p-ISSN 2685-9009; e-ISSN 2685-9580

ASIMTOT: JURNAL KEPENDIDIKAN MATEMATIKA

Volume 6 Nomor 1, Desember 2024 – Mei 2025, halaman 63 – 73

Tersedia Daring pada https://journal.unwira.ac.id/index.php/ASIMTOT

KLASIFIKASI SUPPORT VECTOR MACHINE DAN RANDOM FOREST PADA DATA BIOMEDIS : APLIKASI DALAM ANALISIS DATA PENYAKIT DIABETES

CLASSIFICATION OF SUPPORT VECTOR MACHINE AND RANDOM FOREST ON BIOMEDICAL DATA: APPLICATIONS IN DIABETES DISEASE DATA ANALYSIS

Sefri Imanuel Fallo[1)], Florianus Aloysius Nay[2)]

[1,2] Program Studi Matematika-FMIPA-Universitas San Pedro

Email: *fallosefriimanuel@gmail.com*, *olandnay@unisap.ac.id*

Abstrak: Penelitian ini bertujuan untuk membandingkan kinerja algoritma Support Vector Machine (SVM) dan Random Forest dalam mengklasifikasikan data biomedis terkait penyakit diabetes. Data yang digunakan mencakup informasi tentang deteksi diabetes pada pasien, dengan variabel-variabel seperti Indeks Massa Tubuh (BMI), tingkat HbA1c, dan tingkat glukosa darah. Hasil penelitian menunjukkan bahwa Random Forest memberikan tingkat akurasi yang lebih tinggi dibandingkan SVM. Hasil prediksi random forest menghasilkan pasien terdeteksi diabetes sebanyak 73 kasus dan pasien tidak terdeteksi diabetes 1127 kasus dengan akurasi 97.17%. Evaluasi model menegaskan bahwa Random Forest mencapai nilai kappa sebesar 0.7967, menandakan kemampuannya dalam memprediksi penyakit diabetes dengan lebih efektif. Hasil ini menyiratkan bahwa Random Forest dapat menjadi pilihan yang lebih optimal dalam memodelkan prediksi penyakit diabetes, terutama ketika mempertimbangkan variabel-variabel yang relevan seperti BMI, tingkat HbA1c, dan tingkat glukosa darah dalam analisisnya.

Kata Kunci: Diabetes, Biomedical, Support Vector Machine, Random Forest.

Abstract: *The aim of this study is to compare the performance of the Support Vector Machine (SVM) and Random Forest algorithms in classifying biomedical data related to diabetes mellitus. The dataset comprises information concerning the detection of diabetes in patients, incorporating variables such as Body Mass Index (BMI), HbA1c levels, and blood glucose levels. The findings of the research demonstrate that Random Forest exhibits a higher accuracy rate compared to SVM. The Random Forest prediction results indicate that 73 cases were detected with diabetes while 1127 cases were not, achieving an accuracy of 97.17%. Model evaluation further confirms that Random Forest achieves a kappa value of 0.7967, signifying its effectiveness in predicting diabetes mellitus. These results suggest that Random Forest may represent a more optimal choice for modeling diabetes prediction, particularly when considering relevant variables such as BMI, HbA1c levels, and blood glucose levels in the analysis.*

Keywords*: Diabetes, Biomedical, Support Vector Machine, Random Forest.*

Cara Sitasi: Fallo, S.I., & Nay, F.A. (2024). Klasifikasi Support Vector Machine dan Random Forest pada Data Biomedis: Aplikasi dalam Analisis Data Penyakit Diabetes. *Asimtot: Jurnal Kependidikan Matematika*, "6"("1"), "63-73"

p-ISSN 2685-9009; e-ISSN 2685-9580

ASIMTOT: JURNAL KEPENDIDIKAN MATEMATIKA

Volume 6 Nomor 1, Desember 2024 – Mei 2025, halaman 63 – 73

Tersedia Daring pada https://journal.unwira.ac.id/index.php/ASIMTOT

Penyakit diabetes merupakan salah satu masalah kesehatan global yang signifikan, dengan prevalensi yang terus meningkat di seluruh dunia. Diperkirakan bahwa sekitar 1.3 Milyar orang dewasa menderita diabetes pada tahun 2050, dan angka ini diperkirakan akan terus bertambah dalam beberapa dekade mendatang (Tural Buyuk et al. 2023). Diabetes, baik tipe 1 maupun tipe 2, memiliki dampak yang serius pada kesehatan individu, termasuk risiko tinggi terhadap komplikasi jangka panjang seperti penyakit jantung, stroke, gagal ginjal, dan kehilangan penglihatan. Diabetes melibatkan berbagai faktor yang kompleks dan saling terkait, termasuk faktor genetik, gaya hidup, dan lingkungan (Singh et al. 2024). Pengelolaan penyakit ini membutuhkan pendekatan yang holistik dan personal, yang memungkinkan pencegahan, deteksi dini, dan pengobatan yang tepat waktu. Dalam konteks ini, analisis data biomedis dapat membantu dalam pengembangan model prediktif yang dapat membantu dalam identifikasi individu yang rentan terhadap diabetes, serta dalam pemantauan dan pengelolaan penyakit pada tingkat individual (Agarwal et al. 2024).

Dalam upaya untuk memahami dan mengelola penyakit ini, analisis data biomedis telah menjadi semakin penting (Zhang et al. 2023). Data biomedis, yang mencakup informasi tentang parameter klinis, genetik, dan lingkungan, dapat memberikan wawasan yang berharga tentang faktor risiko, pola perkembangan penyakit, dan respons terhadap perawatan. Di antara berbagai metode analisis data, klasifikasi menggunakan teknik pembelajaran mesin telah menunjukkan keunggulan dalam mengidentifikasi pola kompleks dalam data biomedis (Zhang et al. 2023).

Dalam konteks ini, penelitian ini bertujuan untuk menerapkan dan membandingkan dua teknik klasifikasi yang umum digunakan, yaitu Support Vector Machine (SVM) dan Random Forest, pada data biomedis terkait diabetes (Febrian et al. 2022). SVM dan Random Forest adalah dua algoritma pembelajaran mesin yang kuat dan sering digunakan dalam berbagai aplikasi, termasuk analisis data kesehatan (Mujumdar and Vaidehi 2019). Teknik klasifikasi seperti SVM dan Random Forest telah terbukti efektif dalam menghadapi tantangan yang terkait dengan data biomedis yang kompleks dan berdimensi tinggi (Oikonomou and Khera 2023). SVM bekerja dengan memisahkan dua kelas dengan cara menemukan hyperplane yang paling memaksimalkan margin antara kelas-kelas tersebut dalam ruang fitur, sementara Random Forest menggabungkan prediksi dari beberapa pohon keputusan yang dihasilkan secara acak. Kedua metode ini telah digunakan secara luas dalam berbagai aplikasi, termasuk dalam analisis data kesehatan, karena kemampuan mereka untuk menangani data yang tidak linear, besar, dan berisiko tinggi (Harsa et al. 2023).

Melalui penerapan teknik-teknik ini, kami berharap untuk meningkatkan pemahaman tentang hubungan antara variabel biomedis dan diagnosis diabetes, serta membandingkan kinerja keduanya dalam konteks analisis data ini (Mujumdar and Vaidehi 2019). Penelitian ini dapat memberikan kontribusi signifikan untuk

p-ISSN 2685-9009; e-ISSN 2685-9580

ASIMTOT: JURNAL KEPENDIDIKAN MATEMATIKA

Volume 6 Nomor 1, Desember 2024 – Mei 2025, halaman 63 – 73

Tersedia Daring pada https://journal.unwira.ac.id/index.php/ASIMTOT

pengembangan metode diagnosis dan manajemen penyakit diabetes secara efektif dan efisien.

Metode Penelitian

Data yang digunakan dalam penelitian ini merupakan data sekunder yang diperoleh dari Kaggle.com. Dataset diabetes ini merupakan kumpulan data medis dan demografis dari pasien, beserta status diabetes mereka (positif atau negatif). Data tersebut mencakup fitur seperti usia, jenis kelamin, indeks massa tubuh (BMI), hipertensi, penyakit jantung, riwayat merokok, tingkat HbA1c, dan kadar glukosa darah. Dataset ini kemudian digunakan untuk membangun model machine learning untuk memprediksi diabetes pada pasien berdasarkan riwayat medis dan informasi demografis mereka. Teknik analisis yang digunakan dalam penelitian ini adalah Machine Learning. Machine Learning, sebuah subbidang dari kecerdasan buatan, memungkinkan sistem komputer untuk memperoleh pengetahuan dari data dan menghasilkan prediksi atau keputusan tanpa pemrograman eksplisit (Ponti 2018). Dengan mengenali pola dan tren dalam data, Machine Learning memberdayakan sistem komputer untuk belajar dan beradaptasi secara mandiri. Selama beberapa dekade terakhir, Machine Learning telah muncul sebagai teknologi yang menonjol dan berpengaruh (Tehrani et al. 2022).

1. *Support Vector Machine (SVM)*

   SVM adalah metodologi pembelajaran mesin yang didasarkan pada prinsip Minimalisasi Risiko Struktural (SRM) (Huang, Lu, and Ling 2003). Tujuan utamanya adalah untuk membedakan hyperplane optimal dalam ruang input, yang dapat efektif memisahkan kelas-kelas yang berbeda. Tantangan utama yang dihadapi oleh SVM terletak pada penentuan fungsi yang sesuai yang dapat secara akurat memisahkan kedua kelas, menggunakan informasi yang berasal dari data pelatihan yang tersedia (Pradeep and Naveen 2018). Algoritma SVM dapat diimplementasikan melalui langkah-langkah berikut:

   a. Membagi data menjadi data *testing* dan data *training*
   b. Menetukan *cost value*
   c. Memntukan standar deviasi dan alpha
   d. Mengimplementasikan RBF Kernel

$$\begin{aligned} K\left(x_i,x_j\right) &= e^{-\frac{1}{2\sigma^2}\left\|x_i-x_j\right\|^2} \\ &= \exp\left[-\frac{1}{2\sigma^2}\left(x_i-x_j\right)^T\left(x_i-x_j\right)\right] \\ &= \exp\left[-\frac{1}{2\sigma^2}\left(x_i^T x_i - 2x_i^T x_j + x_i^T x_j\right)\right] \\ &= \exp\left[-\frac{1}{2\sigma^2}\left(x_i^T x_i + x_i^T x_j\right)\right] + \exp\left[x_i^T x_j\right] \end{aligned} \quad (1)$$

   e. Menentukan hasil prediksi dari RBF Kernel

$$f\left(x\right) = sign\left(\sum_{i,j}^{n}\alpha_i y_i \exp\left[-\frac{1}{2\sigma^2}\left(x_i^T x_i + x_i^T x_j\right)\right] + \exp\left[x_i^T x_j\right] + b\right) \quad (2)$$

2. Random Forest

   Random Forest adalah algoritma pembelajaran mesin yang menggabungkan beberapa pohon keputusan untuk membuat prediksi (Tang et al. 2020). Setiap pohon dibangun pada subset acak dari data pelatihan dan fitur, dan algoritma memilih titik pemisahan terbaik berdasarkan kriteria tertentu. Prediksi akhir dibuat dengan menggabungkan prediksi dari semua pohon (Liu et al. 2022). *Random Forest* dikenal

p-ISSN 2685-9009; e-ISSN 2685-9580

ASIMTOT: JURNAL KEPENDIDIKAN MATEMATIKA

Volume 6 Nomor 1, Desember 2024 – Mei 2025, halaman 63 – 73

Tersedia Daring pada https://journal.unwira.ac.id/index.php/ASIMTOT

karena kemampuannya untuk menangani data berdimensi tinggi dan mengurangi *overfitting* (Adnan et al. 2023). Algoritma Random Forest dapat diimplementasikan melalui langkah-langkah berikut:

a. Pemilihan k sampel dari dataset D, secara acak dan dengan penggantian.
b. Memanfaatkan dataset D untuk membangun pohon keputusan ke-i.
c. Dalam membangun pohon keputusan ke-i, metodologi CART dapat digunakan. Metodologi CART menggunakan keuntungan informasi untuk menentukan setiap simpul dalam pohon (Kao et al. 2021). Perhitungan keuntungan informasi dapat dihitung menggunakan persamaan:

$$Gain(A) = Info(D) - Info_A(D) \quad (3)$$

Nilai dari $Info(D)$ dapat ditentukan dengan formula $Info_A(D)$ di bawah ini:

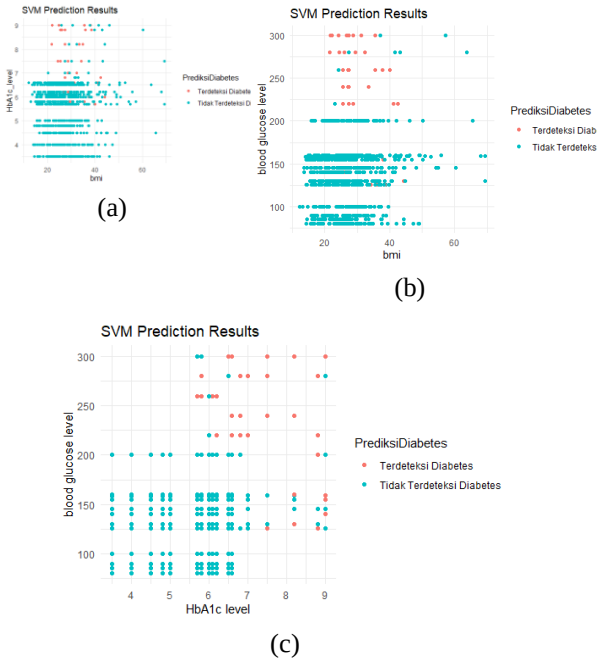
$$Info(D) = -\sum_{i=1}^{m} p_i \log_2(p_i) \quad (4)$$

Hasil Penelitian dan Pembahasan

Hasil

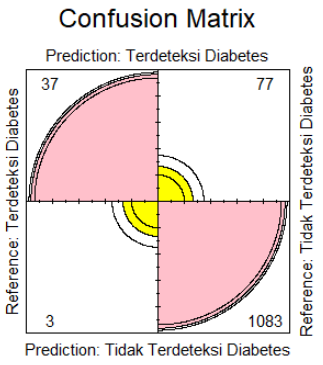
Hasil peneitian ini merupakan evaluasi model machine learning dalam memprediksi penyakit diabetes.

1. *Support Vector Machine(SVM)*
   SVM memprediksi data diabetes dengan kernel RBF, dengan hasil evaluasi model yang disajikan di bawah ini:



Gambar 1. (a)(b)(c) Visualisasi Grafik Hasil Prediksi SVM

Berdasarkan ilustrasi di atas diketahui bahwa hasil prediksi pasien terdeteksi diabetes sebanyak 40 kasus dan pasien tidak terdeteksi diabetes 1160 kasus. Visualisasi hasil prediksi disajikan dengan *confussion matrix* di bawah ini



Gambar 2. Confusion Matriks SVM

Gambar 2 menunjukkan kinerja model klasifikasi dalam mendeteksi diabetes. Berdasarkan matriks ini, terdapat 37 kasus di mana model berhasil memprediksi "Terdeteksi

p-ISSN 2685-9009; e-ISSN 2685-9580

ASIMTOT: JURNAL KEPENDIDIKAN MATEMATIKA

Volume 6 Nomor 1, Desember 2024 – Mei 2025, halaman 63 – 73

Tersedia Daring pada https://journal.unwira.ac.id/index.php/ASIMTOT

Diabetes" secara benar (*True Positives*) dan 1083 kasus di mana model memprediksi "Tidak Terdeteksi Diabetes" secara benar (*True Negatives*). Namun, model juga menghasilkan 77 prediksi salah di mana model mendeteksi diabetes pada kasus yang sebenarnya tidak memiliki diabetes (*False Positives*) dan 3 prediksi salah di mana model tidak mendeteksi diabetes pada kasus yang sebenarnya memiliki diabetes (*False Negatives*). Secara umum, model ini memiliki kinerja yang baik dalam mengidentifikasi kasus tanpa diabetes, terbukti dengan jumlah *true negatives* yang tinggi. Namun, terdapat beberapa kekurangan dalam mendeteksi kasus positif, seperti terlihat dari adanya *false positives* dan *false negatives*. Hasil statistik komprehensif menggunakan algoritma SVM dijelaskan dalam tabel di bawah ini.

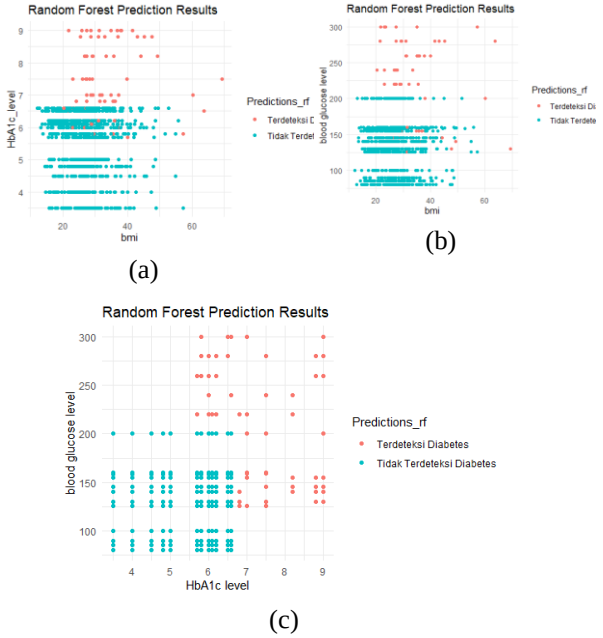
Tabel 1. Statistik Komprehensif SVM

Overall statistics	Nilai
Accuracy	0.9333
95% CI	(0.9177, 0.9468)
No Information Rate	0.9667
P-Value [Acc > NIR]	1
Kappa	0.4536

Hasil prediksi model klasifikasi dalam mendeteksi diabetes menunjukkan bahwa akurasi model mencapai 93,33% dengan interval kepercayaan 95% berada pada rentang 0,9177 hingga 0,9468, yang menunjukkan tingkat ketepatan prediksi yang cukup tinggi. Namun, **No Information Rate (NIR)** sebesar 96,67% lebih tinggi daripada akurasi model, mengindikasikan bahwa kelas dominan adalah "Tidak Terdeteksi Diabetes." P-Value sebesar 1 menunjukkan bahwa akurasi model tidak secara signifikan lebih baik daripada NIR. Nilai **Kappa** sebesar 0,4536 menunjukkan tingkat kesepakatan sedang antara prediksi model dan hasil aktual, yang berarti ada ruang untuk peningkatan dalam membedakan kasus "Terdeteksi Diabetes" dan "Tidak Terdeteksi Diabetes."

2. *Random Forest*

Random Foresr memprediksi data diabetes dengan model CART, dengan hasil evaluasi model yang disajikan di bawah ini:



Gambar 3. (a)(b)(c) Visualisasi Grafik Hasil Prediksi Random Forest

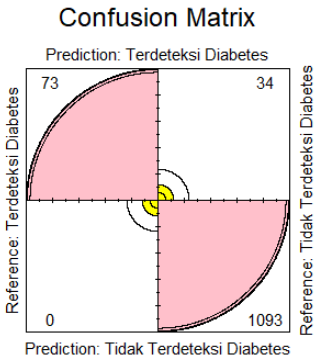
Berdasarkan ilustrasi di atas diketahui bahwa hasil prediksi pasien terdeteksi diabetes sebanyak 73 kasus dan pasien tidak terdeteksi diabetes 1127 kasus. Visualisasi hasil prediksi disajikan dengan confussion matriks di bawah ini

p-ISSN 2685-9009; e-ISSN 2685-9580

ASIMTOT: JURNAL KEPENDIDIKAN MATEMATIKA

Volume 6 Nomor 1, Desember 2024 – Mei 2025, halaman 63 – 73

Tersedia Daring pada https://journal.unwira.ac.id/index.php/ASIMTOT



Gambar 4. Confusion Matriks Random Forest

Gambar 4 menunjukkan hasil prediksi model klasifikasi dalam mendeteksi diabetes. Berdasarkan matriks ini, terdapat 73 kasus di mana model berhasil memprediksi "Terdeteksi Diabetes" dengan benar (*True Positives*) dan 1093 kasus di mana model memprediksi "Tidak Terdeteksi Diabetes" dengan benar (*True Negatives*). Selain itu, terdapat 34 kasus di mana model salah memprediksi "Terdeteksi Diabetes" pada individu yang sebenarnya tidak memiliki diabetes (*False Positives*), namun tidak ada kasus di mana model gagal mendeteksi diabetes pada individu yang sebenarnya memiliki diabetes (*False Negatives*). Matriks ini menunjukkan bahwa model memiliki performa yang baik dalam mendeteksi kasus "Tidak Terdeteksi Diabetes" dengan jumlah *True Negatives* yang tinggi, serta tidak ada kasus yang terlewatkan pada deteksi diabetes (*False Negatives*). Akan tetapi, model masih menghasilkan beberapa *False Positives*, yaitu kesalahan prediksi pada individu yang tidak memiliki diabetes.

Tabel 2. Statistik Komprehensif Random Forest

Overall statistics	Nilai
Accuracy	0.9717
95% CI	(0.9606, 0.9803)
No Information Rate	0.9392
P-Value [Acc > NIR]	1.404e-07
Kappa	0.7964

Hasil prediksi model klasifikasi dalam mendeteksi diabetes menunjukkan bahwa model memiliki akurasi yang sangat tinggi, yaitu 97,17%, dengan interval kepercayaan 95% antara 0,9606 dan 0,9803. *No Information Rate* (NIR) sebesar 93,92% menunjukkan bahwa akurasi model lebih tinggi daripada tingkat akurasi yang hanya berdasarkan kelas dominan. P-Value sebesar 1,404e-07 (sangat kecil) menunjukkan bahwa peningkatan akurasi model ini signifikan dibandingkan dengan NIR. Nilai Kappa sebesar 0,7964 menunjukkan tingkat kesepakatan yang tinggi antara prediksi model dan hasil aktual, yang menunjukkan bahwa model ini cukup andal dalam membedakan antara kasus "Terdeteksi Diabetes" dan "Tidak Terdeteksi Diabetes."

3. Evaluasi Model

Pada fase khusus ini, dilakukan pemeriksaan cermat untuk meneliti kinerja algoritma machine learning yang telah diterapkan. Setelah diterapkan untuk menilai penyakit diabetes untuk tujuan mendeteksi ada tidaknya diabetes pada pasien, evaluasi model yang dihasilkan dijelaskan sebagai berikut

Tabel 3. Evaluasi Model

Algoritma Machine learning	***Accuracy***	***Kappa***
Support Vector Machine	0.9333	0.4536
Random Forest	0.9717	0.7964

Evaluasi model untuk mendeteksi penyakit diabetes menggunakan dua algoritma, yaitu *Support Vector Machine*

p-ISSN 2685-9009; e-ISSN 2685-9580

ASIMTOT: JURNAL KEPENDIDIKAN MATEMATIKA

Volume 6 Nomor 1, Desember 2024 – Mei 2025, halaman 63 – 73

Tersedia Daring pada https://journal.unwira.ac.id/index.php/ASIMTOT

(SVM) dan Random Forest (RF), menunjukkan hasil yang berbeda. Pada metrik *Accuracy*, SVM memiliki nilai 93,33%, yang berarti model ini benar 93,33% dari waktu. Sementara itu, Random Forest menunjukkan nilai yang lebih tinggi, yaitu 97,17%, yang mengindikasikan bahwa model ini lebih sering membuat prediksi yang benar. Di sisi lain, metrik Kappa yang mengukur tingkat kesepakatan antara prediksi model dan kenyataan menunjukkan hasil yang lebih baik pada Random Forest dengan nilai 0,7964, yang menunjukkan kesepakatan yang kuat. SVM memiliki nilai Kappa 0,4536, yang menunjukkan kesepakatan moderat. Secara keseluruhan, Random Forest menunjukkan performa yang lebih baik dibandingkan SVM, baik dalam hal akurasi maupun dalam kemampuan model untuk menghasilkan prediksi yang sesuai dengan kenyataan.

Pembahasan

Penelitian ini menguji dua algoritma klasifikasi, *Support Vector Machine* (SVM) dan Random Forest (RF), untuk mendeteksi penyakit diabetes pada data biomedis. Hasil dari kedua model menunjukkan perbedaan performa yang signifikan, baik dalam hal akurasi maupun Kappa, yang menilai tingkat kesepakatan antara prediksi model dan kenyataan.

Model SVM menunjukkan akurasi 93,33%, yang menunjukkan bahwa model ini cukup andal dalam memprediksi status diabetes, dengan tingkat ketepatan prediksi yang relatif tinggi. Namun, nilai No Information Rate (NIR) sebesar 96,67%, yang lebih tinggi dari akurasi model, mengindikasikan bahwa kelas dominan dalam dataset adalah "Tidak Terdeteksi Diabetes", yang membuat model ini lebih sering memprediksi kasus tanpa diabetes. Meskipun akurasinya cukup baik, tingginya NIR menunjukkan adanya ketidakseimbangan kelas, yang menjadi tantangan bagi model untuk mendeteksi kasus diabetes dengan lebih tepat. Dari matris konfusi, terlihat bahwa SVM berhasil mendeteksi 37 kasus Terdeteksi Diabetes dengan benar (*True Positives*) dan 1083 kasus Tidak Terdeteksi Diabetes dengan benar (*True Negatives*). Namun, ada 77 kasus *False Positives* di mana model salah mendeteksi diabetes pada individu yang sebenarnya tidak mengidapnya, serta 3 kasus *False Negatives* di mana model gagal mendeteksi diabetes pada pasien yang sebenarnya mengidapnya. *False Positives* dan False Negatives ini menunjukkan bahwa meskipun SVM cukup baik dalam mengidentifikasi kasus tanpa diabetes, model ini masih perlu ditingkatkan dalam mendeteksi kasus positif secara akurat. Nilai Kappa 0,4536 menggambarkan bahwa terdapat kesepakatan moderat antara prediksi model dan kenyataan. Meskipun model ini cukup andal untuk kasus negatif (tidak terdeteksi diabetes), ada ruang untuk perbaikan dalam mendeteksi kasus positif. Dengan memperhatikan False Positives dan False Negatives, model ini dapat lebih dioptimalkan untuk meningkatkan kinerja dalam mendeteksi pasien dengan diabetes.

Di sisi lain, model Random Forest menunjukkan performa yang lebih baik dengan akurasi 97,17% dan interval kepercayaan 95% yang sangat tinggi (antara

p-ISSN 2685-9009; e-ISSN 2685-9580

ASIMTOT: JURNAL KEPENDIDIKAN MATEMATIKA

Volume 6 Nomor 1, Desember 2024 – Mei 2025, halaman 63 – 73

Tersedia Daring pada https://journal.unwira.ac.id/index.php/ASIMTOT

96,06% hingga 98,03%). Ini menunjukkan bahwa model RF lebih sering memberikan prediksi yang benar dan lebih dapat diandalkan daripada SVM. No Information Rate (NIR) untuk RF adalah 93,92%, yang lebih rendah dari akurasi model, menandakan bahwa RF lebih mampu membedakan antara kelas Terdeteksi Diabetes dan Tidak Terdeteksi Diabetes.

Conffusion matrix menunjukkan bahwa RF berhasil mendeteksi 73 *kasus True Positives* dan 1093 kasus *True Negatives*. Selain itu, model ini tidak menghasilkan *False Negatives*, yang berarti tidak ada kasus di mana pasien yang sebenarnya mengidap diabetes terlewatkan. Namun, terdapat 34 kasus *False Positives*, yang menunjukkan beberapa kesalahan prediksi pada individu yang tidak memiliki diabetes. Keberhasilan RF dalam mendeteksi semua kasus positif dengan tepat tanpa adanya *False Negatives* sangat menguntungkan dalam konteks diagnosa medis, di mana menghindari *False Negatives* sangat krusial untuk pencegahan dan pengobatan dini. Nilai Kappa 0,7964 menunjukkan kesepakatan yang sangat baik antara prediksi model dan kenyataan, yang menegaskan bahwa RF memiliki kemampuan yang sangat baik dalam membedakan antara kedua kelas dan lebih andal dibandingkan dengan SVM. Dengan tidak adanya *False Negatives* dan hanya beberapa *False Positives*, model RF dapat dianggap lebih unggul dalam hal ketepatan prediksi dan keandalan dibandingkan SVM.

Dari hasil evaluasi, Random Forest jelas lebih unggul dibandingkan dengan *Support Vector Machine* dalam mendeteksi penyakit diabetes. Tidak hanya memiliki akurasi yang lebih tinggi, tetapi RF juga menunjukkan kesepakatan yang lebih baik (Kappa 0,7964) dibandingkan SVM (Kappa 0,4536). Hal ini menunjukkan bahwa RF lebih efektif dalam mengidentifikasi pasien dengan diabetes dan tidak kehilangan kasus positif (*false negatives*).Namun, meskipun Random Forest menunjukkan hasil yang lebih baik secara keseluruhan, kedua model ini menunjukkan pentingnya penanganan terhadap ketidakseimbangan kelas dalam dataset. Baik pada SVM maupun RF, terdapat sejumlah *False Positives*, yang perlu diperhatikan lebih lanjut untuk meningkatkan presisi model, terutama pada aplikasi medis di mana kesalahan prediksi dapat memiliki dampak signifikan.

Simpulan dan Saran

Simpulan

Penelitian ini membandingkan dua algoritma klasifikasi, yaitu Support Vector Machine (SVM) dan Random Forest (RF), dalam mendeteksi penyakit diabetes berdasarkan data biomedis. Hasil penelitian menunjukkan bahwa Random Forest memiliki performa yang lebih unggul dibandingkan dengan Support Vector Machine. RF menghasilkan akurasi yang lebih tinggi (97,17%) dan menunjukkan tingkat kesepakatan yang lebih baik (Kappa 0,7964), yang mengindikasikan kemampuannya dalam membedakan antara pasien dengan diabetes dan yang tanpa diabetes secara lebih akurat. Sementara itu, SVM memiliki akurasi 93,33%

p-ISSN 2685-9009; e-ISSN 2685-9580

ASIMTOT: JURNAL KEPENDIDIKAN MATEMATIKA

Volume 6 Nomor 1, Desember 2024 – Mei 2025, halaman 63 – 73

Tersedia Daring pada https://journal.unwira.ac.id/index.php/ASIMTOT

dan Kappa 0,4536, yang menunjukkan bahwa meskipun model ini cukup andal dalam mendeteksi kasus tanpa diabetes, terdapat ruang untuk perbaikan dalam mendeteksi kasus positif, seperti yang terlihat dari adanya *False Positives* dan *False Negatives*. Secara keseluruhan, Random Forest terbukti lebih efektif dalam mendiagnosis penyakit diabetes dan lebih andal dalam memberikan prediksi yang sesuai dengan kenyataan, terutama dalam menghindari *False Negatives*. Oleh karena itu, Random Forest dapat dianggap sebagai algoritma yang lebih baik untuk aplikasi klasifikasi dalam analisis penyakit diabetes, meskipun penanganan terhadap ketidakseimbangan kelas tetap menjadi perhatian penting untuk meningkatkan ketepatan deteksi lebih lanjut.Simpulan dapat bersifat generalisasi temuan sesuai permasalahan penelitian, dapat pula berupa rekomendatif untuk langkah selanjutnya.

Saran

Berdasarkan hasil penelitian ini, terdapat beberapa saran yang dapat dipertimbangkan untuk pengembangan lebih lanjut dalam penerapan algoritma klasifikasi pada deteksi penyakit diabetes. Salah satu aspek yang perlu diperhatikan adalah penanganan ketidakseimbangan kelas dalam dataset. Meskipun Random Forest menunjukkan kinerja yang lebih baik dibandingkan *Support Vector Machine*, kedua model masih terpengaruh oleh ketidakseimbangan kelas, yang dapat menyebabkan *False Positives* dan *False Negatives*. Oleh karena itu, penerapan teknik *oversampling* atau *undersampling*, serta metode SMOTE (*Synthetic Minority Over-sampling Technique*), dapat membantu meningkatkan kinerja model dalam mendeteksi kasus positif dengan lebih akurat.

Selain itu, penambahan lebih banyak fitur biomedis atau faktor risiko lain, seperti riwayat keluarga, pola makan, dan gaya hidup pasien, dapat meningkatkan kemampuan model dalam membedakan pasien yang berisiko tinggi terhadap diabetes. Dengan demikian, model akan lebih canggih dalam memprediksi penyakit dan mengurangi kesalahan dalam klasifikasi. Tak hanya itu, meskipun Random Forest menunjukkan hasil yang terbaik dalam penelitian ini, penting untuk mengeksplorasi algoritma lain seperti *Gradient Boosting* atau *XGBoost* untuk mengetahui apakah algoritma-algoritma tersebut dapat memberikan akurasi yang lebih tinggi atau menghasilkan lebih sedikit *False Positives* dan *False Negatives*. Perbandingan antara berbagai algoritma ini akan memberikan gambaran yang lebih menyeluruh mengenai performa model dalam mendeteksi diabetes.

Validasi model juga perlu dilakukan dengan menggunakan data yang lebih beragam, termasuk data dari berbagai rumah sakit atau daerah dengan karakteristik pasien yang berbeda. Hal ini penting untuk memastikan bahwa model yang dibangun dapat bekerja dengan baik di berbagai konteks dan populasi, serta meningkatkan generalisasi model. Di sisi lain, hasil yang diperoleh dapat diterapkan dalam sistem pendukung keputusan medis untuk membantu dokter dalam mendiagnosis diabetes secara lebih cepat dan akurat. Penerapan model ini akan sangat

ASIMTOT: JURNAL KEPENDIDIKAN MATEMATIKA

Volume 6 Nomor 1, Desember 2024 – Mei 2025, halaman 63 – 73

Tersedia Daring pada https://journal.unwira.ac.id/index.php/ASIMTOT

bermanfaat dalam upaya deteksi dini penyakit diabetes, yang merupakan langkah penting dalam pencegahan dan pengobatan penyakit tersebut. Dengan mengikuti saran-saran ini, kualitas dan akurasi model dalam mendeteksi penyakit diabetes diharapkan dapat meningkat dan memberikan kontribusi signifikan dalam pengelolaan penyakit diabetes di masyarakat.

Daftar Pustaka

Adnan, Mohammed Sarfaraz Gani et al. 2023. "A Novel Framework for Addressing Uncertainties in Machine Learning-Based Geospatial Approaches for Flood Prediction." *Journal of Environmental Management* 326(PB): 116813. https://doi.org/10.1016/j.jenvman.2022.116813.

Agarwal, Manisha et al. 2024. "Diabetic Retinopathy Screening Guidelines for Physicians in India: Position Statement by the Research Society for the Study of Diabetes in India (RSSDI) and the Vitreoretinal Society of India (VRSI)-2023." *International Journal of Diabetes in Developing Countries* (0123456789): 1–8.

Febrian, Muhammad Exell et al. 2022. "Diabetes Prediction Using Supervised Machine Learning." *Procedia Computer Science* 216(2022): 21–30. https://doi.org/10.1016/j.procs.2022.12.107.

Harsa, Hastuadi et al. 2023. "Machine Learning and Artificial Intelligence Models Development in Rainfall-Induced Landslide Prediction." *IAES International Journal of Artificial Intelligence* 12(1): 262–70.

Huang, Jin, Jingjing Lu, and Charles X. Ling. 2003. "Comparing Naive Bayes, Decision Trees, and SVM with AUC and Accuracy." *Proceedings - IEEE International Conference on Data Mining, ICDM*: 553–56.

Kao, I. Feng, Jia Yi Liou, Meng Hsin Lee, and Fi John Chang. 2021. "Fusing Stacked Autoencoder and Long Short-Term Memory for Regional Multistep-Ahead Flood Inundation Forecasts." *Journal of Hydrology* 598(October 2020): 126371. https://doi.org/10.1016/j.jhydrol.2021.126371.

Liu, Qiang et al. 2022. "Discussion on the Tree-Based Machine Learning Model in the Study of Landslide Susceptibility." *Natural Hazards* 113(2): 887–911. https://doi.org/10.1007/s11069-022-05329-4.

Mujumdar, Aishwarya, and V. Vaidehi. 2019. "Diabetes Prediction Using Machine Learning Algorithms." *Procedia Computer Science* 165: 292–99. https://doi.org/10.1016/j.procs.2020.01.047.

Oikonomou, Evangelos K., and Rohan Khera. 2023. "Machine Learning in Precision Diabetes Care and Cardiovascular Risk Prediction." *Cardiovascular Diabetology* 22(1): 1–16. https://doi.org/10.1186/s12933-023-01985-3.

Ponti, Rodrigo Fernandes de Mello Moacir Antonelli. 2018. 45 Machine Learning *Machine Learning A Practical Approach on the Statistical Learning Theory*. Springer International Publishing AG. https://books.google.ca/books?id=EoYBngEACAAJ&dq=mitchell+machine+learning+1997&hl=en&sa=X&ved=0ahUKEwiomdqfj8TkAhWGslkKHRCbAtoQ6AEIKjAA.

Pradeep, K. R., and N. C. Naveen. 2018. "Lung Cancer Survivability Prediction Based on Performance Using Classification

ASIMTOT: JURNAL KEPENDIDIKAN MATEMATIKA

Volume 6 Nomor 1, Desember 2024 – Mei 2025, halaman 63 – 73

Tersedia Daring pada https://journal.unwira.ac.id/index.php/ASIMTOT

Techniques of Support Vector Machines, C4.5 and Naive Bayes Algorithms for Healthcare Analytics." *Procedia Computer Science* 132: 412–20. https://doi.org/10.1016/j.procs.2018.05.162.

Singh, Awadhesh Kumar et al. 2024. "A Randomized, Double-Blind, Active-Controlled Trial Assessing the Efficacy and Safety of a Fixed-Dose Combination (FDC) of MEtformin Hydrochloride 1000 Mg ER, SItagliptin Phosphate 100 Mg, and DApagliflozin Propanediol 10 Mg in Indian Adults with Type 2 Diabetes: The MESIDA Trial." *International Journal of Diabetes in Developing Countries* (0123456789). https://doi.org/10.1007/s13410-024-01321-9.

Tang, Xianzhe et al. 2020. "Flood Susceptibility Assessment Based on a Novel Random Naïve Bayes Method: A Comparison between Different Factor Discretization Methods." *Catena* 190(March): 104536. https://doi.org/10.1016/j.catena.2020.104536.

Tehrani, Faraz S. et al. 2022. 114 Natural Hazards *Machine Learning and Landslide Studies: Recent Advances and Applications*. Springer Netherlands. https://doi.org/10.1007/s11069-022-05423-7.

Tural Buyuk, Esra, Hatice Uzsen, Merve Koyun, and Reyhan Dönertaş. 2023. "Parental Monitoring Status of the Children with Type 1 Diabetes Mellitus (DM) and Their Quality of Life." *International Journal of Diabetes in Developing Countries* (Dm). https://doi.org/10.1007/s13410-023-01304-2.

Zhang, Lu et al. 2023. "Differences in Nutrition, Handgrip Strength, and Quality of Life in Patients with and without Diabetes on Maintenance Hemodialysis in Xi'an of China." *International Journal of Diabetes in Developing Countries* (0123456789). https://doi.org/10.1007/s13410-023-01282-5.