Artikel Penelitian

Implementasi Machine Learning Untuk Prediksi Potensi Terkena Penyakit Stroke Menggunakan Algoritma Support Vector Machine (SVM)

*, , ,*

*Mahasiswa Jurusan Sistem Informasi Universitas Andalas, Padang 25163, Indonesia*

| Informasi Artikel |  | **ABSTRACT** |
| --- | --- | --- |
| *Sejarah Artikel:*  Diterima Redaksi: 5 Juni 2023  Revisi Akhir: 5 Juni 2023  Diterbitkan *Online*: 5 Juni 2023 | Penelitian ini bertujuan untuk mengimplementasikan algoritma SVM dalam prediksi penyakit stroke. Dataset yang digunakan terdiri dari atribut gender, ever\_married, work\_type, Residence\_type, smoking\_status, age, hypertension, heart\_disease, avg\_glucose\_level, bmi, dan stroke. Data dibagi menjadi 90% data latih dan 10% data uji. Algoritma SVM digunakan dengan teknik cross validation GridSearchCV untuk memaksimalkan akurasi prediksi. Implementasi pemodelan dilakukan menggunakan framework Streamlit. Pada tahap preprocessing, atribut kategorikal dan numerikal dipisahkan untuk data latih dan data uji. Proses preprocessing melibatkan pengkodean ordinal untuk atribut kategorikal dan penskalaan standar untuk atribut numerikal. Data latih dan data uji yang telah diproses kemudian dimasukkan ke dalam pipeline yang mencakup tahap preprocessing dan model SVM. Selanjutnya, dilakukan teknik cross validation GridSearchCV untuk mencari parameter terbaik untuk model SVM. Parameter yang diuji meliputi C, gamma, dan degree. Dengan melakukan evaluasi menggunakan metrik akurasi, ditemukan hasil akurasi sebesar 90% pada data uji. Untuk implementasi pemodelan, digunakan framework Streamlit. Framework ini memungkinkan interaksi yang mudah dengan model yang telah dilatih. Pengguna dapat memasukkan nilai atribut yang relevan dan mendapatkan prediksi mengenai kemungkinan terjadinya penyakit stroke. Penelitian ini menyediakan kontribusi dalam penggunaan algoritma SVM dalam prediksi penyakit stroke. Hasil eksperimen menunjukkan tingkat akurasi yang memadai, sehingga dapat digunakan sebagai alat bantu dalam penanganan penyakit stroke. Implementasi menggunakan framework Streamlit mempermudah penggunaan model dalam aplikasi praktis. |
| Kata Kunci |
| Stroke  Machine Learning  GridSearchCV  Streamlit  Support Vector Machine |
| Korespondensi |
| E-mail:  [2011521019\_daeng@student.unand.ac.id](mailto:2011521019_daeng@student.unand.ac.id)  [2011522022\_muhammad@student.unand.ac.id](mailto:2011522022_muhammad@student.unand.ac.id)  [2011522008\_faiz@student.unand.ac.id](mailto:2011522008_faiz@student.unand.ac.id)  [2011521016\_sefza@student.unand.ac.id](mailto:2011521016_sefza@student.unand.ac.id) |

# PENDAHULUAN

World Health Organization (WHO) mendefinisikan stroke sebagai suatu kondisi dimana ditemukan defisit neurologis fokal dan global dengan tanda klinis yang dapat memburuk dan berlangsung selama kurang lebih 24 jam serta menyebabkan kematian. Menurut statistik prevalensi stroke yang disediakan oleh Organisasi Stroke Dunia, sekitar 13,7 juta kasus stroke baru dan 5,5 juta kematian terkait stroke terjadi setiap tahunnya [1]. Berbagai belahan dunia memiliki tingkat stroke yang berbeda: di Amerika Serikat, penyakit ini menyerang sekitar 7 juta (3,0 persen), sementara di Cina, penyakit ini menyerang antara 1,8% dan 9,4% populasi di daerah perkotaan. Riset Kesehatan Dasar 2018 menemukan bahwa prevalensi stroke di Indonesia meningkat 7% menjadi 10,9% dibandingkan tahun 2013. Menurut Kementerian Kesehatan RI (2018), daerah perkotaan memiliki prevalensi stroke sebesar 63,9%, sedangkan daerah pedesaan memiliki prevalensi stroke sebesar 36,1% [2].

Dalam pemanfaatan teknologi khususnya di bidang ilmu kesehatan dengan menggunakan pemodelan pembelajaran mesin semakin disesuaikan karena dengan pembelajaran mesin dapat mempermudah dalam prediksi dalam melakukan penanganan penyakit [3]. Sebagai contoh di dalam menangani pasien yang sering terlambat dalam melakukan tindakan, peluang pasien bisa diobati jika sebuah sistem penangan secara otomatis itu dilakukan, paling tidak dalam mengambil keputusan awal saat sebelum kejadian. Sehingga dibuat pemodelan pembelajaran mesin yang mampu bekerja secara otomatis agar digunakan secara langsung dalam membantu pasien untuk menekan risiko [4]. Penggunaan pembelajaran mesin, terbukti telah banyak diterapkan dalam topik klasifikasi dan optimasi dalam membuat sistem cerdas untuk meningkatkan penyedia layanan kesehatan [5], [6]

Latar belakang dari project ini yaitu berdasarkan literatur yang telah dilakukan, pada artikel website sehatnegeriku.kemkes.go.id, penyakit stroke menjadi penyebab kematian tertinggi kedua di dunia dan tertinggi di Indonesia. Menurut survei yang dilakukan oleh perusahan asuransi AIA Group, Indonesia menjadi negara dengan pola hidup terburuk di Asia Pasifik. Pola hidup yang buruk merupakan penyebab utama dari penyakit stroke, contohnya pola tidur, pola makan, makanan dan minuman yang dikonsumsi, merokok, alkohol, obat-obatan, dan lain sebagainya. Kebanyakan penderita stroke mengetahui bahwa dia menderita stroke ketika telah terkena stroke. Kita harus mencegah diri kita agar tidak terkena penyakit stroke dengan melakukan pola hidup yang sehat. Project ini bertujuan untuk mengembangkan aplikasi yang dapat mendeteksi apakah dengan pola hidup yang kita lakukan akan berpotensi terkena penyakit stroke atau tidak.

# METODOLOGI

## Objek Penelitian

Penelitian ini difokuskan pada implementasi algoritma Support Vector Machine (SVM) dalam mendeteksi penyakit stroke. Objek penelitian adalah dataset penyakit stroke yang terdapat di Kaggle dan akan diuji menggunakan model machine learning yang dikembangkan.

## **Metode Pengumpulan Data**

Metode pengumpulan data pada penelitian ini menggunakan penelitian kuantitatif yaitu dataset statistik. Dataset statistik merupakan salah satu metode pada penelitian kuantitatif, dimana metode ini merupakan penggunaan informasi yang sudah tersedia secara publik. Metode ini biasanya lebih cepat karena peneliti hanya membutuhkan akses ke data dan tidak perlu menyebarkan kuesioner di lokasi.

Dataset yang digunakan dalam penelitian ini bersumber dari Kaggle.com dengan judul ‘brain stroke dataset’. Data ini dipilih karena memiliki atribut-atribut yang dibutuhkan seperti gender, usia, riwayat penyakit jantung, hipertensi, gula darah, berat badan, status merokok, dan lain sebagainya. Data ini akan diolah menjadi data yang baik digunakan melalui tahapan preprocessing untuk kemudian dilakukan pemrosesan dan evaluasi terhadap model yang dibuat.

## **Algoritma dan Library Machine Learning**

Dalam pembuatan model machine learning untuk memprediksi penyakit stroke, algoritma Support Vector Machine (SVM) dipilih sebagai pendekatan utama. Alasan algoritma SVM dipilih adalah karena SVM adalah salah satu algoritma klasifikasi yang kuat dan populer dalam machine learning. SVM dapat mempelajari batas keputusan yang kompleks dan non-linear untuk memisahkan kelas-kelas yang berbeda dengan baik. Dalam kasus prediksi stroke, SVM dapat membantu dalam membedakan antara pasien yang berisiko tinggi terkena stroke dan yang tidak.

## **Framework dan Tools Pembangunan Aplikasi Website**

Pembangunan aplikasi berbasis website dilakukan menggunakan framework Streamlit. Streamlit adalah framework dari python yang sifatnya open source. Streamlit juga memiliki kemampuan yang dapat mengubah script menjadi sebuah aplikasi web yang bisa ditampilkan dalam hitungan menit. Hal tersebut membuat sangat mudah dalam pembuatan aplikasi berbasis web terutama bagi yang memiliki sedikit pengetahuan tentang desain front–end web development.

Dengan menggunakan kombinasi metode pengumpulan data, algoritma machine learning yang tepat, serta framework dan tools yang sesuai, diharapkan penelitian ini dapat memberikan kontribusi dalam memprediksi penyakit stroke dengan menggunakan algoritma Support Vector Machine dan mengimplementasikannya dalam aplikasi berbasis website menggunakan framework Streamlit.

* 1. **Tahapan Metodologi**

Pada Gambar 2.1 dibawah, akan ditampilkan tahapan-tahapan metodologi dalam mengerjakan pengerjaan penelitian ini yaitu terkait prediksi penyakit stroke menggunakan algoritma Support Vector Machine (SVM).



# HASIL DAN PEMBAHASAN

**3.1 Data Preprocessing**

Tahap preprocessing yang digunakan pada penelitian ini yaitu data cleaning dan transformasi.

* Data Cleaning  
  Langkah ini dilakukan dengan menghapus baris yang memiliki nilai kosong atau null pada dataset. Hal ini dilakukan agar data yang digunakan dalam pemodelan atau analisis tidak mengandung nilai yang hilang.
* Transformasi

Pada langkah ini dilakukan 2 jenis transformasi karena adanya perbedaan tipe data. Data kategorikal ditransformasi menggunakan OneHotEncoder sehingga mendapatkan data numerikal. Data numerikal ditransformasi menggunakan StandarScaler sehingga mendapatkan skala yang sama di setiap atribut yang memiliki data bertipe numerikal.

Tahap preprocessing membantu dalam memastikan data siap untuk pemodelan, termasuk dalam mengatasi nilai kosong, transformasi atribut kategorikal menjadi numerik, dan standarisasi atribut numerik.

**3.2 Perancangan Model**

Model yang dibangun pada penelitian ini dirancang dan diimplementasikan untuk memprediksi penyakit stroke pada pasien. Support Vector Machine dipilih karena algoritma ini efektif dalam mengatasi masalah dengan ruang fitur yang tinggi, yaitu dataset dengan banyak atribut. Dalam kasus prediksi stroke, terdapat beberapa atribut yang relevan seperti usia, tekanan darah, riwayat penyakit jantung, kadar glukosa rata-rata, dan indeks massa tubuh (BMI). Langkah-langkah perancangan model meliputi:

1. Import Library & Dataset

Langkah pertama adalah melakukan import semua library dan dataset yang dibutuhkan. Library-library ini akan digunakan untuk memanipulasi data, membangun model, dan melakukan evaluasi.

1. Cleaning Data  
   Pada langkah ini dilakukan penghapusan pada baris yang kosong atau bernilai null agar tidak memengaruhi akurasi model.
2. Split Data

Data dibagi menjadi 2 subset yaitu data latih dan data uji. Pada penelitian ini, data dibagi menjadi 90% data latih dan 10% data uji.

1. Transformasi Data

Data dibagi menjadi 2 tipe data yaitu data numerikal dan data kategorikal. Data numerikal ditransformasi menggunakan StandarScaler() dan data kategorikal ditransformasi menggunakan OneHotEncoder().

1. Pipeline Model

Pada penelitian ini, preprocessing dibuat dalam bentuk pipeline agar pemodelan yang dibuat menjadi lebih terstruktur secara keseluruhan. Transformasi data numerikal dan kategorikal digabungkan menjadi sebuah pipeline preprocess dan model Support Vector Machine dengan kernel ‘poly’ menjadi pipeline model. Lalu kedua pipeline tersebut digabung menjadi pipeline pipe.

1. GridSearchCV

GridSearchCV digunakan untuk mencari parameter terbaik untuk model SVM. Kombinasi parameter akan diuji menggunakan pemodelan yang telah digabung dalam sebuah pipeline dalam langkah sebelumnya.

1. Evaluasi Model

Kombinasi parameter yang terbaik dari GridSearchCV akan dievaluasi sehingga mendapatkan nilai akurasi.

**3.3 Pengujian Model**

### 

### Gambar 3.1. Import Library

### 

### Gambar 3.2. Import Library

### 

### Gambar 3.3. Preprocessing data

### 

### Gambar 3.4. Memisahkan Atribut dan Label

### 

### Gambar 3.5. Membagi data menjadi data latih dan data uji

### 

### Gambar 3.6. Menampilkan Atribut

### 

### Gambar 3.7. Transformasi data kategorikal dan numerikal

### 

### Gambar 3.8. Membuat pipeline dari transformasi dan model

## 

### 

### Gambar 3.9. Evaluasi Model Menggunakan SVM dan teknik Cross-Validation

### 

### Gambar 3.10. Visualisasi Model

### 

### Gambar 3.11. Menyimpan model

**3.4 Integrasi Model Menggunakan Framework Streamlit**

| import streamlit as st  import pandas as pd  import pickle  import numpy as np  st.title("Aplikasi Prediksi Stroke")  gender = st.selectbox("Apa Jenis Kelamin Anda?", ['Pria', 'Wanita'])  ever\_married = st.selectbox("Apakah Anda Sudah Menikah?", ['Ya', 'Tidak'])  work\_type = st.selectbox("Apa Pekerjaan Anda?", ['Wirausaha', 'Swasta', 'PNS', 'Anak-anak'])  Residence\_type = st.selectbox("Bagaimana Status Permukiman Tempat Tinggal Anda?", ['Pedesaan', 'Perkotaan'])  smoking\_status = st.selectbox("Apakah Anda Seorang Perokok?", ['Tidak Merokok', 'Tidak Diketahui', 'Pernah Merokok', 'Merokok'])  age = st.number\_input("Berapa Umur Anda?")  hypertension = st.number\_input("Apakah Anda Pernah Mengidap Hipertensi? Ketik 1 untuk ya, 0 untuk tidak")  heart\_disease = st.number\_input("Apakah Anda Punya Penyakit Hati? Ketik 1 untuk ya, 0 untuk tidak")  avg\_glucose\_level = st.number\_input("Berapa Level Rata-Rata Gula Anda?")  bmi = st.number\_input("Berapa Indeks Massa Tubuh Anda?")  model = pickle.load(open("model.pkl","rb"))  data\_input = pd.DataFrame([[gender, ever\_married, work\_type, Residence\_type, smoking\_status,  age, hypertension, heart\_disease, avg\_glucose\_level, bmi  ]],  columns=["gender","ever\_married","work\_type","Residence\_type","smoking\_status",  "age","hypertension","heart\_disease","avg\_glucose\_level","bmi"  ])  data\_input.replace({"Pria":"Male","Wanita":"Female",  "Anak-anak":"children", "PNS":"Govt\_job",  "Swasta":"Private", "Wirausaha":"Self-employed",  "Pedesaan":"Rural", "Perkotaan":"Urban",  "Merokok":"smokes", "Pernah Merokok":"formerly smoked",  "Tidak Merokok":"never smoked", "Tidak Diketahui":"Unknown",  "Ya":"Yes","Tidak":"No"}, inplace=True)  hasil = model.predict(data\_input)  if st.button('Submit'):  result = model.predict(data\_input)[0]  if result == 1:  st.text('Anda diprediksi mengidap penyakit stroke.')  else:  st.text('Anda diprediksi tidak mengidap penyakit stroke.') |
| --- |

Penjelasan source kode:

* Pertama, *import library* yang akan digunakan
* Lalu, menentukan judul dari *web application*
* Selanjutnya, form input untuk tiap atribut yang diperlukan didalam pemodelan. Atribut dapat berupa pilihan ganda atau input numerikal.
* Load *data model* untuk pemodelan yang telah dilatih.
* Menyimpan kolom atribut pemodelan pada variabel data\_input dan membuat kolom berdasarkan atribut tersebut
* Karena dataset berupa bahasa inggris, mengganti inputan setiap atribut ke bahasa indonesia menggunakan **.replace** library **pandas**
* Melakukan pemodelan pada **data\_input**
* Menampilkan hasil prediksi

**3.5 Implementasi Tampilan**

### 

### Gambar 3.12. Tampilan aplikasi berbasis web

# KESIMPULAN

Berdasarkan penelitian yang telah dilakukan, dapat disimpulkan bahwa :

1. Implementasi algoritma SVM pada prediksi penyakit stroke: Melalui penelitian ini, telah berhasil mengimplementasikan algoritma Support Vector Machine (SVM) dalam konteks prediksi penyakit stroke. Dalam penelitian ini, SVM digunakan untuk mengklasifikasikan data berdasarkan atribut yang relevan dengan penyakit stroke, seperti jenis kelamin, status pernikahan, jenis pekerjaan, dan faktor-faktor lainnya. Hasilnya menunjukkan potensi SVM dalam memprediksi kasus stroke dengan akurasi yang cukup tinggi.
2. Peningkatan akurasi melalui teknik cross-validation GridSearchCV: Dengan menerapkan teknik cross-validation GridSearchCV, penelitian ini berhasil meningkatkan akurasi prediksi model SVM. Penggunaan teknik ini memungkinkan eksplorasi parameter yang optimal untuk SVM dan membantu menghindari overfitting atau underfitting. Akurasi sebesar 90% yang diperoleh menunjukkan kemampuan model dalam mengklasifikasikan kasus stroke dengan akurasi yang baik.
3. Penggunaan atribut dan data yang relevan: Melalui pemilihan atribut yang tepat seperti jenis kelamin, status pernikahan, jenis pekerjaan, dan faktor-faktor lainnya, penelitian ini berhasil membangun model prediksi yang dapat memberikan wawasan tentang faktor-faktor risiko yang berhubungan dengan penyakit stroke. Penggunaan atribut ini membantu dalam memahami hubungan antara faktor-faktor tersebut dengan risiko stroke dan dapat berguna dalam mengidentifikasi kelompok populasi yang rentan terhadap penyakit ini.
4. Implementasi menggunakan framework Streamlit: Penelitian ini mengimplementasikan pemodelan prediksi menggunakan framework Streamlit. Dengan demikian, model prediksi dapat diakses dan digunakan dengan mudah melalui antarmuka pengguna yang intuitif. Hal ini memungkinkan para pengguna, seperti tenaga medis atau peneliti, untuk memasukkan data yang relevan dan memperoleh prediksi stroke secara real-time, meningkatkan efisiensi dan aplikabilitas model.

Saran yang dapat diberikan berdasarkan penelitian yang telah dilakukan yaitu :

1. Mengumpulkan lebih banyak data tentang kasus stroke yang mencakup beragam atribut yang relevan. Serta menjelaskan bahwa dengan menggunakan jumlah data yang lebih besar, dapat meningkatkan keandalan dan generalisasi model.
2. Dalam penelitian selanjutnya, dapat dipertimbangkan untuk memasukkan atribut tambahan yang dapat memberikan wawasan yang lebih komprehensif tentang penyebab dan risiko stroke. Serta menjelaskan bahwa atribut seperti riwayat keluarga, gaya hidup, dan faktor genetik dapat memberikan kontribusi yang signifikan dalam prediksi stroke.
3. Melakukan analisis lebih lanjut terhadap atribut yang digunakan dalam pemodelan. Jelaskan bahwa dengan menggunakan teknik seleksi fitur, dapat diidentifikasi atribut yang paling berpengaruh terhadap prediksi stroke. Serta menjelaskan manfaat dari pengurangan kompleksitas model dan fokus pada atribut-atribut yang paling relevan.

# UCAPAN TERIMA KASIH

Puji dan syukur saya panjatkan kepada Tuhan Yang Maha Esa, karena atas berkat dan rahmat-Nya, saya dapat menyelesaikan karya tulis ilmiah ini. Penulisan artikel ini dilakukan dalam rangka memenuhi salah satu syarat untuk menyelesaikan Tugas Besar pada mata kuliah Aplikasi Pembelajaran Mesin di Jurusan Sistem Informasi Universitas Andalas.

Penulis mengucapkan terima kasih kepada :

1. Bapak Febby Apri Wenando, M.Eng dan ibuk Aina Hubby Aziira, M.Eng selaku dosen mata kuliah Aplikasi Pembelajaran Mesin
2. Seluruh anggota Kelompok Terakhir (SVM) mata kuliah Aplikasi Pembelajaran Mesin kelas B
3. Seluruh Mahasiswa/i jurusan Sistem Informasi angkatan 2020

# 

# DAFTAR PUSTAKA

[1] “American Heart Association , 2014;,” no. 2010, pp. 1–8, 2015.

[2] “Gambaran Kemampuan Komunikasi Verbal Pada Pasien Stroke Di Rumah Sakit Umum Daerah Buleleng Bali Tahun 2021 - Repository Politeknik Kesehatan Denpasar.” http://repository.poltekkes-denpasar.ac.id/7167/

[3] J. Heo, J. G. Yoon, H. Park, Y. D. Kim, H. S. Nam, and J. H. Heo, “Machine Learning– Based Model for Prediction of Outcomes in Acute Stroke,” Stroke, vol. 50, no. 5, pp. 1263–1265, May 2019, doi: 10.1161/STROKEAHA.118.024293.

[4] Adelina, V., Ratnawati, D. E., & Fauzi, M. A. (2018). Klasifikasi Tingkat Risiko Penyakit Stroke Menggunakan Metode GA-Fuzzy Tsukamoto. Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer e-ISSN, 2548, 964X.

[5] I. Cholissodin, F. Farisuddin, and E. Santoso, “Klasifikasi Tingkat Resiko Stroke Menggunakan Improved Particle Swarm Optimization dan Support Vector Machine,” Konferensi Nasional Sistem & Informasi, pp. 11–13, 2016

[6] K. Shameer, K. W. Johnson, B. S. Glicksberg, J. T. Dudley, and P. P. Sengupta, “Machine learning in cardiovascular medicine: are we there yet?,” Heart, vol. 104, no. 14, pp. 1156–1164, Jul. 2018, doi: 10.1136/heartjnl-2017-311198.

**BIODATA PENULIS**

**Sefza Auma Tiang Alam**

Adalah mahasiswa Jurusan Sistem Informasi Universitas Andalas angkatan 2020 dengan NIM 2011521016. Lahir di Padang pada tanggal 8 Oktober 2002



**Faiz Abdullah**

Adalah mahasiswa Jurusan Sistem Informasi Universitas Andalas angkatan 2020 dengan NIM 2011522008. Lahir di Bukittinggi pada tanggal 27 Agustus 2001.



**Muhammad Farhan Ananda Mirzah**

Adalah mahasiswa Jurusan Sistem Informasi Universitas Andalas angkatan 2020 dengan NIM 2011522022. Lahir di Padang pada tanggal 2 Maret 2002.



**Daeng Febrino**

Adalah mahasiswa Jurusan Sistem Informasi Universitas Andalas angkatan 2020 dengan NIM 2011521019. Lahir di Lubuk Sikaping pada tanggal 8 Maret 2002