

BAB I

PENDAHULUAN

1.1 Latar Belakang

Data medis memainkan peran penting dalam mendukung pengambilan keputusan di bidang Kesehatan, dengan data yang terstruktur, misalnya saja seperti data tabular yang didalamnya memuat data terkait rekam medis pasien, hasil uji laboratorium, maupun status kondisi fisiknya[1]. Data tersebut dapat diolah untuk menghasilkan wawasan yang membantu diagnosis, pengobatan, dan pencegahan penyakit. Salah satu penyakit kronis yang kerap menjadi objek analisis adalah penyakit diabetes, mengingat prevalensinya yang terus meningkat secara global.

Menurut penelitian yang berjudul **Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045 : Results from the International Diabetes Federation Diabetes Atlas, 9th edition**, pada tahun 2019 jumlah penderita diabetes diperkirakan sebanyak 463 juta orang, dan angka ini diproyeksikan meningkat menjadi 578 juta pada 2030 serta mencapai 700 juta pada tahun 2045[2]. Angka ini mencerminkan ancaman serius bagi kesehatan masyarakat global dan menyoroti perlunya perhatian yang lebih besar terhadap pencegahan serta pengelolaan penyakit ini[3], [4]. Beragam faktor dapat memicu terjadinya diabetes, seperti pola makan yang buruk, minimnya aktivitas fisik, serta predisposisi genetik yang turut berperan dalam perkembangan penyakit ini[5], [6]. Gaya hidup modern yang minim aktivitas fisik serta kebiasaan konsumsi makanan tinggi gula dan lemak menjadi penyebab utama meningkatnya jumlah penderita diabetes[7]. Oleh karena itu, upaya deteksi dini dan pengelolaan yang efektif menjadi sangat krusial untuk meningkatkan kesadaran masyarakat akan risiko diabetes dan pentingnya menjalani gaya hidup sehat[4].

Oleh karena itu dalam upaya memprediksi dan mendiagnosis diabetes, model machine learning (ML) telah menjadi suatu alat yang cukup

membantu[8]. Teknologi ini menawarkan kemampuan menganalisis data dalam jumlah besar dengan efisien, serta juga dapat mengidentifikasi pola-pola yang mungkin tidak terlihat dengan menggunakan metode tradisional[9], [10]. Dengan pendekatan ini, model machine learning mampu menghasilkan output yang lebih tepat dan efisien, sebuah kemampuan yang sangat penting dalam sektor kesehatan[11]. Penerapan model ini tidak hanya mempercepat proses diagnosis, tetapi juga mendukung tenaga medis dalam membuat keputusan yang lebih tepat untuk penanganan pasien. Selain itu model ML dapat membantu dalam mengembangkan strategi pencegahan penyakit yang lebih efektif, dengan memberikan wawasan yang mendalam mengenai faktor-faktor risiko perkembangan diabetes.

Namun, meskipun penggunaan model machine learning sangat membantu, terdapat kekhawatiran mengenai bias yang mungkin saja bisa terjadi dalam model prediksi. Bias ini dapat muncul dari data yang tidak seimbang, di mana representasi kelompok tertentu mungkin kurang memadai, ataupun juga bisa dari algoritma yang tidak adil, yang akhirnya malah dapat memperkuat diskriminasi terhadap kelompok tertentu[12], [13]. Masalah ini perlu mendapatkan perhatian serius karena dapat menimbulkan dampak negatif, seperti ketidaksetaraan dalam akses layanan kesehatan serta ketidakakuratan dalam hasil diagnosis[12]. Oleh karena itu, penelitian tentang fairness dalam machine learning harus menjamin bahwa model yang dikembangkan tidak hanya memiliki tingkat akurasi tinggi, tetapi juga menjunjung prinsip keadilan, merupakan hal yang sangat krusial. Dengan memahami dan mengidentifikasi bias dalam model, kita dapat mengurangi risiko diskriminasi yang tidak diinginkan, serta meningkatkan kepercayaan masyarakat terhadap penggunaan teknologi dalam bidang kesehatan[13]. Pendekatan yang bertanggung jawab dalam pengembangan dan penerapan model machine learning akan menciptakan sistem yang lebih adil, dan memberikan manfaat yang merata bagi semua kelompok masyarakat[12].

Sehingga pada penelitian ini, library Dalex akan digunakan sebagai alat untuk menganalisis fairness pada model yang dikembangkan. Library Dalex sendiri dirancang untuk memungkinkan pengguna mengeksplorasi dan memahami bagaimana model berperilaku, serta mendeteksi adanya bias yang mungkin terdapat di dalamnya. Dengan fitur-fitur interaktif yang ditawarkannya, Dalex cukup membantu untuk memperoleh wawasan yang mendalam mengenai keputusan yang diambil oleh model, sehingga dapat diidentifikasi potensi bias yang merugikan[14]. Dengan pemahaman yang lebih baik mengenai bias tersebut, langkah-langkah perbaikan yang tepat dapat diambil nantinya.

Harapannya, penelitian ini dapat mendorong lebih banyak perhatian terhadap pentingnya fairness dalam pengembangan teknologi khususnya pada bidang kesehatan, sehingga kedepannya setiap individu dapat menerima perawatan yang sesuai dengan kebutuhan mereka tanpa terpengaruh oleh faktor-faktor diskriminatif seperti ras, jenis kelamin, atau latar belakang sosial-ekonomi. Ini tidak hanya akan meningkatkan hasil kesehatan secara keseluruhan, tetapi juga membangun kepercayaan masyarakat terhadap teknologi kesehatan yang semakin berkembang.

1.2 Rumusan Masalah

Dengan mempertimbangkan konteks pada latar belakang, dirumuskan beberapa permasalahan utama sebagai berikut:

- a. Bagaimana model machine learning dapat digunakan untuk memprediksi penyakit diabetes, serta algoritma apa yang paling efektif dan akurat?
- b. Apakah terdapat bias dalam model machine learning untuk prediksi diabetes, dan bagaimana dampaknya terhadap keadilan hasil prediksi?

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah yang telah dijabarkan, penelitian ini memiliki beberapa tujuan sebagai berikut:

- a. Menganalisis penggunaan model machine learning untuk memprediksi penyakit khususnya diabetes.
- b. Menganalisis efektivitas berbagai algoritma machine learning dalam prediksi penyakit diabetes.
- c. Mengidentifikasi bias yang mungkin saja muncul dalam model machine learning yang memprediksi penyakit diabetes.
- d. Menilai dampak dari adanya bias dalam model machine learning yang memprediksi penyakit diabetes.

1.4 Batasan Penelitian

Berdasarkan masalah yang telah dijabarkan, penelitian ini dibatasi oleh beberapa aspek penting sebagai berikut:

- a. Penelitian ini terbatas pada penggunaan data penyakit diabetes '**Diabetes Health Indicators**' sebagai sumber data utama untuk pengembangan model machine learning.
- b. Penelitian hanya akan mengeksplorasi algoritma machine learning klasik seperti **Decision Tree, Random Forest, Logistic Regression, dan Naive Bayes**.
- c. Fokus utama penelitian ini adalah **mengecek** apakah dalam sebuah model machine learning yang memprediksi penyakit diabetes **terdapat bias**.
- d. Penelitian ini akan mengandalkan **library Dalex** untuk analisis fairness, sehingga tidak mencakup metode analisis lain di luar library dalex.