ANALISIS PERBANDINGAN ALGORITMA KLASIFIKASI *SUPPORT VECTOR MACHINE, DECISION TREE, NAÏVE BAYES BENCANA BANJIR* BERDASARKAN CURAH HUJAN

Nirmala Saleh[1], Alter Lasarudin[2], Wahyudin Hasyim[3]

[1,2,3] Program Studi Sistem Informasi, Universitas Muhammadiyah Gorontalo, *Jl. Prof. Dr. Mansoer Pateda, Kab.Gorontalo , India Email :*

nirmalasaleh12@gmail.com

ABSTRACT

NIRMALA SALEH, Classification Analysis of Support Vector Machine, Decision Tree, Naive Bayes Flood Disaster Based on Rainfall. Supervised by ALTER LASARUDIN as supervisor 1 and WAHYUDIN HASYIM as supervisor 2.

Floods are one of the deadliest disaters in india, with the frequency of flooding and extreme rainfall events. Projected to recall under an increasingly warmer climate, this resarch aims to classify flood disasters based on rainfall and compare the level of accuracy of the Support Vector Machine, Decision Tree, and Naive Bayes algorithm. The result using the Suppor Vector Machine, Decision Tree, Naive Bayes algorithm with a total of 118 record will result in a Support Vector Machine accuracy value of 58.56%, and a Decision Tree accuracy rate of 52.98%, and a Naive Bayes accuracy rate of 56.33%. This the accuracy value of the three algorithms based on the largest accuracy results are in the Support Vector Machine of 58.56%.

Keywords: Classification, Floods Disaster Based on Rainfall, Support Vector Machine, Decision Tree, Naive Bayes.

ABSTRAK

NIRMALA SALEH, Analisis Klasifikasi *Support Vector Machine, Decision Tree, Naïve Bayes* Bencana Banjir Berdasarkan Curah Hujan. Dibimbing oleh Alter Lasarudin, M.Kom sebagai pembimbing 1 dan Wahyudin Hasyim, M.Kom sebagai pembimbing 2

Banjir adalah salah satu bencana yang paling mematikan di India frekuensi banjir dan kejadian curah hujan ekstream. diproyeksikan akan mengingat di bawah iklim yang semakin hangat. Penelitian ini bertujuan untuk mengklasifikasi bencana banjir berdasarkan curah hujan dan membandingkan tingkat akurasi dari algoritma *Support Vector Machine, Decision Tree,* dan *Naïve Bayes*. Hasil penelitian menggunakan Algoritma *Support Vector Machine, Decision Tree, Naïve Bayes* dengan jumlah *record* 118 maka akan diperoleh nilai akurasi *Support Vector Machine* sebesar 58.56%, dan tingkat akurasi *Decision Tree* sebesar 52.98%, dan tingkat akurasi *Naive Bayes* sebesar 56.33%. Maka nilai akurasi ketiga algoritma berdasarkan hasil akurasi terbesar ada pada *Support Vector Machine* Sebesar 58.56%.

Kata Kunci : Klasifikasi, Bencana Banjir Berdasarkan Curah Hujan, *Support Vector Machine, Decision Tree, Naive Bayes*.

1. Latar Belakang

Banjir adalah salah satu bencana yang paling mematikan di India frekuensi banjir dan kejadian curah hujan ekstream diproyeksikan akan mengingat dibawah iklim yang semakin hangat. Frekuensi banjir di india bervariasi secara geografis karena beberapa wilayah lebih rentan terhadap banjir dari pada yang lain. Di kerala banjir terjadi pada tahun 1923, 1924, 1961, 2018. Dari semua itu, banjir pada tahun 2018 adalah banjir yang paling merusak, meskipun Otoritas Manajemen Bencana Negara Bagian Kerala merumuskan Rencana Manajemen Bencana pada tahun 2016. Kegagalan bendungan adalah penyebab utama banjir tahun 2018. Banjir ini menyebabkan kerusakan ekonomi yang besar, memengaruhi jutaan orang, dan menyebabkan kematian lebih dari 400 orang. Penyebab lain untuk banjir ini mungkin disebabkan oleh penggunaan lahan yang buruk/perubahan tutupan lahan, kondisi hidrologi pendahuluan, penyimpanan dan operasi waduk, penyerbuan daerah banjir, dan faktor alami lainnya. Namun, manajemen banjir memainkan peran penting dalam banjir Kerala 2018 (Naik, 2021).

Curah hujan adalah jumlah air yang turun ke permukaan tanah datar selama periode waktu tertentu, diukur dalam satuan tinggi (mm) di atas permukaan horizontal, tanpa memperhitungkan evaporasi, *runoff*, dan infiltrasi. Curah hujan diukur sebagai jumlah air yang jatuh di permukaan bidang datar dalam periode waktu tertentu, seperti harian, mingguan, bulanan, atau tahunan. Intensitas curah hujan yang tinggi, yang sering disebut sebagai hujan ekstrem, dapat menyebabkan terjadinya banjir. Klasifikasi curah hujan sangat penting karena membuat informasi tentang pola hujan di wilayah tersebut lebih akurat dan khusus. Lama curah hujan intensitas tinggi merupakan faktor penting lain yang menentukan risiko banjir (Ginting, 2021).

2. LANDASAN TEORI

1. Analisis

Analisis data merupakan tahapan penting dalam penelitian yang dilakukan setelah semua data yang diperlukan telah terkumpul secara lengkap. Tingkat keakuratan dan ketepatan dalam penggunaan alat analisis sangat mempengaruhi validitas kesimpulan yang diambil. Oleh karena itu, kegiatan analisis data merupakan aspek yang tidak boleh diabaikan dalam proses penelitian. (Machali, 2021).

2. Data Mining

Menurut (Mabrur,Lubis 2012) dalam jurnal (Noviyanto, 2020) Data mining adalah proses analisis data dari berbagai perspektif dan menghasilkan informasi penting yang dapat digunakan untuk meningkatkan keuntungan, mengurangi biaya pengeluaran, atau bahkan kedua hal tersebut. Secara teknis, data mining merupakan proses untuk menemukan korelasi atau pola dari ratusan atau ribuan bidang data dalam sebuah basis data relasional yang besar.

3. Klasifikasi

Menurut KBBI dalam jurnal (Ramadhani, 2019) Klasifikasi merupakan penyusunan sistematis dalam kelompok atau golongan sesuai dengan aturan atau standar yang berasal dari kata serapan bahasa Belanda, "classificatie", yang juga berasal dari bahasa Prancis, "classification". Maksud dari klasifikasi adalah menemukan suatu model dari kumpulan data latihan yang dapat memisahkan atribut ke dalam kategori atau kelas yang tepat. Setelahnya, model tersebut dimanfaatkan untuk mengklasifikasikan atribut yang tidak memiliki label kelas..(Hasanah et al., 2021).

4. *Rapid Miner*

RapidMiner adalah sebuah perangkat lunak yang terkenal dalam analisis data dan prediksi. Ini adalah salah satu platform yang kuat dan mudah digunakan untuk membangun model prediktif, melakukan analisis data, dan mengekstraksi wawasan dari berbagai sumber data. (Nabila et al., 2021).

Keunggulan *RapidMiner* termasuk kemampuan untuk mengelola dan mengintegrasikan berbagai sumber data, melakukan pemrosesan data yang efisien, dan menyediakan visualisasi yang kuat untuk membantu orang memahami apa yang dihasilkan dari analisis data. Untuk *Rapid Miner*, semua fasilitas sudah tersedia, jadi Anda tidak perlu tahu kode. Rapid Miner didedikasikan untuk data mining. (Sudarsono et al., 2021).

5. *Support Vector Machine*

Menurut (Noble, 2006) berdasarkan jurnal (Febriyanti & Alamsyah, 2022) Algoritma Support Vector Machine mengganti data pelatihan asli ke dimensi yang lebih tinggi dengan menggunakan pemetaan *non linier*. Oleh karena itu, *hyperplane* akan dicari untuk membagi secara linier dengan dimensi baru, dan dengan pemetaan *non linier* yang tepat ke dimensi yang lebih tinggi, data dari dua kelas selalu dapat dibagi dengan *hyperplane* tersebut.

Formulasi yang digunakan yaitu:

$\frac{1}{2}||w||^2 = \frac{1}{2}(w_1 + w_2)$

……..**(2.1)**

Dengan syarat :

$y_1(w.x_i + b) \geq 1, \ldots\ldots \quad i =$ 1,2,3, …. N ……. …..**(2.2)**

$y_i(w_1.x_1 + w_2.x_2 + b) \geq +1$ (Banjir)…….....**(2.3)**

$y_i(w_1.x_1 + w_2.x_2 + b) \geq -1$ (Tidak Banjir)…. ……....**(2.4)**

$|w| =$

$\begin{pmatrix} -1, jika\ w.z + z.b\ x < 0 \\ 1, jika\ w.z + z.b\ x > 0 \end{pmatrix}$……**(2.5)**

6. *Decision Tree*

Menurut (Wahyudin, 2009) berdasarkan jurnal (Fauziningrum, 2021) *Decision Tree* adalah struktur *flowchart* yang terdiri dari *Tree* Pohon, atau pohon, di mana setiap simpul internal menunjukkan suatu tes atribut, setiap cabang menunjukkan hasil tes, dan setiap simpul di daun menunjukkan kelas atau distribusi kelas. *Decision tree* adalah struktur *flowchart* yang mirip dengan pohon; setiap simpul internal menandakan tes pada atribut, setiap cabang menunjukkan hasil tes, dan setiap simpul daun menunjukkan distribusi kelas atau kelas.(Yunial, 2020).

Untuk menghitung nilai *entropy* digunakan persamaan :

$\boldsymbol{Entropy(S)} = \sum_{i=1}^{k} -\boldsymbol{pi} * \boldsymbol{log2\ pi}$

……..(2.6)

Keterangan :

S = Himpunan Kasus

Pi = Proporsi dari Si terhadap S

K = Jumlah kelas data himpunan

Kemudian hitung nilai *gain* menggunakan rumus:

$\boldsymbol{Gain\ (S,\ A)} = \boldsymbol{Entropy(S)} - \sum_{i=1}^{n} \frac{[Si]}{[S]} * \boldsymbol{Entropy(Si)}$

……..(2.7)

Keterangan :

S = Himpunan Kasus

A = Fitur

n = Jumlah partisi atribut A

$|Si|$ = Proporsi Si terhadap S

s = Jumlah kasus dalam S

7. *Naïve Bayes*

Menurut (Pattekari & Parveen, 2012) dalam jurnal (Ramadhani, 2019). Salah satu kelebihan metode *naïve bayes* adalah menghasilkan akurasi yang tinggi dan hanya membutuhkan jumlah data yang sedikit untuk menentukan estimasi parameter yang dibutuhkan dalam proses klasifikasi.

Klasifikasi *naïve bayes* adalah klasifikasi statistik yang berfungsi untuk memprediksi kemungkinan keanggotaan kelas, seperti kemungkinan masuknya tupel ke dalam kelas tertentu. Klasifikasi ini didasarkan pada teorama bayes asumsi *naïve bayes* bahwa pengaruh nilai atribut pada kelas tertentu tidak bergantung pada nilai atribut lainnya.(Triyanto *et al*., 2021)

Adapun persamaan untuk mencari nilai rata-rata hitung (*mean*) adalah seperti dalam persamaan berikut ini :

$$\mu = \frac{1}{n}\sum_i^n Xi \text{ ……..…… } \textbf{(2.8)}$$

Atau

$$\mu = \frac{x1+x2+x3+\cdots+xn}{n} \text{ …… } \textbf{(2.9)}$$

Keterangan :

$\boldsymbol{\mu}$: Nilai rata rata hitung (*mean*)

$\boldsymbol{x_i}$: nilai x ke-i

$\boldsymbol{n}$: jumlah sampel

Sedangkan persamaan untuk menghitung nilai simpangan baku (standar deviasi) dirumuskan sebagai berikut :

$$\boldsymbol{\sigma = \frac{\sqrt{\sum_1^n=1\ (x1-\mu^2}}{n-1}} \textbf{ ……..……. (2.10)}$$

Keterangan :

$\boldsymbol{\sigma}$ **:** Standar Deviasi

$\boldsymbol{x_i}$ **:** Nilai x ke i

$\boldsymbol{\mu}$ **:** Nilai rata-rata hitung

$\boldsymbol{n}$ **:** Jumlah sampel

Mencari Nilai Probabilitas :

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)} \text{ ..(2.11)}$$

Keterangan :

X = Data dengan *class* yang belum diketahui

H = Hipotesis data X merupakan suatu *class* spesifikasi

P(H|X) = Probabilitas hipotesis H berdasarkan kondisi X

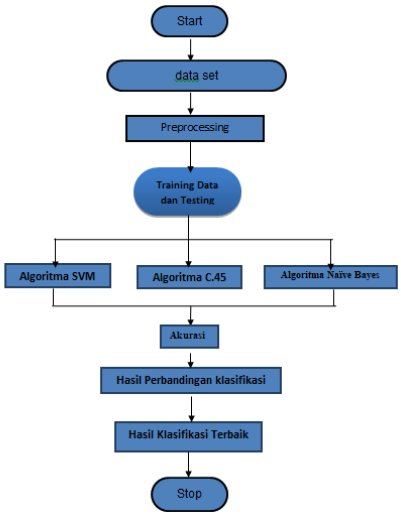
P(H) = Probabilitas hipotesis H

P(X|H) = Probabilitas X berdasarkan Hipotesis H

P(X) = Probabilitas dari X

3. METODOLOGI PENELITIAN

Adapun tahapan penelitian dapat dilihat digambar 1 :



Gambar 1. Tahapan Penlitian

4. HASIL DAN PEMBAHASAN

A. Pengumpulan Data

Data yang digunakan dalam penelitian ini adalah data publik. Data yang digunakan untuk analisis berasal dari India Kerala State. (Naik, 2021). Data ini mencakup jumlah curah hujan selama 115 tahun, dari tahun 1901 hingga 2015.(Naik, 2021). Saat ini, data telah diperbarui dari tahun 1901 hingga

2018, dan label menunjukkan *YES* untuk banjir dan *NO* untuk tidak banjir.
Data ini diambil di kaggle dan di posting oleh Mukul. Dataset secara keseluruhan di unduh dihalaman:
https://www.kaggle.com/code/mukulthakur177/flood-prediction-model/notebook

B. Prepocessing

Sebelum pengolahan data, proses *preprocessing* dilakukan, di mana data dipilih dan diubah menjadi bentuk yang lebih bermanfaat. Karena tidak semua data dibutuhkan untuk proses data mining, proses data seleksi melibatkan pemilihan dan pemilihan data yang relevan dari database untuk dianalisis.

Tabel 1. Kriteria yang dikurangi

SUB DIVISION	ANNUAL RAIN FALL

Dilanjutkan dengan *pre-processing* data, pembuatan model, pelatihan data (*Training)*, dan *Testing*. Saat proses *preprocessing* data dimulai, beberapa hal dilakukan, seperti dengan menghilangkan data *Sub Division* yang berupa data *text* nama Kerala dan *Annual Rainfall.*
data *training* dibagi 80%, yaitu memiliki 94 *record*, sedangkan data *testing* dibagi 20%, yaitu memiliki 24 *record* data dari jumlah keseluruhan data 118.

a) Analisis Kinerja *Support Vector Machine*

Tabel 2. Data Testing

Tahun	Jun	Jul	Aug	Sep
1906	625.4	1048.5	1398.9	423.6
1907	340.5	1027.6	356.4	100.5
1908	663.1	1025.1	320.6	222.4

Klasifikasi sebagai keterangan Banjir apabila **nilai lebih dari 0** apabila **kurang dari nol** di keterangan Tidak banjir.

$$|w| = \begin{cases} -1, jika\ w.z + z.b\ x < 0 \\ 1, jika\ w.z + z.b\ x > 0 \end{cases}$$

Dihitung benar didata testing yang pertama tidak banjir, input data atau masukkan nilai masing-masing bobot yang telah diperoleh yaitu:

Tabel 3. Hasil Pengujian

Tahun	Jun	Jul	Aug	Sep
1906	625.4	1048.5	1398.9	423.6
1907	340.5	1027.6	356.4	100.5
1908	663.1	1025.1	320.6	222.4

1906 = -625.4(0.000406931) -1048.5(3971440303) -1398.9(-2.51254) -423.6(8.80231) -4.55326
= 0.254494616 -4.16752 +5.5723 -7.07147 -4.55326
= -3.75

1907 = 340.5(0.000406931) + 1027.6(3971440303) + 356.4(-2.51254) + 100.5(8.80231) -4.55326
= 0.138559988 + 4.08105 -8.95469 +8.84633 -4.55326
= 2.45

1908 = 663.1(0.000406931) + 1025.1(3971440303) + 320.6(-2.51254) + 222.4(8.80231) -4.55326
= 0.269835912 + 4.07112 -8.0552 +1.95763 - 4.55326
= -2.46

Setelah kita lihat dikategori dikelas positif atau negative ternyata hasil > 0, klasifikasi banjir

1. Pengujian Akurasi *Support Vector Machine*

X-Validation	Accuracy
2	43.62%
3	54.37%
4	52.08%
5	48.89%
6	57.36%
7	57.22%
8	57.48%
9	56.46%
10	58.56%

Tabel 4. Pengujian Accuracy SVM

Berdasarkan tabel 12. Terlihat bahwa percobaan dengan 10 variabel periode dengan menggunakan *cross validation* dan mengganti nilai *input number of validation* = 2-10. Dari hasil percobaan yang dilakukan dengan menggunakan algoritma SVM, maka diperoleh nilai tertinggi pada *x-validation* = 10 *accuracy* 58.56%+/- 10.22%(mikro: 58.51%).

2. Pengujian Akurasi *Decision Tree*

X-Validition	Accuracy
2	46.81%
3	52.12%
4	45.83%
5	45.56%
6	49.93%
7	52.98%
8	49.91%
9	46.67%
10	48.78%

Tabel 5. Pengujian Accuracy C4.5

Berdasarkan tabel 39. Terlihat bahwa percobaan dengan 10 variabel periode dengan menggunakan *cross validation* dan mengganti nilai *input number of validation* = 2-10. Dari hasil percobaan yang dilakukan dengan menggunakan algoritma C4.5, maka diperoleh nilai tertinggi pada *x-validation* = 7 *accuracy* 52.98%+/- 7.11%(mikro: 53.19%).

3. Pengujian Accuracy *Naïve Bayes*

X-Validition	Accuracy
2	48.94%
3	40.52%
4	54.35%
5	49.88%
6	49.86%
7	51.96%
8	54.92%
9	49.90%
10	56.33%

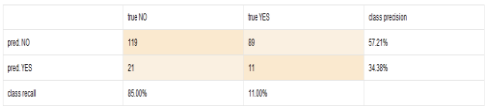
Tabel 6. Pengujian Accuracy *Naïve Bayes*

Berdasarkan Tabel 6. Terlihat bahwa percobaan dengan 10 variabel periode dengan menggunakan *cross validation* dan mengganti nilai *input number of validation* = 2-10. Dari hasil percobaan yang dilakukan dengan menggunakan algoritma *Naïve Bayes*, maka diperoleh nilai tertinggi pada *x-validation* = 10 *accuracy* 56.33%+/- 9.57%%(mikro: 56.38%)

Keterangan :
Dari ketiga pengujian yang dilakukan dengan algoritma *Support Vector Machine, Decision Tree*, dan *Naïve Bayes* maka diperoleh nilai *accuracy* terbaik terdapat pada Algoritma *Support Vector Machine* dengan nilai akurasi 56.33%+/- 9.57%%(mikro: 56.38%) *class precision* 55.38% dan *class recall* 75.00%.

C. Hasil Klasifikasi

Pada tahapan penelitian menggunakan data training sebanyak 94 data dan data testing 24. Dengan X-Validation = 10. Maka hasil klasifikasi menggunakan ketiga algoritma untuk klasifikasi bencana banjir dapat dilihat pada gambar dibawah ini.



	true NO	true YES	class precision
pred. NO	119	89	57.21%
pred. YES	21	11	34.38%
class recall	85.00%	11.00%	

Gambar 2. Hasil Klasifikasi SVM

Untuk hasil klasifikasi SVM menggunakan data testing 20% atau 24 data. Dan menggunakan operator *Cross Validation* dengan *Number Fold* 10 kali pengujian, jadi total data testing menjadi 240 data testing.

	true NO	true YES	class precision
pred. NO	18	1	94.74%
pred. YES	122	99	44.80%
class recall	12.86%	99.00%	

Gambar 3. Hasil Klasifikasi C4.5

Untuk hasil klasifikasi C4.5 menggunakan data testing 20% atau 24 data. Dan menggunakan operator *Croos Validation* dengan *Number Fold* 10 kali pengujian. Jadi total data testing menjadi 240 data testing.

	true NO	true YES	class precision
pred. NO	62	23	72.94%
pred. YES	78	77	49.68%
class recall	44.29%	77.00%	

Gambar 4. Hasil Klasifikasi Naïve Bayes

Untuk hasil klasifikasi Naïve Bayes menggunakan data testing 20% atau 24 data. Dan menggunakan operator *Croos Validation* dengan *Number Fold* 10 kali pengujian. Jadi total data testing menjadi 240 data testing.

5. KESIMPULAN

Berdasarkan hasil penelitian yang sudah dilakukan, maka dapat diambil kesimpulan sebagai berikut:

1. Dengan menganalisa hasil perhitungan perbandingan dengan menggunakan algoritma *Support Vector Machine*, *Decision Tree*, dan *Naïve Bayes* maka dari pengujian hasil klasifikasi bencana banjir menggunakan 118 data sampel yang telah dilakukan diperoleh nilai akurasi yang berbeda jauh, tingkat akurasi *Support Vector Machine* sebesar 58.56%, dan tingkat akurasi *Decision Tree* sebesar 52.98%, dan tingkat akurasi algoritma *Naïve Bayes* sebesar 56.33%.
2. Nilai keakurasian tertinggi hasil klasifikasi yang ditinjau dari parameter algoritma *Support Vector Machine* 58.56%
3. Dari ketiga akurasi berhasil divalidasi dan di implementasi dapat mengklasifikasikan bencan banjir. Berdasarkan data *training* dan pengujian langkah berikutnya melakukan klasifikasi sehingga menghasilkan kesimpulan bahwa jumlah data *training* mempengaruhi pada presentase keakurasian.
4. Penerapan algoritma *Support Vector Machine*, *Decision Tree*, dan *Naïve Bayes* untuk mengklasifikasi data bencana banjir mendapatkan hasil akurasi yang belum memuaskan dikarenakan dengan jumlah *record* data yang sedikit.

DAFTAR PUSTAKA

Fauziningrum. (2021). Penerapan Data Mining Metode Decision Tree Untuk Mengukur Penguasaan Bahasa Inggris Maritim (Studi Kasus Di Universitas Maritim Amni). *Jurnal Sains Dan Teknologi Maritim*, *22*(1), 41–50.

Febriyanti, A., & Alamsyah, A. (2022). Peningkatan Akurasi Algoritma Support Vector Machine Menggunakan Dual-Tree Complex Wavelet Transform Pada Klasifikasi Citra Katarak dan Normal. *Indonesian Journal of Mathematics and Natural Sciences*, *45*(2), 51–58.

Ginting, S. (2021). Analisis Curah Hujan Penyebab Banjir Bandang Di Ujung Berung, Bandung. *Akselerasi : Jurnal Ilmiah Teknik Sipil*, *2*(2), 9–17.

Hasanah, M. A., Soim, S., & Handayani, A. S. (2021). Implementasi CRISP-DM Model Menggunakan Metode Decision Tree dengan Algoritma CART untuk Prediksi Curah Hujan Berpotensi Banjir. *Journal of Applied Informatics and Computing*, *5*(2), 103–108.

Mabrur, L. (2012). PENERAPAN DATA

MINING UNTUK MEMPREDIKSI Program Studi Teknik Informatika Jurnal Komputer dan Informatika (KOMPUTA). *Jurnal Komputer Dan Informatika (KOMPUTA)*, *1*, 53–57.

Machali. (2021). Metode Penelitian Kuantitatif: Panduan Praktis Merencanakan, Melaksanakan dana Analisis dalam Penelitian Kuantitatif. In Abdau Quraini Habib (Ed.), *Fakultas Imu Tarbiyah dan Keguruan. Universitas Islam Negeri* (Issue April). UIN.

Nabila, Z., Rahman Isnain, A., & Abidin, Z. (2021). Analisis Data Mining Untuk Clustering Kasus Covid-19 Di Provinsi Lampung Dengan Algoritma K-Means. *Jurnal Teknologi Dan Sistem Informasi (JTSI)*, *2*(2), 100–108.

Naik, S. (2021). *Flood Prediction using Logistic Regression for Kerala State*. *9*(3), 2020–2022.

Noble. (2006). Support vector machine learning from heterogeneous data: An empirical analysis using protein sequence and structure. *Bioinformatics*, *22*(22), 2753–2760.

Noviyanto. (2020). Penerapan Data Mining dalam Mengelompokkan Jumlah Kematian Penderita COVID-19 Berdasarkan Negara di Benua Asia. *Paradigma - Jurnal Komputer Dan Informatika*, *22*(2), 183–188.

Ramadhani, N. D. (2019). *Analisis Kinerja Algoritma Naïve Bayes Classifier dengan seleksi fitur CHI-SQUARE DAN CONFIX STRIPPING STREMMER PADA SISTEM KLASIFIKASI TERJEMAHAN HADIST*. Universitas Islam Negeri Syarief Hidayatullah Jakarta.

Sudarsono, B. G., Leo, M. I., Santoso, A., & Hendrawan, F. (2021). Analisis Data Mining Data Netflix Menggunakan Aplikasi Rapid Miner. *JBASE - Journal of Business and Audit Information Systems*, *4*(1), 13–21.

Triyanto, S., Sunyoto, A., & Arief, M. R. (2021). Analisis Klasifikasi Bencana Banjir Berdasarkan Curah Hujan Menggunakan Algoritma Naïve Bayes. *JOISIE (Journal Of Information Systems And Informatics Engineering)*, *5*(2), 109–117.

Wahyudin. (2009). Metode Iterative Dichotomizer 3 (ID3) Untuk. *Indonesian Community on Information and Communication Technology*, *2*(2), 5–15.

Yunial. (2020). Analisa Perbandingan Algoritma Klasifikasi Support Vector Machine, Decession Tree Dan Naive Bayes. *Prosiding Seminar Informatika Dan Sistem Informasi*, *5*(2), 138–156.