

Perbandingan Metode K-Nearest Neighbor Dan Gaussian Naive Bayes Untuk Klasifikasi Penyakit Stroke

Comparison Of K-Nearest Neighbor And Gaussian Naive Bayes Methods For Stroke Disease Classification

Dinda Ulfatul Maula Rachmad¹, Hardian Oktavianto^{2*}, Miftahur Rahman³

¹ Mahasiswa Fakultas Teknik, Universitas Muhammadiyah Jember

Email: dindaulfatul13@gmail.com

²Fakultas Teknik, Universitas Muhammadiyah Jember* Koresponden Author

Email : hardianoktavianto@unmuhjember.ac.id

³Fakultas teknik, universitas muhammadiyah jember

email : miftahurrahman@unmuhjember.ac.id

Abstrak

Stroke adalah kondisi bahaya yang perlu ditangani secepatnya, karena sel otak dapat mati dalam hitungan menit. Kematian secara mendadak mungkin terjadi apabila seorang pasien mengalami yang sangat parah. Penanganan yang cepat dapat mengurangi tingkat kerusakan pada otak dan kemungkinan timbulnya komplikasi. Maka perlu dilakukan prediksi orang tersebut terkena penyakit stroke atau tidak. Salah satu cara untuk memprediksi penyakit stroke yaitu menggunakan klasifikasi. Penyakit stroke perlu diklasifikasi agar dapat memprediksi penyakit dengan akurat. Hasil prediksi yang akurat membantu praktisi kesehatan dalam mengambil keputusan dengan tepat. Metode klasifikasi yang digunakan pada penelitian ini yaitu membandingkan algoritma K-Nearest Neighbor dan Gaussian Naive Bayes. Hasil yang diperoleh dari penelitian ini pada algoritma *K-Nearest Neighbor* didapatkan dengan hasil akurasi sebesar 68,30%, presisi sebesar 67,20% dan recall sebesar 73,34%, sedangkan pada algoritma Gaussian Naive Bayes mendapatkan hasil akurasi sebesar 74,45%, presisi sebesar 74,01% dan recall sebesar 75,71%. Dari perbandingan akurasi, presisi dan recall dapat dilihat bahwa terdapat peningkatan akurasi sebesar 6,15%, presisi sebesar 6,81% dan recall sebesar 2,37%, sehingga membuktikan bahwa kinerja algoritma Gaussian Naive Bayes lebih baik.

Keywords: *Klasifikasi penyakit, Stroke,*

Gaussian Naive Bayes, K-Nearest Neighbor.

Abstract

Stroke is a dangerous condition that needs to be treated as soon as possible, because brain cells can die in a matter of minutes. Sudden death is possible if a patient has a very severe condition. Prompt treatment can reduce the level of damage to the brain and the possibility of complications. So it is necessary to predict whether the person has a stroke or not. One way to predict stroke is to use classification. Stroke needs to be classified in order to predict the disease accurately. Accurate prediction results help health practitioners in making the right decisions. The classification method used in this study is to compare the K-Nearest Neighbor and Gaussian Naive Bayes algorithms. The results obtained from this study on the K-Nearest Neighbor algorithm were obtained with an accuracy of 68.30%, a precision of 67.20% and a recall of 73.34%, while the Gaussian Naive Bayes algorithm obtained an accuracy of 74.45%, precision is 74.01% and recall is 75.71%. From the comparison of accuracy, precision and recall, it can be seen that there is an increase in accuracy of 6.15%, precision of 6.81% and recall of 2.37%, thus proving that the performance of the Gaussian Naive Bayes algorithm is better.

Keywords: *Disease classification, Stroke, Gaussian Naive Bayes, K-Nearest Neighbor.*

1. PENDAHULUAN

Stroke adalah kondisi bahaya yang perlu ditangani secepatnya, karena sel otak dapat mati dalam hitungan menit. Menurut data Riskedasa, prevalensi stroke di Indonesia di tahun 2013 meningkat dibandingkan dengan data 2007 dengan nilai 8,3%, naik hingga 12,1% per 1.000 orang (RI, 2013). Gejala yang biasanya terjadi pada penyakit stroke antara lain pada bagian wajah, lengan yang sering merasah lemas. Efek dari stroke bermacam – macam tergantung pada bagian otak mana yang mengalami gangguan dan seberapa parah gangguan tersebut. Kematian secara mendadak mungkin terjadi apabila seorang pasien mengalami yang sangat parah (Woto, 2014).

Penanganan yang cepat dapat mengurangi tingkat kerusakan pada otak dan kemungkinan timbulnya komplikasi. Maka perlu dilakukan prediksi orang tersebut terkena penyakit stroke atau tidak. Salah satu cara untuk memprediksi penyakit stroke yaitu menggunakan klasifikasi. Penyakit stroke perlu diklasifikasi agar dapat memprediksi penyakit dengan akurat. Hasil prediksi yang akurat membantu praktisi kesehatan dalam mengambil keputusan dengan tepat. K-Nearest Neighbor merupakan teknik yang sederhana, efisiensi dan efektif dalam bidang pengenalan pola, kategori teks, pengolahan objek dan lain lain, karena sederhana dalam pengolahannya dan dapat melakukan training data dalam jumlah yang besar. Penerapan algoritma K-Nearest Neighbor dan Gaussian Naive Bayes sebagai proses klasifikasi yang dirancang untuk mendiagnosa penyakit dengan mengklasifikasi gejala yang dialami pasien dengan memperoleh nilai akurasi tertinggi. Maka dari itu, penulis memutuskan menggunakan algoritma K-Nearest Neighbor dan Gaussian Naive Bayes pada proses klasifikasi dalam melakukan perbandingan dan menentukan algoritma yang memiliki akurasi lebih tinggi dalam proses klasifikasi

2. TINJAUAN PUSTAKA

A. STROKE

Stroke merupakan gangguan yang timbul disebabkan oleh masalah peredaran darah di otak yang mengakibatkan masalah pada jaringan otak sehingga seseorang mengalami kelumpuhan, defisit neurologis yang menyerang mendadak yang diakibatkan oleh *cardiovascular disease* (Agustina, 2019).

Penyakit stroke dapat diminimalisir dengan melakukan pencegahan faktor -faktor yang menyebabkan stroke. Faktor – faktor tersebut dibagi menjadi dua yaitu faktor resiko yang tidak bisa diubah dan faktor resiko yang dapat diubah meliputi :

- a. Faktor resiko yang tidak dapat diubah yaitu:
 1. Usia
 2. Jenis Kelamin
 3. Perempuan
 4. Hereditas atau riwayat keluarga
- b. Faktor resiko yang dapat diubah yaitu:
 1. Hipertensi
 2. Hiperkolesterol dan lemak
 3. Riwayat Diabetes Militus
 4. Riwayat Penyakit Jantung
 5. *Transient Ischemic Attack* (TIA)
 6. Merokok
 7. Obesitas
 8. Kehamilan
 9. Penyalahgunaan obat
 10. Konsumsi alkohol

B. DATA MINING

Data mining dapat mengatasi masalah dengan penganalisisan objek yang telah ada dalam dataset. Sering juga disebut dengan *knowledge discovery in database* (KDD) merupakan aktivitas meliputi pengumpulan dan penggunaan data untuk mendapatkan pola yang teratur atau pola hubungan pada kumpulan data yang berukuran besar (Santoso, 2007). Hasil dari proses *data mining* dapat dijadikan bantuan dalam pengambilan keputusan di masa mendatang.

C. KLASIFIKASI

Klasifikasi adalah proses penemuan model yang menggambarkan dan membedakan kelas

data atau konsep dengan tujuan supaya bisa digunakan untuk memprediksi kelas dari objek yang label kelasnya tidak diketahui (Han & Kamber, 2006). Model yang diturunkan didasarkan pada analisis dari training data. Pada proses klasifikasi terbagi menjadi dua fase yaitu, learning dan testing. Pada fase learning, sebagian data yang telah diketahui kelas datanya (training set) digunakan untuk membentuk model.

D. K-NEAREST NEIGHBOR

K-Nearest Neighbor merupakan salah satu metode yang dapat diterapkan dalam melakukan klasifikasi terhadap suatu data, dengan mencari data yang mempunyai jarak terdekat dengan suatu objek penelitian, sesuai dengan jumlah tetangga terdekatnya yang diinisialisasikan dengan K . Pencarian jarak terdekat biasanya dihitung menggunakan jarak *Euclidean*. Jarak *Euclidean* memiliki persamaan sebagai berikut.

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

Keterangan:

$d(x, y)$ = Jarak Euclidean

x_i = Data Training ke- i

y_i = Data Testing ke- i

E. NAIVE BAYES

Naive Bayes adalah metode klasifikasi populer/umum atau teknik pembelajaran mesin Digunakan untuk klasifikasi teks, ia bekerja dengan baik di banyak bidang, sederhana dan efisien. Namun, Naive Bayes sangat sensitif terhadap pemilihan fitur, terlalu banyak fitur akan menyebabkan Tingkatkan waktu perhitungan dan kurangi akurasi klasifikasi (Azman, 2019).

$$P(H|X) = \frac{P(X|H) \cdot P(H)}{P(X)}$$

Keterangan :

X : Data dengan kelas yang belum diketahui.

H : Hipotesis data X adalah suatu kelas spesifik.

$P(H|X)$: Probabilitas hipotesis H sesuai kondisi X (*posteriori probability*).

$P(H)$: Probabilitas hipotesis H (prior probability).

$P(X|H)$: Probabilitas X sesuai kondisi terhadap hipotesis H .

$P(X)$: Probabilitas X

F. GAUSSIAN NAIVE BAYES

Jika data ditemukan kontinu, distribusi Gaussian Naive Bayes akan digunakan. Berikut ini adalah bentuk distribusi dari Gaussian Naive Bayes.

1. Baca data latih
2. Hitung jumlah dan probabilitas, namun apabila data numerik maka
 - a. Cari nilai mean dan standar deviasi dari masing-masing parameter yang merupakan data numerik.
 - b. Cari nilai probabilitistik dengan cara menghitung jumlah data yang sesuai dari kategori yang sama dibagi dengan jumlah data pada kategori tersebut. Menggunakan persamaan dibawah ini :

$$P(X_i = x_i | Y = y_j) = \frac{1}{\sqrt{2\pi\sigma_{ij}}} e^{-\frac{(x_i - \mu_{ij})^2}{2\sigma_{ij}^2}}$$

Keterangan :

P : Peluang

X_i : Atribut ke i

x_i : Nilai atribut ke i

Y : Kelas yang dicari

y_j : Sub kelas Y yang dicari

μ : Mean, menyatakan rata rata dari seluruh atribut

σ : Deviasi standar, menyatakan varian dari seluruh atribut

G. MEAN

Persamaan yang digunakan untuk menghitung *mean* atau rata-rata hitung dapat dilihat sebagai berikut (Reza Ade, 2018):

$$= \mu \frac{x_1 + x_2 + x_3 + \dots + x_n}{n}$$

Keterangan:

μ : Rata-rata hitung (*mean*)

x_1 : Nilai sample ke $-i$

n : Jumlah sampel

H. STANDAR DEVIASI

Persamaan yang digunakan untuk menghitung standar deviasi sebagai berikut (Reza Ade, 2018):

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \mu)^2}{n - 1}}$$

Keterangan:

σ : standar deviasi

x_i : nilai x ke $-i$

μ : rata-rata hitung

n : jumlah sampel

I. CONFUSION MATRIX

Matriks konfusi merupakan metode yang dapat digunakan untuk pengevaluasian suatu model klasifikasi. *Confusion matrix* berisi nilai dari kelas prediksi yang didapatkan dari model yang dibandingkan dengan kelas yang asli dari dataset dengan kata lain berisi informasi jumlah. Dari kelas aktual dan kelas prediksi pada klasifikasi (Abdillah, 2015). Untuk menghitung nilai akurasi, presisi, recall menggunakan persamaan sebagai berikut :

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN} \times 100\%$$

$$Presisi = \frac{TP}{TP + FP} \times 100\%$$

$$Recall = \frac{TP}{TP + FN} \times 100\%$$

J. CROSS VALIDATION

Cross Validaton, adalah teknik validasi dengan cara membagi data secara acak sesuai k bagian dan proses klasifikasi akan diterapkan pada tiap bagian (Kamber, 2006).

Dalam metode *k-fold*, data disegmentasi secara random ke dalam k partisi yang berukuran sama. Selama proses, salah satu dari partisi dipilih untuk menjadi data testing, sedangkan sisanya digunakan untuk data training (Eko, 2012).

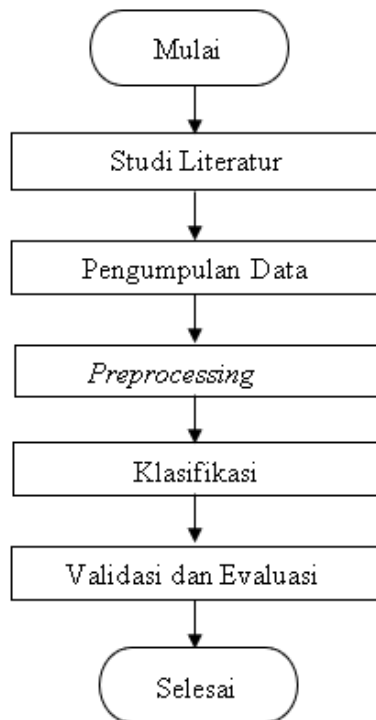
K. PHYTON

Python mendukung berbagai paradigma pemrograman, termasuk Pemrograman Berorientasi Objek, Pemrograman Imperatif, dan Pemrograman Fungsional. Python memiliki sejumlah fitur, salah satunya adalah bahasa pemrograman dinamis yang didukung oleh manajemen memori otomatis. Python sering digunakan sebagai bahasa script, mirip dengan bahasa lain. Namun, dalam praktiknya, python lebih sering digunakan, terutama dalam konteks di mana bahasa skrip tidak sesuai. Python dapat digunakan untuk berbagai tujuan, dan dapat berjalan di berbagai platform dan sistem operasi (Nur, 2018).

L. JUPYTER NOTEBOOK

Jupyter Notebook adalah alat berbasis browser open source yang digunakan sebagai notebook laboratorium virtual untuk mendukung alur kerja, kode, data, dan visualisasi proses penelitian yang terperinci. Dapat dibaca oleh mesin dan manusia, yang kondusif untuk interoperabilitas dan pertukaran akademik (Bernadette M dkk, 2017).

3. METODOLOGI PENELITIAN



Gambar 1 Alur Penelitian

Sumber : Hasil gambar menurut alur penilitan sendiri.

A. STUDI LITERATUR

Tujuan dilakukan study literatur adalah untuk mendapatkn landasan- landasan yang digunakan untuk penelitian ini, study literartur ini bisa didapatkan melalui buku dan jurnal yang berhubungan dengan penelitian pengklasifikasian.

B. PENGUMPULAN DATA

Data yang digunakan diambil dari situs kaggle. Data yang digunakan yaitu 360 data, melakukan pengujian hasil prediksi meliputi 2 klasifikasi yaitu stroke dan tidak stroke. Karena atribut yang ada dalam data set memiliki berbagai macam value. Maka untuk atribut yang memiliki value bukan numerik diubah menjadi numerik.

C. PREPROCESSING DATA

Setelah melakukan pengumpulan data, tahap selanjutnya yaitu preprocessing data sebelum data tersebut pada proses pengujian.

Dalam proses ini yaitu dilakukan pembersihan nilai kosong. Tidak semua atribut digunakan, hanya menggunakan atribut yang dianggap relevan atau atribut yang berpengaruh saja yang akan digunakan pada penelitian ini.

4. HASIL DAN PEMBAHASAN

A. Hasil Klasifikasi K-Nearest Neighbor

Berikut ini akurasi, presisi, recall diperoleh dari hasil pengujian metode *K-Nearest Neighbor* dengan menggunakan k-fold 2 dan 10 skenario pertama dengan nilai K = 3.

Tabel 1 *Confusion Matrix* Algoritma *K-Nearest Neighbor 2-fold*

KELAS AKTUAL	KELAS PREDIKSI	
	Stroke	Tidak stroke
Stroke	59	45
Tidak stroke	23	53

Sumber : Hasil Perhitungan

Perhitungan akurasi, presisi, *recall* :

$$\begin{aligned}
 \text{Akurasi} &= \frac{59 + 53}{59 + 53 + 45 + 23} \times 100\% \\
 &= \frac{112}{180} \times 100\% = 62,22\%
 \end{aligned}$$

$$\begin{aligned}
 \text{Presisi} &= \frac{59}{59 + 45} \times 100\% = \frac{59}{104} \times 100\% \\
 &= 56,73\%
 \end{aligned}$$

$$\begin{aligned}
 \text{Recall} &= \frac{59}{59 + 23} \times 100\% = \frac{59}{82} \times 100\% \\
 &= 71,95\%
 \end{aligned}$$

Tabel 2 *Confusion Matrix* Algoritma *K-Nearest Neighbor 10-fold*

KELAS AKTUAL	KELAS PREDIKSI	
	Stroke	Tidak stroke
Stroke	14	8
Tidak stroke	5	9

Sumber : Hasil Perhitungan

Perhitungan akurasi, presisi, *recall* :

$$Akurasi = \frac{14 + 9}{14 + 9 + 8 + 5} \times 100\% = \frac{23}{36} \times 100\% = 63,89\%$$

$$Presisi = \frac{14}{14 + 5} \times 100\% = \frac{14}{19} \times 100\% = 73,68\%$$

$$Recall = \frac{14}{14 + 8} \times 100\% = \frac{14}{22} \times 100\% = 63,64\%$$

Setelah menghitung akurasi, presisi dan recall dengan menggunakan pengujian k-fold 2 sampai 10 dan K yang digunakan yaitu 3,5,7,9,11. Kemudian setiap iterasi di rata-rata untuk mendapatkan akurasi model seperti pada tabel 3.

Tabel 3 Nilai Akurasi, Presisi, Recall pada Percobaan k

Nilai K	Akurasi	Presisi	Recall
3	66,41%	66,17%	69,53%
5	67,23%	66,45%	70,94%
7	69,78%	67,88%	74,41%
9	68,89%	66,85%	75,46%
11	69,17%	68,65%	76,36%
Nilai rata-rata	68,30%	67,20%	73,34%

Sumber : Hasil Perhitungan

B. Hasil Klasifikasi Gaussian Naive Bayes

Berikut ini perhitungan matrix beserta perhitungan akurasi, presisi, dan recall pada *Gaussian Naive Bayes* dengan menggunakan *Cross Validation*.

Tabel 4 *Confusion Matrix* pada Algoritma GNB dengan k-fold 2

KELAS AKTUAL	KELAS PREDIKSI	
	Stroke	Tidak stroke
Sroke	60	25
Tidak stroke	22	73

Sumber : Hasil Perhitungan

Perhitungan akurasi, presisi, *recall* :

$$Akurasi = \frac{60 + 73}{60 + 73 + 25 + 22} \times 100\% = \frac{133}{180} \times 100\% = 73,89\%$$

$$Presisi = \frac{60}{60 + 25} \times 100\% = \frac{60}{85} \times 100\% = 70,59\%$$

$$Recall = \frac{60}{60 + 22} \times 100\% = \frac{60}{82} \times 100\% = 73,17\%$$

Setelah menghitung akurasi, presisi dan recall dengan menggunakan pengujian k-fold 2 sampai 10, kemudian setiap iterasi di rata-rata untuk mendapatkan akurasi model seperti pada tabel 5.

Tabel 5 Hasil Nilai Rata-Rata Klasifikasi pada Algoritma GNB

k-fold	Akurasi	Presisi	Recall
2	74,45%	73,68%	75,36%
3	70,55%	72,21%	73,12%
4	73,89%	73,25%	75,71%
5	72,22%	71,58%	74,44%
6	73,89%	73,52%	75,12%
7	73,61%	72,72%	74,95%
8	73,65%	74,01%	74,49%
9	73,06%	71,95%	74,16%
10	73,89%	72,45%	74,47%

Sumber : Hasil Perhitungan

Tabel 5 Perbandingan Akurasi, Presisi dan Recall

	Akurasi	Presisi	Recall
<i>K-Nearest Neighbor</i>	68,30%	67,20%	73,34%
<i>Gaussian Naive Bayes</i>	74,45%	74,01%	75,71%
Peningkatan	6,15%	6,81%	2,37%

Sumber : Hasil Perhitungan

Tabel perbandingan hasil akurasi, presisi, dan recall berdasarkan kumpulan data penyakit stroke menggunakan algoritma *K-Nearest Neighbor* dan *Gaussian Naive Bayes*. Berdasarkan tabel hasil akurasi, presisi, dan

recall untuk dataset penyakit stroke dengan algoritma K-Nearest Neighbor didapatkan dengan hasil akurasi sebesar 68,30%, presisi sebesar 67,20% dan recall sebesar 73,34%, sedangkan pada algoritma Gaussian Naive Bayes mendapatkan hasil akurasi sebesar 74,45%, presisi sebesar 74,01% dan recall sebesar 75,71%. Dari perbandingan akurasi, presisi dan recall dapat dilihat bahwa terdapat peningkatan akurasi sebesar 6,15%, presisi sebesar 6,81% dan recall sebesar 2,37%, sehingga membuktikan bahwa kinerja algoritma Gaussian Naive Bayes lebih baik.

5. KESIMPULAN DAN SARAN

A. KESIMPULAN

Berdasarkan penelitian ini untuk melakukan klasifikasi penyakit stroke menggunakan Algoritma *K-Nearest Neighbor* dan Gaussian Naive Bayes maka dapat disimpulkan bahwa :

1. Hasil pengujian akurasi tertinggi pada klasifikasi penyakit stroke diantara algoritma *K-Nearest Neighbor* dan *Gaussian Naive Bayes* adalah pada algoritma *Gaussian Naive Bayes* sebesar 74,45%.
2. Hasil presisi tertinggi pada klasifikasi penyakit stroke diantara algoritma *K-Nearest Neighbor* dan *Gaussian Naive Bayes* adalah pada algoritma *Gaussian Naive Bayes* sebesar 74,01%, sehingga dalam memberikan informasi prediksi pasien terkena penyakit stroke yang pada kenyataannya pasien tidak terkena stroke Algoritma Gaussian Naive Bayes lebih baik dibandingkan K-Nearest Neighbor (KNN).
3. Hasil *recall* tertinggi pada klasifikasi penyakit stroke diantara algoritma *K-Nearest Neighbor* dan *Gaussian Naive Bayes* adalah pada algoritma *K-Nearest Neighbor* sebesar 75,71% sehingga dalam memberikan informasi prediksi pasien tidak terkena penyakit stroke tetapi sebenarnya pasien terkena penyakit stroke algoritma *Gaussian Naive Bayes* lebih baik dibandingkan dengan *K-Nearest Neighbor*.

4. Kinerja algoritma *Gaussian Naive Bayes* lebih baik dibandingkan algoritma *K-Nearest Neighbor*

B. SARAN

Berdasarkan penelitian yang dilakukan, adapun saran yang diberikan untuk penelitian selanjutnya adalah sebagai berikut :

1. Dapat menggunakan metode yang berbeda untuk membandingkan nilai akurasi. Metode yang digunakan bisa dengan konsep algoritma yang sama.
2. Menambahkan seleksi fitur untuk mengetahui attribut yang berpengaruh untuk penyakit stroke.
3. Mengimplementasikan sistem dalam bentuk website maupun android, sehingga dapat membantu dalam mengklasifikasi penyakit.

6. DAFTAR PUSTAKA

- Abdillah, S. (2015), "Penerapan Algoritma Decision Tree C4.5 Untuk Diagnosa Penyakit Stroke Dengan Klasifikasi Data Mining Pada Rumah Sakit Santra Maria Pematang," *Jurnal Teknik Informatika*, Vol 3(4), 1–12.
- Adzhana, A Mahran. Kembang,R. Nugroho,H. (2020), "Penerapan Naive Bayes Gaussian Pada Klasifikasi Jenis Jamur Berdasarkan Ciri Statistik Orde Pertama,"*Jurnal Ilmiah NERO*, Vol.5 No.2
- Ade, P, Reza. (2018). "Penerapan Naive Bayes Classifier dengan Gaussian Function Untuk Menentukan Kelompok UKT". *Jurnal Ilmiah Informatika Global*, Vol.09, No.2, Desember 2018.
- Adelina, V, Eka.R, Ali.F.M. (2018), "Klasifikasi Tingkat Risiko Penyakit Stroke Menggunakan Metode GA-Fuzzy Tsukamoto," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, Vol. 2, No. 9, September2018, hlm. 3015-3021.
- Agustina, B.D. (2019), "Asuhan Keperawatan Gawat Darurat Pada Pasien Dengan Stroke Diinstalasi Gawat Darurat Rumah Sakit

Umum Daerah Sleman Yogyakarta,”
Skripsi, A.Md.Kep.,

Bernadette, M, Irene, V, Milena, S, Christine, L.
(2017), “*Using the Jupyter Notebook as a
Tool for Open*”

Science Jayanti, A.A. (2015), “Hubungan
Hipertensi dengan kejadian stroke di
Sulawesi Selatan tahun 2013,” Skripsi,
S.Kes, Universitas Islam Negeri Syarif
Hidayatullah, Jakarta

Han, J, Kamber, M (2006). *Data Mining:
Concepts and Techniques Second*. Morgan
Kaufman Publishers.

Y. A. Gerhana, I. Fallah, W. B. Zulfikar, D. S.
Maylawati, and M. A. Ramdhani,
“*Comparison of naive Bayes classifier and
C4.5 algorithms in predicting student study
period*,” J. Phys. Conf. Ser., vol. 1280, no.
2, 2019