

BAB I

PENDAHULUAN

I.1 Latar Belakang

Preeklampsia adalah terjadinya hipertensi dan proteinuria yang terjadi setelah usia kehamilan 20 minggu pada pasien yang sebelumnya memiliki tekanan darah normal (Rana et al., 2019). Preeklampsia merupakan kondisi hipertensi yang menjadi penyebab utama kematian ibu, menyumbang 70.000 kematian ibu dan 500.000 bayi yang dilaporkan setiap tahun. *International Society for the Study of Hypertension in Pregnancy* (ISSHP) merekomendasikan agar wanita hamil dengan hipertensi didiagnosis preeklampsia (Fox et al., 2019). Gangguan Hipertensi Kehamilan (HDP) dan khususnya preeklampsia adalah penyebab utama kematian di dunia (Navaratnam et al., 2013).

Preeklampsia merupakan salah satu penyumbang utama kematian ibu secara global (Zhang et al., 2019). Meskipun insiden bervariasi dari satu daerah ke daerah lain, preeklampsia menyumbang 40% kematian di beberapa negara. Oleh karena itu Preeklampsia harus dideteksi dan ditangani dengan tepat sebelum timbulnya kejang (eklampsia) dan komplikasi lain yang mengancam jiwa. Preeklampsia saat ini tetap menjadi salah satu penyebab utama kematian ibu, meskipun prevalensinya masih diremehkan di beberapa tempat karena tidak dilaporkan. Oleh karena itu para profesional kesehatan harus tahu bagaimana menangani dan mengambil tindakan pencegahannya (Mayrink et al., 2018).

Preeklampsia terjadi secara mendadak dan sering tanpa disadari. Hal ini diperburuk dengan fenomena bahwa sebagian besar ibu hamil dengan preeklampsia datang ke pusat pelayanan kesehatan dalam kondisi yang sudah berat (Suprihatin & Wuryaningsih, 2019). Untuk mengurangi kematian akibat penyakit ini, pencegahan menjadi tujuan utama (Mikat et al., 2012).

Usaha untuk menyelesaikan masalah preeklampsia telah dilakukan, misalnya menggunakan parameter klinis seperti tekanan darah dan *protein uriene* sebagai salah satu fitur untuk dilakukan penelitian. Di beberapa negara telah melakukan

penelitian untuk menekan jumlah preeklamsia dengan berbagai cara seperti dengan mengembangkan sistem deteksi dini menggunakan pembelajaran mesin (Marić et al., 2020). Peneliti lain mengembangkan model menggunakan data rekam medis elektronik rumah sakit berbasis pembelajaran mesin dengan metode statistik konvensional (Jhee et al., 2019), dan pengembangan sistem monitoring tekanan darah (*blood pressure*) untuk preeklamsia (Musyoka et al., 2019), prediksi preeklamsia menggunakan *small-for-gestational-age (SGA) screening* (J. Zhang et al., 2019) dan juga menggunakan logika *fuzzy* dan *wearable devices* (Macarena et al., 2017), selain itu pengembangan model prediksi dengan menggunakan dataset BPJS dengan membandingkan beberapa model klasifikasi (Sufriyana et al., 2020)

Kelemahan dari penelitian yang sudah dilakukan oleh para peneliti sebelumnya adalah belum mempertimbangkan masalah kelas yang tidak seimbang (*imbalance class*).

Pembelajaran mesin telah mencapai peningkatan yang signifikan dalam banyak aplikasi pembelajaran mesin, seperti deteksi objek, klasifikasi gambar, diagnosis medis, kendaraan otonom, pemulihan gambar, dan segmentasi gambar. Pembelajaran mesin membutuhkan daya komputasi yang signifikan dan data yang banyak untuk melatih model. Dengan kata lain, peningkatan signifikan oleh teknologi telah dicapai pembelajaran mesin dengan kekuatan data. Namun, mendapatkan kumpulan data besar merupakan tantangan dalam aplikasi dunia nyata, terutama jika berkaitan dengan area baru dalam pembelajaran mesin. Untuk melatih pembelajaran mesin, kuantitas data sama pentingnya dengan kualitas data karena pelatihan dengan hanya kumpulan data kecil sering mengakibatkan penurunan kinerja pembelajaran mesin.

Belajar dari kumpulan data kecil masih menjadi tantangan dan menjadi masalah yang lebih sulit ketika ditambah dengan masalah menantang lainnya, misalnya ketidakseimbangan kelas, variasi dalam kelas, dan kesamaan antar kelas. Kinerja algoritma pembelajaran mesin menurun dalam kondisi data yang tidak seimbang karena hasil klasifikasi menunjukkan bias pada sebagian besar kelas data. Selain itu, terjadinya *overfitting* ditunjukkan dengan akurasi yang tinggi pada data pelatihan (*training*), namun pada saat pengujian (*testing*) dapat menunjukkan perbedaan akurasi. Oleh karena itu, metode teknik resampling digunakan yang

dapat mencegah terjadinya *overfitting*. Namun, bukan hanya ketidakseimbangan kelas dan *overfitting* yang berkontribusi terhadap penurunan kinerja model. Ada juga masalah tumpang tindih (*overlapping*) yang perlu dipertimbangkan.

I.2 Identifikasi Masalah

Beberapa penelitian telah menerapkan pembelajaran mesin untuk diagnosis medis, namun tidak menggunakan seleksi fitur untuk memilih fitur yang paling relevan untuk melatih model, yang dapat meningkatkan kinerja model, dan juga tidak menggunakan *parameter tuning* dan teknik resampling terutama ketika menangani masalah ketidakseimbangan kelas. Satu masalah penting saat menerapkan pembelajaran mesin adalah biasanya adanya kelas tidak seimbang hal ini menyebabkan model kinerjanya buruk (Drosou dan Koukouvinos, 2017).

Berdasarkan uraian di atas ada beberapa permasalahan penelitian yang bisa kita jabarkan:

- 1 Model pembelajaran mesin tidak dapat diandalkan untuk menangani kasus kelas tidak seimbang (Haixiang et al., 2017).
- 2 Memiliki kinerja yang sangat baik pada kumpulan kelas yang seimbang, tapi kinerja yang buruk pada kelas yang tidak seimbang, terutama untuk deteksi kategori minoritas (Nwe & Lynn, 2020).
- 3 Biasanya mengasumsikan kesamaan sampel di setiap kelas (Xu et al., 2020).

I.3 Perumusan Masalah

Dan berdasarkan pernyataan masalah di atas maka dapat dibuat perumusan masalah sebagai berikut:

- 1 Bagaimana mengembangkan model pembelajaran mesin yang andal untuk menangani dataset dengan kelas tidak seimbang?
- 2 Bagaimana meningkatkan kinerja model dalam mendeteksi kategori minoritas pada dataset yang tidak seimbang? Berdasarkan penelitian oleh Nwe & Lynn (2020), model pembelajaran mesin sering menunjukkan kinerja yang buruk pada kategori minoritas ketika berhadapan dengan dataset tidak seimbang.

- 3 Bagaimana mengatasi asumsi kesamaan jumlah sampel antar kelas dalam model pembelajaran mesin pada dataset tidak seimbang?

Rumusan masalah ini menjadi dasar untuk penelitian yang lebih mendalam tentang penanganan kelas tidak seimbang dalam pembelajaran mesin. Oleh karena itu dibutuhkan model yang dapat menangani kelas tidak seimbang (Chawla et al., 2004).

I.4 Maksud Dan Tujuan Penelitian

1. Mengembangkan model pembelajaran mesin untuk mengatasi kelas yang tidak seimbang menggunakan seleksi fitur dan teknik *resampling*.
2. Mengembangkan model untuk mengatasi kelas yang tidak seimbang pada kasus preeklampsia menggunakan kombinasi seleksi fitur, teknik *resampling*, dan pembelajaran mesin.

I.5 Kebaruan Penelitian

Penelitian ini menggunakan model KNN dengan *parameter tuning* dan dengan menggunakan seleksi fitur dan teknik *resampling* yang digunakan untuk pemilihan fitur dan menyeimbangkan data sehingga dapat menghasilkan klasifikasi yang akurat. Pengembangan FHR-KNN untuk meningkatkan kinerja klasifikasi pada kelas tidak seimbang. Disertasi ini memperkenalkan metode baru yaitu FHR-KNN, yang dirancang untuk mengatasi tantangan ketidakseimbangan kelas dalam pembelajaran mesin, khususnya pada kasus preeklampsia.

1. Optimalisasi pembelajaran mesin melalui *parameter tuning*. Disertasi ini juga berkontribusi pada optimalisasi pembelajaran mesin menggunakan *parameter tuning* agar kinerjanya lebih optimal.
2. Penerapan seleksi fitur dalam pengolahan data (*data preprocessing*) untuk memilih fitur yang relevan agar efisien sehingga model lebih akurat. Penelitian ini menunjukkan pentingnya seleksi fitur dalam tahap *preprocessing*, yang secara signifikan mengurangi dimensi data dan meningkatkan akurasi.

3. Pengembangan model untuk mengatasi masalah kelas yang tidak seimbang dengan kombinasi *parameter tuning*, seleksi fitur, teknik *resampling*, dan pembelajaran mesin.

1.6 Manfaat Penelitian

Manfaat dari penelitian ini antara lain:

1. Peningkatan akurasi klasifikasi dini preeklampsia.
2. Penelitian ini bertujuan untuk mengembangkan model pembelajaran mesin yang lebih akurat dalam mengklasifikasi preeklampsia, terutama dalam kondisi kelas yang tidak seimbang. Hal ini akan berkontribusi pada peningkatan kualitas diagnosis dan pencegahan dini kondisi preeklampsia.
3. Mengatasi tantangan ketidakseimbangan kelas.
 Dengan fokus pada pengolahan kelas yang tidak seimbang, penelitian ini akan menyediakan solusi atau metode baru untuk mengatasi salah satu tantangan utama dalam data kesehatan, yaitu ketidakseimbangan kelas. Ini akan membuka jalan bagi penelitian serupa di bidang kesehatan lainnya.
4. Kontribusi pada bidang kesehatan ibu dan anak.
 Dengan mengembangkan metode yang lebih efektif dalam mendeteksi preeklampsia, penelitian ini akan berkontribusi pada peningkatan kesehatan ibu dan anak, mengurangi risiko komplikasi yang serius yang dapat terjadi akibat preeklampsia.
5. Pendekatan berbasis data dalam pengambilan keputusan medis.
 Penelitian ini akan menjelaskan bagaimana analisis data berbasis pembelajaran mesin dapat digunakan untuk membuat keputusan medis yang lebih tepat, khususnya dalam konteks kondisi kesehatan yang kompleks seperti preeklampsia.
6. Inovasi dalam teknologi Kesehatan.
 Hasil dari penelitian ini dapat membuka peluang untuk pengembangan alat atau aplikasi kesehatan yang inovatif, yang dapat digunakan oleh praktisi kesehatan untuk meningkatkan efektivitas dan efisiensi dalam mendiagnosis dan mengelola preeklampsia.

Penelitian ini tidak hanya penting dalam konteks akademis dan ilmiah, tetapi juga memiliki implikasi praktis yang signifikan dalam meningkatkan kualitas perawatan kesehatan bagi pasien.

1.7 Pembatasan Masalah

Untuk memastikan fokus penelitian dan menghindari ruang lingkup yang terlalu luas, penelitian ini dibatasi pada:

1. Dataset yang digunakan adalah dataset medis dengan fokus pada kasus preeklampsia.
2. Teknik yang digunakan untuk penanganan kelas tidak seimbang mencakup *oversampling*, *undersampling*, dan *hybrid resampling*.
3. Model pembelajaran mesin yang digunakan dalam penelitian ini adalah *K-Nearest Neighbor* (KNN) dengan kombinasi *parameter tuning* dan seleksi fitur.
4. Evaluasi kinerja model akan dilakukan berdasarkan metrik *accuracy*, *precision*, dan *sensitivity (recall)*.
5. Penelitian dilakukan menggunakan data terstruktur dengan fitur-fitur yang relevan, seperti tekanan darah, usia kehamilan, dan parameter klinis lainnya.

1.8 Sistematika Penulisan

Untuk memudahkan pemahaman, sistematika penulisan dalam penelitian ini disusun sebagai berikut:

Bab I : Pendahuluan

Bab ini membahas latar belakang penelitian, identifikasi masalah, perumusan masalah, maksud dan tujuan penelitian, kebaruan penelitian, manfaat penelitian, serta pembatasan masalah dan sistematika penulisan. Pendahuluan memberikan gambaran umum terkait pentingnya penelitian ini dalam mengatasi masalah ketidakseimbangan kelas pada dataset medis preeklampsia.

Bab II : Landasan Teori

Bab ini menjelaskan teori-teori yang mendukung penelitian, meliputi konsep preeklampsia, dasar-dasar pembelajaran mesin, masalah kelas tidak seimbang, algoritma *K-Nearest Neighbor* (KNN), teknik seleksi fitur, teknik resampling, serta

parameter tuning. Tinjauan pustaka dari penelitian terdahulu juga disajikan untuk menunjukkan perkembangan penelitian terkait.

Bab III : Metode Penelitian

Bab ini berisi penjelasan metodologi yang digunakan dalam penelitian, mencakup bahan dan alat, prosedur penelitian, eksplorasi data awal, preprocessing data, seleksi fitur, penerapan hybrid resampling, hyperparameter tuning, dan evaluasi kinerja model. Setiap tahapan dijelaskan secara sistematis agar penelitian dapat direplikasi.

Bab IV : Hasil Penelitian dan Pembahasan

Bab ini menyajikan hasil penelitian yang diperoleh dari eksperimen yang dilakukan. Analisis dilakukan terhadap hasil eksplorasi data, seleksi fitur, penerapan teknik resampling, dan tuning parameter pada model KNN. Perbandingan kinerja metode tradisional dengan metode resampling dan metode ensemble turut dibahas secara mendalam.

Bab V : Kesimpulan dan Saran

Bab ini berisi kesimpulan dari hasil penelitian yang telah dilakukan, yang menjawab rumusan masalah penelitian. Selain itu, disampaikan saran-saran untuk penelitian selanjutnya yang diharapkan dapat mengembangkan metode atau model yang lebih optimal dalam menangani data medis dengan kelas tidak seimbang.

Daftar Pustaka

Berisi daftar referensi yang digunakan dalam penelitian ini, baik dari jurnal ilmiah, buku, maupun sumber terpercaya lainnya yang relevan dengan topik penelitian.

Lampiran

Bagian ini memuat data tambahan yang mendukung hasil penelitian, seperti hasil pengujian model, dokumentasi tahapan penelitian, serta informasi lain yang relevan sebagai pendukung naskah utama.