

BAB I

PENDAHULUAN

1.1 Latar Belakang

Machine learning dapat didefinisikan sebagai aplikasi komputer dan algoritma matematika yang diadopsi dengan cara pembelajaran yang berasal dari data dan menghasilkan prediksi di masa yang akan datang. Adapun proses pembelajaran yang dimaksud adalah suatu usaha dalam memperoleh kecerdasan yang melalui dua tahap antara lain latihan (*training*) dan pengujian (*testing*). Bidang *machine learning* berkaitan dengan pertanyaan tentang bagaimana membangun program komputer agar kecerdasan program meningkat secara otomatis berdasarkan dari pengalaman [1].

Penelitian tentang penggunaan *machine learning* pada bidang kesehatan sudah pernah dilakukan oleh para peneliti – peneliti sebelumnya. Seperti penelitian yang dilakukan oleh (Marios Anthimopoulos, stargios chistodoulidis, Lucas ebner, Adreas christe dan Stavroula mougiakakou, 2016) dengan judul “*Lung Pattern Classification For Interstitial Lung Disiases Using a Deep Convulutional Neural Network*”, penelitian menggunakan metode CNN dalam mengklasifikasi pola dan dievaluasi dari 120 kasus dari dua rumah sakit yang berbeda. Metode ini menghasilkan pola yang unggul dari metode lainnya dan lebih cepat dalam mengklasifikasi penyakit yang ada di paru paru dalam bentuk gambar yang lebih besar dan jelas. Dalam penelitian lain yang dilakukan oleh (Rifqi Hammad, Julia Kurniasih, Nur Fitrianiingsih Hasan, Christin Nandari Dengen, Kusrini. Tahun terbit 2019) dengan judul “*Machine learning pada bidang kesehatan demensia (penurunan fungsi otak)*” penelitian

ini menggunakan metode *forward chaining* digunakan untuk mengembangkan *prototipe machine learning* untuk prognosis penyakit demensia, dengan menggunakan data literatur pemeriksaan pasien yang telah didiagnosis penyakit demensia, meliputi pemeriksaan tekanan darah, kadar lemak darah, kadar gula darah, vesicular dan inspeksi abdomen. Dalam penelitian ini, berhasil dijalankan dan mendapatkan solusi prognosis penyakit demensia yang sesuai dengan hasil pemeriksaan yang optimal. Dalam penelitian yang lain dilakukan oleh Sana Rebbah, daniel Delahaye, Stepane Puechmorel, Pierre Marechal, Florenco Nicol. OHBM 2019, pertemuan tahunan ke 25 organisasi pemetaan otak manusia, juni 2019, Roma, Italia. Dengan judul "*Classification of Multiple Sclerosis Patients Using a Histogram based KNN Algorithm*" dalam penelitian ini menggunakan algoritma KNN dalam menyelesaikan masalah, penelitian ini mendapat akurasi Dalam tingkat akurasi dalam membedakan pasien HC dan MS mencapai 80 %. dalam penelitian ini mencakup 111 subjek, 71 kontrol sehat dari database alzheimer disease neuroimaging initiative (ADNI) dan menggunakan 40 pasien yang mengidap penyakit *multiple sclerosis* [2].

Penelitian ini dibuat dengan tujuan untuk memperbaiki penelitian sebelumnya. Pada penelitian sebelumnya, data yang digunakan berasal dari data **Pima Indians Diabetes Dataset**. Data Pima Indian sendiri berasal dari **National Institute of Diabetes and Digestive and Kidney Disease** dengan mengambil sampel terhadap 768 wanita keturunan pima indian usia 21 tahun ke atas. Dari 768 sampel data, 500 sampel tidak terdiagnosis diabetes sedangkan 268 sampel terdiagnosis positif diabetes. Data pada penelitian tersebut terbagi menjadi 75%

untuk data *training* dan 25% untuk data *testing*. Pada penelitian tersebut menghasilkan akurasi prediksi dengan keakuratan sebesar 75,97%.

Pada penelitian ini dilakukan proses penambahan data dengan menggunakan teknik *One Point Cross Over*. Penambahan data dilakukan dengan tujuan agar jumlah pasien yang terdeteksi diabetes dan tidak diabetes menjadi seimbang. Penambahan data dilakukan dengan cara mengambil sampel pasien yang terdeteksi diabetes sebanyak 232 sampel, kemudian dari 232 sampel tersebut dihasilkan sampel baru melalui proses *One Point Cross Over*, Sehingga jumlah sampel menjadi 1000, dengan 500 sampel terdeteksi diabetes dan 500 sampel tidak diabetes.

Kesehatan adalah sesuatu yang mahal dan diinginkan oleh manusia. Oleh karena itu setiap manusia ingin selalu menjaga kesehatannya agar bisa menikmati hidup dengan nyaman. Menurut organisasi kesehatan dunia, dalam konstitusinya, mendefinisikan kesehatan sebagai **“keadaan kesehatan fisik, mental, dan sosial yang lengkap”**. Saat ini, definisi itu terus digunakan sebagai definisi kesehatan yang mencakup segalanya. Hal ini menjadi standar keadaan ideal setiap manusia [3]. Sayangnya banyak manusia yang tidak bisa menjaga kesehatannya, sehingga banyak bermunculan penyakit yang dapat merusak kondisi tubuh, bahkan di era modern saat ini banyak manusia yang menderita penyakit yang beresiko menyebabkan kematian.

Diabetes adalah salah satu penyakit yang paling ditakuti oleh manusia, karena penyakit diabetes dapat merusak bagian anggota tubuh bahkan dapat menyebabkan kematian. Definisi diabetes adalah salah satu penyakit yang disebabkan oleh ketidak mampuan pankreas untuk menghasilkan insulin yang

mengontrol gula dalam darah. Hal ini jika dibiarkan akan berbahaya dikarenakan dapat terjadi komplikasi seperti jantung, stroke, glaukoma ataupun kanker. Di Indonesia, berdasarkan data yang dikeluarkan oleh *World Health Organization* (WHO) pada tahun 2016, kematian yang disebabkan oleh diabetes sebanyak 99.400 jiwa, dimana 48.300 kasus kematian terjadi pada usia produktif [4]. Diabetes menempati peringkat salah satu penyakit paling kronis di dunia. Statistik menunjukkan bahwa pada 2013 sekitar 382,8 juta orang berusia antara 20 sampai 79 tahun telah didiagnosis mengidap diabetes di seluruh dunia. Selama tahun itu terjadi 4,6 juta kematian dan telah menghabiskan biaya pengobatan sebesar \$ 548 miliar [5].

Berdasarkan permasalahan yang telah diuraikan, maka diperlukan teknik klasifikasi untuk menyelesaikan permasalahan pada bidang *data analitic*. Definisi dari klasifikasi menurut Kamus Besar Bahasa Indonesia (KBBI) yaitu, “penyusunan bersistem dalam kelompok atau golongan menurut kaidah atau standar yang ditetapkan” [6]. Ada banyak jenis klasifikasi, antara lain *Linier Regression*, *Support Vector Machine*, *Decission Tree*, *Random Forest*, dan lain-lain. Pada Penelitian ini akan digunakan metode *Logistic Regression*, untuk menyelesaikan permasalahan klasifikasi penyakit diabetes.

Metode *Logistic Regression*, juga dikenal sebagai analisis *Logistic Regression*, metode ini sering digunakan dalam berbagai bidang, yaitu digunakan dalam penambangan data, diagnosis penyakit otomatis, prediksi ekonomi dan bidang lainnya [7]. Oleh karena itu metode ini bisa digunakan untuk klasifikasi penyakit diabetes. Pada penelitian ini telah dilakukan lima kali skenario percobaan dengan menggunakan metode *Logistic Regression*, dari percobaan tersebut

diperoleh akurasi tertinggi yaitu sebesar 80%. Diharapkan hasil dari penelitian ini dapat berguna untuk analisis data pada klasifikasi penyakit diabetes.

1.2 Rumusan Masalah

1. Bagaimana mengimplementasikan metode *Logistic Regression* untuk penerapan klasifikasi penyakit diabetes?
2. Bagaimana keakuratan metode *Logistic Regression* untuk klasifikasi penyakit diabetes?
3. Bagaimana cara memperbaiki kekurangan pada penelitian sebelumnya, terutama pada nilai akurasi?

1.3 Tujuan

1. Mengimplementasikan metode *Logistic Regression* untuk penerapan klasifikasi penyakit diabetes.
2. Menguji keakuratan metode *Logistic Regression* pada implementasi klasifikasi penyakit diabetes.
3. Memperbaiki kekurangan pada penelitian sebelumnya, terutama pada nilai akurasi prediksi.

1.4 Ruang Lingkup

Agar pengerjaan tugas akhir ini menjadi lebih terarah dan mendapatkan hasil yang lebih spesifik, maka sistem yang dirancang dibatasi pada ruang lingkup pembahasan sebagai berikut:

1. Data yang tersedia merupakan data yang diambil dari internet, data tersebut diterbitkan oleh “**National Institute of Diabetes and Digestive and Kidney Diseases**”.
2. Pasien yang menjadi sampel adalah pasien berusia 21 tahun ke atas.

3. Data pada penelitian ini terdiri dari sembilan kolom dan 769 baris.

1.5 Metode Penelitian

Metodologi penelitian merupakan cara untuk menguji sesuatu dengan menggunakan fikiran untuk mencapai tujuan. Dengan metodologi penelitian, didapatkan kesimpulan dan solusi dari suatu permasalahan. Berikut ini adalah metodologi penelitian yang digunakan dalam menyelesaikan tugas akhir, yang terdiri dari:

1. Pengumpulan data

Mengambil, dan mengumpulkan data klasifikasi penyakit diabetes. Adapun data tersebut merupakan data sekunder yang diambil dari internet.

2. Studi Literatur

Mencari dan mempelajari referensi-referensi yang berkaitan dengan *machine learning*, *Logistic Regression*, dan data penyakit diabetes, yang didapat dari jurnal, buku, maupun media lain seperti internet.

3. Analisis Kebutuhan dan Perancangan Perangkat Lunak

Melakukan analisis terhadap kebutuhan sistem dan mengidentifikasi kebutuhan informasi berdasarkan hasil observasi yang telah dilakukan. Sistem diimplementasikan menggunakan bahasa pemrograman Python.

4. Pengujian dan Evaluasi

Pengujian meliputi uji coba data klasifikasi penyakit diabetes dengan metode *Logistic Regression*. Hasil pengujian dianalisis untuk mendapatkan kesimpulan dari seluruh rangkaian penelitian.

1.6 Sistematika Penulisan

Tujuan sistematika penulisan adalah memberikan gambaran singkat mengenai laporan yang dibuat secara garis besar dalam penyusunan laporan penelitian ini, dimana terbagi dalam beberapa bab antara lain, sebagai berikut:

BAB I PENDAHULUAN

Bab ini berisi latar belakang, rumusan masalah, tujuan, ruang lingkup, metodologi dan sistematika penulisan.

BAB II LANDASAN TEORI

Pada bab ini diuraikan tentang teori yang berhubungan dengan judul tugas akhir, seperti *machine learning*, klasifikasi, *Logistic Regression* dan lain-lain.

BAB III ANALISIS DAN PERANCANGAN

Pada bab ini dijelaskan mengenai analisis dan perancangan yang akan dilakukan untuk memabangun sistem.

BAB IV IMPL EMENTASI DAN PENGUJIAN

Pada bab ini dipaparkan mengenai sistem untuk klasifikasi penyakit diabetes. Metode yang digunakan dalam mengimplementasikan sistem adalah metode *Logistic Regression*. Dijelaskan juga hasil pengujian sistem yang meliuti uji pelatihan menggunakan metode *Logistic Regression*, serta uji keakuratan metode *Logistic Regression*.

BAB V PENUTUP

Pada bab ini dipaparkan hasil kesimpulan dari penelitian klasifikasi penyakit diabetes dengan metode *Logistic Regression* dan saran-saran untuk pengembangan sistem yang dibangun.