

BAB 2

TINJAUAN PUSTAKA

2.1 Kajian Pustaka

2.1.1 Penyakit Tidak Menular

Penyakit tidak menular (PTM) merupakan kelompok penyakit yang tidak disebabkan oleh patogen (virus, bakteri, atau jamur) dan tidak dapat ditularkan dari satu individu ke individu lain [20]. PTM mencakup berbagai penyakit kronis seperti penyakit kardiovaskular, diabetes mellitus, kanker, dan penyakit pernapasan kronis. PTM menjadi masalah kesehatan utama di banyak negara, termasuk Indonesia, karena menyebabkan angka kematian yang tinggi dan mempengaruhi kualitas hidup penderitanya. PTM menyumbang lebih dari 70% dari seluruh kematian global, atau sekitar 41 juta orang setiap tahunnya [20]. Penyakit ini tidak hanya berdampak pada kualitas hidup individu tetapi juga memberikan beban ekonomi yang besar karena biaya perawatan kesehatan dan hilangnya produktivitas tenaga kerja [21]. Sebagian besar PTM disebabkan oleh faktor risiko yang dapat dimodifikasi seperti pola makan yang buruk, kurangnya aktivitas fisik, penggunaan tembakau, dan konsumsi alkohol yang berlebihan [22]. Faktor-faktor ini dapat diintervensi melalui pendekatan kesehatan masyarakat yang efektif untuk mengurangi prevalensi PTM.

2.1.1.1 Hipertensi

Hipertensi, atau tekanan darah tinggi, adalah gangguan hemodinamik yang ditandai dengan peningkatan resistensi pembuluh darah perifer, yang dapat menyebabkan penyakit serius seperti infark miokard, gagal ginjal, stroke, dan bahkan kematian jika tidak didiagnosis dan diobati dengan benar [23]. Secara global, hipertensi adalah penyebab utama penyakit kardiovaskular dan kematian dini. Penggunaan obat antihipertensi secara luas telah membantu menjaga rata-rata tekanan darah global tetap konstan atau sedikit menurun dalam empat dekade terakhir. Namun, prevalensi hipertensi meningkat, terutama di negara-negara berpenghasilan rendah dan

menengah. Diperkirakan bahwa 31,1% orang dewasa (1,39 miliar) di seluruh dunia mengalami hipertensi pada tahun 2010. Prevalensi hipertensi di negara-negara berpenghasilan rendah dan menengah adalah 31,5% dibandingkan dengan 28,5% di negara berpenghasilan tinggi [23].

Hipertensi dapat dikategorikan menjadi beberapa jenis seperti hipertensi esensial (EHT), hipertensi paru (PHT), dan hipertensi arteri paru (PAHT). Hipertensi esensial adalah jenis yang paling umum dan mencakup 95% dari semua kasus [24]. Diagnosis hanya dapat dilakukan melalui pengukuran tekanan darah. Pasien dengan hipertensi sering mengalami gejala seperti sakit kepala, muntah, pusing, dan mimisan lebih sering [25]. Selain faktor genetik dan lingkungan, patofisiologi hipertensi juga melibatkan faktor-faktor multifaktorial seperti disregulasi sistem renin-angiotensin-aldosteron, aktivasi sistem saraf simpatik, dan stres oksidatif. Semua faktor ini mempengaruhi kontraksi otot polos vaskular dan menyebabkan peningkatan resistensi vaskular yang berkontribusi pada perkembangan hipertensi [26].

2.1.1.2 Faktor Risiko Penyakit Tidak Menular

Penyakit tidak menular muncul dari kombinasi faktor risiko yang tidak dapat dimodifikasi dan faktor risiko yang dapat dimodifikasi. Faktor risiko yang tidak dapat dimodifikasi oleh individu adalah usia, jenis kelamin, dan genetika. Sedangkan faktor risiko yang dapat dimodifikasi adalah faktor yang dapat diubah melalui kesadaran individu itu sendiri dan intervensi sosial. Faktor-faktor yang dapat dimodifikasi tersebut adalah:

1. Merokok

Efek berbahaya dari merokok terhadap kematian yang disebabkan oleh kanker, penyakit kardiovaskuler, dan penyakit pernapasan kronis telah lama diketahui. Selain itu, paparan asap rokok pada perokok pasif seperti ibu hamil, anak-anak, dan orang dewasa yang tidak hamil di rumah maupun di tempat-tempat umum menyebabkan hasil kelahiran yang merugikan, penyakit pernapasan pada masa kanak-kanak, dan penyakit lainnya seperti yang diderita oleh perokok aktif. Penelitian di Vietnam menunjukkan bahwa

lebih dari 50% pria adalah perokok aktif, sementara prevalensi di kalangan wanita jauh lebih rendah yaitu 1,5% [27]. Selain pergeseran pola prevalensi merokok, telah terjadi perubahan dalam jenis rokok yang tersedia, seperti rokok rendah tar dan rokok elektrik. Namun, hasil tinjauan menyimpulkan bahwa selama lima dekade desain rokok berkembang tidak mengurangi risiko penyakit di kalangan perokok. Satu-satunya tindakan yang efektif untuk mencegah bahaya merokok adalah dengan pencegahan dan penghentian merokok.

2. Konsumsi Alkohol

Alkohol merupakan zat psikoaktif dengan memproduksi substansi yang membuat ketergantungan pengkonsumsinya. Dampak alkohol ditentukan oleh volume alkohol yang dikonsumsi, pola minum, dan kualitas alkohol yang dikonsumsi. Pada tahun 2012, sekitar 3.3 juta kematian, atau sekitar 5.9% dari seluruh kematian global disebabkan oleh konsumsi alkohol. Konsumsi Alkohol sangat umum di seluruh dunia meskipun membawa risiko yang merugikan bagi kesehatan dan konsekuensi sosial terkait efek memabukkan, sifat beracun, dan ketergantungan. Konsumsi alkohol merupakan faktor risiko utama untuk beban penyakit di negara berkembang berkaitan dengan berbagai penyakit dan cedera, termasuk kecelakaan lalu lintas, kekerasan, dan bunuh diri. Secara keseluruhan, 5.1% dari beban penyakit global dan cedera disebabkan oleh alkohol (diukur dalam Disability-Adjusted Life Years, DALYs). Konsumsi alkohol yang berlebihan tidak hanya meningkatkan risiko cedera secara substansial, tetapi juga memperburuk penyakit kardiovaskuler dan hati. Konsumsi alkohol terus meningkat di Jepang, Cina, dan banyak negara lain di Asia yang sebelumnya rendah.

Faktor lingkungan meliputi pembangunan, ekonomi, budaya, ketersediaan alkohol, serta kelengkapan tingkat pelaksanaan dan penegakkan kebijakan alkohol mempengaruhi pola konsumsi alkohol dan besarnya masalah yang berhubungan dengan alkohol dalam populasi.

3. Pola Makan yang Buruk

Sekitar 16 juta (1%) DALYs (ukuran potensial kehilangan kehidupan karena kematian dini dan tahun-tahun produktif yang hilang karena cacat) dan 1.7 juta (2.8%) dari kematian di seluruh dunia disebabkan oleh kurangnya konsumsi buah dan sayur. Konsumsi cukup buah dan sayur mengurangi risiko penyakit kardiovaskular, kanker perut, dan kanker kolorektal. Konsumsi makanan tinggi kalori seperti makanan olahan yang tinggi lemak dan gula cenderung menyebabkan obesitas dibandingkan makanan rendah kalori seperti buah dan sayuran.

Jumlah garam yang dikonsumsi merupakan faktor penentu penting dari tingkat tekanan darah dan risiko kardiovaskuler secara keseluruhan. Diperkirakan bahwa mengurangi asupan garam dari konsumsi rata-rata 9-12 gram per hari menjadi 5 gram per hari memiliki dampak besar pada tekanan darah dan penyakit kardiovaskuler.

Konsumsi makanan tinggi lemak jenuh dan trans fatty acid terkait dengan penyakit jantung; minyak nabati tak jenuh ganda dapat menjadi pengganti untuk menurunkan risiko penyakit jantung coroner dan diabetes mellitus tipe 2.

4. Kurangnya Aktivitas Fisik

Aktivitas fisik yang tidak memadai merupakan satu dari sepuluh faktor risiko utama kematian global. Orang yang kurang aktif secara fisik memiliki 20%- 30% peningkatan faktor risiko penyebab kematian dibandingkan dengan mereka yang setidaknya melakukan aktivitas fisik selama 150 menit per minggu, atau setara seperti yang direkomendasikan WHO.

Aktivitas fisik yang teratur mengurangi risiko penyakit jantung iskemik, diabetes, kanker payudara, dan kanker kolon. Selain itu, aktivitas yang cukup mengurangi risiko stroke, hipertensi, dan depresi. Aktivitas fisik juga merupakan penentu utama dari pengeluaran energi dan dengan demikian penting untuk keseimbangan energi dan control berat badan.

Empat perilaku umum diatas (merokok, konsumsi alkohol, pola makan yang buruk, dan kurangnya aktivitas fisik) menyebabkan gangguan metabolik berupa peningkatan tekanan darah, kelebihan berat badan/obesitas, tingginya kadar glukosa darah, dan peningkatan kadar kolesterol yang berpengaruh terhadap kejadian penyakit tidak menular.

5. Peningkatan Tekanan Darah

Peningkatan tekanan darah merupakan faktor risiko utama untuk penyakit jantung koroner, iskemik, dan stroke hemoragik. Tingkat tekanan darah telah terbukti berhubungan dengan risiko tersebut. Dikatakan dalam beberapa kelompok usia, setiap kenaikan 20/10 mmHg tekanan darah, mulai dari 115/75 mmHg meningkatkan risiko dua kalilipat terkena penyakit kardiovaskuler. Selain penyakit jantung koroner, iskemik, dan stroke, komplikasi peningkatan tekanan darah dapat menyebabkan gagal jantung, penyakit pembuluh darah perifer, gangguan ginjal, dan gangguan penglihatan. Mengontrol tekanan darah sampai kurang dari 140/90 mmHg dikaitkan dengan penurunan komplikasi kardiovaskuler.

6. Kelebihan Berat Badan

Obesitas memiliki efek metabolik yang buruk pada tekanan darah, kolesterol, trigliserida, dan resistensi insulin. Risiko penyakit jantung koroner, stroke iskemik, dan diabetes mellitus tipe 2 terus meningkat seiring dengan meningkatnya indeks massa tubuh (IMT). IMT yang meningkat juga meningkatkan risiko kanker payudara, kanker kolon, kanker prostat, kanker endometrium, kanker ginjal, dan kanker hati. Untuk mencapai kesehatan optimal, IMT rata-rata untuk populasi dewasa harus berada pada isaran 21-23 kg/m², sedangkan bagi individu harus menjaga IMT dalam kisaran 18.5-24.9 kg/m². Terdapat peningkatan risiko penyakit penyerta untuk orang dengan IMT 25-29.9 kg/m² dan komorbiditas yang parah untuk IMT lebih dari 30 kg/m².

7. Kadar Glukosa Darah yang Tinggi

Diabetes bertanggung jawab untuk kematian 1,5 juta jiwa pada tahun 2012 dan 89 juta DALYs. Toleransi glukosa yang terganggu, dan gangguan gula darah puasa adalah kategori risiko untuk diabetes dan penyakit kardiovaskuler. Orang dengan diabetes memiliki risiko dua kali lipat terkena stroke. Diabetes juga menyebabkan kegagalan ginjal pada banyak populasi. Amputasi tungkai bawah meningkat 10 kali lebih umum pada orang dengan diabetes. Diabetes juga merupakan penyebab utama gangguan penglihatan dan kebutaan. Prevalensi hiperglikemi bergantung pada kriteria diagnostik epidemiologi, dikatakan nilai gula darah puasa ≥ 7.0 mmol/L (126 mg/dL) sudah cukup untuk mendiagnosis diabetes.

8. Peningkatan Kadar Kolesterol

Kadar kolesterol yang tinggi meningkatkan risiko penyakit jantung dan stroke. Secara umum, sepertiga dari penyakit jantung iskemik disebabkan oleh kadar kolesterol yang tinggi. Kolesterol yang tinggi diperkirakan menyebabkan 2.6 juta kematian (4.5% dari total kematian) dan 2.0% dari total DALYs.

2.1.2 Kementerian Kesehatan Republik Indonesia

Kementerian Kesehatan Republik Indonesia (Kemenkes RI) adalah lembaga pemerintah yang bertanggung jawab atas penyelenggaraan urusan kesehatan di Indonesia. Kemenkes RI memiliki misi untuk meningkatkan derajat kesehatan masyarakat melalui berbagai program kesehatan, kebijakan, dan regulasi. Lembaga ini bekerja untuk memastikan bahwa seluruh warga negara Indonesia memiliki akses yang adil dan merata ke layanan kesehatan yang berkualitas.

Kemenkes RI memainkan peran penting dalam merumuskan kebijakan kesehatan nasional, mengatur standar layanan kesehatan, mengawasi pelaksanaan program kesehatan, serta memberikan dukungan teknis dan finansial kepada fasilitas kesehatan di seluruh Indonesia. Kementerian ini juga bertanggung jawab atas berbagai program kesehatan nasional, termasuk imunisasi, penanggulangan penyakit menular,

kesehatan ibu dan anak, gizi masyarakat, serta pengendalian penyakit tidak menular (PTM).

2.1.2.1 Pos Pembinaan Terpadu Penyakit Tidak Menular

Pos Pembinaan Terpadu Penyakit Tidak Menular (Posbindu PTM) adalah salah satu program yang diinisiasi oleh Kemenkes RI untuk mengendalikan PTM di tingkat komunitas. Program ini bertujuan untuk meningkatkan kesadaran masyarakat tentang pentingnya pencegahan dan pengendalian PTM seperti diabetes, hipertensi, kanker, dan penyakit jantung. Melalui Posbindu PTM, masyarakat dapat menjalani pemeriksaan kesehatan rutin untuk mendeteksi dini PTM serta mendapatkan edukasi tentang gaya hidup sehat.

Kegiatan di Posbindu PTM meliputi pengukuran tekanan darah, kadar gula darah, berat badan, tinggi badan, lingkar perut, dan kadar kolesterol. Selain itu, Posbindu PTM memberikan penyuluhan mengenai pola makan sehat, pentingnya aktivitas fisik, bahaya merokok, dan manajemen stres. Program ini didukung oleh kader kesehatan yang terlatih serta tenaga kesehatan dari puskesmas setempat, yang bekerja sama untuk memberikan layanan dan edukasi kepada masyarakat. Kemenkes RI menyediakan pelatihan, alat pemeriksaan, serta melakukan evaluasi dan monitoring untuk memastikan kualitas layanan di Posbindu PTM.

2.1.3 Data Mining

Knowledge Discovery in Database (KDD) atau penemuan pengetahuan melalui database ditandai dengan kemajuan dalam strategi dan prosedur untuk memanfaatkan informasi. Proses dalam penemuan pengetahuan melalui database diantaranya adalah Data Mining[15]. Secara teknis, data mining adalah proses yang memanfaatkan teknik-teknik statistik, matematika, dan kecerdasan buatan untuk mengekstrak dan mengidentifikasi informasi dan knowledge selanjutnya (atau pola-pola) yang berasal dari sekumpulan data yang sangat besar. Berbagai macam pola tersebut bisa dalam bentuk aturan bisnis, kesamaan-kesamaan, korelasi, trend, atau model-model prediksi. Kebanyakan literatur mendefinisikan data mining sebagai “proses yang rumit untuk mengidentifikasi pola-pola yang valid, baru, memiliki potensi bermanfaat, dan bisa

dipahami, terhadap data yang disimpan di dalam database yang terstruktur”, dimana data diorganisir dalam baris-baris yang terstruktur menurut kategori, ordinal/berurutan, dan variable-variabel yang berkesinambungan.

2.1.3.1 Data Mining dalam Dunia Kesehatan

Data mining telah digunakan dalam berbagai aplikasi seperti pemasaran, manajemen hubungan pelanggan, teknik, dan analisis obat-obatan, prediksi ahli, penambangan web, dan komputasi seluler. Akhir-akhir ini, data mining telah berhasil diterapkan dalam penipuan perawatan kesehatan dan mendeteksi kasus penyalahgunaan. Keputusan klinis sering dibuat berdasarkan intuisi dan pengalaman dokter daripada berdasarkan data yang kaya akan pengetahuan yang tersembunyi dalam database. Praktik ini menimbulkan bias yang tidak diinginkan, kesalahan dan biaya medis yang berlebihan yang mempengaruhi kualitas pelayanan yang diberikan kepada pasien. Data mining memiliki potensi untuk menghasilkan lingkungan yang kaya pengetahuan yang dapat membantu meningkatkan kualitas keputusan klinis secara signifikan. Aplikasi data mining yang sukses telah memberikan dorongan bagi pihak-pihak terkait untuk sepenuhnya memanfaatkannya karena mereka telah menyadari bahwa data mining sangat penting dalam perolehan informasi berharga untuk semua sektor yang terlibat dalam industri terkait perawatan kesehatan [28].

2.1.3.2 *Random Forest*

Random Forest adalah salah satu algoritma pembelajaran mesin yang digunakan untuk tugas klasifikasi, regresi, dan pengelompokan data. *Random Forest* termasuk dalam kategori *ensemble learning*, yang berarti ia menggabungkan prediksi dari beberapa model untuk meningkatkan performa dan kestabilan. *Random Forest* bekerja dengan membangun banyak pohon keputusan secara simultan. Setiap pohon dibangun dari subset acak dengan cara melakukan *bootstrapping* dari dataset pelatihan, dan hanya menggunakan sebagian kecil fitur yang dipilih secara acak [23]. Dengan demikian, *Random Forest* dapat mencapai variasi di antara pohon-pohon untuk meningkatkan keakuratan dan kestabilan model.

2.1.3.3 *Gradient Boosting*

Gradient Boosting adalah algoritma pembelajaran mesin yang termasuk dalam kategori *ensemble learning*, mirip seperti *Random Forest*. Namun, berbeda dengan *Random Forest* yang membangun beberapa pohon keputusan secara independen, *Gradient Boosting* membangun pohon-pohon keputusan secara sekuensial, di mana setiap pohon dibangun untuk mengatasi kesalahan prediksi pohon sebelumnya [25], [26].

Cara kerja *Gradient Boosting* dimulai dengan membangun pohon pertama, yang memberikan prediksi awal. Selanjutnya, algoritma menghitung kesalahan prediksi pada data pelatihan, yang disebut sebagai residual. Pohon berikutnya kemudian dibangun untuk memprediksi residual dari model sebelumnya, mencoba mengoreksi kesalahan tersebut. Proses ini berlanjut dengan pembangunan pohon-pohon tambahan, masing-masing berfokus pada residual dari prediksi sebelumnya. Prediksi dari semua pohon dijumlahkan, dan kontribusi setiap pohon dapat diatur oleh *learning rate*, yang mengontrol seberapa besar pengaruh setiap pohon terhadap model akhir. Dengan membangun secara iteratif dan mengoreksi kesalahan prediksi sebelumnya, *Gradient Boosting* menghasilkan model *ensemble* yang kuat dengan kemampuan adaptasi yang baik terhadap data pelatihan. Proses ini dapat diulang hingga mencapai jumlah pohon yang ditentukan atau ketika tidak ada peningkatan yang signifikan dalam performa model.

2.1.3.4 *Extreme Gradient Boosting*

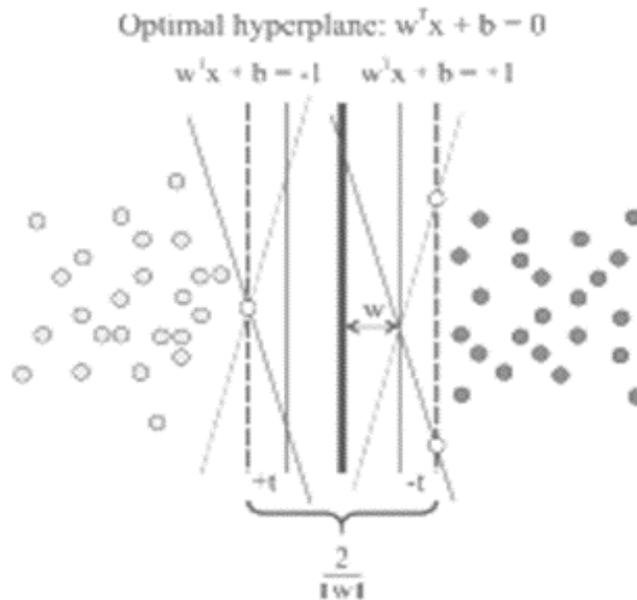
Extreme Gradient Boosting (XGBoost) adalah implementasi khusus dari algoritma *Gradient Boosting* yang dirancang untuk memberikan performa tinggi dan efisiensi komputasional. XGBoost memperkenalkan unsur regularisasi pada fungsi objektifnya untuk mengendalikan kompleksitas model dan mencegah *overfitting*. Regularisasi L1 (*Lasso*) dan L2 (*Ridge*) yang terintegrasi dalam XGBoost membantu mengontrol kompleksitas model dan mencegah *overfitting*, terutama ketika bekerja dengan data yang kompleks atau ketika jumlah fitur sangat besar. Meskipun *Gradient Boosting* juga dapat menggunakan parameter untuk mengurangi *overfitting*, namun

gradient boosting tidak memiliki regularisasi eksplisit seperti XGBoost, sehingga XGBoost sering kali lebih mampu menghasilkan model yang lebih generalis [29].

Keunggulan lain dari XGBoost adalah efisiensi komputasinya. Algoritma ini dirancang dengan optimasi tingkat rendah yang memungkinkan penggunaan sumber daya komputasi secara lebih efektif. XGBoost memanfaatkan pemrograman paralel dan teknik pruning yang canggih, yang mempercepat proses pelatihan tanpa mengorbankan akurasi [29]. Efisiensi ini menjadikan XGBoost pilihan yang lebih baik ketika bekerja dengan dataset yang besar dan kompleks, di mana kecepatan dan kinerja menjadi faktor penting.

2.1.3.5 Support Vector Machine

Support Vector Machine (SVM) merupakan teknik klasifikasi. Konsep dari SVM adalah memaksimalkan margin, yang merupakan jarak pemisah antara kelas data dengan mencari hyperplane terbaik [20],[21],[22]. Support Vector Machine (SVM) adalah metode untuk memisahkan (mendiskriminasi) kumpulan data ke dalam beberapa kelompok. Pemisahan kelompok dilakukan dengan membangun garis pemisah yang tegas yang memisahkan suatu kelompok dengan kelompok yang lain. Pada umumnya, garis pemisah tersebut adalah linier.



Gambar 2.1. Konsep SVM untuk mencari *optimal hyperplane* [30]

2.1.3.6 *Confusion Matrix*

Confusion matrix adalah alat penting yang digunakan untuk mengevaluasi kinerja model klasifikasi dengan cara membandingkan hasil prediksi model terhadap nilai sebenarnya dalam dataset pengujian [30]. Dalam kasus klasifikasi biner, confusion matrix terdiri dari empat komponen utama yaitu: *True Positive* (TP), *False Positive* (FP), *True Negative* (TN), dan *False Negative* (FN). TP mewakili jumlah prediksi benar untuk kelas positif, sementara TN mewakili prediksi benar untuk kelas negatif. Sebaliknya, FP adalah jumlah kasus di mana model salah memprediksi kelas positif, padahal seharusnya negatif, dan FN adalah jumlah kasus di mana model salah memprediksi kelas negatif, padahal seharusnya positif. Melalui matrix ini, kita bisa menghitung metrik evaluasi seperti akurasi, *precision*, *recall*, dan *F1-score*, yang memberikan gambaran tentang seberapa baik model mampu membedakan antara dua kelas tersebut.

Ketika diterapkan pada kasus multi kelas, *confusion matrix* diperluas menjadi tabel yang lebih besar, dengan jumlah baris dan kolom sesuai dengan jumlah kelas yang ada. Setiap elemen dalam matriks menunjukkan bagaimana model memprediksi suatu kelas

dibandingkan dengan kelas aktualnya. Diagonal utama dari matriks ini menggambarkan jumlah prediksi yang benar untuk masing-masing kelas, sementara elemen di luar diagonal utama menunjukkan jumlah kesalahan prediksi yang terjadi antar kelas. Metrik evaluasi seperti *precision*, *recall*, dan *F1-score* dapat dihitung secara individual untuk setiap kelas, memberikan wawasan yang lebih mendalam tentang performa model pada setiap kelas secara spesifik [31]. *Confusion matrix* multi kelas memungkinkan analisis yang lebih rinci terhadap kesalahan model, membantu mengidentifikasi kelas-kelas mana yang lebih sulit untuk diprediksi dengan benar oleh model. Dengan demikian, confusion matrix merupakan alat yang sangat berguna dalam memahami kinerja model klasifikasi, baik dalam skenario biner maupun multi kelas.

1. Akurasi merupakan adalah ukuran untuk mengukur ketepatan prediksi pengklasifikasi pada kelas tertentu.

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

2. Precision merupakan ukuran untuk mengukur ketepatan prediksi pengklasifikasi pada kelas tertentu.

$$Precision = \frac{TP}{TP + FP} \quad (5)$$

3. Recall adalah ukuran untuk mengukur berapa banyak data dari kelas tertentu yang dapat diprediksikan secara benar.

$$Recall = \frac{TP}{TP + FN} \quad (6)$$

4. Specificity adalah metrik evaluasi yang digunakan untuk mengukur kemampuan model dalam mengidentifikasi kelas negatif dengan benar.

$$Specificity = \frac{TN}{TN + FP} \quad (7)$$

5. F1-score menghitung rata-rata harmonis dari precision dan recall untuk setiap kelas, memberikan keseimbangan antara kedua metrik tersebut.

$$F1 - score = 2 \times \frac{precision \times recall}{precision + recall} \quad (8)$$

2.1.3.7 ROC-AUC

Receiver Operating Characteristic (ROC) dan *Area Under the Curve* (AUC) adalah alat yang sering digunakan untuk mengevaluasi kinerja model klasifikasi, khususnya dalam kasus di mana kelas-kelas dalam data tidak seimbang. ROC adalah kurva grafis yang menunjukkan kemampuan model dalam membedakan antara kelas positif dan negatif pada berbagai threshold prediksi. Kurva ROC dibuat dengan memplot True Positive Rate (TPR) atau Recall di sumbu y terhadap False Positive Rate (FPR) di sumbu x [32]. TPR menunjukkan proporsi dari semua kasus positif yang benar-benar diidentifikasi oleh model, sedangkan FPR menunjukkan proporsi dari semua kasus negatif yang secara salah diklasifikasikan sebagai positif oleh model.

ROC memberikan gambaran yang jelas tentang trade-off antara sensitivitas (kemampuan mendeteksi kasus positif) dan spesifisitas (kemampuan mendeteksi kasus negatif) pada berbagai threshold [32]. Dengan mengamati kurva ROC, kita dapat menilai bagaimana perubahan threshold mempengaruhi kinerja model. Model yang baik akan memiliki kurva ROC yang mendekati sudut kiri atas grafik, menunjukkan bahwa model mampu mendeteksi hampir semua kasus positif dengan sangat sedikit kesalahan klasifikasi positif palsu.

AUC adalah nilai numerik yang menggambarkan seberapa baik model dapat membedakan antara kelas positif dan negatif secara keseluruhan. AUC diukur dari 0 hingga 1, dengan nilai 1 menunjukkan bahwa model memiliki kemampuan yang sempurna dalam membedakan antara dua kelas, sedangkan nilai 0,5 menunjukkan bahwa model tidak lebih baik dari tebakan acak [32]. AUC memberikan ukuran kinerja model yang lebih komprehensif karena mencakup semua kemungkinan threshold. Model dengan AUC yang tinggi menunjukkan bahwa model tersebut memiliki kemampuan prediksi yang baik, baik dalam mendeteksi kasus positif maupun dalam menghindari kesalahan deteksi kasus negatif.

Dalam praktiknya, AUC sering digunakan sebagai metrik utama untuk menilai kinerja model, terutama dalam skenario di mana distribusi kelas tidak seimbang. Misalnya, dalam kasus deteksi penyakit yang jarang terjadi, di mana kelas negatif jauh lebih banyak daripada kelas positif, AUC dapat memberikan gambaran yang lebih

akurat tentang kinerja model daripada metrik sederhana seperti akurasi. Ini karena AUC memperhitungkan kemampuan model untuk mempertahankan tingkat deteksi positif yang tinggi sambil meminimalkan kesalahan klasifikasi positif palsu, yang sangat penting dalam aplikasi medis dan banyak aplikasi lain di mana kesalahan klasifikasi dapat memiliki konsekuensi serius.

2.2 Penelitian Terdahulu

Terdapat beberapa artikel yang melakukan prediksi hipertensi untuk populasi umum berskala nasional dan multi negara menggunakan faktor resiko yang relatif mudah diperoleh (*socio-demografi, anthropometry, life style*).

Tabel 2.1 Penelitian Terdahulu

No	Referensi	Jumlah responden	Algoritma yang digunakan	Algoritma Terbaik
1.	[23] Zhao et al. 2021	29,700 (China)	1. <i>Random forest</i> 2. <i>CatBoost</i> 3. <i>Multi-layer Perceptron</i> 4. <i>Logistic regression</i>	<i>Random forest</i>
2.	[24] Alkaabi et al. 2020	987 (Qatar)	1. <i>Decision tree</i> 2. <i>Random forest</i> 3. <i>Logistics regression</i>	<i>Random Forest</i>
3.	[25] Islam et al. 2022	818,603 (Bangladesh, Nepal, and India)	1. <i>Decision tree</i> 2. <i>Random forest</i> 3. <i>Gradient boosting</i> 4. <i>Extreme gradient boosting</i> 5. <i>Logistic regression</i>	<i>Gradient Boosting</i>

			<i>6. Linear discriminant analysis</i>	
4.	[26] Kurniawan et al. 2023	30,320 (Indonesia)	<ol style="list-style-type: none"> 1. <i>Decision tree</i> 2. <i>Random forest</i> 3. <i>Gradient boosting</i> 4. <i>Logistic regression</i> 	<i>Logistic regression</i>
5.	[33] Ambika et al. 2020	599 (India)	<ol style="list-style-type: none"> 1. SVM 2. <i>Artificial Neural Network</i> 3. <i>K-Nearest Neighbor</i> 4. <i>Naive Bayes</i> 	

Namun dari semua artikel tersebut tidak ada artikel yang melakukan klasifikasi hipertensi menjadi 4 kelas: normal, prehipertensi, hipertensi tahap 1, dan hipertensi tahap 2.