



Komparasi Algoritma KNN dan SVM dalam Memprediksi Penyakit Stroke

Rahel Lina Smanjuntak

Universitas Negeri Medan

rahelsimanjuntak12@gmail.com

Rizki Agung Ramadhan

Universitas Negeri Medan

rizkiagungramdhan@mhs.unimed.ac.id

Theresia Romauli Siagian

Universitas Negeri Medan

theresiasiangian11@gmail.com

Vina Anggriani

Universitas Negeri Medan

vinaanggriani81@gmail.com

Jl. William Iskandar Ps. V, Kenangan Baru, Kec. Percut Sei Tuan, Kabupaten Deli Serdang, Sumatera Utara 20221

Abstract. *Stroke is a serious medical condition that affects many people around the world. The ability to predict a person's stroke risk can help in effective prevention, treatment and care. In this study, a comparison between the K-Nearest Neighbors (KNN) and Support Vector Machine (SVM) algorithms was conducted to predict stroke risk. The KNN algorithm is a method that searches for the nearest neighbors among the data points to be predicted and assigns the most common label among its neighbors. Experimental results show that both KNN and SVM can provide fairly accurate stroke predictions. However, from an operational point of view, SVM consistently performed better than KNN in terms of accuracy and precision. This research provides insight into the differences between KNN and SVM algorithms in the context of stroke prediction. The results can provide guidance for researchers and practitioners in choosing the right algorithm to predict stroke risk based on the characteristics of the available datasets.*

Keywords: *stroke disease, prediction, K-Nearest Neighbors, SVM*

Abstrak. Penyakit stroke adalah kondisi medis serius yang mempengaruhi banyak orang di seluruh dunia. Kemampuan untuk memprediksi risiko stroke seseorang dapat membantu dalam mencegah, melakukan pengobatan, dan perawatan yang efektif. Pada penelitian ini dilakukan perbandingan antara algoritma K-Nearest Neighbors (KNN) dan Support Vector Machine (SVM) untuk memprediksi risiko stroke. Algoritma KNN adalah metode yang mencari tetangga terdekat di antara titik-titik data yang akan diprediksi dan memberikan label yang paling umum di antara tetangga-tetangganya. Hasil percobaan menunjukkan bahwa KNN dan SVM dapat memberikan prediksi stroke yang cukup akurat. Namun, dari sudut pandang operasional, SVM secara konsisten memberikan kinerja yang lebih baik daripada KNN dalam hal akurasi dan presisi.

Penelitian ini memberikan wawasan tentang perbedaan antara algoritma KNN dan SVM dalam konteks prediksi stroke. Hasilnya dapat memberikan panduan bagi para peneliti dan praktisi dalam memilih algoritma yang tepat untuk memprediksi risiko stroke berdasarkan karakteristik dataset yang tersedia.

Kata kunci: penyakit stroke, prediksi, K-Nearest Neighbors, SVM

LATAR BELAKANG

Stroke, menurut Organisasi Kesehatan Dunia (WHO), diartikan sebagai hilangnya fungsi atau tugas otak secara tiba-tiba yang terjadi dalam waktu 24 jam atau lebih. Pada tahun 2015, Stroke menempati peringkat kedua sebagai penyebab kematian tertinggi di seluruh dunia, sementara pada tahun 2014, Stroke menjadi penyebab kematian utama di Indonesia. Prevalensi stroke di Indonesia pada tahun 2018, berdasarkan diagnosis dokter pada penduduk yang berusia 15 tahun ke atas, mencapai 10,9% atau diperkirakan sekitar 2.120.362 orang. Berdasarkan data dari World Stroke Organization (WSO), setiap tahunnya terjadi sekitar 13,7 juta kasus stroke baru, dan sekitar 5,5 juta orang meninggal akibat stroke. Kejadian stroke cenderung meningkat seiring bertambahnya usia, dengan sekitar 60% dari seluruh kasus stroke terjadi pada individu yang berusia di bawah 70 tahun, dan sekitar 8% terjadi pada mereka yang berusia di bawah 44 tahun (Lindsay dkk., 2019).

Di Indonesia, hasil riset kesehatan dasar tahun 2020 menunjukkan kecenderungan peningkatan kasus stroke, dengan jumlah kasus mencapai 1,7 juta orang menurut penelitian oleh (Beyer dkk.,2020). Prevalensi kejadian stroke di Provinsi Bali, berdasarkan kelompok usia, menunjukkan bahwa kasus tertinggi terjadi pada mereka yang berusia 75 tahun ke atas dengan prevalensi sebesar 40,1%, sementara kasus terendah terdapat pada kelompok usia 25-34 tahun dengan prevalensi 1,1%. Dilihat dari jenis kelamin, kasus stroke lebih banyak dialami oleh laki-laki, dengan prevalensi sebesar 12,3%, dibandingkan dengan perempuan yang memiliki prevalensi sebesar 9,0%.(Beyer dkk.,2020)

Maka dari itu perlunya pencegahan yang dilakukan untuk menghindari stroke yang lebih parah. Dalam hal itu stroke dapat diprediksi dengan akurat menggunakan Machine Learning. Penelitian sebelumnya stroke dapat diprediksi menggunakan teknik Case Base Reasoning dan Algoritma Probabilistik Symetric yang dapat membantu mendiagnosa gejala awal berbasis Web aplikasi dengan akurasi sebesar 66.66%. Tanpa harus menemui

dokter pasien atau masyarakat dapat memprediksi gejala awal stroke. Pada penelitian ini peneliti mencoba membangun program dengan tujuan membandingkan akurasi yang pada program yang dibangun.(Rachman,2021)

Menggunakan program Machine Learning Support SVM (Support Vector Machine) dan KNN (K-Nearest Neighbors) peneliti ingin melihat seberapa akurat akurasi yang dapat diprediksi. Hasil akurasi ini diharapkan dapat menjadi rujukan untuk memprediksi penyakit stroke.

KAJIAN TEORITIS

A. Stroke

Stroke adalah jenis penyakit serebrovaskular, juga dikenal sebagai kerusakan otak, yang membatasi aliran darah dan merusak kemampuan tubuh untuk mengganti darah yang hilang. Secara umum, pembuluh darah mencegah nutrisi dan oksigen masuk ke telinga yang berlawanan (Nia,2020).Organisasi Kesehatan Dunia (WHO) melaporkan pada tahun 2012 bahwa 51% kasus stroke global disebabkan oleh tekanan darah tinggi, faktor risiko yang paling signifikan untuk Stroke berulang adalah sama dengan faktor risiko Stroke pertama. (Ulfa dkk,2022).

Organisasi Kesehatan Dunia melaporkan bahwa kelemahan wajah yang tiba-tiba dan bahkan mati rasa di lengan, kaki, atau di satu sisi tubuh adalah gejala yang paling sering dialami. Selain itu, fungsi sensorik tubuh hilang, dan sakit kepala yang sangat parah dirasakan, yang dapat menyebabkan pingsan atau tidak sadarkan diri(Yufiz dkk,2022).

B. Data Mining

Data mining ialah sesuatu yang bertujuan menyelesaikan suatu permasalahan dengan menerapkan model matematika agar menemukan pola dari data yang sudah ada sebelumnya secara otomatis. Data mining sering dikatakan pula Knowledge Discovery in Databases (KDD), yang menggambarkan semua proses mengubah sekumpulan data biasa menjadi pengetahuan yang bermanfaat (Bahtiar dkk,2020).

Data mining didefinisikan sebagai penggalian data yang menggunakan metode tertentu untuk mengekstrak sekumpulan data tertentu. Algoritma metodologis, atau teknik data mining yang tepat, secara signifikan berdampak pada data mining dengan memanfaatkan

bidang pengetahuan lain seperti matematika, statistik, dan polagraf (Renaldy dkk,2018). Selain itu, data mining mendefinisikan analisis otomatis pada data dalam jumlah besar untuk mengidentifikasi pola-pola penting yang secara konsisten menghindari deteksi. (Sugarwanto,2018)(Ulfa dkk,2022).

C. K-Nearest Neighbours (KNN)

Metode K-Nearest Neighbor digunakan untuk mengklasifikasikan objek berdasarkan data pembelajaran yang berdekatan atau memiliki selisih nilai yang kecil dengan objek yang dimaksud. Memanfaatkan karakteristik, data uji, dan data objek yang baru diperoleh, algoritma ini mampu mengklasifikasikan data. mengenai karakteristik yang ditentukan oleh jarak Euclidean. Jika contoh pertama adalah $(a_1, a_2, a_3, \dots, a_n)$ dan contoh kedua adalah $(b_1, b_2, b_3, \dots, b_n)$, maka perbandingan berikut ini dapat digunakan untuk menentukan hubungan antara kedua contoh tersebut:

$$d = \sqrt{(b_1 - a_1)^2} + \sqrt{(b_2 - a_2)^2} + \dots + \sqrt{(a_n - b_n)^2}$$

Penjelasan rumus :

A: data uji, yang digunakan untuk mengevaluasi model yang dibangun menggunakan data latih.

B: Model terbaik dipilih menggunakan data pelatihan data ke-n (Yufiz dkk,2022)

D. Support Vector Machine (SVM)

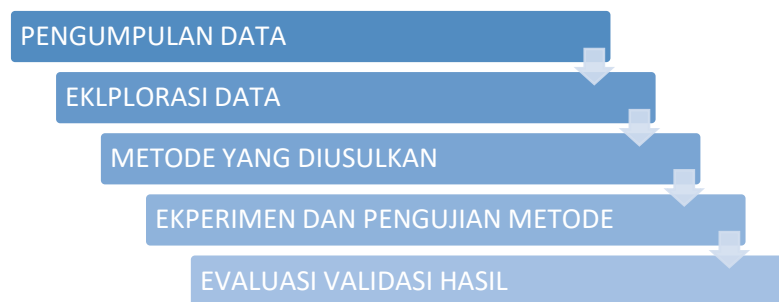
Support Vector Machine (SVM) adalah sebuah sistem pembelajaran yang didasarkan pada teori fungsi linear pada fitur tertentu yang dilatih menggunakan algoritma yang diturunkan dari teori optimal (Rahmad dkk,2021). Boser, Guyon, dan Vapnik mengembangkan SVM yang pertama kali dipublikasikan pada tahun 1992 di Annual Workshop on Computational Learning Theory. Landasan teori SVM berasal dari teori komputer yang sudah ada sebelumnya. Prinsip dasar algoritma ini adalah klasifikasi linier, yang kemudian dimodifikasi agar dapat berfungsi dalam klasifikasi non-linier. Prinsip SVM merupakan pengembangan yang mengklasifikasikan data linier sehingga dapat diselesaikan untuk masalah non-linier. Prinsip dasar ini menerapkan tipu daya kernel pada fitur berdimensi tinggi (Ulfa dkk,2022). Keakuratan data yang diperoleh algoritma SVM ditentukan oleh parameter dan fungsi kernel yang digunakan. Beberapa jenis kernel yang tersedia tercantum pada Tabel 2.1 (Rahmad dkk,2021).

Tabel 1 Tabel Rumus Kernel SVM

Nama Kernel	Fungsi Kernel
Linear (Dot)	$G(x_1, x_2) = x_1'x_2$
Radial Basis Function (RBF)	$G(x_1, x_2) = exp$
Polynomial	$G(x_1, x_2) = \gamma x_1 x_2 + c$

METODE PENELITIAN

Penelitian ini dimulai dengan mengambil data dari Kagle yang berjudul *healthcare-dataset-stroke-data*. Berikutnya data diproses menggunakan Python melalui Google Colab. Pemrosesan data ini dimulai dengan *preprocessing data*. Setelah itu, penerapan Machine Learning menggunakan Algoritma SVM (Support Vector Machine) dan KNN (K-Nearest Neighbors). Data yang sudah yang diklasifikasi menggunakan dua algoritma tersebut dianalisis akurasi dengan tujuan algoritma mana yang paling optimal akurasi.



Gambar 1 Preprocessing data

Dataset yang digunakan adalah data *healthcare-dataset-stroke-data* data yang berasal dari kagle. Situs ini merupakan tempat kumpulan dataset yang siap dipakai untuk mengimplementasi Machine Learning. Data ini berisikan seperti dijelaskan dalam Tabel 2.

Tabel 2. Dataset

No	Atribut	Deskripsi
1.	ID	ID Pasien
2.	Gender	Male/Female
3.	Age	Umur dalam tahun (0-100)

4.	Hypertension	Memiliki hipertensi ya/tidak (0,1)
5.	Heart Disease	Memiliki penyakit jantung - tidak (0,1)
6.	Ever Married	Pernah menikah - ya/tidak (0,1)
7.	Work Type	Jenis pekerjaan – (Private, Self Employed, Govt Job, children, never worked)
8.	Residence Type	Jenis tempat tinggal – (Rural, Urban)
9.	Average Glucose Level	Rata glukosa dalam darah –(0-250 mg/dl)
10.	BMI	Indeks Massa Badan (Kg)
11.	Smoking Status	Status Merokok (formerly smoked, never smoked, smokes or Unknown)
12.	Stroke	Status Stroke ya/tidak (0,1)

HASIL DAN PEMBAHASAN

Pada hasil dan pembahasan peneliti akan merangkum hasil data.

1. Preprocessing Data

Pada tahap ini dilakukan preprocessing data yang bertujuan agar dataset dapat dimodelkan menggunakan Machine Learning. Tahapan Preprocessing data antara lain seperti data pembersihan data, transformasi data, mencari *missing values*, standarisasi hingga *oversampling data*.

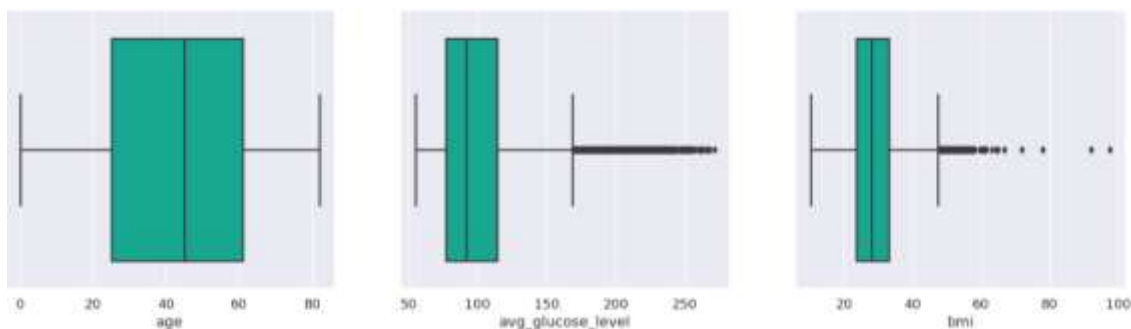
Dalam dataset ini terdapat data yang tidak memiliki pengaruh baik terhadap analisis ini yakni data id. Peneliti membuang data id karena memengaruhi kinerja permodelan sehingga 0,4% KNN. Terlihat hasil dataset yang baru pada gambar dibawah ini.

```

RangeIndex: 5110 entries, 0 to 5109
Data columns (total 11 columns):
#   Column              Non-Null Count  Dtype
---  ---             
0   gender              5110 non-null   object
1   age                 5110 non-null   float64
2   hypertension        5110 non-null   int64
3   heart_disease       5110 non-null   int64
4   ever_married        5110 non-null   object
5   work_type           5110 non-null   object
6   Residence_type      5110 non-null   object
7   avg_glucose_level   5110 non-null   float64
8   bmi                 4909 non-null   float64
9   smoking_status      5110 non-null   object
10  stroke              5110 non-null   int64
dtypes: float64(3), int64(3), object(5)
memory usage: 439.3+ KB
    
```

Gambar 2 Dataset Baru

Setelah membersihkan data kita dapat melihat digambar 2 pada kolom Hypertension, Heart Disease, dan Stroke memiliki tipe data integer tetapi diketahui bahwa data tersebut merupakan variabel kategorikal. Maka dari itu jenis data-data terkait diubah menjadi jenis data object. Setelah itu, dataset kami coba eksplorasi lagi dan peneliti melihat ada *Outliers*(pencilan). Pencilan terdapat di variabel Average Glucose Level dan BMI (Terlihat Di Gambar 3).



Gambar 3 Data Outliers (Pencilan)

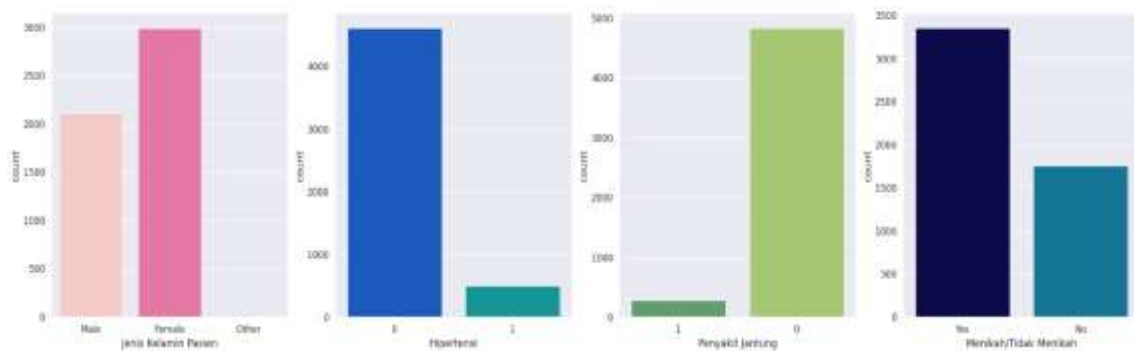
Untuk menangani pencilan dilakukan transformasi logaritma natural terhadap variable Average Glucose Level dan BMI. Transformasi logaritma sendiri dipilih karena teknik umum dalam preprocessing data terutama digunakan untuk mengatasi skewness atau asimetri dalam distribusi data. Terlihat pada kolom yang sudah dilakukan transformasi logaritma pada gambar di bawah (Gambar 4).

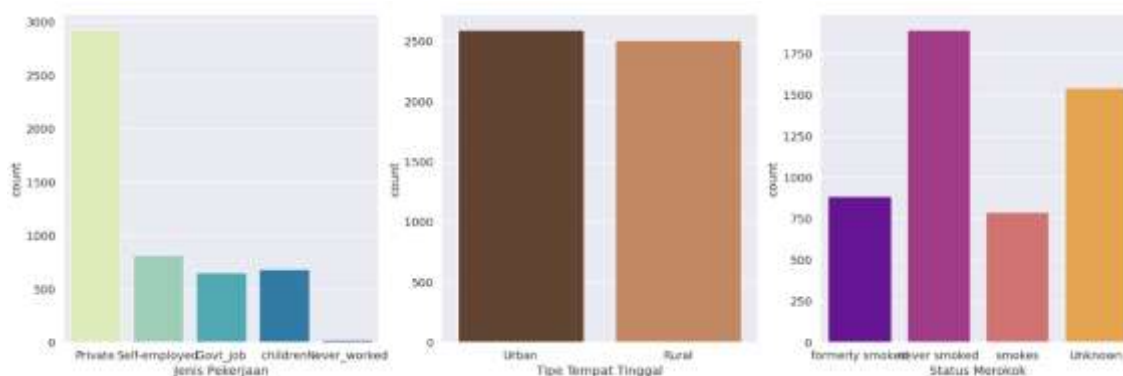
	age	avg_glucose_level	bmi
count	5110.000000	5110.000000	4909.000000
mean	43.226614	4.592465	3.328423
std	22.612647	0.361985	0.265064
min	0.080000	4.009513	2.332144
25%	25.000000	4.346982	3.157000
50%	45.000000	4.520538	3.335770
75%	61.000000	4.736988	3.499533
max	82.000000	5.604846	4.580877

Gambar 4 Transformasi Algoritma

Berdasarkan gambar 4 pencilaan sudah ditangani dengan baik maka selanjutnya peneliti akan menampilkan countplot untuk variable target 'stroke' untuk melihat seimbang atau tidaknya dataset. Didapatkan hasil pada data ini sangat sedikit pasien yang menderita stroke. Setelah melihat countplot dari variabel target dilakukan juga countplot terhadap fitur kategorial.

Diagram Batang Jumlah untuk Berbagai Fitur Kategorikal

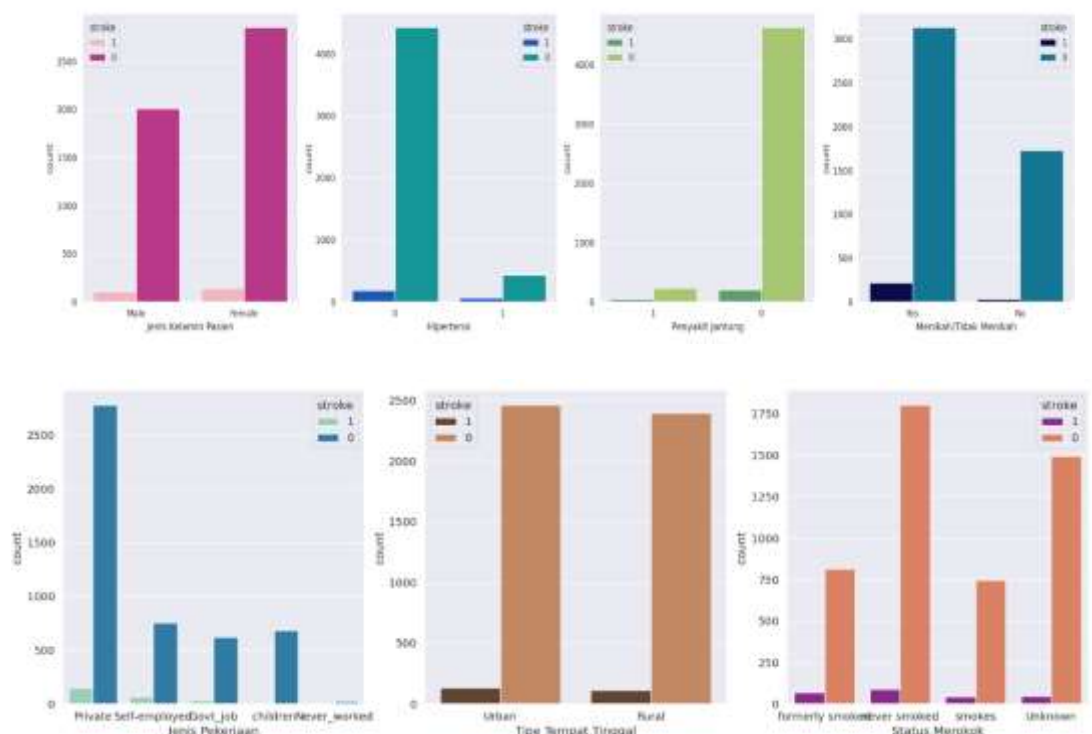




Gambar 5 Countplot Fitur Kategorial

Terdapat kategori “other” di kolom jenis kelamin yang berjumlah 1 dan akan membuat analisis ambigu pada fitur kategorial. Maka dari itu kategori other akan dihapus pada kolom jenis kelamin. Setelah itu dilakukan visualisasi berbagai fitur kategorial terhadap variabel target. Salah satu yang dapat kita lihat orang yang tinggal di perkotaan lebih beresiko terkena penyakit stroke. Terlampir dalam gambar di bawah.

Diagram Batang Jumlah untuk Berbagai Fitur Kategorial Berdasarkan Variabel Target



Gambar 6 Diagram Batang Jumlah untuk Berbagai Fitur Berdasarkan Variabel Target

Setelah itu peneliti mencari *missing values*. Terlihat melalui tabel 2 terdapat nilai yang hilang di kolom bmi. Untuk mengatasi ini peneliti akan menggunakan imputasi KNN. Imputasi KNN digunakan untuk mengisi nilai yang hilang dengan memperkirakan nilai berdasarkan nilai-nilai yang ada di sekitarnya. Dalam konteks ini, model K-Nearest Neighbors Regressor digunakan untuk memahami pola hubungan antara fitur-fitur numerik yang sudah ada dan nilai target yang sedang diimputasi. Model ini memperkirakan nilai yang hilang dengan melihat pada data pelatihan yang memiliki fitur-fitur serupa. Pendekatan ini dapat memberikan hasil yang lebih akurat dibandingkan dengan metode imputasi sederhana seperti rata-rata atau median, terutama jika ada hubungan yang kompleks antara fitur-fitur tersebut.

Tabel 2. Pencarian Missing Values

gender	0
age	0
Hypertension	0
Heart_disease	0
Ever_married	0
Work_type	0
Residence_type	0
Avg_glucose_level	0
Bmi	201
Smoking status	0
Stroke	0

Setelah itu dilakukan pengecekan pada nilai yang hilang. Terlihat pada Gambar dibawah variabel bmi sudah tidak memiliki nilai yang hilang.

gender	0
age	0
hypertension	0
heart_disease	0
ever_married	0
work_type	0
Residence_type	0
avg_glucose_level	0
bmi	0
smoking_status	0
stroke	0

Gambar 7 Pengecekan nilai yang hilang

Dari pengimputan data yang hilang telah diinput, setelah itu memasuki proses encoding. Encoding adalah proses mengubah data dari suatu bentuk atau format ke bentuk atau format lain, terutama saat bekerja dengan data kategorikal. One-hot encoding adalah sebuah metode untuk mengonversi variabel kategorikal menjadi bentuk yang dapat digunakan untuk pemodelan, khususnya dalam konteks machine learning. Dengan one-hot encoding, setiap nilai kategori diubah menjadi kolom baru dan diberi nilai 0 atau 1 sesuai dengan keberadaan nilai tersebut.

Setelah dilakukan encoding berikutnya adalah tahap scaling. Scaling adalah proses mengubah rentang nilai dari suatu variabel sehingga variabel tersebut memiliki skala yang serupa atau dapat dibandingkan. Dilakukakannya scaling karna terdapat skala yang berbeda antara fitur kategorial.

	age	avg_glucose_level	bmi
0	1.051242	2.320709	1.027679
1	0.785889	1.980714	0.781547
2	1.626174	0.194204	0.574693
3	0.255182	1.521257	0.791320
4	1.581949	1.567499	-0.581283

Gambar 8 Scaling

Setelah dimilikinya rentang yang sama seperti terlihat di gambar maka data di terapkannya tekni oversampling dengan menggunakan RandomOverSampler dikelas

minoritas. Setelah itu data sudah siap dibagi untuk di modelkan dalam algoritma Machine Learning.

2. Data Splitting

Setelah melakukan oversampling, peneliti memisahkan data yang sudah di oversampling menjadi data pelatihan dan data testing. Dalam hal ini 80 % digunakan sebagai data pelatihan dan 20% data testing. Pembagian ini dengan acak dengan seed yang ditetapkan 42 untuk reproduktibilitas. Pada data pelatihan terdapat 3892 sampel yang termasuk dalam kelas 0 dan 3884 sampel yang termasuk dalam kelas 1. Sedangkan pada data testing terdapat 976 sampel yang termasuk 0 dan 968 sampel yang termasuk dalam kelas 1.

3. Penerapan Algoritma Machine Learning

a.) KNN (K-Nearest Neighbors)

Pada klasifikasi ini peneliti menetapkan tetangga terdekat senilai 2. Nilai yang akan dihitung adalah nilai akurasi, presisi, f-1, recall hingga matrik konfusi agar dapat dilihat seberapa baik algoritma ini dapat diterapkan. Hasil dari dari permodelan ini adalah melalui gambar berikut

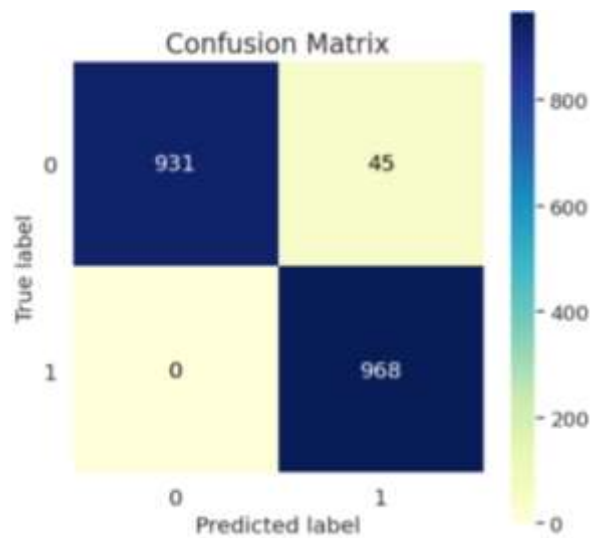


The image shows two screenshots of machine learning results. The left screenshot displays the overall accuracy and ROC AUC score. The right screenshot shows a confusion matrix and summary metrics for the K-Nearest Neighbors classifier.

	precision	recall	f1-score	support
0	1.00	0.95	0.98	976
1	0.96	1.00	0.98	968
accuracy			0.98	1944
macro avg	0.98	0.98	0.98	1944
weighted avg	0.98	0.98	0.98	1944

Gambar 9 Hasil Akurasi

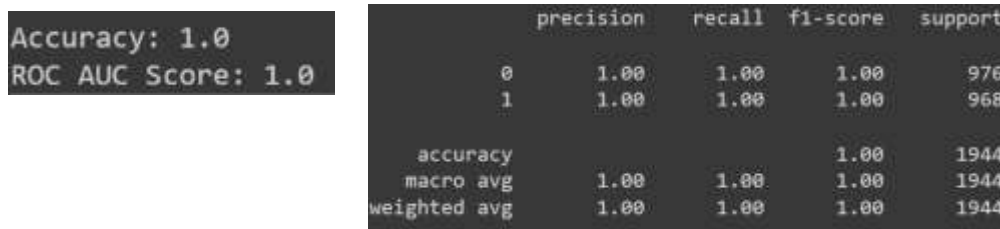
Dapat dilihat akurasi yang didapatkan adalah 0,97% dan juga nilai ROC AUC sebesar 0.97%. Sedangkan untuk konfusi matriks didapat sebanyak 911 nilai positif dan 45 nilai negative.



Gambar 10 Confusion Matrix

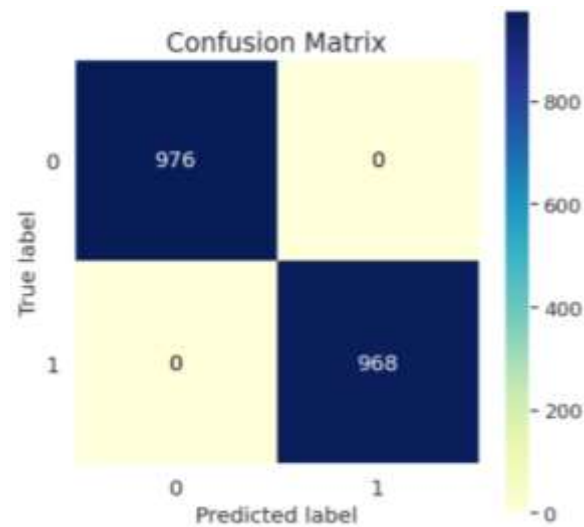
b.) SVM

Sedangkan pada klasifikasi SVM peneliti juga menghitung nilai akurasi, presisi, f-1, recall hingga matrik konfusi agar dapat dilihat seberapa baik algoritma ini dapat diterapkan. Dengan menggunakan Tuning Parameter dengan Metod RandomizeSearchCV Hasil dari dari permodelan ini adalah melalui gambar dibawah. Dapat dilihat akurasi dari permodelan SVM sebesar 100%.



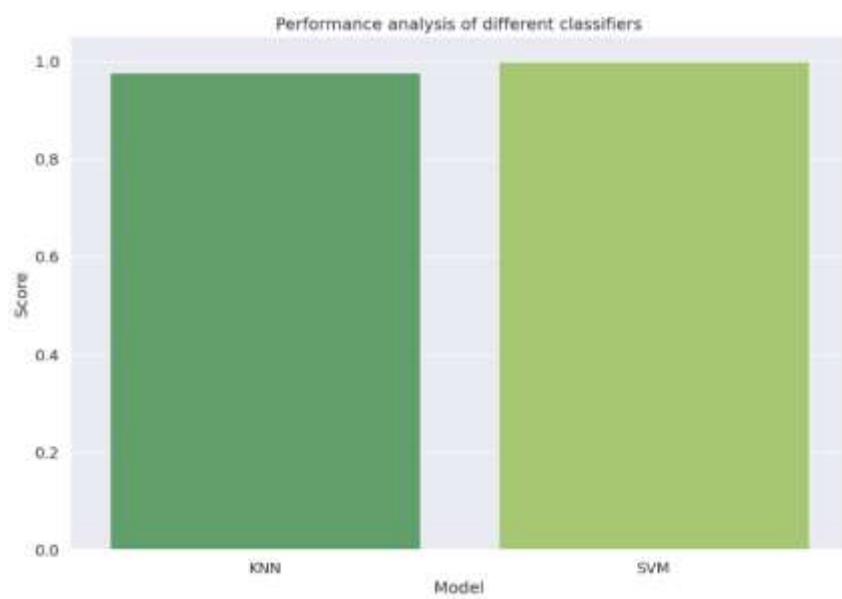
Gambar 11 Hasil Akurasi

Nilai matriks konfusi yang didapat 976 nilai positif dan 0 nilai negative.



Gambar 12 Confusion Matrix

4. Komparasi Model



Gambar 13 Perbandingan Diagram Batang dari Kedua Algoritma

Dari kedua algoritma ini menunjukkan algoritma SVM memiliki akurasi yang paling baik dengan akurasi 100%. Sedangkan KNN, memiliki akurasi 97%

KESIMPULAN DAN SARAN

Berdasarkan hasil uji coba yang dilakukan menggunakan algoritma Support Vector Machine (SVM) dan K-Nearest Neighbors (KNN) dalam bahasa pemrograman Python di Google Colab, dapat disimpulkan bahwa SVM memberikan hasil terbaik dalam hal

akurasi dan kinerja untuk memprediksi penyakit stroke. Algoritma SVM mencapai akurasi 100%, sementara KNN hanya mencapai 97%, menghasilkan perbedaan sebesar 3% lebih rendah dibandingkan SVM. Oleh karena itu, disarankan penggunaan Support Vector Machine (SVM) sebagai pilihan terbaik dalam konteks ini, dibandingkan dengan KNN.

DAFTAR PUSTAKA

Akmal, K., Faqih, A., & Dikananda, F. (2023). PERBANDINGAN METODE ALGORITMA NAÏVE BAYES DAN K-NEAREST NEIGHBORS UNTUK KLASIFIKASI PENYAKIT STROKE. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 7(1), 470-477.

Amelia, U., Indra, J., & Masruriyah, A. (2022). Implementasi Algoritma Support Vector Machine (SVM) untuk Prediksi Penyakit Stroke dengan Atribut Berpengaruh. *Scientific Student Journal for Information, Technology and Science*, 3(2), 254-259.

Atmaja, S. (2016). *SISTEM PENDUKUNG KEPUTUSAN UNTUK DETEKSI DINI RISIKO PENYAKIT STROKE MENGGUNAKAN LEARNING VECTOR QUANTIZATION* (Doctoral dissertation, Universitas Muhammadiyah Gresik).

Azhar, Y., Firdausy, A. K., & Amelia, P. J. (2022). Perbandingan Algoritma Klasifikasi Data Mining Untuk Prediksi Penyakit Stroke. *SINTECH (Science and Information Technology) Journal*, 5(2), 191-197.

Hutama, R. S., Hidayat, N., & Santoso, E. (2018). Sistem Pakar Deteksi Dini Penyakit Stroke Menggunakan Metode Naïve Bayes-Certainty Factor. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 2(11), 4333-4339.

Permatasari, N. (2020). Perbandingan stroke non hemoragik dengan gangguan motorik pasien memiliki faktor resiko diabetes melitus dan hipertensi. *Jurnal Ilmiah Kesehatan Sandi Husada*, 9(1), 298-304.

Rachman, R (2021). Implementasi Case Based Reasoning Mendiagnosa Penyakit Stroke Menggunakan Algoritma Probabilistic Symetric. *Jurnal Informatika*, 8(1),10-16

Yusra, R. N., Sitompul, O. S., & Sawaluddin, S. (2021). Kombinasi K-Nearest Neighbor (KNN) dan Relief-F Untuk Meningkatkan Akurasi Pada Klasifikasi Data. *InfoTekJar: Jurnal Nasional Informatika dan Teknologi Jaringan*, 6(1), 16-21.