PREDIKSI RISIKO DIABETES DENGAN METODE NAIVE BAYES: IDENTIFIKASI FAKTOR RISIKO UTAMA DAN EVALUASI AKURASI MODEL

Alrayssa Davinka Sembiring Depari, Cha Cha Kirana,
Choirun Nissa Oktariana, Fahmi Akbar, Fathoni *
Fakultas Ilmu Komputer, Program Studi Sistem Informasi,
Universitas Sriwijaya, Kota Palembang, Indonesia
fathoni@unsri.ac.id

ABSTRAK

Diabetes merupakan salah satu penyakit kronis yang terus meningkat secara global, termasuk di Indonesia. Pendeteksian dini terhadap risiko diabetes sangat penting untuk mencegah komplikasi serius yang dapat terjadi di masa depan. Permasalahan utama dalam prediksi diabetes terletak pada keterbatasan data klinis, distribusi kelas yang tidak seimbang, serta rendahnya akurasi dalam mendeteksi pasien yang benar-benar mengidap diabetes. Tujuan dari penelitian ini adalah untuk mengevaluasi kemampuan algoritma Naive Bayes dalam memprediksi risiko diabetes, mengidentifikasi faktor risiko utama, serta meningkatkan akurasi model melalui teknik seleksi fitur dan evaluasi performa model. Metode yang digunakan meliputi pengumpulan data dari platform Kaggle, pra-pemrosesan data (transformasi dan pemilihan atribut), pembagian data menjadi data training dan testing, serta penerapan algoritma Naive Bayes menggunakan aplikasi RapidMiner. Hasil penelitian menunjukkan bahwa model Naive Bayes mampu mencapai akurasi sebesar 76,05%, dengan recall sebesar 57,34% untuk pasien yang mengidap diabetes dan precision sebesar 68,05%. Meskipun model cukup andal dalam mengidentifikasi pasien sehat, performa dalam mendeteksi pasien diabetes masih dapat ditingkatkan melalui pemilihan fitur yang tepat dan penggunaan dataset yang lebih seimbang. Penelitian ini memberikan gambaran mengenai potensi dan keterbatasan algoritma Naive Bayes dalam klasifikasi risiko diabetes, serta dapat dijadikan dasar untuk pengembangan sistem prediksi kesehatan yang lebih akurat dan aplikatif di masa depan.

Kata kunci : *Diabetes, Naive Bayes, Klasifikasi, Deteksi Dini, RapidMiner, Machine Learning.*

1. PENDAHULUAN

Diabetes adalah penyakit kronis yang paling umum di dunia, dan kejadiannya meningkat dengan cepat. Menurut laporan 2021 oleh *International Diabetes Union*, sekitar 537 juta orang telah didiagnosis di seluruh dunia. Sekitar 783 juta pasien diperkirakan akan didiagnosis pada tahun 2045. Sangat penting untuk memeriksa prediktor risiko diabetes dan penyebab utama penyakit.

Naive Bayes adalah metode klasifikasi populer untuk pembelajaran mesin untuk AI dan analisis data medis karena dapat menangani catatan data berukuran kecil hingga menengah dan efisiensi aritmatika tinggi. Jika ada banyak kasus diabetes, keakuratan memprediksi model dan diagnosis kondisi ini juga meningkat. Selain itu, akurasi diagnostik dan prediksi diabetes meningkat menggunakan algoritma pembelajaran *ensemble* mesin baru [1]. RNA juga digunakan untuk meningkatkan gen diabetes yang terkait dengan kadar penetrasi genetik. Naive Bayes sering digunakan untuk mengidentifikasi prevalensi dan prediksi penyakit diabetes. Mengingat faktor - faktor yang digunakan oleh manusia, keakuratannya di wilayah ini dapat bersaing [2].

Prediksi risiko diabetes tunduk pada berbagai gangguan, termasuk data. Namun, metode pembelajaran mesin dalam diagnosis klasifikasi membutuhkan peningkatan, terutama dalam hal fitur klinis yang terbatas dan akurasi prediktif. Model ini memprediksi kelas mayoritas daripada kelas minoritas. Ini karena data sering mengandung distribusi kelas yang tidak seimbang. Keterbatasan dalam mengekstraksi pola karakteristik klinis yang kompleks adalah faktor dalam pengembangan model prediktif spesifik untuk belajar motor.

Banyak penelitian telah menggunakan *Naive Bayes* untuk mengklasifikasikan diabetes, seperti: *K-Nearest Neighbor* mencapai akurasi sekitar 99%, sedangkan Naive Bayes mencapai akurasi 75%. Selain itu, Naive Bayes mencapai akurasi 66%, dan ketika digunakan dalam set pengujian dengan 30 titik data, *K-Nearest Neighbor* mencapai akurasi 53%. Sebaliknya, penelitian lain telah menunjukkan bahwa penilaian klasifikasi menghasilkan data yang mencatat diabetes dengan akurasi 90,20% [3]. Ini menunjukkan bahwa pemodelan klasifikasi menggunakan algoritma Naive Bayes mencapai hasil yang masuk akal, tetapi akurasi membutuhkan peningkatan lebih lanjut melalui teknik optimasi seperti pembelajaran ensemble dan metode lainnya.

Penelitian tentang prediksi diabetes semakin berkembang seiring dengan meningkatnya jumlah penderita penyakit ini di seluruh dunia. Salah satu metode yang sering digunakan dalam klasifikasi risiko diabetes adalah algoritma Naïve Bayes. Studi yang menggunakan data pasien dari *Sylhet Diabetes Hospital* di Bangladesh untuk mengembangkan model prediksi berbasis Naïve Bayes. Dengan memanfaatkan 16 atribut klinis sebagai faktor prediktor, model ini mampu mencapai tingkat akurasi sekitar 91%, yang menunjukkan efektivitas algoritma ini dalam deteksi dini diabetes[4].

Selanjutnya, penelitian yang menggunakan dataset dari UCI *Machine Learning Repository* untuk memprediksi risiko diabetes tahap awal dengan metode pembelajaran mesin. Hasil penelitian menunjukkan bahwa model berbasis Naïve Bayes dapat mengklasifikasikan risiko diabetes dengan akurasi 87,88% dan tingkat kesalahan sebesar 12,12% [5]. Hal ini membuktikan bahwa metode ini cukup andal dalam deteksi dini diabetes, terutama ketika diterapkan pada data dengan jumlah sampel yang besar.

Penelitian ini unik karena menggunakan diabetes dataset Kaggle, yang memiliki karakteristik berbeda dari catatan data yang berbeda. Tujuan dari penelitian ini adalah untuk melihat seberapa baik metode Naive Bayes memprediksi potensi diabetes, menentukan faktor risiko utama untuk penyakit, dan membuat prediksi lebih akurat dengan meningkatkan model menggunakan teknik seleksi karakteristik dan estimasi probabilitas. Selain itu, penelitian ini mengevaluasi keakuratan metode Naive Bayes dibandingkan dengan metode lain yang digunakan dalam penelitian sebelumnya.

2. TINJAUAN PUSTAKA

2.1. Penelitian Terdahulu

Penelitian tentang prediksi diabetes semakin berkembang seiring dengan meningkatnya jumlah penderita penyakit ini di seluruh dunia. Salah satu metode yang sering digunakan dalam klasifikasi risiko diabetes adalah algoritma Naïve Bayes. Beberapa penelitian terkait prediksi penyakit diabetes sudah banyak dilakukan dengan metode ini, seperti penelitian yang menggunakan sumber data dari Pima diabetes dataset dengan banyak data sekitar 768 data, dimana dengan menggunakan metode Naïve Bayes mereka memperoleh akurasi sebesar 80% [6]

Ada pula penelitian yang juga menggunakan dataset bersumber dari UCI Repository, dimana mereka membandingkan metode Naïve Bayes dan Greedy Forward Selection dalam prediksi diabetes tahap awal, metode Naïve Bayes dalam penelitian ini mendapatkan hasil akurasi sebesar 87,69% [7]

Berikutnya penelitian yang menggunakan dataset dari Kaggle dimana mereka membandingkan metode Naïve Bayes dan K-Nearest Neighbor (KNN) dalam prediksi penyakit diabetes. Hasil penelitian menunjukkan bahwa algoritma Naïve Bayes mampu mengidentifikasi pasien diabetes dengan pendekatan probabilistik, sementara KNN menunjukkan hasil prediksi yang konsisten terhadap pasien positif diabetes pada beberapa nilai K, khususnya K=3 dan K=5, meskipun nilai akurasi tidak disebutkan secara eksplisit. Penelitian ini menegaskan bahwa kedua metode dapat digunakan secara efektif sesuai karakteristik data yang digunakan. [8]

2.2. Prediksi

Prediksi merupakan upaya untuk memperkirakan apa yang kemungkinan akan terjadi di masa depan dengan memanfaatkan informasi-informasi relevan dari masa lalu. Proses ini biasanya dilakukan dengan pendekatan ilmiah agar dapat menghasilkan estimasi angka atau kejadian yang akurat di waktu mendatang.[9]

2.3. Diabetes

Diabetes merupakan penyakit yang setiap tahunnya akan terus meningkat. Jumlah diabetes tahun 2021 mencapai 537 juta di dunia dan diprediksi akan mencapai 643 juta di tahun 2030 dan 783 juta pada tahun 2045. [10]

2.4. Naive Bayes

Proses klasifikasi data berdasarkan pendekatan probabilistic dilakukan dengan memanfaatkan algoritma Naive Bayes. Teknik ini mengandalkan perhitungan peluang, dan secara luas dikenal karena tingkat ketelitiannya yang tinggi. Metode ini kerap diterapkan dalam menyelesaikan permasalahan prediksi yang berbasis *classification* [11]. Algoritma Naive Bayes merupakan metode klasifikasi terkini dalam *machine learning* yang berlandaskan pada asumsi bahwa setiap atribut memiliki pengaruh secara terpisah terhadap kelas yang diamati, tanpa bergantung pada atribut lainnya. Pendekatan Naive Bayes dikembangkan dari penerapan teorema Bayes, yang menganggap bahwa setiap fitur bersifat bebas dan tidak saling terkait satu sama lain [12]

2.5. *Rapidminer*

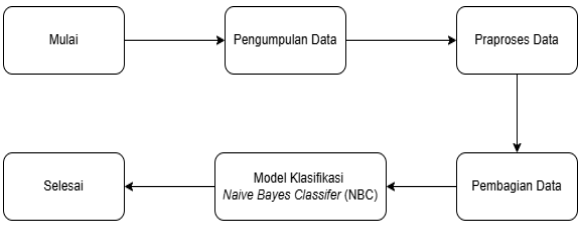
RapidMiner adalah program berbasis open source. RapidMiner adalah mesin penambangan data yang dapat diintegrasikan ke dalam produk-produknya sendiri dan juga tersedia sebagai perangkat lunak mandiri untuk melakukan analisis data, dengan tujuan memberikan informasi terbaik kepada pengguna untuk membuat keputusan terbaik.[13]

2.6. Risiko

Menurut Kamus Besar Bahasa Indonesia, risiko diartikan sebagai akibat yang tidak menyenangkan atau merugikan, baik secara finansial maupun nonfinansial, dari suatu tindakan atau keputusan. Sementara itu, Arthur J. Keown menjelaskan risiko sebagai ketidakpastian atas hasil yang tidak diinginkan, yang bisa diukur melalui deviasi standar. Hanafi dalam tulisannya berjudul *Risiko Kredit* menyebut bahwa risiko sering kali merujuk pada perbedaan antara harapan terhadap tingkat pengembalian dan hasil aktual yang terjadi. Manajemen risiko sendiri merupakan proses yang dirancang untuk merespons risiko-risiko yang telah diidentifikasi, dengan tujuan mengurangi kemungkinan dan dampaknya. Melalui pendekatan ini, potensi kerugian bisa dikenali sejak awal, sehingga strategi yang tepat dapat disiapkan untuk menghadapinya di masa mendatang [14]

3. METODE PENELITIAN

Adapun Tahapan dalam melaksanakan penelitian ini, terdiri dari pengumpulan data, pra proses data, pembagian data dan model klasifikasi. Alur penelitian ini dapat dilihat pada Gambar 1 berikut.



Gambar 1. Alur Penelitian

3.1. Pengumpulan Data

Pengumpulan data dalam Dataset Prediksi Diabetes ini dilakukan dengan mengakses sumber daya yang tersedia di *platform* Kaggle. Dataset ini mencakup informasi yang dikumpulkan dari berbagai survei kesehatan dan studi klinis yang bertujuan untuk memahami faktor-faktor risiko diabetes. Data ini diperoleh melalui pengukuran medis standar, seperti tes toleransi glukosa, tekanan darah, ketebalan kulit, insulin serum, dan indeks massa tubuh (IMT), yang semuanya berhubungan erat dengan kondisi kesehatan individu. Setiap entri data dalam dataset ini mewakili informasi dari individu yang terlibat dalam studi, dengan atribut yang mencakup variabel seperti usia, jumlah kehamilan, dan fungsi pedigree diabetes. Kaggle menyediakan akses terbuka untuk dataset ini, memungkinkan peneliti dan ilmuwan data untuk melakukan eksplorasi lebih lanjut dan mengembangkan model prediksi diabetes dengan menggunakan teknik-teknik *machine learning*. Dengan berbagi data ini, Kaggle berkontribusi pada kolaborasi antara peneliti dalam meningkatkan pemahaman dan pengelolaan diabetes secara lebih efektif.

3.2. Pra Proses Data

Pra-proses atau *Preprocessing* merupakan salah satu proses pengolahan data agar data lebih mudah dan akurat untuk di analisis. Pra-proses memiliki banyak tahap, dalam penelitian ini tahap yang dilalui hanyalah dua, yaitu Tahap transformasi, pemilihan atribut, dan pembagian data. Semua tahap Pra proses ini terdapat dalam aplikasi RapidMiner.

3.3. Transformasi Data

Mengubah isi data, dimana pada proses ini fokusnya kepada perubahan isi data pada suatu atribut. Transformasi melibatkan konversi atau modifikasi data awal menjadi data yang lebih terstruktur untuk mempermudah proses analisis.

3.4. Pemilihan Atribut

Pemilihan atribut merupakan tahap penting dalam proses pengolahan data, karena menentukan variabel mana yang akan dijadikan acuan dalam analisis atau pemodelan. Dalam konteks dataset ini, atribut yang dipilih sebagai label atau target klasifikasi adalah *Outcome*, yang merepresentasikan hasil akhir dari kondisi yang diamati.

3.5. Pembagian Data

Proses ini dilakukan untuk memisahkan data ke dalam dua bagian, yaitu data *training* dan data *testing*, yang keduanya memiliki fungsi yang sangat penting dalam model prediksi. Data *training* digunakan untuk mengajari algoritma bagaimana membaca pola dan hubungan antara fitur dengan data yang diberikan tentang apakah suatu berkas adalah malware, sedangkan data testing digunakan untuk mengevaluasi kinerja dari model yang sudah dibuat. Tujuannya adalah agar model dapat memprediksi bagus atau tidaknya apabila melihat data yang belum pernah dilihat, Pemisahan ini sangat penting untuk menghindari *overfitting* dan memastikan bahwa model memiliki kemampuan generalisasi yang baik saat diterapkan pada data nyata di luar sampel pelatihan.

3.6. Model Klasifikasi *Naive Bayes Classifier* (NBC)

Model klasifikasi Naive Bayes dibuat menggunakan aplikasi RapidMiner Studio versi 2025.0.0, dengan memanfaatkan operator Naïve Bayes untuk membangun model. Pada proses pembentukan model ini, digunakan parameter *laplace correction* sebagai solusi untuk mengatasi kemungkinan munculnya nilai probabilitas nol, yang seringkali menjadi masalah dalam perhitungan probabilitas pada algoritma tersebut

4. HASIL DAN PEMBAHASAN

4.1. Pengumpulan Data

Dalam penelitian ini, data yang digunakan diperoleh dari *platform* berbagi *dataset* terbuka, yaitu Kaggle. Dataset ini bersifat publik sehingga dapat diakses secara bebas oleh peneliti, praktisi, maupun pengembang untuk keperluan analisis dan pengembangan model prediktif. Data tersebut terdiri dari 2768 entri yang mencakup berbagai atribut penting terkait kesehatan.

4.2. Transformation

Salah satu fungsi dari transformasi adalah mengubah jenis atribut nominal menjadi teks, di mana dalam konteks penelitian ini dilakukan dengan mengubah semua nilai integer pada atribut *Outcome* menjadi nilai berupa string. Hasil dari perubahan ini adalah kolom baru yang berisi nilai teks dari data nominal sebelumnya. Peneliti melakukan proses transformasi data ini menggunakan fitur *Text to Columns* di Excel, yang memungkinkan konversi data numerik ke bentuk teks dengan lebih mudah dan terstruktur, sehingga memudahkan dalam proses analisis lebih lanjut. Contoh hasil dari transformasi tersebut ialah sebagai berikut :

Tabel 1. Hasil Transformasi Data

Atribut	Hasil
Pregnancies	6
Glucose	148
Blood Pressure	72
Skin Thickness	35
Insulin	0
BMI	33.6
Diabetes Pedigree Function	0,627
Age	50
Outcome	Mengidap Diabetes

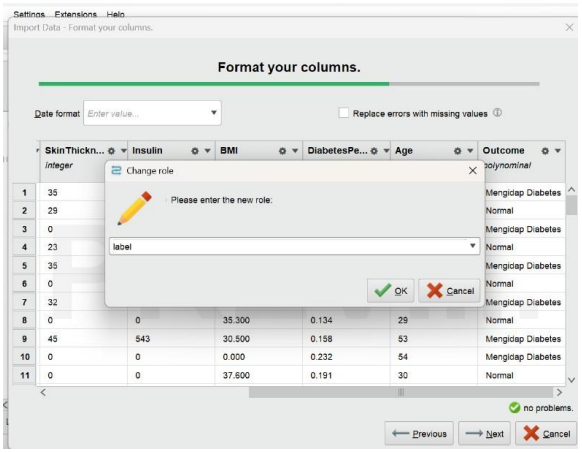
4.3. Pemilihan Atribut

Terdapat beberapa atribut penting yang menjadi fokus utama, yaitu Kehamilan (*Pregnancies*), Glukosa (*Glucose*), Tekanan Darah (*Blood Pressure*), Ketebalan Kulit (*Skin Thickness*), Insulin, Indeks Masa Tubuh (IMT), *Diabetes Pedigree Function*, Usia (*Age*), dan *Outcome*.

Tabel 2. Tabel Atribut

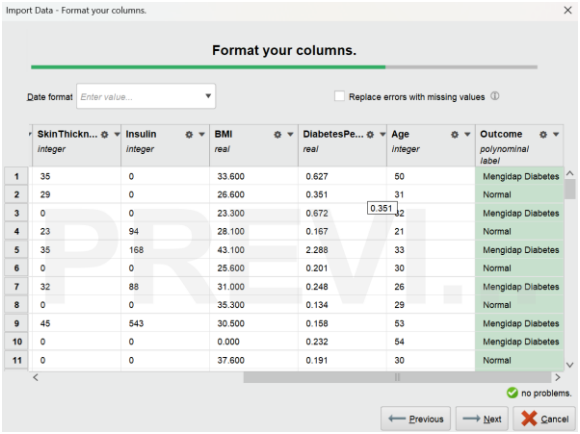
Atribut	*Type*
Pregnancies	Integer
Glucose	Integer
Blood Pressure	Integer
Skin Thickness	Integer
Insulin	Integer
IMT	Integer
Diabetes Pedigree Function	Integer
Outcome	Polynomial

Berikutnya setelah meng-import data kedalam RapidMiner, diperlukan penambahan peran yang perlu disesuaikan. Peran target atau biasa diketahui dengan Label ini digunakan sebagai peran utama dalam penelitian ini.



Gambar 2. Fitur Label

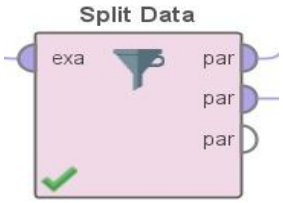
Dalam penelitian ini atribut yang diubah ialah *Outcome*. Hasil dari perubahan peran tersebut adalah sebagai berikut:



Gambar 3. Hasil Penglabelan

4.4. Pembagian Data

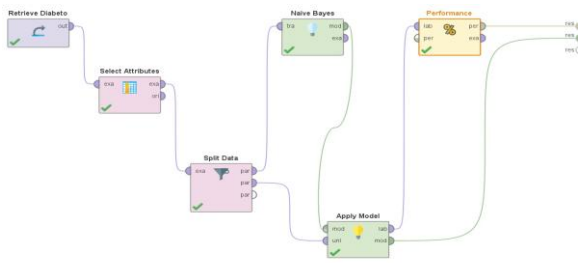
Pembagian Data dalam RapidMiner dapat dilakukan melalui fitur *Split Data*. Pada penelitian ini, data dibagi menjadi 70:30, yang dimana artinya 70% data merupakan data *training* dan 30% merupakan data *testing*. Proses ini menggunakan fitur *Split Data* pada RapidMiner, yang berbentuk sebagai berikut :



Gambar 4. Split Data

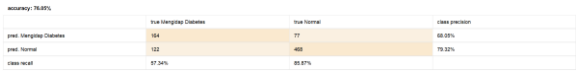
4.5. Model Klasifikasi Naive Bayes Classifier

Tahap berikutnya merupakan tahap klasifikasi menggunakan algoritma Naive Bayes, yang berperan penting dalam proses analisis prediksi penyakit diabetes. Algoritma Naive Bayes kemudian dilatih menggunakan data latih pada node Naive Bayes, menghasilkan model klasifikasi yang kemudian diterapkan pada data uji melalui *node Apply Model.* Hasil klasifikasi ini akhirnya dievaluasi menggunakan *node Performance*, yang memberikan matrik evaluasi seperti akurasi, presisi, dan *recall* untuk menilai seberapa baik model mampu melakukan prediksi terhadap data diabetes. Seluruh proses ini dilakukan secara sistematis untuk memastikan hasil klasifikasi yang akurat dan dapat diandalkan.



Gambar 5. Klasifikasi Naive Bayes

Berikut merupakan hasil dari pengolahan data menggunakan metode Naive Bayes pada RapidMiner :


accuracy: 76.05%

	true Mengidap Diabetes	true Normal	class precision
pred. Mengidap Diabetes	164	77	68.05%
pred. Normal	122	468	79.32%
class recall	57.34%	85.87%	

Gambar 6. Hasil Naive bayes pada Rapidminer

Berdasarkan hasil evaluasi model Naive Bayes pada data uji, diperoleh nilai akurasi sebesar 76.05%, yang menunjukkan bahwa sebanyak 76.05% dari seluruh prediksi yang dilakukan oleh model adalah benar. Nilai ini mengindikasikan bahwa model memiliki performa klasifikasi yang cukup baik dalam membedakan antara pasien yang mengidap diabetes dan yang tidak. Tabel *confusion matrix* menunjukkan bahwa dari total pasien yang benar-benar mengidap diabetes, sebanyak 164 pasien berhasil diprediksi dengan benar, sementara 122 pasien salah diprediksi sebagai tidak mengidap diabetes. Di sisi lain, dari pasien yang sebenarnya tidak mengidap diabetes, sebanyak 468 berhasil diklasifikasikan dengan benar, dan 77 pasien salah diklasifikasikan sebagai mengidap diabetes.

Dari segi *recall*, kelas "Mengidap Diabetes" memiliki nilai sebesar 57.34%, yang berarti model hanya mampu mengenali sekitar 57% dari total pasien yang benar-benar mengidap diabetes. Ini menunjukkan bahwa masih terdapat cukup banyak kasus positif yang tidak terdeteksi (*false negative*). Sementara itu, *recall* untuk kelas "Normal" sebesar 85.87%, memperlihatkan bahwa model lebih baik dalam mengenali pasien yang sehat. Untuk *precision*, kelas "Mengidap Diabetes" memiliki nilai 68.05%, yang menunjukkan bahwa dari semua pasien yang diprediksi mengidap diabetes oleh model, sekitar 68% benar-benar positif. *Precision* untuk kelas "Normal" tercatat 79.32%, menunjukkan hasil prediksi yang cukup dapat diandalkan untuk pasien tanpa diabetes.

5. KESIMPULAN DAN SARAN

Penerapan algoritma Naive Bayes terbukti mampu melakukan klasifikasi risiko diabetes berdasarkan atribut-atribut klinis seperti kadar glukosa, tekanan darah, indeks massa tubuh, dan lainnya. Meskipun memiliki keunggulan dari segi kesederhanaan dan efisiensi komputasi, akurasi *Naive Bayes* masih tergolong lebih rendah dibandingkan algoritma lain seperti *K-Nearest Neighbor* yang mampu memberikan hasil prediksi lebih optimal. Untuk meningkatkan akurasi, disarankan untuk mengkombinasikan algoritma ini dengan metode *ensemble learning* seperti *bagging* atau *boosting*. Pemilihan fitur yang relevan juga penting agar model hanya menggunakan variabel yang signifikan terhadap risiko diabetes. Penggunaan dataset yang lebih besar dan seimbang dapat membantu mengurangi bias klasifikasi terhadap kelas mayoritas. Selain akurasi, evaluasi performa model sebaiknya melibatkan metrik tambahan seperti *precision*, *recall*, dan *F1-score* guna memberikan gambaran yang lebih menyeluruh. Dukungan tools seperti *RapidMiner* mempermudah proses *preprocessing*, klasifikasi, dan analisis data. Visualisasi data serta eksplorasi korelasi antar atribut juga dapat membantu memahami karakteristik penderita diabetes secara lebih mendalam dan aplikatif.

DAFTAR PUSTAKA

[1] Rahman, M. M., & Hasan, M. (2020). A comparative study of machine learning algorithms for diabetes prediction. International Journal of Computer Applications, 975, 8887. Retrieved from https://www.ijcaonline.org/archives/volume183/number14/nar-2021-ijca-921468.pdf

[2] Ikhromr, F. N., et al. (2023). Implementasi data mining untuk memprediksi penyakit diabetes menggunakan algoritma Naïve Bayes dan K-Nearest Neighbor. INTECOMS: Journal of Information Technology and Computer Science, 6(1), 416-428.

[3] Ridwan, A. (2020). Penerapan algoritma Naïve Bayes untuk klasifikasi penyakit diabetes mellitus. Jurnal SISKOM-KB (Sistem Komputasi dan Kecerdasan Buatan), 4(1), 15-21.

[4] Pradhani, P. C., Indrayani, A. S., Azzahra, N., Aflikha, S. E., Zahra, F., & Kalifia, A. D. (2025). Prediksi diabetes mellitus berdasarkan data pasien Sylhet Diabetes Hospital dengan metode Naïve Bayes. Jurnal Ilmiah Sain dan Teknologi, 3(3), 342-353.

[5] Muhtajuddin, & Muhidin, A. (2023). Analisis prediksi risiko diabetes tahap awal menggunakan algoritma Naïve Bayes. Jurnal Teknologi Informatika dan Komputer MH. Thamrin, 9(2), 1443-1450.

[6] Rahman, M. M., & Hasan, M. (2020). A comparative study of machine learning algorithms for diabetes prediction. International Journal of Computer Applications, 975, 8887.

[7] Fitriyani. (2021). *Prediksi diabetes menggunakan algoritma Naïve Bayes dan Greedy Forward Selection*. Jurnal Nasional Teknologi dan Sistem Informasi, 7(2), 61-69.

[8] N. Rahmansyah, S. A. Lusinia, dan Ilmawati (2024). "Analisa Prediksi Penyakit Diabetes Menggunakan Metode Naive Bayes dan K-NN," INNOVATIVE: Journal Of Social Science Research, vol. 5, no. 1, pp. 5311–5323.

[9] C. A. Rahayu, R. Hartono, dan A. Sudiarjo (2023). "Prediksi Penderita Diabetes Menggunakan Metode Naïve Bayes," Jurnal Informatika dan Teknik Elektro Terapan (JITET), vol. 11, no. 3, pp. 261–266, doi: 10.23960/jitet.v11i3.3055.

[10] International Diabetes Federation, "Facts & figures," International Diabetes Federation, 2025. [Online].

[11] Maulidah, N., et al. (2021). Prediksi Penyakit Diabetes Melitus Menggunakan Metode Support Vector Machine dan Naive Bayes. Indonesian Journal on Software Engineering (IJSE), 7(1), 63-68. p-ISSN: 2461-0690, e-ISSN: 2714-9935.

[12] M. Y. Putra and D. I. Putri, "Pemanfaatan Algoritma Naïve Bayes dan K-Nearest Neighbor Untuk Klasifikasi Jurusan Siswa Kelas XI," J. Tekno Kompak, vol. 16, no. 2, pp. 176–187, 2022.

[13] D. Pascalina, R. Widhiastono, and C. Juliane, "Pengukuran Kesiapan Transformasi Digital Smart City Menggunakan Aplikasi Rapid Miner," Technomedia J., vol. 7, no. 3, pp. 293 302, 2022, doi: 10.33050/tmj.v7i3.1914.

[14] N. Hamonangan, N. Maelisa, dan R. Serang, "Analisa Risiko Pada Proyek Rehabilitasi Gedung Arsip Unit Hidrologi Balai Sungai Wilayah Maluku," *Jurnal Manumata*, vol. 8, no. 2, pp. 167–176, 2022.