

BAB I

PENDAHULUAN

Dalam berbagai analisis dan pemrosesan data, ketidakseimbangan informasi sering kali menjadi tantangan yang mempengaruhi akurasi dan keandalan hasil yang diperoleh. Data yang tidak merata dapat menyebabkan bias dalam pemodelan, di mana kategori yang lebih dominan cenderung memiliki pengaruh lebih besar dibandingkan kategori yang kurang terwakili. Untuk mengatasi masalah ini, diperlukan pendekatan yang tidak hanya mampu menyeimbangkan distribusi tetapi juga memastikan bahwa data tetap bersih dan representatif. Dengan strategi yang tepat, proses analisis dapat dilakukan secara lebih optimal, memungkinkan pengambilan keputusan yang lebih akurat serta peningkatan kualitas hasil yang diperoleh.

1.1. Latar Belakang

Dalam beberapa dekade terakhir, diabetes telah berkembang menjadi salah satu isu kesehatan yang krusial di Indonesia. Kondisi ini terutama disebabkan oleh gaya hidup masyarakat yang berubah dan pola makan mereka. Menurut *International Diabetes Federation* (IDF), jumlah penderita diabetes di seluruh dunia pada tahun 2019 mencapai 463 juta orang. Indonesia menempati peringkat ke-7 di dunia. Pada tahun 2030, diperkirakan ada 587 juta penderita diabetes di Indonesia [1]. Dampak yang disebabkan oleh tingginya angka diabetes antara lain, adalah biaya pengobatan diabetes yang sangat tinggi. Hal tersebut mencakup biaya perawatan kesehatan dan komplikasi yang mungkin akan timbul. Masyarakat dengan penghasilan rendah rentan terhadap komplikasi akibat keterbatasan akses terhadap perawatan kesehatan yang memadai. Komplikasi serius seperti penyakit jantung, stroke, kerusakan ginjal, dan kebutaan dapat muncul jika diabetes tidak dikelola dengan baik.

Dengan peningkatan jumlah kasus yang tercatat sejak awal tahun 1980-an, diabetes telah lama diidentifikasi menjadi salah satu isu utama dalam bidang kesehatan masyarakat di Indonesia. Kondisi ini tidak hanya memengaruhi kualitas hidup individu yang mengidapnya, tetapi juga membebani sistem kesehatan negara secara keseluruhan. Dengan lebih dari 10 juta orang yang menderita diabetes,

Indonesia adalah negara terbesar di Asia Tenggara dengan prevalensi 6,2% di antara orang dewasa. [2]. Tantangan ini semakin berat karena banyak kasus yang tidak teridentifikasi atau tidak ditangani dengan baik, terutama di daerah pedesaan dengan layanan kesehatan yang terbatas. Penanganan yang tepat dan berkelanjutan diperlukan untuk diabetes, yang dapat diketahui oleh kadar gula darah tinggi serta komplikasi kesehatan yang dapat mempengaruhi berbagai organ tubuh. Risiko penyakit jantung, gagal ginjal, dan kerusakan saraf akan meningkat jika tidak ada pengobatan yang tepat. Akibatnya, Indonesia harus meningkatkan kesadaran masyarakat akan pentingnya pencegahan dan pengelolaan diabetes sekaligus memperkuat sistem kesehatannya untuk menangani peningkatan kasus diabetes. Diabetes terdiri dari diabetes tipe 1, diabetes tipe 2, dan diabetes gestasional. Diabetes tipe 2 adalah yang paling umum, hal ini dikarenakan kebiasaan hidup yang kurang sehat, termasuk pola makan yang tidak tepat serta minimnya aktivitas fisik. Diabetes tipe 1 terjadi karena sistem imun menyerang sel-sel di pankreas yang menghasilkan insulin. Wanita yang menderita diabetes gestasional mengalami kondisi ini selama kehamilan dan biasanya hilang setelah kelahiran.

Penelitian sebelumnya, "Klasifikasi Penderita Penyakit Diabetes menggunakan Algoritma C4.5", menganalisis kemampuan algoritma C4.5 untuk mengkategorikan pasien diabetes berdasarkan kumpulan data yang dikumpulkan. Dataset yang dimanfaatkan penelitian ini terdiri dari 768 baris dan sepuluh kolom dan mencakup berbagai fitur yang terkait dengan faktor risiko diabetes, seperti usia, kadar glukosa, tekanan darah, dan sejumlah variabel medis lainnya. Hasil penelitian menunjukkan bahwa algoritma C4.5 dapat mengklasifikasikan seseorang dengan tingkat akurasi sebesar 74,08% dalam menentukan apakah mereka menderita diabetes atau tidak. Angka-angka ini menunjukkan kemampuan algoritma untuk menangani data medis dan menunjukkan bahwa mereka dapat digunakan dalam sistem pendukung keputusan untuk membantu diagnosis awal diabetes [3].

Dalam upaya meningkatkan akurasi dalam klasifikasi penyakit diabetes, penelitian ini mengusulkan pendekatan yang memadukan dua teknik utama, yaitu metode balancing data menggunakan *SMOTETomek* dan algoritma *XGBoost*. Untuk menyeimbangkan distribusi data, kami mengadopsi algoritma *SMOTETomek*, yang menggabungkan metode oversampling *SMOTE* dan metode

undersampling Tomek-Link [4]. Sementara itu, *XGBoost* (*Extreme Gradient Boosting*), sebagai algoritma pembelajaran mesin berbasis pohon keputusan yang sangat kuat, akan digunakan untuk mengklasifikasi penyakit diabetes berdasarkan data yang telah seimbang. Algoritma ini dikenal karena kemampuan untuk menyelesaikan masalah klasifikasi yang kompleks dan meningkatkan efisiensi dan akurasi komputasi. Penelitian ini bertujuan untuk menciptakan model prediksi yang lebih andal untuk mendiagnosis penyakit diabetes, terutama untuk data yang tidak seimbang, dengan mengintegrasikan *SMOTETomek* dan *XGBoost*.

Penerapan metode data balancing, seperti teknik pengambilan sampel berlebih minoritas sintetis yang dikombinasikan dengan *Tomek-link* (*SMOTETomek*), memiliki potensi untuk meningkatkan kinerja model pembelajaran mesin pada kelas minoritas. Model pembelajaran mesin yang bias dan kinerja buruk pada kelas minoritas dapat disebabkan oleh dataset yang tidak seimbang. Data tersebut akan menjalani serangkaian langkah pemrosesan, termasuk penanganan nilai yang hilang, duplikasi, normalisasi, dan implementasi *SMOTETomek* untuk mengatasi ketidakseimbangan data [5]. Metode penyeimbangan data seperti teknik pengambilan sampel berlebih minoritas sintetis ditambah dengan *Tomek-link* (*SMOTETomek*) dapat membantu meningkatkan kinerja model ML pada kelas minoritas.

Pengembangan model klasifikasi yang lebih akurat dan andal untuk mendeteksi serta memprediksi penyakit diabetes diharapkan dapat memberikan manfaat signifikan bagi berbagai kalangan masyarakat, mulai dari tenaga medis hingga individu yang berisiko tinggi. Model klasifikasi seperti implementasi algoritma *XGBOOST* yang lebih canggih akan memungkinkan identifikasi dini terhadap individu yang berpotensi mengidap diabetes, sehingga intervensi medis dapat dilakukan lebih cepat dan tepat sasaran. *XGBOOST* adalah algoritma pembelajaran ensemble yang menawarkan sejumlah keunggulan, termasuk fleksibilitas tinggi, prediktabilitas yang kuat, kemampuan generalisasi yang kuat, skalabilitas tinggi, efisiensi pelatihan model yang tinggi, dan ketahanan yang tinggi [6]. Oleh karena itu, perawatan yang lebih baik dan terarah dapat mengurangi risiko penyakit jantung, kerusakan ginjal, dan gangguan penglihatan.

Selain itu, model klasifikasi yang andal juga diharapkan dapat berfungsi sebagai alat bantu bagi para profesional kesehatan untuk mengantisipasi perkembangan penyakit pada pasien. Penelitian dalam bidang kecerdasan buatan, khususnya dalam bidang medis, sedang mengalami perkembangan yang cepat dan dinamis. Contoh yang menonjol dari bidang yang berkembang ini adalah *Computer Aided Diagnosis* (CAD). CAD mewakili integrasi yang signifikan dari kemajuan dalam ilmu komputer dengan teknologi medis di bidang medis [7]. Dengan teknologi ini, para dokter dan tenaga kesehatan dapat menggunakan hasil klasifikasi yang lebih akurat sebagai dasar dalam merancang rencana perawatan yang lebih *personalized*, yang disesuaikan dengan kebutuhan dan kondisi spesifik setiap pasien. Hal ini akan sangat berguna dalam mencegah perkembangan diabetes.

Sistem klasifikasi yang dikembangkan dalam penelitian ini dirancang untuk membantu masyarakat dalam mengenali penyakit [8]. Sebaliknya, masyarakat umum juga akan mendapat manfaat dari model klasifikasi ini, terutama dalam hal meningkatkan kesadaran dan edukasi tentang pentingnya manajemen risiko dan deteksi dini. Teknologi yang semakin berkembang ini memungkinkan penggunaan alat-alat diagnostik berbasis data yang lebih terjangkau dan dapat diakses secara luas, yang pada akhirnya akan membantu mengurangi beban kesehatan akibat diabetes di Indonesia. Dalam jangka panjang, pengembangan model klasifikasi yang lebih akurat ini diharapkan tidak hanya menurunkan prevalensi diabetes, tetapi juga meningkatkan kualitas hidup individu yang berisiko atau sudah terdiagnosis dengan penyakit ini.

1.2. Rumusan Masalah

Berdasarkan penjabaran pada latar belakang, diperoleh rumusan masalah sebagai berikut:

1. Bagaimana cara mengimplementasikan teknik *SMOTETomek* untuk *balancing data* dalam klasifikasi penyakit diabetes menggunakan algoritma *XGBoost*
2. Bagaimana hasil implementasi teknik *SMOTETomek* untuk *balancing data* dalam klasifikasi penyakit diabetes menggunakan algoritma *XGBoost*

1.3. Tujuan Penelitian

Tujuan dari penelitian ini mengacu pada rumusan masalah yang telah diidentifikasi, yakni sebagai berikut:

1. Terbentuknya model implementasi teknik *SMOTETomek* untuk *balancing data* dalam klasifikasi penyakit diabetes menggunakan algoritma *XGBoost*.
2. Mengetahui hasil implementasi teknik *SMOTETomek* untuk *balancing data* dalam klasifikasi penyakit diabetes menggunakan algoritma *XGBoost*.

1.4. Manfaat Penelitian

Manfaat yang diharapkan sesuai dengan tujuan penelitian yang telah ditetapkan, adalah sebagai berikut:

1. Menghasilkan model klasifikasi yang lebih akurat untuk deteksi penyakit diabetes yang dapat diimplementasikan dalam sistem pendukung keputusan medis.
2. Membantu tenaga medis dalam melakukan diagnosis dini penyakit diabetes melalui model prediksi yang lebih andal.
3. Memperkaya literatur ilmiah tentang efektivitas kombinasi teknik *SMOTETomek* dengan algoritma *XGBoost* dalam penanganan data tidak seimbang pada kasus medis.

1.5. Batasan Masalah

Dalam penelitian ini, terdapat batasan yang perlu diperhatikan untuk memahami cakupan dan kendala penelitian. Batasan-batasan tersebut meliputi sebagai berikut:

1. Penelitian ini menggunakan set data 2.000 yang dikumpulkan melalui Kaggle..
2. Dalam penelitian ini, teknik *SMOTETomek* akan digunakan untuk *Balancing data*.
3. Klasifikasi menggunakan 2 kelas, yaitu Iya dan Tidak.
4. Hasil dari penelitian ini merupakan hasil akurasi dari model implementasi teknik *SMOTETomek* untuk *balancing data* dalam klasifikasi penyakit diabetes menggunakan algoritma *XGBoost*.

Halaman ini sengaja dikosongkan