

## OPTIMIZATION OF THE K-NEAREST NEIGHBORS ALGORITHM USING THE ELBOW METHOD ON STROKE PREDICTION

Febri Sutomo<sup>1</sup>, Daffa Ammar Muaafii<sup>2</sup>, Daffa Naufaldi Al Rasyid<sup>3</sup>, Yogie Indra Kurniawan<sup>\*4</sup>, Lasmedi Afuan<sup>5</sup>, Teguh Cahyono<sup>6</sup>, Eddy Maryanto<sup>7</sup>, Dadang Iskandar<sup>8</sup>

<sup>1,2,3,4,5,6,7,8</sup>Informatika, Fakultas Teknik, Universitas Jenderal Soedirman, Indonesia

Email: <sup>1</sup>[febri.sutomo@mhs.unsoed.ac.id](mailto:febri.sutomo@mhs.unsoed.ac.id), <sup>2</sup>[daffa.muaafii@mhs.unsoed.ac.id](mailto:daffa.muaafii@mhs.unsoed.ac.id), <sup>3</sup>[daffa.rasyid@mhs.unsoed.ac.id](mailto:daffa.rasyid@mhs.unsoed.ac.id),  
<sup>4</sup>[yogie@unsoed.ac.id](mailto:yogie@unsoed.ac.id), <sup>5</sup>[lasmedi.afuan@unsoed.ac.id](mailto:lasmedi.afuan@unsoed.ac.id), <sup>6</sup>[teguh.cahyono@unsoed.ac.id](mailto:teguh.cahyono@unsoed.ac.id),  
<sup>7</sup>[eddy.maryanto@unsoed.ac.id](mailto:eddy.maryanto@unsoed.ac.id), <sup>8</sup>[dadang.iskandar@unsoed.ac.id](mailto:dadang.iskandar@unsoed.ac.id)

(Naskah masuk: 16 Januari 2023, Revisi : 01 Februari 2023, diterbitkan: 10 Februari 2023)

### Abstract

Stroke is the second most deadly disease in the world according to WHO. The sufferer has an injury to the nervous system. Because of this, health experts, especially in the field of nursing, need special attention. Technological advances continue to change over time so that information needs are needed in life. Currently the data on stroke sufferers is extensive enough so that adequate information presentation techniques are needed so that the information received is very accurate and in accordance with user needs. Therefore, it is necessary to process data mining on stroke patient data to obtain useful information for users. This study aims to prove the performance of the Elbow Method to produce the optimum k value in the stroke prediction data using the K-Nearest Neighbors (KNN) algorithm. The optimum k value is generated from the Elbow Method which is executed with the Google Collaboratory using the Python programming language. The test results show that the Elbow Method produces the optimum k value at  $k = 7$ . The KNN model that uses the optimum k value from the Elbow Method can increase the accuracy and precision values reaching 6% and 0.12, respectively.

**Keywords:** Data Mining, Elbow Method, K-Nearest Neighbors, Stroke prediction.

## OPTIMASI ALGORITMA K-NEAREST NEIGHBORS MENGGUNAKAN METODE ELBOW PADA PREDIKSI STROKE

### Abstrak

Stroke merupakan penyakit paling mematikan nomor dua di dunia menurut WHO. Penderitanya mengalami cedera pada sistem saraf. Karena hal inilah para pakar kesehatan khususnya dibidang keperawatan memerlukan perhatian khusus.. Kemajuan teknologi terus berubah seiring waktu sehingga kebutuhan informasi sangat dibutuhkan dalam kehidupan. Saat ini data mengenai penderita stroke sudah cukup luas sehingga teknik penyajian informasi yang memadai sangatlah dibutuhkan supaya informasi yang diterima sangat akurat dan sesuai dengan kebutuhan pengguna. Maka dari itu, perlu dilakukan proses data mining pada data penderita stroke untuk mendapatkan informasi yang bermanfaat bagi para pengguna. Penelitian ini bertujuan untuk membuktikan kinerja *Elbow Method* untuk menghasilkan nilai k yang optimum pada data prediksi penyakit stroke menggunakan algoritma *K-Nearest Neighbour* (KNN). Nilai k optimum dihasilkan dari *Elbow Method* yang dieksekusi dengan *Google Colaboratory* yang menggunakan bahasa pemrograman *Python*. Hasil pengujian memperlihatkan bahwa *Elbow Method* menghasilkan nilai k optimum pada  $k = 7$ . Model KNN yang menggunakan nilai k optimum dari hasil *Elbow Methode* dapat meningkatkan nilai *accuracy* dan *precision* masing-masing mencapai 6% dan 0.12..

**Kata kunci:** Data Mining, K-Nearest Neighbors, metode Elbow, Prediksi stroke.

### 1. PENDAHULUAN

Stroke merupakan penyebab utama kematian dan kecacatan secara global. Diagnosis tergantung pada fitur klinis dan pencitraan otak untuk membedakan antara stroke iskemik dan perdarahan intraserebral [1]. Stroke, atau kecelakaan serebrovaskular, melibatkan cedera pada sistem saraf

pusat sebagai akibat dari penyebab vaskular, dan merupakan penyebab utama kecacatan di seluruh dunia [2]. Stroke merupakan kematian tertinggi kedua di seluruh dunia dengan penyebab utama adalah kecacatan. Dari WHO sendiri 70% secara global disebabkan oleh stroke, 87% kematian akibat stroke terjadi di negara-negara berpenghasilan rendah dan

menengah [3]. Stroke menjadi penyebab pertama kematian yang diprediksi pada tahun 2030 (14,4% dari total kematian) dan penyebab ketiga DALY lost (6% dari total DALY) di negara-negara berpenghasilan menengah; dan penyebab kematian yang ketiga (8,2% dari total kematian) dan 8 penyebab utama DALY lost (2,8% dari total DALY) di negara-negara berpenghasilan rendah [1], [4]. Masalah penyakit stroke di Indonesia memerlukan perhatian yang serius karena jumlah kasus yang terus meningkat dan mempunyai angka kematian yang tinggi [4].

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 [5]. 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 [6]. Penggunaan pembelajaran mesin, terbukti telah banyak diterapkan dalam topik klasifikasi dan optimasi dalam membuat sistem cerdas untuk meningkatkan penyedia layanan kesehatan [7], [8].

Banyak cara telah dilakukan untuk memprediksi berbagai penyakit dengan membandingkan kinerja teknologi Data Mining prediktif. sebagai proses pemilihan fitur, algoritma analisis komponen prinsip digunakan untuk mengurangi dimensi dan mengadopsi algoritma klasifikasi dalam membangun model klasifikasi [9]. Klasifikasi adalah teknik untuk membentuk model data yang belum diklasifikasikan, maka model dapat digunakan untuk mengklasifikasikan data baru.

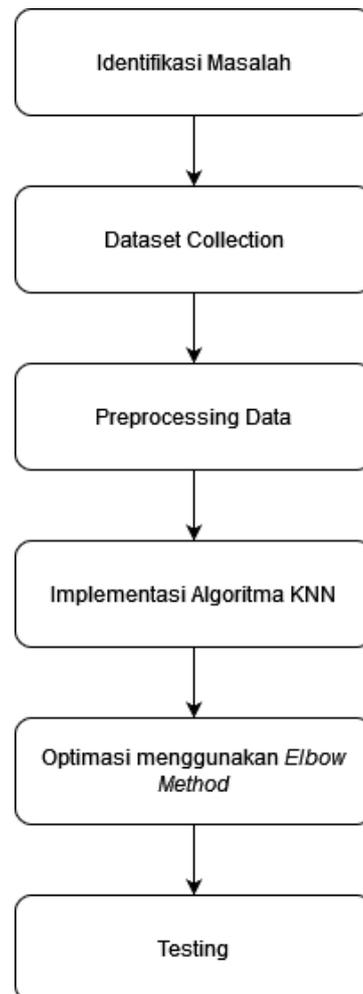
Salah satu algoritma klasifikasi yang cukup populer yaitu algoritma *K-Nearest Neighbour* (KNN). Algoritma KNN adalah sebuah metode untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek yang diuji [10]–[13]. Algoritma KNN akan mengelompokkan hasil perhitungan dengan data latih yang mempunyai kerabat terbanyak dalam nilai jangkauan yang ditentukan. Jarak antara data latih dan data uji dihitung menggunakan persamaan Euclidean.

Dalam meningkatkan tingkat akurasi pada suatu algoritma digunakan suatu teknik optimasi. Salah satu teknik optimasi yang dapat digunakan yaitu *Elbow Method*. Beberapa penelitian sebelumnya yang menggunakan *Elbow Method* sebagai metode optimasi yaitu sebagai berikut [11], [12], [14]. Dari ketiga penelitian tersebut, *Elbow Method* digunakan untuk optimasi algoritma *K-Means* dengan menghasilkan jumlah *cluster* yang optimal.

Penelitian ini bertujuan untuk membuktikan kinerja *Elbow Method* untuk menghasilkan nilai *k* yang optimum pada data prediksi penyakit stroke menggunakan algoritma *K-Nearest Neighbour* (KNN). Nilai *k* optimum dihasilkan dari *Elbow Method* yang dieksekusi dengan *Google Colaboratory* yang menggunakan bahasa pemrograman *Python*.

## 2. METODE PENELITIAN

Pada penelitian ini, menggunakan langkah metodologi yang terdiri dari beberapa tahap dan dapat ditunjukkan oleh Gambar 1.



Gambar 1. Metode Penelitian

### 2.1. Identifikasi Masalah

Identifikasi masalah dalam penelitian ini adalah bagaimana cara melakukan optimasi algoritma *K-Nearest Neighbors* (KNN) menggunakan *Elbow Method* untuk menentukan nilai *k* optimum pada dataset *stroke prediction*.

### 2.2. Pengumpulan Data

Penelitian menggunakan dataset *stroke prediction* yang diunduh dari website kaggle [15].

Data tersebut terdiri atas 5110 *instance* dengan 12 atribut. Atribut dan penjelasannya disajikan pada Tabel 1. Berdasarkan penjelasan tersebut, dataset memiliki enam atribut binary, delapan atribut dalam bentuk kategorikal dan tiga atribut dalam bentuk numerik.

Tabel 1. Deskripsi Dataset *Stroke Prediction*

No	Atribut	Deskripsi
1	<i>ID</i>	Nomor Ide pasien
2	<i>Gender</i>	Jenis kelamin pasien
3	<i>Age</i>	Umur pasien
4	<i>Hypertension</i>	0 – tidak hipertensi 1 - hipertensi
5	<i>Heart disae</i>	0 – tidak memiliki riwayat penyakit jantung 1 – memiliki riwayat penyakit jantung
6	<i>Martial status</i>	Menikah atau tidak menikah
7	<i>Work type</i>	Jenis pekerjaan pasien
8	<i>Residence area</i>	Wilayah tempat tinggal pasien
9	<i>Avg glukose</i>	Rata-rata kadar gula dalam darah
10	<i>BMI</i>	<i>Body Mass Indeks</i> pasien
11	<i>Smoking status</i>	Status merokok pasien
12	<i>Stroke status</i>	0 – stroke 1 – tidak stroke

### 2.3. Preprocessing Data

Data Preprocessing adalah teknik penambahan data awal untuk mengubah data mentah menjadi format dan informasi yang lebih efisien dan berguna [13]. Preprocessing data harus dilakukan dalam proses data mining, karena tidak semua data atau atribut data dalam data digunakan dalam proses data mining. Proses ini dilakukan agar data yang akan digunakan sesuai kebutuhan [16].

Pada tahap ini dilakukan beberapa proses untuk mengolah data sebelum dilakukan klasifikasi menggunakan algoritma KNN. Prosesnya meliputi:

a) *Data Cleaning*

Proses *data cleaning* mencakup antara lain membuang duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data.

b) *Feature Selection*

*Feature Selection* bertujuan untuk memilih *feature* yang berpengaruh dan mengesampingkan *feature* yang tidak berpengaruh dalam suatu kegiatan pemodelan atau penganalisaan data.

c) *Feature Encoding*

*Feature Encoding* adalah proses dimana data yang telah dipilih akan diubah kedalam bentuk dimana data bisa diproses dalam data mining..

### 2.4. Implementasi *K-Nearest Neighbors (KNN)*

Metode *K-Nearest Neighbors* atau biasa disebut KNN merupakan algoritma klasifikasi yang bekerja dengan mengambil sejumlah *K* data terdekat (tetangganya) sebagai acuan untuk menentukan kelas dari data baru. Algoritma ini mengklasifikasikan data berdasarkan *similarity* atau kemiripan atau kedekatannya terhadap data lainnya. Untuk

mendefinisikan jarak antara dua objek *x* dan *y*, digunakan rumus jarak Euclidean seperti pada persamaan 1.

$$d = \sqrt{\sum (X_1 + X_2)} \quad (1)$$

Keterangan:

*d* = jarak Euclidean

*X1* = data training

*X2* = data testing

Kemudian dilakukan algoritma KNN sebagai berikut:

1. Menentukan nilai *K* sebagai parameter (jumlah tetangga paling dekat)
2. Untuk sebuah data testing, hitung jarak Euclidean pada setiap data training.
3. Urutkan semua data berdasarkan jarak Euclidean terkecil secara ascending
4. Eliminasi data sesuai nilai *K*
5. Klasifikasi ditentukan dengan target terbanyak yang ada pada data hasil eliminasi.

### 2.5. Pengujian *Precision, Recall dan Accuracy*

Pada tahap ini, dilakukan pengujian dengan menghitung nilai *precision, recall serta accuracy* dari KNN asli dan KNN optimasi *Elbow Method*. Langkah awal dalam tahap ini adalah dengan membagi data di setiap kasus menjadi 2, yaitu data training atau data latih dan data testing atau data uji. Data training digunakan sebagai data rujukan dalam perhitungan setiap algoritma, sedangkan data testing digunakan untuk menilai prediksi maupun penentuan yang dilakukan oleh setiap algoritma sudah tepat atau tidak

*Precision* merupakan perhitungan terhadap perkiraan proporsi kasus positif yang benar dan dirumuskan dalam persamaan 2:

$$precision = \frac{TP}{TP+FP} \quad (2)$$

Keterangan:

*TP* = *True Positive*

*FP* = *False Positive*

Hasil presisi merupakan jumlah data yang benar dan tepat (*true positive*) dibagi dengan data yang benar dan tepat dan data hasil tidak terduga (*true positive* dan *false positive*).

*Recall* merupakan perhitungan terhadap perkiraan proporsi kasus positif yang diidentifikasi benar dan dirumuskan dalam persamaan 3:

$$recall = \frac{TP}{TP+FN} \quad (3)$$

Keterangan:

*TP* = *True Positive*

*FN* = *False Negative*

*Recall* merupakan jumlah data yg benar dan tepat (*true positive*) dibagi dengan data yang benar dan tepat dan data hasil hilang (*true positive* dan *false negative*).

*Accuracy* merupakan perhitungan terhadap proporsi dari jumlah total prediksi yang benar dan dirumuskan dalam persamaan 4:

$$accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (4)$$

Keterangan:

TP = *True Positive*

TN = *True Negative*

FP = *False Positive*

FN = *False Negative*

Hasil akurasi klasifikasi merupakan jumlah data yang tepat (*true positive* dan *true negative*) dibagi dengan total data.

### 3. HASIL DAN PEMBAHASAN

#### 3.1. Preprocessing Data

Sebelum dilakukan proses data mining, dataset terlebih dahulu dilakukan *preprocessing* data yang meliputi *data cleaning*, *feature selection* dan *feature encoding*.

Pada proses data cleaning dari 1000 *instance* yang digunakan terdapat 58 data yang bernilai *null* pada atribut BMI seperti yang terlihat pada Gambar 1. Data yang bernilai *null* tersebut kemudian dihapus sehingga tersisa 942 *instance*,

```

id          0
gender      0
age         0
hypertension 0
heart_disease 0
ever_married 0
work_type   0
Residence_type 0
avg_glucose_level 0
bmi         58
smoking_status 0
stroke      0
dtype: int64
    
```

Gambar 1. Jumlah Data yang Bernilai *Null*

Setelah dilakukan *data cleaning* selanjutnya dilakukan *feature encoding*. Dari 12 atribut yang ada didapatkan 6 atribut yang merupakan data binary yang terlihat pada Gambar 2 dan didapatkan 3 atribut yang merupakan data *categorical* yang tersaji pada Gambar 3. Atribut bertipe *binary* kemudian ditransformasi menggunakan library *label binarizer* sedangkan atribut bertipe *categorical* ditransformasi menggunakan library *label encoder*

```

['gender',
 'hypertension',
 'heart_disease',
 'ever_married',
 'Residence_type',
 'stroke']
    
```

Gambar 2. Atribut yang Bertipe *Binary*

```

['work_type', 'smoking_status']
    
```

Gambar 3. Atribut yang Bertipe *Categorical*

*Preprocessing* data yang terakhir yaitu *feature selection*. Dari seluruh atribut yang ada ditentukan bahwa atribut *Id* memiliki nilai yang unik untuk setiap data. Atribut tersebut tidak akan mempengaruhi target class. Oleh karena itu atribut *Id* tidak digunakan sehingga tersisa 11 atribut seperti pada Gambar 4.

```

Index(['gender', 'age', 'hypertension', 'heart_disease', 'ever_married',
       'work_type', 'Residence_type', 'avg_glucose_level', 'bmi',
       'smoking_status', 'stroke'],
      dtype='object')
    
```

Gambar 4. Atribut Setelah *Feature Selection*.

#### 3.2. Hasil Implementasi KNN

Dataset yang telah dilakukan *preprocessing* data kemudian dilakukan *splitting* untuk data testing dan data training. Dalam penelitian ini dibuat 3 model data dengan *splitting* data test/data training masing-masing sebesar 20/80, 25/75 dan 30/70.

Pengujian pertama dilakukan dengan menggunakan klasifikasi KNN tanpa menggunakan *Elbow Methode*. Nilai *k* yang digunakan pada percobaan ini yaitu nilai *k* terkecil yaitu 3.

Tabel 2. Precision, Recall dan Accuracy Sebelum Menggunakan *Elbow Methode*

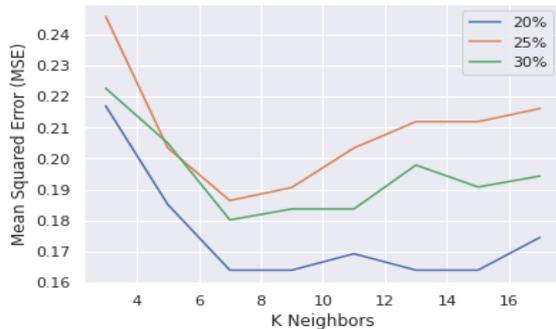
Split	Precision	Recall	Accuracy
20-80	0.43	0.51	78%
25-75	0.42	0.49	75%
30-70	0.44	0.51	78%

Hasil pengujian Knn dengan menggunakan nilai *k* = 3 tersaji pada Tabel 2. Dari tabel tersebut didapatkan nilai akurasi tertinggi sebesar 78 % yang dimiliki oleh model *splitting* data 20-80 dan model *splitting* data 30-70. Untuk model *splitting* data 20-80 didapatkan nilai *precision* sebesar 0.43 dan *recall* sebesar 0.51. Untuk model *splitting* data 30-70 didapatkan nilai *precision* sebesar 0.44 dan *recall* sebesar 0.51. Kemudian pada model *splitting* data 25-75 diperoleh nilai akurasi sebesar 75% dengan *precision* sebesar 0.42 dan *recall* sebesar 0.49.

#### 3.3. Hasil Optimasi Menggunakan *Elbow Method*

Untuk mengoptimasi knn digunakan *Elbow Method* untuk menentukan nilai *k* yang optimum. Nilai *k* optimum diperoleh dari perbandingan nilai *k*

dengan nilai *Mean Squared Error* (MSE). Nilai k dengan nilai MSE minimum nantinya yang ditentukan sebagai k optimum. Dari percobaan ini, nilai k yang digunakan pada *Elbow Method* merupakan bilangan ganjil dengan rentang 3 sampai 20.



Gambar 5. Grafik Nilai MSE Terhadap Nilai K

Grafik nilai MSE terhadap nilai k untuk ketiga model disajikan pada Gambar 5. Dari grafik yang ada pada Gambar 5. terlihat ketiga model tersebut memiliki titik *elbow* (siku) yang sama terhadap nilai k yaitu 7. Pada titik k = 7 ketiga model memiliki nilai MSE minimum. Untuk model splitting data 30-70 memiliki nilai MSE minimum 0.165, model splitting data 25-75 memiliki nilai MSE 0.18 dan model splitting data 20-80 memiliki nilai MSE 0.178. Maka dari itu dapat ditentukan bahwa nilai k optimum untuk model ketiga model tersebut yaitu 7.

Nilai k optimum yang didapat kemudian diaplikasikan pada ketiga model sebelumnya. Dari pengujian yang dilakukan diperoleh nilai precision, recall dan akurasi seperti pada Tabel 3.

Tabel 3. Precision, Recall dan Accuracy Setelah Menggunakan Elbow Methode

Split	Precision	Recall	Accuracy
20-80	0.56	0.51	84%
25-75	0.56	0.47	81%
30-70	0.54	0.49	82%

Dari Tabel 3. terlihat bahwa setelah menggunakan Elbow Method, ketiga model mengalami kenaikan akurasi. Akurasi tertinggi yang didapatkan yaitu sebesar 84% yang dimiliki pada data dengan model splitting data 20-80. Dengan model yang sama didapatkan precision sebesar 0.56 dan recall sebesar 0.51. Untuk model splitting data 25-75 diperoleh akurasi sebesar 81%, dengan precision sebesar 0.56 dan recall sebesar 0.51. Kemudian untuk model split data 30-70 diperoleh nilai akurasi sebesar 82% dengan precision sebesar 0.54 dan recall sebesar 0.49.

#### 4. DISKUSI

Hasil perbandingan optimasi sebelum dan sesudah menggunakan Elbow Method dapat ditunjukkan oleh tabel 4.

Tabel 4. Perbandingan Sebelum dan Sesudah Menggunakan Elbow Method

Split	Precision		Recall		Accuracy	
	KNN	KNN + Elbow	KNN	KNN + Elbow	KNN	KNN + Elbow
20-80	0.43	0.56	0.51	0.51	78%	84%
25-75	0.42	0.56	0.49	0.47	75%	81%
30-70	0.44	0.54	0.51	0.49	78%	82%

Dari data yang tersaji pada Tabel 4. setelah menggunakan nilai k yang didapatkan dari *elbow method*, untuk split data 20/80 akurasi meningkat 6%, precision meningkat 0.13 dan recall bernilai tetap yaitu 0.51. Untuk split data 25/75 nilai akurasi meningkat sebesar 6%, precision meningkat 0.14 dan recall mengalami penurunan sebesar 0.02. Dan terakhir untuk split data 30/70 akurasi meningkat sebesar 5%, precision meningkat 0.10 dan recall mengalami penurunan sebesar 0.02

#### 5. KESIMPULAN

Berdasarkan hasil pengujian yang dilakukan terhadap dataset prediksi penyakit stroke, penggunaan Elbow Method dapat meningkatkan nilai akurasi dan precision. Setelah menggunakan *Elbow Methode* nilai akurasi rata-rata naik sekitar 6% dan *precision* rata-rata naik sekitar 0.12. Selain itu jumlah data training juga mempengaruhi nilai akurasi, model dengan data training terbanyak mendapatkan nilai akurasi paling tinggi.

#### DAFTAR PUSTAKA

- [1] A. Byna and M. Basit, "Penerapan Metode Adaboost Untuk Mengoptimasi Prediksi Penyakit Stroke Dengan Algoritma Naïve Bayes," *Jurnal Sisfokom (Sistem Informasi dan Komputer)*, vol. 9, no. 3, pp. 407–411, 2020, doi: 10.32736/sisfokom.v9i3.1023.
- [2] J. Kissane, J. A. Neutze, and H. Singh, *Radiology fundamentals: Introduction to imaging & technology*. Springer Nature, 2020.
- [3] J. S. Brown, "Ending the doctor-nurse game by enhancing the role of the ward nurse," *BMJ*, p. l698, Feb. 2019, doi: 10.1136/bmj.l698.
- [4] A. Faisal and A. Subekti, "Deep Neural Network untuk Prediksi Stroke," *Jurnal Edukasi dan Penelitian Informatika (JEPIN)*, vol. 7, no. 3, p. 443, 2021, doi: 10.26418/jp.v7i3.50094.
- [5] 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.
- [6] V. Adelina, "Klasifikasi Tingkat Risiko Penyakit Stroke Menggunakan Metode GA-Fuzzy Tsukamoto," Universitas Brawijaya, 2018.
- [7] 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.
- [8] 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.
- [9] M. S. Singh and P. Choudhary, "Stroke prediction using artificial intelligence," in *2017 8th Annual Industrial Automation and Electromechanical Engineering Conference (IEMECON)*, Aug. 2017, pp. 158–161. doi: 10.1109/IEMECON.2017.8079581.
- [10] F. Dhewayani, D. Amelia, D. Alifah, B. Sari, and M. Jajuli, "Implementasi K-Means Clustering untuk Pengelompokan Daerah Rawan Bencana Kebakaran Menggunakan Model CRISP-DM," *Jurnal Teknologi dan Informasi*, vol. 12, no. 1, pp. 64–77, 2022, doi: 10.34010/jati.v12i1.6674.
- [11] A. Winarta and W. J. Kurniawan, "Optimasi cluster k-means menggunakan metode elbow pada data pengguna narkoba dengan pemrograman python," *Jurnal Teknik Informatika Kaputama (JTIK)*, vol. 5, no. 1, pp. 113–119, 2021.
- [12] N. T. Hartanti, "Metode Elbow dan K-Means Guna Mengukur Kesiapan Siswa SMK Dalam Ujian Nasional," *Jurnal Nasional Teknologi dan Sistem Informasi*, vol. 6, no. 2, pp. 82–89, 2020, doi: 10.25077/teknosi.v6i2.2020.82-89.
- [13] H. Said *et al.*, "Penerapan Algoritma K-Nearest Neighbor Untuk Memprediksi Kualitas Air Yang Dapat Dikonsumsi," vol. 21, no. 2, pp. 256–267, 2022.
- [14] D. A. I. C. Dewi and D. A. K. Pramita, "Analisis Perbandingan Metode Elbow dan Silhouette pada Algoritma Clustering K-Medoids dalam Pengelompokan Produksi Kerajinan Bali," *Matrix : Jurnal Manajemen Teknologi dan Informatika*, vol. 9, no. 3, pp. 102–109, Nov. 2019, doi: 10.31940/matrix.v9i3.1662.
- [15] Fedesoriano, "Stroke Prediction Dataset | Kaggle." <https://www.kaggle.com/datasets/fedesorian/stroke-prediction-dataset/code?datasetId=1120859> (accessed Jun. 10, 2022).
- [16] R. R. Rerung, "Penerapan Data Mining dengan Memanfaatkan Metode Association Rule untuk Promosi Produk," *Jurnal Teknologi Rekayasa*, vol. 3, no. 1, p. 89, 2018, doi: 10.31544/jtera.v3.i1.2018.89-98.10.26634/jse.13.3.15515..