

**PENERAPAN METODE NAÏVE BAYES UNTUK KLASIFIKASI
PENYAKIT STROKE**

(Skripsi)

Oleh

**NABILLA YOLANDA PARAMITHA
NPM. 1917031079**



**JURUSAN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2023**

ABSTRACT

APPLICATION OF NAÏVE BAYES METHOD FOR STROKE DISEASE CLASSIFICATION

By

NABILLA YOLANDA PARAMITHA

Classification is a method to determine a member into a certain predetermined class. The method used in the classification in this research is Naïve Bayes. This study aims to classify stroke patients using the Naive Bayes method for several different training and testing data proportions. The data used is stroke patient data with the class affected by stroke or not affected by stroke and using ten independent variables. Classification results with the largest accuracy value of 80% were obtained when the proportion of training and testing data was 80:20.

Keywords: Classification, Naïve Bayes, Stroke

ABSTRAK

PENERAPAN METODE NAÏVE BAYES UNTUK KLASIFIKASI PENYAKIT STROKE

Oleh

NABILLA YOLANDA PARAMITHA

Klasifikasi adalah metode untuk menentukan sebuah anggota masuk ke dalam suatu kelas tertentu yang telah ditentukan sebelumnya. Metode yang digunakan dalam pengklasifikasian pada penelitian ini adalah Naïve Bayes. Penelitian ini bertujuan untuk mengklasifikasikan pasien penyakit stroke menggunakan metode Naive Bayes untuk beberapa proporsi data *training* dan data *testing* yang berbeda. Data yang digunakan adalah data pasien penyakit stroke dengan kelas terkena stroke atau tidak terkena stroke dan menggunakan 10 variabel bebas. Hasil klasifikasi dengan nilai akurasi terbesar sebesar 80% diperoleh ketika proporsi data *training* dan data *testing* 80:20.

Kata kunci: Klasifikasi, Naïve Bayes, Stroke

**PENERAPAN METODE NAÏVE BAYES UNTUK KLASIFIKASI
PENYAKIT STROKE**

Oleh

NABILLA YOLANDA PARAMITHA

Skripsi

**Sebagai salah satu syarat untuk Mencapai Gelar
SARJANA MATEMATIKA**

Pada

**Jurusan Matematika
Fakultas Matematika dan Ilmu Pengetahuan Alam**



**JURUSAN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2023**

Judul Skripsi : **PENERAPAN METODE NAÏVE BAYES
UNTUK KLASIFIKASI PENYAKIT
STROKE**

Nama Mahasiswa : **Nabilla Yolanda Paramitha**

Nomor Pokok Mahasiswa : **1917031079**

Progam Studi : **Matematika**

Falkutas : **Matematika dan Ilmu Pengetahuan Alam**



MENYETUJUI,

1. Komisi Pembimbing

A handwritten signature in blue ink, appearing to read 'Aang Nuryaman'.

Dr. Aang Nuryaman, S.Si., M.Si.
NIP. 197403162005011001

A handwritten signature in black ink, appearing to read 'Ahmad Faisol'.

Dr. Ahmad Faisol, S.Si., M.Sc.
NIP.198002062003121003

2. Ketua Jurusan Matematika

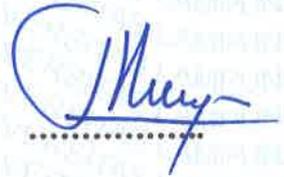
A handwritten signature in blue ink, appearing to read 'Aang Nuryaman'.

Dr. Aang Nuryaman, S.Si., M.Si.
NIP. 197403162005011001

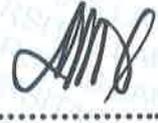
MENGESAHKAN

1. Tim Penguji

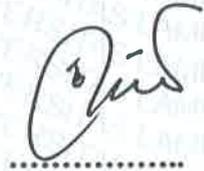
Ketua : **Dr. Aang Nuryaman, S.Si., M.Si.**



Sekretaris : **Dr. Ahmad Faisol, S.Si., M.Sc.**



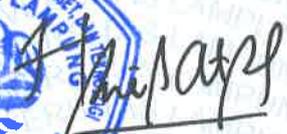
Penguji
Bukan Pembimbing : **Drs. Eri Setiawan, M.Si.**



2. **Plt. Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Lampung**



Dr. Eng. Heri Satria, S.Si., M.Si.
NIP. 197110012005011002



Tanggal Lulus Ujian Skripsi : **29 Maret 2023**

PERNYATAAN

Yang bertanda tangan di bawah ini :

Nama : **Nabilla Yolanda Paramitha**
Nomor Pokok Mahasiswa : **1917031079**
Jurusan : **Matematika**
Judul Skripsi : **PENERAPAN METODE NAÏVE BAYES
UNTUK KLASIFIKASI PENYAKIT
STROKE**

Dengan ini menyatakan bahwa penelitian ini adalah hasil pekerjaan saya sendiri dan apabila kemudian hari terbukti bahwa skripsi ini merupakan hasil salinan atau dibuat oleh orang lain, maka saya bersedia menerima sanksi sesuai dengan ketentuan akademik yang berlaku.

Bandar Lampung, 11 April 2023
Yang menyatakan



Nabilla Yolanda Paramitha
NPM. 1917031079

RIWAYAT HIDUP

Penulis dilahirkan pada tanggal 21 Agustus 2002 di Muara Enim. Terlahir dari keluarga pasangan Bapak Fahmy Xandius dan Ibu Yeni Andriyani yang merupakan anak kedua dari empat bersaudara.

Penulis menyelesaikan pendidikan di Taman Kanan-Kanak Perwanida pada tahun 2007. Pendidikan Sekolah Dasar di SD Negeri 6 Muara Enim pada tahun 2013. Pendidikan Sekolah Menengah Pertama di SMP Negeri 1 Muara Enim pada tahun 2016. Pendidikan Sekolah Menengah Atas di SMA Negeri 1 Muara Enim pada tahun 2019. Kemudian penulis melanjutkan Pendidikan S1 Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam (FMIPA) Universitas Lampung pada tahun 2019 melalui jalur SBMPTN.

Selama menjadi mahasiswa penulis merupakan anggota Generasi Muda Penerus Himatika (GEMATIKA) pada periode 2019. Pada periode 2019, penulis menjadi anggota bidang Dana dan Usaha Ikatan Mahasiswa Sumatera Selatan. Sebagai bentuk aplikasi bidang ilmu pengetahuan, penulis telah mengikuti Studi Independen dalam program Merdeka Belajar Kampus Merdeka (MBKM) di PT. Hacktivate Teknologi Indonesia selama empat bulan mulai dari bulan Agustus sampai Desember 2022.

Sebagai bentuk aplikasi bidang ilmu kepada masyarakat, penulis telah melakukan Kuliah Kerja Nyata (KKN) selama empat puluh hari di Kelurahan Sindur, Kecamatan Cambai, Kota Prabumulih, Provinsi Sumatera Selatan pada bulan Januari sampai Februari 2022. Dan sebagai bentuk aplikasi bidang ilmu di dunia

kerja, penulis telah melaksanakan Kerja Praktik (KP) selama empat puluh hari di SAMSAT Bandar Lampung pada bulan Juni 2022 sampai Juli 2022.

KATA INSPIRASI

“Every cloud has a silver lining”

“Orang-orang yang ada di sekitarmu dapat dijadikan inspirasi, atau bahkan menguras tenagamu. Jadi, pilihlah secara baik-baik.”

(Hans F. Hanson)

“Jangan mencoba membuat diri cocok dengan orang di sekitar kita. Hidup yang kita jalani adalah milik kita dan kitalah yang seharusnya mengendalikan hidup kita, bukan orang lain.”

(Byun Baekhyun of EXO)

PERSEMBAHAN

Puji dan syukur kepada Allah SWT, karena berkah, rahmat dan karunia-Nya sehingga aku masih diberi kesempatan untuk menyelesaikan studi ini.

Ku persembahkan karya sederhana perjuangan dan kesabaran ini kepada kedua orangtua ku. Terima kasih untuk segala kasih sayang, pengorbanan serta doa dan sujud yang selalu menantikan keberhasilanku dengan sabar dan penuh pengertian. Untuk kakak dan adik ku tercinta yang selalu memberikan semangat dan motivasi dalam pengerjaan skripsi ini. Untuk almamaterku Universitas Lampung tempatku menimba ilmu. Terima kasih untuk pengalaman dan kenangan yang sangat berharga.

SANWACANA

Puji syukur penulis panjatkan kepada Allah SWT karena berkat limpahan rahmat dan izin-Nya kepada penulis sehingga penulisan dapat menyelesaikan skripsi dengan judul **“Penerapan Metode Naïve Bayes Untuk Klasifikasi Penyakit Stroke”**. Skripsi ini disusun sebagai salah satu syarat untuk mendapatkan gelar Sarjana Matematika pada Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.

Tidak sedikit kendala yang dihadapi penulis dalam pelaksanaan serta dalam proses penulisan skripsi ini tapi berkat ridha Allah SWT dan bantuan dari berbagai pihak sehingga penulis dapat menyelesaikan skripsi ini. Menyadari bahwa dalam menyusun penulisan skripsi ini tidak terlepas dari bimbingan, bantuan dan kerjasama dari berbagai pihak. Oleh karena itu, pada kesempatan ini penulis ingin mengucapkan terima kasih kepada :

1. Bapak Dr. Aang Nuryaman, S.Si., M.Si. selaku Ketua Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung dan dosen pembimbing utama yang telah memberikan bimbingan, arahan dan saran kepada penulis dalam menyelesaikan skripsi.
2. Bapak Dr. Ahmad Faisol, M.Si., M.Sc. selaku dosen pembimbing kedua yang telah memberikan waktu, arahan dan nasihat yang membantu penulis menyelesaikan skripsi ini.
3. Bapak Drs. Eri Setiawan, M.Si. selaku dosen pembahas dan dosen pembimbing akademik. Terima kasih telah memberikan saran, arahan dan bantuan kepada penulis pada seminar proposal dan hasil penelitian. Dan terima kasih untuk bimbingan, motivasi dan memberikan perhatian penuh kepada penulis selama perkuliahan.

4. Bapak Dr. Eng. Heri Satria, S.Si., M.Si. selaku Plt. Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.
5. Seluruh dosen, staff dan karyawan Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.
6. Kepada orang tua penulis, Bapak Fahmy Xandius dan Ibu Yeni Andriyani yang selalu memberikan semangat, dukungan serta doa kepada penulis. Terima kasih untuk segala hal yang telah diberikan dan semoga penulis dapat selalu membanggakan kalian.
7. Kakak dan adik penulis Bima Rizky Xanditama, Fairuz Khansa Amalia dan Arsyaf Aliyu Haq yang juga memberikan doa dan perhatian serta semangat kepada penulis.
8. Sahabat-sahabat semasa perkuliahan, Citra Puspa Tria, Dinda Ayu Rachmadina, Fadhilah Gustriandini, Putri Aisyah Kusuma Wardani, Sund Grace Yuni Hutagalung, Zidny Ilma Zain, Yeni Suci Paramitha dan Anom Wisnu Syaputra yang selalu memberikan semangat, membantu serta memberikan keceriaan dalam masa perkuliahan hingga penyusunan skripsi ini.
9. Sahabat-sahabat yang selalu menemani hingga saat ini, Adela Oza Azza, Anisa Indriani, Belma Asha Rimadini, Cut Feblita Putri, Dhiya Ulhaq Fatiyah, Fachra Iswarini, Mutiara Fatonah, Rizky Lukisyafirah, Rosalina Febriani dan Vina Wahyuningsih yang selalu memotivasi dan mendoakan penulis.
10. Teman-teman Matematika 2019 atas kebersamaan serta keceriaan yang telah diberikan kepada penulis selama menempuh pendidikan di Universitas Lampung.
11. Seluruh pihak yang telah membantu dalam penyusunan skripsi ini yang tidak dapat disebutkan satu persatu.

Penulis menyadari skripsi ini masih jauh dari kata sempurna. Oleh karenanya saran dan kritik yang membangun akan penulis terima dengan senang hati demi menyempurnakan skripsi ini. Semoga atas segala dukungan dan bantuan yang

telah diberikan kepada penulis mendapatkan balasan dari Allah SWT serta penulis berharap semoga skripsi ini dapat bermanfaat bagi semua pihak yang membutuhkan baik diri penulis sendiri maupun bagi pembaca.

Bandar Lampung, 11 April 2023

Penulis

Nabilla Yolanda Paramitha

DAFTAR ISI

DAFTAR TABEL	vii
DAFTAR GAMBAR.....	ix
I. PENDAHULUAN	1
1.1 Latar Belakang dan Masalah	1
1.2 Tujuan Penelitian.....	4
1.3 Manfaat Penelitian.....	5
II. Tinjauan Pustaka.....	6
2.1 <i>Data Mining</i>	6
2.2 <i>Machine Learning</i>	7
2.3 Klasifikasi.....	7
2.4 <i>Naïve Bayes Classifier</i>	8
2.5 <i>Preprocessing Data</i>	11
2.6 <i>Data Training dan Data Testing</i>	11
2.7 <i>Confusion Matrix</i>	12
2.8 Stroke.....	14
III. Metodologi Penelitian.....	15
3.1 Waktu dan Tempat Penelitian	15
3.2 Data Penelitian	15
3.3 Metode Penelitian	17
IV. HASIL DAN PEMBAHASAN	19
4.1 Visualisasi Data	19
4.2 <i>Preprocessing Data</i>	22
4.2.1 Pembersihan Data.....	23
4.2.2 Transformasi Data.....	23
4.3 Pembagian Data.....	24
4.4 Membangun Model <i>Naïve Bayes</i>	26
4.5 Evaluasi Model <i>Naïve Bayes</i>	39

V. KESIMPULAN	46
5.1 Kesimpulan.....	46
5.2 Saran.....	46
DAFTAR PUSTAKA	47
LAMPIRAN.....	50

DAFTAR TABEL

Tabel	Halaman
1. <i>Confussion Matrix</i>	13
2. Data Variabel Penyakit Stroke	16
3. Jenis Kelamin Pasien.....	20
4. Status Pernikahan Pasien.....	21
5. Tabel Statistika Deskriptif	22
6. <i>Encoding Label</i> Variabel.....	23
7. Interval Kelas Variabel	24
8. Probabilitas <i>Prior</i> Pada Setiap Kelas	27
9. Probabilitas Variabel Jenis Kelamin	28
10. Probabilitas Variabel Usia.....	29
11. Probabilitas Variabel Hipertensi	30
12. Probabilitas Variabel Penyakit Jantung	31
13. Probabilitas Variabel Pernikahan	32
14. Probabilitas Variabel Pekerjaan	33
15. Probabilitas Variabel Tempat Tinggal	34
16. Probabilitas Variabel Gula Darah	35
17. Probabilitas Variabel Indeks Masa Tubuh	36

18. Probabilitas Variabel Rokok	37
19. <i>Confusion Matrix</i> 60:40	39
20. <i>Confusion Matrix</i> 70:30	41
21. <i>Confusion Matrix</i> 80:20	42
22. <i>Confusion Matrix</i> 90:10	44
23. Perbandingan Nilai Evaluasi Model.....	45

DAFTAR GAMBAR

Gambar	Halaman
1. Diagram Alir Pengklasifikasian Naïve Bayes.....	18
2. Diagram Lingkaran Presentase Jumlah Pasien Terkena Penyakit Stroke.....	19
3. Diagram Batang Jenis Kelamin Pasien	20
4. Status Pernikahan Pasien.....	21

I. PENDAHULUAN

1.1 Latar Belakang dan Masalah

Dengan perkembangan zaman yang kian hari semakin canggih dan berkembang pertumbuhan teknologi informasi juga semakin berkembang. Dengan bantuan teknologi manusia dipermudahkannya dalam melakukan banyak hal untuk memecahkan masalah yang begitu banyak dan kompleks. Seiringan itu *data mining* menjadi salah satu bidang yang sangat berkembang juga. *Data mining* didefinisikan sebagai salah satu teknik untuk mengolah data yang digunakan untuk menemukan hubungan dari suatu data yang tidak diketahui dan ditampilkan kedalam bentuk yang mudah dimengerti yang nantinya hubungan itu akan dijadikan acuan untuk mengambil keputusan (Ridwan et al., 2013).

Data mining disebutkan sebagai rangkaian tindakan yang melibatkan teknik statistik, matematik, kecerdasan buatan, maupun *machine learning* untuk mengekstraksi dan mengidentifikasi informasi dan pengetahuan berbasis data besar. *Data Mining* ini merupakan berbagai proses untuk mengali nilai tambah dari beberapa data pengetahuan yang tidak dapat diketahui secara manual. Salah satu teknik pengenalan pola dalam *data mining* adalah menggunakan *machine learning*. (Bramer, 2016).

Machine learning adalah suatu metode yang digunakan dalam membuat program dimana program tersebut belajar dari data, *machine learning* ini adalah penerapan dari kecerdasan buatan (Igawa & Ohashi, 2009). *Machine learning* membuat komputer dapat mempelajari beberapa data sehingga mendapatkan suatu model

yang mana nantinya bisa digunakan tanpa harus memasukan kode programnya secara terus menerus dalam melakukan proses *input-output* (Yuliani et al., 2019). Dalam *machine learning* terdapat 2 teknik pembelajaran yaitu *supervised learning* dan *unsupervised learning* (Yang et al., 2019). *Supervised learning* adalah teknik pembelajaran yang digunakan apabila hasil yang sedang dicari atau diperkirakan telah diketahui. Salah satu kategori dalam *supervised learning* adalah klasifikasi (Nurhayati et al., 2019).

Klasifikasi adalah metode untuk menentukan suatu anggota ke dalam kelas tertentu yang telah ditetapkan sebelumnya. Anggota tersebut dimasukkan ke dalam kelas-kelas tertentu berdasarkan persamaan karakter data tersebut (Turban et al., 2005). Definisi lain mengatakan klasifikasi ialah suatu proses mengevaluasi objek data untuk memuatnya ke kelas tertentu dari beberapa kelas yang ada. Ada dua hal yang dilakukan dalam proses klasifikasi, yaitu yang pertama adalah membuat model yang akan disimpan sebagai memori dan yang kedua menggunakan model tersebut untuk mengklasifikasi objek data lain untuk mengetahui kelas mana yang dimiliki objek data dalam model yang disimpannya (Ariyanti & Iswardani, 2020). Ada beberapa macam pengklasifikasian dalam *data mining* salah satunya adalah Naïve Bayes.

Naïve Bayes adalah metode klasifikasi yang menggunakan peluang, Naïve Bayes ditemukan oleh ilmuwan Inggris yaitu Thomas Bayes. Metode ini memprediksi peluang di masa depan berdasarkan pengalaman di masa lalu sehingga disebut dengan teorema Bayes. Teorema ini dipadukan dengan naif yang menganggap bahwa kondisi antar variabel saling bebas (Bustami, 2013). Definisi lain mengatakan bahwa Naïve Bayes ialah metode yang memanfaatkan nilai probabilitas dari data. Dalam menangani perkara ketidakpastian Naïve Bayes mampu menangani perkara tersebut, dimana Naïve Bayes adalah bagian dari teknik probabilitas (Rohmana & Arifudin, 2014). Naïve Bayes memiliki keuntungan dalam pemakaiannya yaitu metode ini hanya membutuhkan data *training* yang kecil dalam menentukan perkiraan parameter yang diharapkan dalam klasifikasi (Watraton et al., 2020).

Beberapa penelitian sebelumnya yang telah dilakukan terkait penggunaan metode Naïve Bayes yaitu, penelitian yang dilakukan oleh Putra tahun 2019 untuk menentukan kelompok UKT menggunakan Naïve Bayes *Classifier* dengan fungsi Gaussian. Akurasi yang dihasilkan dari metode Naïve Bayes ini sebesar 86,67% (R. A. Putra, 2019). Selanjutnya, penelitian yang dilakukan Handoko dan Neneng pada tahun 2021. Naïve Bayes digunakan untuk mendiagnosis penyakit yang terjadi selama masa kehamilan. Hasil dari penelitian ini mendapat akurasi sebesar 77% (Handoko & Neneng, 2021). Penelitian lainnya yaitu penelitian yang dilakukan oleh Wulandari, dkk. pada tahun 2020. Mereka menggunakan Naïve Bayes sebagai metode untuk mengklasifikasi diagnosa penyakit ISPA. Akurasi yang didapatkan dengan menggunakan metode Naïve Bayes sebesar 86,5% (Wulandari et al., 2020). Selanjutnya, penelitian yang dilakukan oleh Putro, dkk. pada tahun 2020. Penelitian ini menerapkan metode Naïve Bayes untuk klasifikasi pelanggan mana yang berpotensi membeli ayam dan tidak berpotensi membeli ayam. Akurasi yang didapatkan dari metode Naïve Bayes adalah sebesar 92% (Putro et al., 2020).

Di Indonesia penyakit stroke telah menjadi salah satu penyumbang penyakit paling mematikan. Stroke adalah penyakit kronis yang memiliki akibat berbahaya, stroke ini disebabkan oleh gangguan peredaran darah otak akibat penyumbatan pembuluh darah arteri dari penumpukan darah pada pembuluh darah, pecahnya pembuluh darah akibat melemahnya dinding pembuluh darah atau kelainan pada kondisi darah itu sendiri yang mengakibatkan kurangnya cadangan oksigen dan nutrisi ke otak yang menimbulkan kerusakan di jaringan otak (Sulaiman & Anggriani, 2017). Penyakit stroke dapat disebabkan oleh beberapa faktor, faktor-faktor itu termasuk faktor risiko yang tidak dapat diubah maupun faktor risiko yang dapat diubah. Faktor risiko yang tidak dapat diubah antara lain usia, jenis kelamin, ras serta genetik. Sedangkan faktor risiko yang dapat diubah antara lain hipertensi, diabetes mellitus, merokok, obesitas (Wayunah & Saefulloh, 2016).

Terdapat beberapa penelitian yang terkait dengan penyakit stroke diantaranya, penelitian yang dilakukan oleh Puspitawuri, dkk. pada tahun 2019. Mereka mengkombinasikan metode *K-Nearest Neighbor* dan Naïve Bayes untuk melakukan diagnosis tingkat risiko penyakit stroke dengan menggunakan sebaran kelas yang seimbang dan tidak seimbang. Akurasi yang didapatkan dari mengkombinasikan kedua metode ini adalah untuk kelas seimbang sebesar 96,67% dan kelas tidak seimbang sebesar 100% (Puspitawuri et al., 2019). Penelitian selanjutnya dilakukan oleh Maskuri, dkk. pada tahun 2022. Pada penelitian ini menggunakan algoritma *K-Nearest Neighbor* untuk memprediksi penyakit stroke. Akurasi yang didapat dengan menggunakan metode ini sebesar 95% (Maskuri et al., 2022). Hal yang membedakan penelitian saya dengan penelitian Maskuri, dkk. adalah penggunaan metode dalam pengklasifikasian dan juga data yang digunakan. Penelitian selanjutnya dilakukan oleh Pambudi, dkk. pada tahun 2022. Pada penelitian ini menggunakan algoritma *Decision Tree C.45* untuk mengklasifikasikan penyakit stroke, dan akurasi yang dihasilkan sebesar 99,07% (Pambudi et al., 2022). Hal yang membedakan penelitian saya dengan penelitian Pambudi, dkk. adalah penggunaan metode dalam pengklasifikasian dan juga data yang digunakan. Oleh karena itu, penelitian ini akan dilakukan klasifikasi penyakit stroke dengan menerapkan metode Naïve Bayes.

1.2 Tujuan Penelitian

Penelitian ini memiliki tujuan sebagai berikut:

1. Menerapkan metode Naïve Bayes dalam pengklasifikasian penyakit stroke.
2. Mendapatkan nilai akurasi model terbaik dalam pengklasifikasian penyakit stroke menggunakan metode Naïve Bayes.

1.3 Manfaat Penelitian

Penelitian ini memiliki manfaat sebagai berikut:

1. Sebagai bahan pengembangan ilmu matematika dan statistika mengenai metode Naïve Bayes *Classifier*.
2. Sebagai bahan pertimbangan dan informasi tambahan bagi pembaca jika ingin melakukan penelitian lanjut terkait klasifikasi dan model yang telah didapatkan.

II. Tinjauan Pustaka

2.1 Data Mining

Data mining adalah proses dimana data berukuran besar dikumpulkan dengan mengekstrasi dan mengidentifikasi pola-pola penting atau mencari data yang berada di basis data (Vulandari, 2017). *Data mining* juga dapat diartikan sebagai proses yang menggunakan teknik statistika, matematika, kecerdasan buatan dan *machine learning* untuk mengidentifikasi dan memperoleh informasi yang berguna dan pengetahuan yang berhubungan dengan berbagai basis data besar (Turban et al., 2005).

Data mining juga memiliki beberapa sebutan yang memiliki makna yang sama yaitu *Knowledge Discovery in Databases* (KDD). Pada umumnya dalam *data mining* memiliki bentuk-bentuk yaitu deskripsi, prediksi, estimasi, klasifikasi, *clustering*, dan asosiasi (Larose, 2005). Di dalam *Knowledge Discovery in Database* (KDD) terdapat tahapan-tahapan yaitu pembersihan data (*data cleaning*), integrasi data (*data integration*), seleksi data (*data selection*), transformasi data (*data transformation*), proses *mining*, evaluasi pola (*pattern evaluation*), presentasi pengetahuan (Han et al., 2012).

2.2 *Machine Learning*

Machine Learning merupakan bagian dari kecerdasan buatan yang bertujuan untuk mempelajari identifikasi pola untuk memperoleh solusi yang maksimal. *Machine learning* ini memungkinkan komputer untuk mendapatkan solusi data yang maksimal secara otomatis (Barboza et al., 2017). *Machine learning* membuat komputer dapat mempelajari beberapa data sehingga mendapatkan suatu model untuk melakukan proses *input-output* tanpa harus memasukan kode program secara berulang (Yuliani et al., 2019). Dalam *machine learning* terdapat 2 teknik pembelajaran yaitu *supervised learning* dan *unsupervised learning*.

a. *Supervised Learning*

Supervised learning adalah teknik pembelajaran yang digunakan apabila hasil yang sedang dicari atau diperkirakan telah diketahui. Salah satu kategori dalam *supervised learning* adalah klasifikasi (Nurhayati et al., 2019).

b. *Unsupervised Learning*

Unsupervised learning berfungsi pada objek data yang tidak berlabel. Jenis pembelajaran ini sering digunakan untuk ekstrasi data, menemukan pola dan struktur penting, mencocokkan objek terkait, dan tujuan praktis. Jenis pembelajaran ini cocok digunakan untuk klastering (Yang et al., 2019).

2.3 **Klasifikasi**

Klasifikasi adalah proses menilai suatu objek data dimana objek tersebut dimasukkan ke dalam kelas tertentu dari jumlah kelas yang tersedia. Klasifikasi ini membentuk suatu model yang didasari oleh data latih yang ada, kemudian melakukan klasifikasi pada data yang baru dengan menggunakan model tersebut (Utomo & Mesran, 2020). Klasifikasi juga difenisikan sebagai proses untuk mendapatkan suatu model atau fungsi yang menjelaskan dan membedakan data ke dalam beberapa kelas. Klasifikasi ini menyertakan proses pemeriksaan karakteristik objek dan memasukan objek ke dalam salah satu kelas yang telah

didefinisikan (Larose, 2005). Ada beberapa algoritma dalam pengklasifikasian diantaranya *Logistic Regression*, *Support Vector Learning*, *Random Forest Classifier*, dan Naïve Bayes (Soofi & Awan, 2017).

2.4 Naïve Bayes Classifier

Naïve Bayes pertama kali diusulkan oleh Revered Thomas Bayes, antara tahun 1702 dan 1761 penggunaan Naïve Bayes mulai dikenalkan. Naïve Bayes Classifier adalah suatu algoritma di dalam data mining yang menerapkan teorema Bayes untuk klasifikasi. Penggunaan metode Naïve Bayes dalam pengklasifikasian memiliki keuntungan dimana penggunaan metode ini hanya membutuhkan jumlah data *training* yang kecil dalam menentukan estimasi parameter selama proses klasifikasi (Santosa, 2007).

Definisi lain mengatakan Naïve Bayes adalah suatu pengklasifikasian probabilitik sederhana yang menghitung peluang dari menambahkan frekuensi dan kombinasi nilai dari dataset. Metode ini didasarkan pada asumsi bahwa nilai variabel saling bebas jika diberikan nilai *output* (Riduwan, 2013). Persamaan teorema Bayes dituliskan sebagai berikut:

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)} \quad (2.1)$$

dimana:

- $P(A|B)$: Peluang terjadinya A berdasarkan kondisi B (*posteriori prob*)
- $P(A)$: Peluang terjadinya A (*prior prob*)
- $P(B|A)$: Peluang terjadinya B berdasarkan kondisi pada hipotesis A
- $P(B)$: Peluang terjadinya B

Dalam proses klasifikasi Naïve Bayes diperlukan beberapa petunjuk yang digunakan dalam penentuan kelas yang tepat untuk sampel yang dianalisis sehingga metode Naïve Bayes dapat disesuaikan sebagai berikut:

$$P(Y|X_1, \dots, X_n) = \frac{P(Y) \cdot P(X_1, \dots, X_n|Y)}{P(X_1, \dots, X_n)} \quad (2.2)$$

dengan:

$P(Y|X_1, \dots, X_n)$: Peluang masuknya sampel variabel tertentu dalam kelas Y
(*posterior*)

$P(Y)$: Peluang munculnya kelas Y sebelum masuknya sampel
(*prior*)

$P(X_1, \dots, X_n|Y)$: Peluang kemunculan variabel-variabel sampel pada kelas Y
(*likelihood*)

$P(X_1, \dots, X_n)$: Peluang kemunculan variabel-variabel sampel secara umum (*evidence*), dengan n dibuat dari sampel n atribut, berturut-turut A_1, \dots, A_n .

Karena itu, persamaan (2.2) bisa juga ditulis secara sederhana sebagai berikut:

$$Posterior = \frac{Prior \times Likelihood}{Evidence} \quad (2.3)$$

Nilai *posterior* itu akan dibandingkan dengan nilai-nilai *posterior* kelas lainnya untuk menentukan di kelas mana sampel harus ditempatkan. Nilai *evidence* selalu tetap untuk setiap kelas pada suatu sampel yaitu bernilai 1 maka nilai *evidence* dapat diabaikan. Sehingga *posterior* dapat dihitung dengan mengalikan nilai *prior* dan *likelihood*. Nilai *prior* yang merupakan peluang kelas Y muncul sebelum sampel masuk dapat dihitung menggunakan persamaan sebagai berikut (Watraton et al., 2020) :

$$P(Y_n) = \frac{d_n}{d} \quad (2.4)$$

dengan:

- d : Banyaknya pengamatan
 d_n : Banyaknya pengamatan pada kelas Y_n

Untuk nilai *likelihood* yang merupakan peluang munculnya variabel-variabel sampel pada kelas Y . Nilai *likelihood* dapat dihitung dengan menggunakan persamaan sebagai berikut:

$$P(X_i|Y_n) = \frac{d_{ni}}{d_n} \quad (2.5)$$

dengan:

- d_{ni} : Banyaknya pengamatan pada kelas Y_n yang mempunyai nilai X_i untuk A_i
 d_n : Banyaknya pengamatan pada kelas Y_n

Penjabaran $P(Y|X_1, \dots, X_n)$ lebih lanjut dengan menggunakan definisi peluang bersyarat sebagai berikut:

$$\begin{aligned} P(Y|X_1, \dots, X_n) &= P(Y) \cdot P(X_1, \dots, X_n|Y) \\ &= P(Y) \cdot P(X_1|Y) \cdot P(X_2, \dots, X_n|Y, X_1) \\ &= P(Y) \cdot P(X_1|Y) \cdot P(X_2|Y, X_1) \cdot P(X_3, \dots, X_n|Y, X_1, X_2) \\ &= P(Y) \cdot P(X_1|Y) \cdot P(X_2|Y, X_1) \dots P(X_n|Y, X_1, X_2 \dots X_{n-1}) \end{aligned} \quad (2.6)$$

Terlihat bahwa hasil penjabaran tersebut semakin banyak variabel dari data maka semakin rumit faktor-faktor syarat yang mempengaruhi nilai peluangnya. Maka dengan menggunakan asumsi independen pada Naïve Bayes membuat masing - masing variabel (X_1, \dots, X_n) saling bebas satu sama lain. Dengan adanya asumsi ini, maka berlaku juga persamaan sebagai berikut:

$$P(X_i|X_j) = \frac{P(X_i \cap X_j)}{P(X_j)} = \frac{P(X_i) \cdot P(X_j)}{P(X_j)} = P(X_i) \quad (2.7)$$

untuk $i \neq j$ maka

$$P(X_i|X_j) = P(X_i|Y) \quad (2.8)$$

Persamaan (2.7) menunjukkan bahwa asumsi independensi yang kuat membuat syarat peluang menjadi lebih sederhana. Sehingga $P(Y|X_1, \dots, X_n)$ dapat sederhanakan, persamaan ini nantinya akan digunakan dalam pengklasifikasian.

$$\begin{aligned} P(Y|X_1, \dots, X_n) &= P(Y) \cdot P(X_1|Y) \cdot P(X_2|Y) \dots P(X_n|Y) \\ &= P(Y) \prod_{i=1}^n P(X_i|Y) \end{aligned} \quad (2.9)$$

2.5 Preprocessing Data

Preprocessing data adalah tahapan dalam data mining untuk mempersiapkan data untuk digunakan dalam proses selanjutnya. Dalam tahapan *preprocessing data* hal pertama yang dilakukan ialah pembersihan data dimana pada tahap ini dilakukan pengecekan data hilang dan data duplikat (Fikri et al., 2020). Jika ditemukanya data hilang maka bisa melakukan imputasi dengan mean dan median atau juga bisa dengan menghapus data hilang. Jika terdapat data duplikat maka bisa dilakukan penghapusan data duplikat. Tahapan selanjutnya ialah pengubahan bentuk data agar dapat mempermudah dalam proses pengklasifikasian. Data yang diubah yaitu data yang awalnya bertipe numerik diubah menjadi data bertipe kategorik (Adi, 2018).

2.6 Data Training dan Data Testing

Dalam pengujian klasifikasi, data yang digunakan dibagi menjadi data latih dan data uji. Pembagian data sangat penting dilakukan karena akan menentukan tingkat akurasi dari penerapan suatu metode. Data *training* atau data latih ialah

data yang sebelumnya sudah diketahui untuk label kelompok dan digunakan untuk membangun model klasifikasi. Data *testing* atau data uji ialah data yang belum diketahui label kelompoknya. Data *training* dapat digunakan untuk mencari model yang sesuai dan dapat digunakan untuk melatih algoritma. Data *testing* digunakan untuk mengukur keakuratan *classifier* berhasil melakukan klasifikasi dengan benar (Witten et al., 2011). Dalam pembagian proporsi data *training* dan data *testing* biasanya dibagi menggunakan 60:40, 70:30, 80:20 dan 90:10. Penjelasan pembagian data *training* dan data *testing* adalah 60:40 yang berarti 60% data *training* dan 40% data *testing* adapun 70:30 yang berarti 70% data *training* dan 30% data *testing*. Untuk menghitung jumlah data *training* dan data *testing* dapat menggunakan persamaan sebagai berikut.

$$\text{Jumlah data } training = \text{proporsi data } training \times N \quad (2.10)$$

$$\text{Jumlah data } testing = N - \text{jumlah data } training \quad (2.11)$$

dengan:

N = jumlah seluruh data (data *training* + data *testing*)

2.7 Confusion Matrix

Tahap pengujian model dilakukan dengan menggunakan metode *confusion matrix* yang menampilkan hasil dari penilaian model melalui tabel matrik. Jika dataset memiliki dua kelas maka kelas yang pertama disebut positif dan kelas lainnya disebut negatif (D. Putra & Wibowo, 2020). *Confusion matrix* ini menggunakan tabel matriks seperti pada Tabel 1 dimana terdapat *record* perbandingan hasil klasifikasi data *testing* atau data uji berdasarkan data *training* atau data latih dengan data sebenarnya (Utomo & Mesran, 2020).

Tabel 1. *Confusion Matrix*

Kelas Asli	Kelas Prediksi	
	Positif	Negatif
Positif	<i>True Positive (TP)</i>	<i>False Negative (FN)</i>
Negatif	<i>False Positive (FP)</i>	<i>True Negative (TN)</i>

True positive adalah jumlah dari *record* positif yang diklasifikasikan sebagai positif, *false positive* adalah jumlah dari *record* negatif yang diklasifikasikan sebagai positif, *false negative* adalah jumlah dari *record* positif yang diklasifikasikan sebagai negatif, *true negative* adalah jumlah dari *record* negatif yang diklasifikasikan sebagai negatif (D. Putra & Wibowo, 2020). Untuk mengukur performa suatu model dapat menggunakan empat metode sebagai berikut (Saputro & Sari, 2020):

a. Akurasi

Metode pertama ialah akurasi, akurasi adalah metode yang didasari tingkat kedekatan antara nilai prediksi dengan nilai sebenarnya. Akurasi adalah hasil dari penjumlahan nilai diagonal dibagi dengan jumlah total keseluruhan data dan selanjutnya dikalikan 100%. Rumus akurasi dijabarkan pada persamaan (2.12).

$$Akurasi = \frac{TP + TN}{TP + FP + TN + FN} \times 100\% \quad (2.12)$$

b. Presisi

Presisi adalah metode yang dipakai untuk menghitung nilai proporsi kelas positif yang berhasil diprediksi dengan benar dari keseluruhan hasil prediksi kelas positif. Presisi menunjukkan jumlah data kategori positif yang diklasifikasi secara benar dibagi dengan total data yang diklasifikasi positif. Rumus presisi dijabarkan pada persamaan (2.13).

$$Presisi = \frac{TP}{TP + FP} \times 100\% \quad (2.13)$$

c. *Recall*

Recall adalah metode yang digunakan untuk menghitung presentase kelas data positif yang berhasil diprediksi benar dari keseluruhan data kelas positif. Rumus *recall* dijabarkan pada persamaan (2.14).

$$Recall = \frac{TP}{TP + FN} \times 100\% \quad (2.14)$$

d. *F1-score*

Metode *F1-score* ialah rata-rata harmonik dari presisi dan recall. *F1-score* juga biasa disebut *F1-measure*. Rumus *F1-score* dijabarkan pada persamaan (2.15).

$$F1 - score = 2 \times \frac{Presisi \times Recall}{Presisi + Recall} \quad (2.15)$$

2.8 Stroke

Stroke adalah gangguan fungsi otak yang disebabkan oleh gangguan aliran darah ke otak akibat pecahnya atau tersumbatnya pembuluh darah arteri yang menuju ke otak, sehingga nutrisi dan oksigen yang dibutuhkan otak tidak terpenuhi dengan baik, sehingga menyebabkan hilangnya sensasi dan bagian tubuh (Noviyanti Dewi, 2013). Sumber lain juga menyebutkan bahwa stroke didefinisikan sebagai penyakit yang disebabkan dari gangguan peredaran darah yang terjadi di otak yang dipengaruhi oleh berbagai faktor risiko baik yang tidak dapat diubah seperti usia dan jenis kelamin maupun yang dapat diubah seperti hipertensi, diabetes mellitus, konsumsi rokok, dan pekerjaan (Laily, 2017).

III. Metodologi Penelitian

3.1 Waktu dan Tempat Penelitian

Penelitian ini dilakukan pada semester ganjil tahun akademik 2022/2023 bertempat di Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.

3.2 Data Penelitian

Data yang digunakan merupakan data sekunder yang diperoleh dari situs *Kaggle* yang berjudul “*Brain Stroke Prediction Dataset*” dengan link sebagai berikut: <https://www.kaggle.com/datasets/zzetrkalpakbal/full-filled-brain-stroke-dataset>. Dengan jumlah data sebanyak 200 dan variabel sebanyak 11. Data penelitian ini memiliki 10 variabel independen sebagai atribut dan 1 variabel dependen sebagai variable target. Dari 10 variabel ini terdapat 7 variabel bertipe data kategorik dan 3 data bertipe data numerik. Pada penelitian ini, data akan dibagi menjadi 2 bagian yaitu data *training* untuk pembentukan model dan data *testing* untuk pemilihan model terbaik. Untuk pembagian data dibagi menjadi 4 bagian. Pertama, 60% data *training* dan 40% data *testing*. Kedua, 70% data *training* dan 30% data *testing*. Ketiga, 80% data *training* dan 20% data *testing*. Keempat, 90% data *training* dan 10% data *testing*.

Tabel 2. Data Variabel Penyakit Stroke

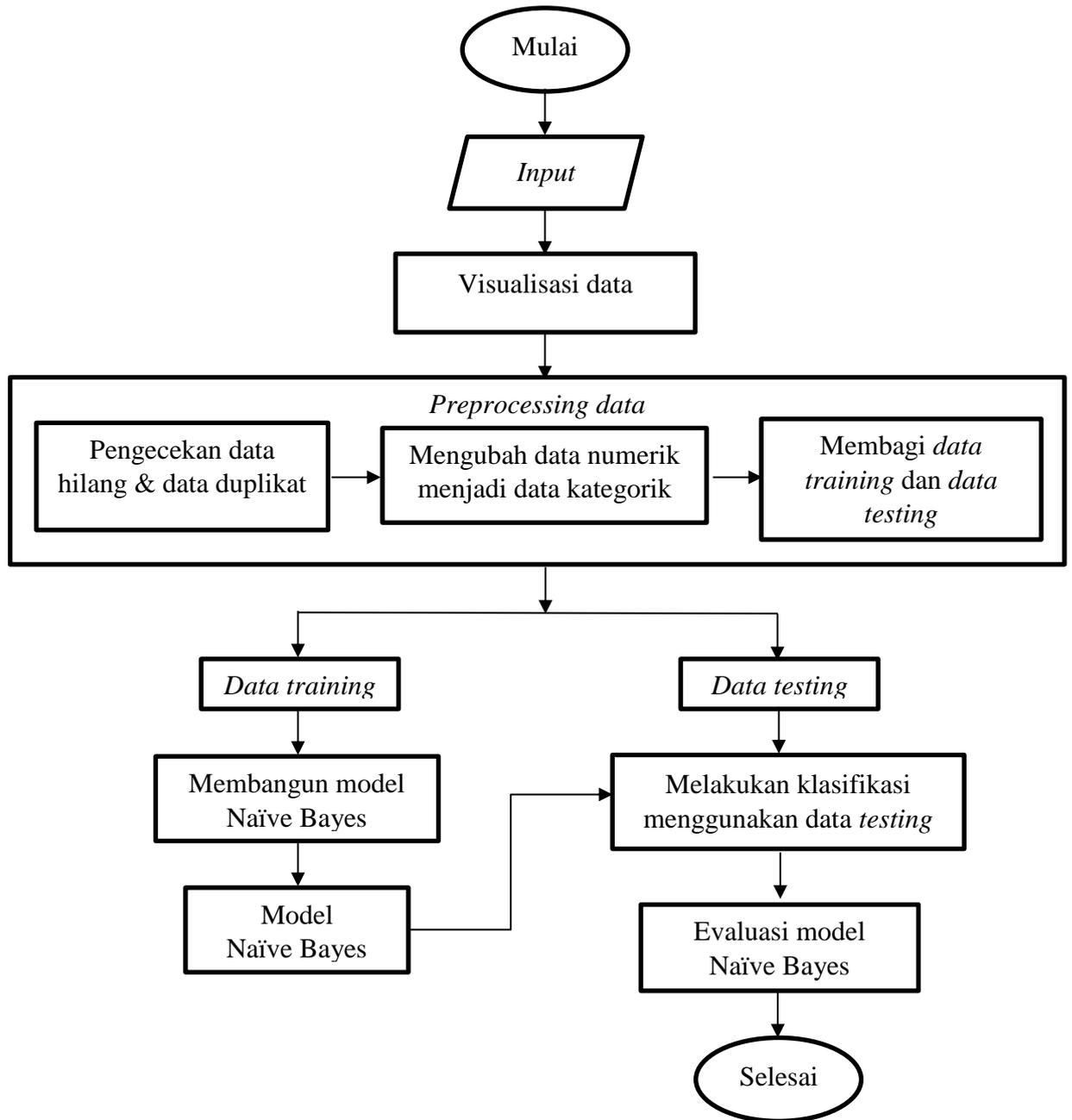
No	Variabel	Kategori	Keterangan
1	<i>Gender</i> (Jenis kelamin)	Kategorik	Jenis kelamin pasien, laki-laki atau perempuan
2	<i>Age</i> (Usia)	Numerik	Usia pasien (tahun)
3	<i>Hypertension</i> (Hipertensi)	Kategorik	Memiliki penyakit darah tinggi atau tidak
4	<i>Heart disease</i> (Penyakit Jantung)	Kategorik	Memiliki riwayat penyakit jantung atau tidak
5	<i>Ever married</i>	Kategorik	Pernah menikah atau tidak
6	<i>Work type</i>	Kategorik	Tipe pekerjaan pasien seperti swasta, PNS, wiraswasta, anak-anak
7	<i>Residence type</i>	Kategorik	Tipe tempat tinggal seperti perkotaan atau perdesaan
8	<i>Average glucose level</i>	Numerik	Rata-rata tingkat kadar gula dalam darah (mg/dl)
9	BMI	Numerik	Indeks masa tubuh
10	<i>Smoking status</i>	Kategorik	Status merokok seperti sebelumnya merokok, merokok, tidak pernah merokok, tidak diketahui
11	<i>Stroke</i>	Kategorik	Terkena stroke atau tidak

3.3 Metode Penelitian

Penelitian ini menggunakan metode Naïve Bayes untuk mengklasifikasi penyakit *brain stroke* dengan bantuan pemrograman *Python*. Langkah-langkah yang dilakukan dalam penelitian ini sebagai berikut:

1. Memasukan data *brain stroke prediction* yang didapat dari situs *kaggle* ke dalam *Python* menggunakan *Jupyter Notebook*.
2. Melakukan visualisasi data *brain stroke prediction* ke dalam bentuk diagram.
3. Melakukan *preprocessing* data, diantaranya:
 - Melakukan pengecekan jika terdapat data hilang dan data duplikat.
 - Melakukan penghapusan data hilang dan data duplikat jika terdapat.
 - Mengubah data numerik menjadi data kategorik menjadi beberapa interval.
 - Melakukan pemberian label pada data kategorik.
 - Melakukan pembagian data *training* dan data *testing* dengan rasio 60:40, 70:30, 80:20, dan 90:10.
4. Membangun model menggunakan Naïve Bayes.
5. Melakukan klasifikasi menggunakan data *testing*.
6. Melakukan evaluasi model menggunakan *confusion matrix* yang didapatkan dari hasil klasifikasi.

Adapun proses dalam klasifikasi metode Naïve Bayes dengan menggunakan Bahasa pemograman *python* dapat ditampilkan secara singkat dengan menggunakan diagram alir dibawah ini.



Gambar 1. Diagram Alir Pengklasifikasian Naïve Bayes

V. KESIMPULAN

5.1 Kesimpulan

Berdasarkan hasil analisis dan pembahasan yang telah dilakukan terkait pembangunan model Naïve Bayes dalam mengklasifikasi pasien penyakit stroke, diperoleh kesimpulan sebagai berikut:

1. Hasil perhitungan evaluasi model menggunakan *confusion matrix* pada data penyakit stroke dengan proporsi pembagian data 80:20 menghasilkan tingkat akurasi yang lebih tinggi dibandingkan proporsi lainnya yaitu sebesar 80% yang mana ini menjadikan proporsi 80:20 mendapatkan model terbaik.
2. Penentuan proporsi pembagian data *training* dan data *testing* dapat mempengaruhi hasil pengujian presentase tingkat akurasi, karena pola data *training* akan dijadikan acuan dalam menentukan kelas pada data *testing*.

5.2 Saran

Saran yang dapat diberikan pada penelitian ini yaitu:

1. Pada penelitian selanjutnya dapat menambahkan variabel riwayat keluarga dan dapat menggunakan data yang lebih banyak, sehingga memperoleh hasil akurasi yang lebih baik.
2. Penelitian ini dapat dikembangkan dengan menggabungkan metode lain ataupun membandingkan dengan metode lainnya dengan menggunakan data yang sama pada penelitian ini, sehingga dapat melihat metode mana yang lebih baik dalam melakukan klasifikasi.

DAFTAR PUSTAKA

- Adi, S. 2018. Implementasi Algoritma Naive Bayes Classifier Untuk Klasifikasi Penerima Beasiswa PPA di Universitas Amikom Yogyakarta. *Jurnal Mantik Penusa*. **22**(1): 11–16.
- Argina, A. M. 2020. Penerapan Metode Klasifikasi K-Nearest Neighbor pada Dataset Penderita Penyakit Diabetes. *Indonesian Journal of Data and Science*. **1**(2): 29–33.
- Ariyanti, D., & Iswardani, K. 2020. Teks Mining untuk Klasifikasi Keluhan Masyarakat Pada Pemkot Probolinggo Menggunakan Algoritma Naive Bayes. *Jurnal IKRA-ITH Informatika*. **4**(3): 125–132.
- Barboza, F., Kimura, H., & Altman, E. 2017. Machine learning models and bankruptcy prediction. *Expert Systems with Applications*. **83**: 405–417.
- Bramer, M. 2016. Principles of Data Mining. In *Drug Safety* (Third Edit, Vol. 30, Issue 7). Springer Nature.
- Bustami. 2013. Penerapan Algoritma Naive Bayes Untuk Mengklasifikasi Data Nasabah Asuransi. *Jurnal Penelitian Teknik Informatika*. **3**(2): 127–146.
- Fikri, M. I., Sabrila, T. S., & Azhar, Y. 2020. Perbandingan Metode Naive Bayes dan Support Vector Machine pada Analisis Sentimen Twitter. *Smatika Jurnal*. **10**(02): 71–76.
- Han, J., Kamber, M., & Pei, J. 2012. Data Mining Concepts and Techniques. In *Proceedings - 2013 International Conference on Machine Intelligence Research and Advancement, ICMIRA 2013* (3rd ed). In Elsevier Inc.
- Handoko, M. R., & Neneng. 2021. Sistem Pakar Diagnosa Penyakit Selama Kehamilan Menggunakan Metode Naive Bayes Berbasis Web. *Jurnal Teknologi Dan Sistem Informasi (JTSI)*. **2**(1): 50–58.
- Igawa, K., & Ohashi, H. 2009. A negative selection algorithm for classification and reduction of the noise effect. *Applied Soft Computing Journal*. **9**(1): 431–438.

- Laily, R. S. 2017. Hubungan Karakteristik Penderita dan Hipertensi dengan Kejadian Stroke Iskemik. *Jurnal Berkala Epidemiologi*. **5**(1): 48–59.
- Larose, D. T. 2005. *Discovering Knowledge In Data: An Introduction to Data Mining*. John Wiley & Sons.
- Maskuri, M. N., Harlina, Sukerti, K., & Bhakti, R. 2022. Penerapan Algoritma K-Nearest Neighbor (KNN) untuk Memprediksi Penyakit Stroke Stroke Disease Predict Using KNN Algorithm. *Jurnal Ilmiah Intech : Information Technology Journal of UMUS*. **4**(1): 130–140.
- Noviyanti, R. 2013. Faktor Risiko Penyebab Meningkatnya Kejadian Stroke Pada Usia Remaja Dan Usia Produktif. *Profesi*. **10**: 52–56.
- Nurhayati, Busman, & Pradono, R. 2019. Pengembangan Algoritma Unsupervised Learning Technique Pada Big Data Analisis di Media Sosial Sebagai Media Promosi Online Bagi Masyarakat. *Jurnal Teknik Informatika*. **12**(1): 79–96.
- Pambudi, R. E., Sriyanto, S., & Firmansyah, F. 2022. Klasifikasi Penyakit Stroke Menggunakan Algoritma Decision Tree C.45. *Teknika*. **16**(2): 221–226.
- Puspitawuri, A., Santoso, E., & Dewi, C. 2019. Diagnosis Tingkat Risiko Penyakit Stroke Menggunakan Metode K-Nearest Neighbor dan Naïve Bayes. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*. **3**(4): 3319–3324.
- Putra, D., & Wibowo, A. 2020. Prediksi Keputusan Minat Penjurusan Siswa SMA Yadika 5 Menggunakan Algoritma Naïve Bayes. *Prosiding Seminar Nasional Riset Dan Information Science (SENARIS)*. **2**: 84–92.
- Putra, R. A. 2019. Penerapan Naïve Bayes Classifier dengan Gaussian Function Untuk Menentukan Kelompok UKT. *Jurnal Ilmiah Informatika Global*. **9**(2): 112–117.
- Putro, H. F., Vuldari, R. T., & Saptomo, W. 2020. Penerapan Metode Naive Bayes Untuk Klasifikasi Pelanggan. *Jurnal Teknologi Informasi Dan Komunikasi (TIKOMSiN)*. **8**(2): 19–24.
- Riduwan. 2013. *Skala Pengukuran Variabel-Variabel Penelitian*. ALFABETA.
- Ridwan, M., Suryono, H., & Sarosa, M. 2013. Implementasi Data Mining untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier. *EECCIS*. **7**(1): 59–64.
- Rohmana, I., & Arifudin, R. 2014. Perbandingan Jaringan Syaraf Tiruan Dan Naive Bayes Dalam Deteksi Seseorang Terkena Penyakit Stroke. **37**(2): 178–191.

- Santosa, B. 2007. *Data Mining: Teknik Pemanfaatan Data untuk Keperluan Bisnis*. Garah Ilmu.
- Saputro, I. W., & Sari, B. W. 2020. Uji Performa Algoritma Naïve Bayes untuk Prediksi Masa Studi Mahasiswa. *Creative Information Technology Journal*. **6**(1): 1.
- Soofi, A. A., & Awan, A. 2017. Classification Techniques in Machine Learning: Applications and Issues. *Journal of Basic & Applied Sciences*. **13**: 459–465.
- Sulaiman, & Anggriani. 2017. Sosialisasi Pencegahan Kasus Stroke Pada Lanjut Usia Di Desa Hampan Perak Kecamatan. *Amaliah: Jurnal Pengabdian Kepada Masyarakat*. **1**(2): 70–74.
- Turban, E., Aronson, J., & Liang, T. P. 2005. *Decision Support Systems and Intelligent Systems*. Prentice Hall.
- Utomo, D. P., & Mesran, M. 2020. Analisis Komparasi Metode Klasifikasi Data Mining dan Reduksi Atribut Pada Data Set Penyakit Jantung. *Jurnal Media Informatika Budidarma*. **4**(2): 437.
- Watratan, A. F., & Moeis, D. 2020. IMPLEMENTASI AIGORITMA NAIVE BAYES UNTUK MEMREDIKSI TINGKAT PENYEBARAN COVID. *Jurnal of Applied Computer Science and Technology*. **1**(1): 7–14.
- Wayunah, W., & Saefulloh, M. 2016. Analisis Faktor Yang Berhubungan Dengan Kejadian Stroke Di Rsud Indramayu. *Jurnal Pendidikan Keperawatan Indonesia*. **2**(2): 65–76.
- Witten, I. H., Frank, E., & Hall, M. 2011. *Data Mining: Partical Machine Learning Tools and Techniques* (Third Edit). Morgan Kaufmann Publishers.
- Wulandari, F., Jusia, P. A., & Jasmir. 2020. Klasifikasi Data Mining Untuk Mendiagnosa Penyakit ISPA Menggunakan Metode Naïve Bayes Pada Puskesmas Jambi Selatan. *Jurnal Manajemen Teknologi Dan Sistem Informasi (JMS)*. **2**(3): 214–227.
- Yang, J., Liu, Y., Qian, M., Guan, C., & Yuan, X. 2019. Information Extraction From Electronic Medical Records Using Multitask Recurrent Neural Network With Contextual Word Embedding. *Applied Sciences*. **9**(18).
- Yuliani, E., Aini, A. N., & Khasanah, C. U. 2019. Perbandingan Jumlah Epoch Dan Steps Per Epoch Pada Convolutional Neural Network Untuk Meningkatkan Akurasi Dalam Klasifikasi Gambar. *Jurnal INFORMA Politeknik Indonusa Surakarta*. **5**: 2442–7942.