

## Komparasi Metode Klasifikasi terhadap Data Penderita Penyakit Diabetes Menggunakan Python 3

Ayu Okta Pratiwi<sup>1\*</sup>, Tri Basuki Kurniawan<sup>2</sup>, Edi Surya Negara<sup>3</sup> and Yesi Novaria Kunang<sup>4</sup>

Magister Teknik Informatika, Universitas Bina Darma, Palembang, Sumatera Selatan,  
e-mail: <sup>1\*</sup>ayuoktapratiwi1@gmail.com, <sup>2</sup>tribasukikurniawan@binadarma.ac.id,  
<sup>3</sup>e.s.negara@binadarma.ac.id, <sup>4</sup>yesinovariakunang@binadarma.ac.id

Submitted Date: July 06<sup>th</sup>, 2023  
Revised Date: September 01<sup>st</sup>, 2023

Reviewed Date: August 11<sup>th</sup>, 2023  
Accepted Date: September 01<sup>st</sup>, 2023

### Abstract

*Diabetes is a serious challenge in the world of health, with broad impacts. In an effort to overcome this problem, it is important to analyze the classification of diabetes data to provide valuable insights. This study focuses on the comparison of the two main classification methods, namely Naive Bayes and Support Vector Machine (SVM), in analyzing diabetes data. We use the Python 3 programming language for implementation. The initial study involved the characterization of the dataset, including parameters such as blood pressure and blood glucose levels, which were important factors in the analysis. The preprocessing process is carried out to ensure data quality by overcoming missing or invalid values. After that, the dataset is divided into training and testing subsets. The Naive Bayes and SVM methods are implemented using the scikit-learn library in Python 3. Both models are trained using a training subset and tested on a test subset. The test results show that both methods have good performance in classifying diabetes data, but SVM stands out with higher accuracy. SVM has the ability to handle complex data and find optimal decision boundaries. The Naive Bayes model achieves the highest accuracy of 78.13% on 70% training data and 30% testing data, while the SVM model achieves 79.63% on 90% training data and 10% testing data. Overall, this study provides an in-depth understanding of the effectiveness of both methods in the context of classifying data on diabetics.*

*Keyword: Classification; Comparison; Diabetes; Python 3; Naive Bayes; SVM (Support Vector Machine)*

### Abstrak

Penyakit diabetes adalah tantangan serius dalam dunia kesehatan, dengan dampak yang luas. Dalam usaha untuk mengatasi masalah ini, analisis klasifikasi data penderita diabetes menjadi penting untuk memberikan wawasan yang berharga. Penelitian ini fokus pada perbandingan dua metode klasifikasi utama, yaitu Naive Bayes dan Support Vector Machine (SVM), dalam menganalisis data diabetes. Kami menggunakan bahasa pemrograman Python 3 untuk implementasi. Awal penelitian melibatkan karakterisasi dataset, termasuk parameter seperti tekanan darah dan kadar glukosa darah, yang menjadi faktor penting dalam analisis. Proses preprocessing dilakukan untuk memastikan kualitas data dengan mengatasi nilai yang hilang atau tidak valid. Setelah itu, dataset dibagi menjadi subset pelatihan dan pengujian. Metode Naive Bayes dan SVM diterapkan menggunakan pustaka scikit-learn dalam Python 3. Kedua model ini diberi pelatihan menggunakan subset pelatihan dan diuji pada subset pengujian. Hasil pengujian menunjukkan bahwa kedua metode memiliki performa yang baik dalam mengklasifikasikan data diabetes, namun SVM menonjol dengan akurasi lebih tinggi. SVM memiliki kemampuan untuk menangani data kompleks dan menemukan batas keputusan yang optimal. Model Naive Bayes mencapai akurasi tertinggi 78.13% pada 70% data training dan 30% data pengujian, sedangkan model SVM mencapai 79.63% pada 90% data training dan 10% data pengujian. Secara keseluruhan, penelitian ini memberikan pemahaman mendalam tentang efektivitas kedua metode dalam konteks klasifikasi data penderita diabetes.



Keywords: *Klasifikasi; Komparasi; Diabetes; Python 3; Naive Bayes; SVM (Support Vector Machine)*

## 1. Pendahuluan

Penyakit diabetes merupakan penyakit degeneratif yang setiap tahun presentasi penderitanya selalu meningkat (Elfaladonna & Rahmadani, 2019). Penyakit diabetes merupakan masalah kesehatan di seluruh dunia dengan perkiraan dari 120 juta penderita. Angka tersebut akan semakin meningkat jika adanya ketidaktahuan masyarakat umum tentang faktor-faktor yang dapat memicu terkena penyakit diabetes (Putri et al., 2021). Diabetes dapat memicu berbagai komplikasi penyakit berbahaya lainnya, jika terlambat ditangani (Widodo et al., 2021).

Dalam era digital dan perkembangan teknologi informasi, penggunaan metode klasifikasi dalam menganalisis data penderita diabetes menjadi lebih relevan dan efektif. Metode klasifikasi memungkinkan identifikasi pola dan hubungan antara atribut-atribut yang berkaitan dengan penyakit diabetes, sehingga dapat membantu dalam diagnosis dini, pengobatan yang tepat, dan pencegahan komplikasi yang berpotensi. Klasifikasi merupakan metode dalam ranah data mining yang berguna untuk mendukung perkiraan hasil klasifikasi penyakit diabetes (Nurdiana & Algifari, 2020).

Guna melakukan klasifikasi gejala diabetes secara efisien dan akurat, diperlukan data yang sah dan metode yang handal untuk meminimalkan kesalahan dalam proses klasifikasi. Data ini perlu melalui tahap pemrosesan sebelum dapat digunakan untuk mendiagnosis. Dalam konteks ini, dipilih pendekatan teknologi Naive Bayes dan Support Vector Machine untuk memungkinkan program untuk belajar dari data yang ada dan melakukan klasifikasi kondisi diabetes pada individu. Implementasi teknologi ini dilakukan dengan menggunakan bahasa pemrograman Python 3 dan memanfaatkan alat bantu *Jupyter Notebook* untuk mengolah data dengan lebih efektif.

*Python 3* sebagai salah satu bahasa pemrograman yang populer dan memiliki ekosistem yang kuat, menawarkan berbagai pustaka dan alat untuk analisis data dan implementasi metode klasifikasi. Dengan menggunakan *Python 3*, peneliti dapat mengimplementasikan metode klasifikasi yang efisien dan melakukan komparasi antara metode-metode tersebut.

Salah satu motivasi utama penelitian ini muncul dari studi terkait yang membahas perbandingan algoritma klasifikasi dalam data mining untuk memprediksi kasus penyakit jantung. Dalam penelitian tersebut, ditemukan bahwa tingkat akurasi tertinggi yang berhasil dicapai adalah sebesar 80,38%. Hasil penelitian tersebut mengindikasikan bahwa algoritma Random Forest dan Decision Stump mampu memberikan performa yang superior dalam mengklasifikasikan data pada dataset yang diteliti. Selain itu, algoritma C4.5 dan Naive Bayes juga menunjukkan kinerja yang memuaskan. Namun, metode k-NN tidak mampu memberikan hasil yang memadai dalam implementasinya pada dataset yang sama (Annisa, 2019).

Kemudian, dalam penelitian mengenai klasifikasi penderita diabetes menggunakan algoritma decision tree C4.5. Dalam penelitian ini, algoritma C4.5 diidentifikasi sebagai salah satu pendekatan yang memiliki potensi dalam meramalkan penyakit diabetes. C4.5 Decision Tree merujuk pada metode yang mampu menghasilkan keputusan melalui pembentukan struktur pohon keputusan. Hasil uji coba menunjukkan tingkat akurasi yang signifikan, mencapai 97,12%, dengan tingkat Precision mencapai 93,02%, serta Recall mencapai 100,00% (Hana, 2020).

Dalam situasi ini, penelitian ini bertujuan untuk membandingkan dua pendekatan klasifikasi, yakni Naive Bayes dan SVM (Support Vector Machine), terhadap dataset penderita diabetes. Dengan melakukan perbandingan antara keduanya, tujuannya adalah untuk mengidentifikasi metode yang paling optimal dalam mengklasifikasikan data tersebut.

Naive Bayes adalah sebuah metode klasifikasi probabilistik yang sederhana, menghitung probabilitas dari sejumlah nilai dalam dataset dengan menggabungkan frekuensi dan kombinasi. Algoritma ini menerapkan prinsip teorema Bayes serta mengasumsikan bahwa semua atribut bersifat mandiri atau tidak memiliki ketergantungan yang signifikan, berdasarkan nilai pada variabel kelas (Saleh, 2015). Sementara itu, pendekatan SVM adalah singkatan dari Support Vector Machine. SVM beroperasi dengan mempartisi dua kelompok data kelas melalui penerapan fungsi linear di dalam ruang ciri yang

memiliki dimensi yang tinggi. Metode ini berusaha menemukan garis pemisah optimal (*hyperplane*) yang memaksimalkan margin antara ruang input dan ruang ciri dengan memanfaatkan konsep kernel (Alita et al., 2020)

Riset ini memiliki peluang untuk berkontribusi secara signifikan dalam pengembangan sistem pendukung keputusan untuk mengatasi penyakit diabetes. Dengan memahami dan menggunakan data dengan efisien, penanganan diabetes dapat ditingkatkan, diagnosis awal dapat diimplementasikan, dan tindakan pencegahan yang tepat dapat diambil.

## 2. Metodologi

### 2.1. Naïve Bayes

Klasifikasi Bayesian merujuk pada pendekatan statistik yang berguna untuk mengantisipasi probabilitas inklusi dalam suatu kelas. Metode ini berdasarkan pada teorema Bayes dan menunjukkan kemampuan klasifikasi yang mirip dengan *decision tree* dan jaringan saraf (*neural network*). Secara praktis, klasifikasi Bayesian telah terbukti memiliki kecepatan dan akurasi yang tinggi saat diterapkan pada database berukuran besar. Pendekatan statistik Bayes ini merupakan metode yang memungkinkan untuk melakukan inferensi dan deduksi dalam konteks klasifikasi (Annur, 2018).

Pengklasifikasi Bayes merupakan salah satu bentuk klasifikasi statistik, yang memiliki kapabilitas untuk meramalkan probabilitas keanggotaan suatu data tuple dalam kelas tertentu, melalui perhitungan probabilitas yang relevan. Dasar dari pengklasifikasi Bayes terletak pada teorema Bayes yang pertama kali ditemukan oleh Thomas Bayes pada abad ke-18. Dalam konteks perbandingan algoritma klasifikasi, ditemukan metode sederhana yaitu *Bayesian Naïve* atau yang lebih dikenal sebagai *Naïve Bayes classifier*. Penggunaan *Naïve Bayes classifier* terbukti memberikan tingkat akurasi dan kecepatan yang tinggi terutama ketika diaplikasikan pada dataset yang besar. Karena tingkat akurasi yang tinggi dan perhitungan yang relatif sederhana, metode ini sering digunakan dalam pemecahan masalah dalam bidang pembelajaran mesin.

### 2.2. Support Vector Machine

Support Vector Machine (SVM) adalah suatu metode yang tergolong dalam pembelajaran

berbimbing (*supervised learning*), yang menganalisis data serta mengidentifikasi pola guna keperluan klasifikasi. SVM mengharuskan teks diubah menjadi vektor sebelum dimanfaatkan untuk klasifikasi. Konsep inti dari SVM adalah mencari permukaan keputusan (*Hyperplane*) yang optimal bagi setiap titik data. Untuk melatih mesin dengan dukungan vektor, atau yang lebih dikenal sebagai Support Vector Machine (SVM), diperlukan penyelesaian masalah *Quadratic Programming (QP)* yang berskala besar. *Quadratic Programming* adalah persoalan matematika yang bertujuan menemukan vektor  $x$  yang meminimalkan fungsi kuadrat. Dengan memanfaatkan pemisahan kelas melalui hiperplane, setiap kelas positif, netral, dan negatif dapat diisolasi berdasarkan wilayah masing-masing. Hal ini memungkinkan penentuan kelas data baru berdasarkan wilayah positif, netral, atau negatif yang sesuai (Supriyatna and Mustika 2018).

*Support Vector Machine (SVM)* merupakan salah satu metode yang sudah banyak diterapkan untuk berbagai jenis penelitian dibidang data dan text mining karena telah mampu menunjukkan performa yang lebih baik. Guna penelitian ini adalah mendapatkan model klasifikasi yang mempunyai akurasi tinggi atau error yang kecil dalam melakukan klasifikasi kualitas pengelasan. (Qisthiano et al., 2023).

### 2.3. Klasifikasi

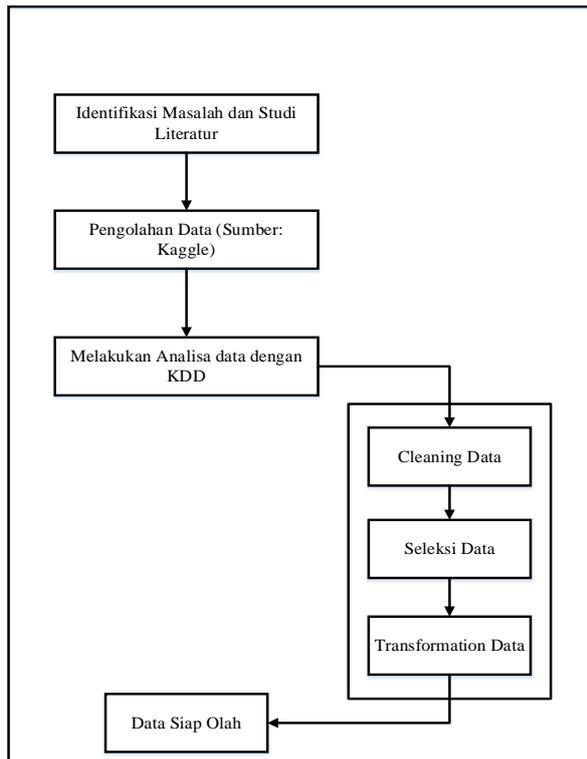
Klasifikasi merupakan salah satu bagian dari data mining. Algoritma klasifikasi dalam data mining bermacam-macam model. Karena setiap model yang ada di algoritma klasifikasi tidak sama, maka akurasinya tentu akan berubah. Untuk mengetahui baik tidaknya sebuah algoritma klasifikasi, indikatornya adalah tingkat akurasi (Faid et al., 2019).

### 2.4. Python

*Python* merupakan contoh bahasa pemrograman tingkat tinggi (*High-Level Language*). Bahasa pemrograman ini menempati peringkat kelima dalam daftar bahasa pemrograman paling banyak diadopsi secara global. Seperti halnya bahasa pemrograman lainnya, Python juga memiliki keunggulan dan keterbatasan yang tidak dapat diabaikan (Trisno, 2016).

## 2.5. Kerangka Pemikiran

Alur penelitian ini dimulai dengan melakukan pengumpulan data yang dibutuhkan untuk penelitian. Data tersebut diperoleh dari berbagai sumber. Setelah data terkumpul, maka dilakukan pengolahan data. Jika data yang telah diolah sudah cukup, maka dilakukanlah analisis. Namun jika data dirasa belum cukup, maka akan dilakukan kembali pengumpulan data.



Gambar 1. Kerangka Pemikiran

1. Proses pertama peneliti akan melakukan identifikasi masalah serta studi pustaka terhadap penelitian terdahulu mengenai metode dan jenis penelitian yang digunakan.
2. Selanjutnya peneliti mulai melakukan pengumpulan data ialah dataset mengenai penderita penyakit diabetes.
3. Dataset tersebut lalu diolah dan dianalisa dengan tahapan KDD yang dimulai dari cleaning, interation, selection, transformation hingga ke tahapan minning.
4. Ketika data siap diolah maka tahapan selanjutnya adalah proses prediksi dengan metode Naïve Bayes, untuk diambil akurasi dan output prediksi.
5. Hasil klasifikasi yang sudah diolah dengan Bahasa Pemrograman Python 3 serta tools

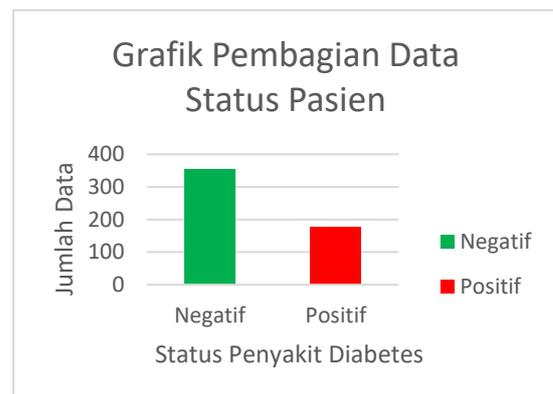
jupyter notebook, serta dijadikan perbandingan untuk mendapatkan akurasi terbaik antara metode Naïve Bayes dan Support Vector Machine.

6. Hasil akhir merupakan hasil perbandingan dari klasifikasi antara metode Naïve Bayes dan Support Vector Machine.
7. Kesimpulan dan saran.

## 3. Hasil Penelitian

### 3.1 Data Penelitian

Dataset yang diambil dipilih sesuai dengan keperluan peneliti, terdiri dari informasi penderita penyakit diabetes yang akan diproses dan digunakan sebagai data uji serta data latih dari seluruh dataset. Dari dataset tersebut, sejumlah data akan diambil untuk mendapatkan informasi tentang kadar gula, tekanan darah, ketebalan kulit, insulin, berat badan, riwayat diabetes dalam keluarga, usia, dan label data. Label ini akan digunakan untuk memprediksi data yang telah diberi label positif dan negatif.



Gambar 2. Grafik Pembagian Data

Total 768 *record data* yang peneliti kumpulkan dan melewati proses cleaning data maka didapat 532 *record data* yang dapat diolah. Untuk pembagian data terhadap data pasien yang positif diabetes berjumlah 177 pasien sedangkan pasien yang negatif penyakit berjumlah 355 pasien.

### 3.2 Klasifikasi Naïve Bayes

Pemilihan algoritma Naïve Bayes diputuskan berdasarkan kinerja yang memadai dan waktu pemrosesan yang efisien. Pada tahap ini, akan dijelaskan bagaimana cara melakukan klasifikasi data pasien yang sebelumnya telah diproses untuk penyakit diabetes. Data ini akan diolah kembali menggunakan algoritma Naïve

Bayes dengan menggunakan alat bantu Jupyter Notebook dan bahasa pemrograman Python.

```

Accuracy: 0.7812
Confusion matrix: {}
[[95 14]
 [21 30]]
precision recall f1-score support
Negatif 0.82 0.87 0.84 109
Positif 0.68 0.59 0.63 51
accuracy 0.78 160
macro avg 0.75 0.73 0.74 160
weighted avg 0.78 0.78 0.78 160
    
```

Gambar 3. Hasil Pengujian dengan *Naïve Bayes*

Gambaran hasil dalam Gambar 3 mencerminkan akurasi terunggul dari lima percobaan yang menggunakan metode naïve bayes. Dengan menggunakan 532 rekaman data dan membaginya dalam data pelatihan sebanyak 70% serta data pengujian sebanyak 30%, hasil akurasi mencapai 0.7812 atau setara dengan 78.12%. Pembagian data ini diterapkan secara acak melalui sistem dan ditegaskan bahwa dari 532 rekaman data yang terlibat, penerapan metode naïve bayes memberikan hasil akurasi tertinggi sebesar 78.12% berdasarkan lima uji coba yang dilakukan.

### 3.3 Klasifikasi *Support Vector Machine*

Proses ini hampir sama dengan proses sebelumnya, yaitu setelah dataset dari pasien penderita penyakit diabetes telah siap untuk digunakan, maka proses selanjutnya adalah penggunaan *library* untuk melakukan prediksi serta menguji dan mengevaluasi model yang akan digunakan, berikut merupakan *source code* yang akan digunakan dalam memanggil *library* dari *algoritma Support Vector Machine*.

```

Accuracy: 0.7963
Confusion matrix: {}
[[31 5]
 [ 6 12]]
precision recall f1-score support
0 0.84 0.86 0.85 36
1 0.71 0.67 0.69 18
accuracy 0.80 54
macro avg 0.77 0.76 0.77 54
weighted avg 0.79 0.80 0.79 54
    
```

Gambar 4. Hasil Pengujian dengan *Support Vector Machine*

Hasil Gambar 4 merupakan hasil akurasi tertinggi dari 5 kali uji coba dengan menggunakan *support vector machine*, diketahui dengan menggunakan 532 *record data* dan di uji dengan menggunakan data *training 90%* dan data *testing 10%* didapati hasil akurasi sebesar 0.7963 atau 79.63%. Pembagian data dilakukan secara acak melalui sistem dan diketahui bahwa dengan menggunakan 532 *record data* serta menggunakan metode *support vector machine* memiliki akurasi tertinggi sebesar 79.63% dari 5 kali pengujian

### 3.4 Hasil Pengujian

Pada tahap merupakan proses dari evaluasi model pada *source code* sebelumnya terhadap evaluasi model dengan *confussion matrix*. Serta melakukan pengujian terhadap *performance* dari *algoritma Naïve bayes* dan *algoritma support vector machine* pada bahasa pemrograman *python 3* dan data pasien penderita penyakit diabetes yang digunakan.

Hasil pengujian Gambar 3 dan Gambar 4 di atas disusun kedalam tabel hasil *confussion matrix* sebagai berikut ini

Tabel 1. Tabel Hasil Pengujian *Naïve Bayes*

Record Data	Pembagian Data	Hasil Akurasi	Hasil Presentase
532	90:10	0.7222	72.22%
532	80:20	0.7290	72.90%
532	70:30	0.7813	78.13%
532	60:40	0.7746	77.46%
532	50:50	0.7594	75.94%

Tabel 2. Tabel Hasil Pengujian *Support Vector Machine*

Record Data	Pembagian Data	Hasil Akurasi	Hasil Presentase
532	90:10	0.7963	79.63%
532	80:20	0.7196	71.96%
532	70:30	0.7563	75.63%
532	60:40	0.7653	76.53%
532	50:50	0.7669	76.69%

Dari Tabel 1 dan Tabel 2 dapat dilihat bahwa hasil dari *confusion matrix* dari algoritma *naïve bayes* dan *support vector machine* dari hasil tersebut didapati dengan menggunakan algoritma *naïve bayes* memiliki akurasi tertinggi pada pembagian data training 70% dan *data test* 30% hasilnya 78.13%. Sedangkan algoritma *support vector machine* memiliki akurasi tertinggi pada pembagian data training 90% dan data test 10% hasilnya 79.63%.

Pembagian data yang berbeda juga dapat mempengaruhi hasil dari kedua metode ini. Ketika data dibagi secara acak, perbedaan dalam hasil prediksi antara SVM dan Naive Bayes mungkin tidak signifikan. Karakteristik data penyakit diabetes juga dapat mempengaruhi kinerja kedua metode. Jika data memiliki hubungan yang kompleks atau non-linear antara fitur-fitur, SVM mungkin lebih baik dalam menemukan batas keputusan yang optimal. Namun, jika hubungan antara fitur-fitur cenderung sederhana atau linear, Naive Bayes dapat memberikan hasil yang baik.

Ukuran sampel dalam dataset dapat berdampak terhadap hasil akurasi. Jika dataset penyakit diabetes memiliki ukuran sampel yang terbatas, Naive Bayes cenderung memberikan hasil akurasi yang tinggi karena kemampuannya dalam situasi sampel yang kecil. Di sisi lain, SVM mungkin membutuhkan volume data lebih besar untuk membangun model yang presisi.

Pemisahan data yang berbeda juga dapat menghasilkan variasi dalam akurasi antara Naive Bayes dan SVM. Dalam pembagian data acak, perbedaan mungkin tidak begitu mencolok. Namun, jika pembagian data menyebabkan subset yang tak seimbang, seperti proporsi yang tidak merata antara kelas positif dan negatif, akurasi dapat terpengaruh.

Dari hasil-hasil ini, diperoleh simpulan bahwa dalam klasifikasi data penyakit diabetes, akurasi yang dicapai oleh algoritma support vector machine (SVM) lebih unggul dibandingkan dengan metode *naïve bayes*. Hasil pengujian menunjukkan bahwa akurasi tertinggi metode Naive Bayes dicapai ketika menggunakan pembagian data pelatihan 70% dan pengujian 30%. Sementara itu, metode SVM menghasilkan akurasi tertinggi dengan pembagian data pelatihan 90% dan pengujian 10%. Perbedaan ini bisa diakibatkan oleh faktor-faktor seperti jumlah data pelatihan

yang dimanfaatkan untuk melatih model dan jumlah data uji yang digunakan. Metode Naive Bayes mungkin memiliki performa lebih baik dengan dataset pelatihan dan uji yang lebih besar, sedangkan SVM mungkin memerlukan data lebih banyak untuk membangun model yang akurat. Selain itu, perbandingan dalam pembagian data juga memainkan peranan dalam hasil akurasi. Pembagian data 70% untuk pelatihan dan 30% untuk pengujian memberikan lebih banyak data untuk melatih model Naive Bayes, yang mungkin berkontribusi pada kinerja yang lebih baik. Sementara itu, pembagian data 90% untuk pelatihan dan 10% untuk pengujian memberikan lebih banyak data pelatihan bagi SVM, yang dapat mempengaruhi kinerjanya.

#### 4. Kesimpulan

Kesimpulannya, dalam penelitian ini kami berhasil membandingkan metode klasifikasi Naive Bayes dan SVM dalam analisis data penderita diabetes. Meskipun keduanya memberikan hasil yang baik, metode SVM menunjukkan keunggulan dalam mengklasifikasikan data tersebut, dengan menggunakan algoritma *naïve bayes* memiliki akurasi tertinggi pada pembagian data training 70% dan *data test* 30% hasilnya 78.13%. Sedangkan algoritma *support vector machine* memiliki akurasi tertinggi pada pembagian data training 90% dan data test 10% hasilnya 79.63%. Hasil penelitian ini dapat memberikan kontribusi penting dalam pengembangan sistem pendukung keputusan untuk penanganan penyakit diabetes.

#### References

- Alita, D., Fernando, Y., & Sulistiani, H. (2020). Implementasi Algoritma Multiclass Svm Pada Opini Publik Berbahasa Indonesia Di Twitter. *Jurnal Tekno Kompak*, 14(2), 86. <https://doi.org/10.33365/jtk.v14i2.792>
- Annisa, R. (2019). Analisis Komparasi Algoritma Klasifikasi Data Mining Untuk Prediksi Penderita Penyakit Jantung. *Jurnal Teknik Informatika Kaputama (JTik)*, 3(1), 22–28. <https://jurnal.kaputama.ac.id/index.php/JTik/article/view/141/156>
- Annur, H. (2018). Klasifikasi Masyarakat Miskin Menggunakan Metode Naive Bayes. *ILKOM Jurnal Ilmiah*, 10(2), 160–165. <https://doi.org/10.33096/ilkom.v10i2.303.160-165>
- Elfaladonna, F., & Rahmadani, A. (2019). Analisa Metode Classification-Decission Tree Dan



- Algoritma C.45 Untuk Memprediksi Penyakit Diabetes Dengan Menggunakan Aplikasi Rapid Miner. *SINTECH (Science and Information Technology) Journal*, 2(1), 10–17. <https://doi.org/10.31598/sintechjournal.v2i1.293>
- Faid, M., Jasri, M., & Rahmawati, T. (2019). Perbandingan Kinerja Tool Data Mining Weka dan Rapidminer Dalam Algoritma Klasifikasi. *Teknika*, 8(1), 11–16. <https://doi.org/10.34148/teknika.v8i1.95>
- Hana, F. M. (2020). Klasifikasi Penderita Penyakit Diabetes Menggunakan Algoritma Decision Tree C4.5. *Jurnal SISKOM-KB (Sistem Komputer Dan Kecerdasan Buatan)*, 4(1), 32–39. <https://doi.org/10.47970/siskom-kb.v4i1.173>
- Nurdiana, N., & Algifari, A. (2020). Studi Komparasi Algoritma Id3 Dan Algoritma Naive Bayes Untuk Klasifikasi Penyakit Diabetes Mellitus. *Infotech Journal*, 6(2), 18–23.
- Putri, S., Irawan, E., & Rizky, F. (2021). Implementasi Data Mining Untuk Prediksi Penyakit Diabetes Dengan Algoritma C4.5. *Januari*, 2(1), 39–46.
- Qisthiano, M. R., Ruswita, I., & Armilia, P. (2023). Implementasi Metode SVM dalam Analisis Sentimen Mengenai Vaksin dengan Menggunakan Python 3. *13(1)*, 1–7.
- Saleh, A. (2015). Implementasi Metode Klasifikasi Naive Bayes Dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga. *Journal of Informatics, Information System, Software Engineering and Applications (INISTA)*, 2(3). <https://doi.org/10.20895/inista.v1i2.73>
- Trisno, I. B. (2016). *Belajar Pemrograman Sulit ? Coba Python* (Y. Hari (ed.)). Ubhara Manajemen Press.
- Widodo, Y. B., Anggraeini, S. A., & Sutabri, T. (2021). Perancangan Sistem Pakar Diagnosis Penyakit Diabetes Berbasis Web Menggunakan Algoritma Naive Bayes. *Jurnal Teknologi Informatika Dan Komputer MH. Thamrin*, 7(1), 112–123.