakan algoritma *K-Nearest Neighbor* (KNN), diantaranya:

1. Penelitian ini dapat berguna bagi tenaga medis dalam mendeteksi serta mengklasifikasikan tingkat obesitas pasien secara cepat dan akurat.
2. Penelitian ini dapat menjadi titik acuan bagi penelitian selanjutnya dalam mengembangkan model klasifikasi kesehatan lainnya.
3. Penelitian ini diharapkan dapat meningkatkan kesadaran masyarakat dalam menjaga kesehatan serta mencegah kondisi obesitas lebih lanjut.

1.5 SISTEMATIKA PENULISAN

Penulis membuat sistematika penulisan agar penulis menjadi lebih terarah dan mencapai tujuannya. Berikut ini sistematika penulisan yang dibuat, yaitu:

BAB I : PENDAHULUAN

Pada bab ini menjelaskan tentang latar belakang, rumusan masalah, batasan masalah, tujuan dan manfaat penelitian serta sistematika penulisan

BAB II : LANDASAN TEORI

Pada bab ini menjelaskan beberapa teori yang mendasari pembahasan laporan secara khusus diantaranya berupa pengertian dasar mengenai pembelajaran mesin (*machine learning*), klasifikasi, *python*, obesitas, algoritma *K-Nearest Neighbor* (KNN) dan tabel penelitian sejenis yang didapat melalui studi literatur sebagai dasar dalam penelitian.

BAB III : METODOLOGI PENELITIAN

Pada bab ini menjelaskan tentang kerangka kerja penelitian, metode pengumpulan data, metode klasifikasi, serta alat bantu yang digunakan terhadap penelitian ini.

BAB IV : ANALISIS

Pada bab ini menjelaskan tentang gambaran umum objek penelitian, serta menganalisis dataset yang didapat dari web *kaggle.*com dengan melakukan persiapan data, kemudian penulis akan melakukan analisis data eksploratori (*Exploratory*

Data Analysis) untuk mengetahui prediksi tingkat akurasi data menggunakan metode klasifikasi algoritma *K-Nearest Neighbor* (KNN) dengan menggunakan alat bantu (*tools*) yaitu *google colaboratory* atau biasa disebut dengan *google colab*. Hasil dari analisis yang telah dilakukan diinterpretasikan menjadi informasi mengenai tingkat obesitas menggunakan alat bantu yaitu *google colab*.

BAB V : KESIMPULAN

Pada bab ini merupakan bagian penutup dari penelitian ini yang berisi kesimpulan dari pembahasan beberapa bab sebelumnya serta beberapa saran yang dapat berguna bagi beberapa pihak yang bersangkutan dengan penelitian ini.