

Penerapan Metode Komparasi Probabilitas Penentuan Penyakit Jantung

Vincencius Agung Wibowo Joyopranoto¹⁾, Mikha El Charisto Rahardjo²⁾, Bernaditus Yordan Baraga³⁾, Diwahana Mutiara Candrasari Hermanto⁴⁾

^{1,2,3,4)} Program Studi Teknik Informatika, Sekolah Tinggi Ilmu Komputer Yos Sudarso
Jl. SMP 5, Windusara, Karangklesem, Kec. Purwokerto Sel., Kabupaten Banyumas, Jawa Tengah 53144

Info Articles

Keywords:
Data Analysis; Heart Disease; K-Nearest Neighbor; Naïve Bayes; Probability.

Abstrak

Penyakit jantung, sebagai salah satu penyebab utama kematian global, memerlukan pendekatan diagnostik yang cepat dan akurat untuk pencegahannya dan pengelolaannya yang efektif. Dengan berkembangnya era teknologi informasi dan analisis data, metode komparasi probabilitas menawarkan perspektif baru dalam evaluasi risiko dan diagnosis penyakit jantung, memanfaatkan pendekatan probabilistik untuk menilai kemungkinan individu terkena penyakit ini berdasarkan analisis berbagai faktor risiko. Penelitian ini bertujuan untuk mengimplementasikan dan membandingkan algoritma Naïve Bayes dan KNN untuk mengetahui klasifikasi penyakit jantung. Dataset tersebut terdiri dari 300 data dengan 14 atribut yaitu keberadaan penyakit jantung. metode KNN dan Naïve Bayes digunakan untuk menentukan klasifikasi penyakit jantung, dengan mengeksplorasi efektivitas kombinasi atribut dalam meningkatkan akurasi diagnosa. Hasil penelitian menunjukkan bahwa Naïve Bayes, baik dengan atau tanpa normalisasi, mencapai akurasi sebesar 92.67%, sedangkan KNN normalisasi menghasilkan akurasi lebih tinggi yaitu 93.33% dibandingkan tanpa normalisasi yang hanya menghasilkan 79.33%. Penelitian ini menyimpulkan bahwa metode komparasi probabilitas, melalui algoritma Naïve Bayes dan KNN, memiliki potensi signifikan dalam meningkatkan akurasi klasifikasi risiko penyakit jantung, dengan penggunaan normalisasi pada KNN memberikan peningkatan akurasi yang signifikan. Kesimpulan ini mendukung potensi integrasi metode analisis data probabilitas dalam sistem kesehatan untuk deteksi dini dan manajemen penyakit jantung yang lebih efisien.

Abstract

Heart disease, a major global cause of mortality, necessitates a swift and precise diagnostic approach for effective prevention and management. In the era of information technology, probabilistic comparison methods such as Naïve Bayes and KNN offer a fresh perspective on assessing the risk and diagnosis of heart disease. This research, based on a dataset of 300 records with 14 attributes indicating the presence of heart disease, implements and compares these algorithms. The study reveals that Naïve Bayes, with or without normalization, achieves an accuracy of 92.67%, while normalized KNN outperforms with 93.33% accuracy, compared to 79.33% without normalization. Conclusively, the study supports the significant potential of probabilistic data analysis methods, emphasizing the integration of these techniques in the healthcare system for more accurate risk classification, early detection, and efficient management of heart disease.

✉ Alamat Korespondensi

E-mail: ¹yincensius.202203008@student.stikomvos.ac.id, ²mikha.202203011@student.stikomvos.ac.id,
³bernaditus.202203024@student.stikomvos.ac.id, ⁴andrasari5860@stikomvos.ac.id

p-ISSN 2621-9484

e-ISSN 2620-8415

PENDAHULUAN

Penyakit jantung merupakan salah satu penyakit yang memiliki dampak serius terhadap kesehatan masyarakat global. Menurut *World Health Organization* (WHO), penyakit jantung dan pembuluh darah merupakan penyebab kematian nomor satu di dunia. Oleh karena itu, penentuan penyakit jantung dengan cepat dan akurat sangat krusial dalam upaya pencegahan dan pengelolaan penyakit tersebut [1][2][3]. Jantung merupakan organ manusia yang berperan dalam sistem peredaran darah. Penyakit jantung adalah sebuah kondisi dimana jantung tidak dapat melaksanakan tugasnya dengan baik. Secara global, penyebab kematian nomor satu setiap tahunnya adalah penyakit kardiovaskuler. Penyakit kardiovaskuler adalah penyakit yang disebabkan gangguan fungsi jantung dan pembuluh darah[4][5].

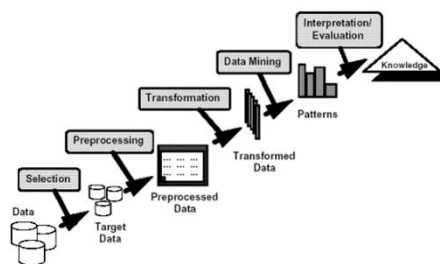
Dalam era teknologi informasi dan perkembangan metode analisis data, penerapan metode komparasi probabilitas menjadi suatu pendekatan yang menarik dalam menentukan risiko dan diagnosis penyakit jantung. Metode ini memungkinkan penggunaan probabilistik untuk mengukur sejauh mana suatu individu memiliki kemungkinan terkena penyakit jantung berdasarkan sejumlah faktor risiko yang dianalisis [2][7]. Hasran membahas penelitiannya menggunakan algoritma *K-Nearest Neighbor* terhadap dataset penyakit jantung yang didapat dari pusat dataset UCI Machine Learning Repository dan memperoleh hasil nilai $K=6$ dengan akurasi sebesar 85% [3].

Berdasarkan atas hal tersebut, penelitian ini bertujuan untuk Komporasi antara *K-Nearest Neighbor* dan *Naïve Bayes*.

METODE

1. Data mining

Data mining adalah analisa terhadap data untuk menemukan hubungan yang jelas serta menyimpulkannya yang belum diketahui sebelumnya dengan cara terkini dipahami dan berguna bagi pemilik data tersebut [4]. Data mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan machine learning untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar [5]. Metode *data Mining* dalam penelitian ini berguna untuk mengetahui pola *dataset* sehingga dapat dijadikan sebagai faktor pengambilan keputusan dalam memprediksi resiko penyakit jantung. Pada gambar 1 adalah proses data meaning.



gambar 1 Metode Data Meaning

2. K-Nearest Neighbor

K-Nearest Neighbor (KNN) adalah sebuah metode supervised yang berarti membutuhkan data training untuk mengklasifikasikan objek yang jaraknya paling dekat. Prinsip kerja *K-Nearest Neighbor* adalah mencari jarak terdekat antara data yang akan di evaluasi dengan k tetangga (*neighbor*) dalam data penyakit jantung [6]. Kelebihan *K-Nearest Neighbor* adalah tidak terlalu terpengaruh jika banyak *data noise* [12].

3. Naïve Bayes

Naïve Bayes merupakan metode pengklasifikasian probabilistik sederhana. Metode ini akan menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset yang diberikan. Metode naive bayes menganggap semua atribut pada setiap kategori tidak memiliki ketergantungan satu sama lain [7]. Kelemahan penggunaan *Naive Bayes* yaitu lebih efesien jika data yang digunakan berjumlah sedikit latih untuk menentukan *accuracy* dari *dataset* penyakit jantung [14].

4. Klasifikasi

Klasifikasi merupakan proses mengelompokkan suatu data menjadi kelas-kelas tertentu dengan memanfaatkan struktur data. Pengklasifikasian data dapat diselesaikan dengan berbagai metode, contohnya *K-Nearest Neighbor* dan *Naïve Bayes* [15]. Penggunaan metode dalam klasifikasi berguna untuk mempermudah proses atribut *dataset* penyakit jantung. Untuk mempermudah estimasi, penelitian ini menggunakan *dataset* yang sudah tersedia untuk umum dari situs *Kaggle*. *Dataset* yang digunakan memiliki judul *Heart Disease* dan memuat 300 data dengan 12 atribut, dimana 11 atribut berupa input dan 1 atribut berupa target. Target yang dihasilkan adalah '0' untuk data pasien yang tidak mengalami penyakit jantung dan '1' untuk data pasien yang mengalami penyakit jantung. Pada tabel 1 adalah klasifikasi masing-masing atribut.

Tabel 1 Informasi Atribut

NO	Atribut	Keterangan
1	Umur	-

2	Jenis Kelamin	-
		ATA (1)
		NAP (2)
3	Tipe Nyeri Dada	ASY (3)
		TA (4)
4	Tekanan Dada Istirahat	Tekanan Darah Istirahat
5	Kolesterol	Kolestrol
6	Gula Darah Puasa	Gula Darah Puasa (+1)
		NORMAL (1)
7	Elektrokardiogram Istirahat	ST (2)
		LVH (3)
8	Denyut Nadi Maksimun	Denyut Nadi Maksimum
	Nyeri Dada	N (1)
9		Y (2)
10	Pengukuran Tes Jantung	Pengukuran Tes Jantung (+1)
	Kemiringan Elektrokardiogram	DOWN (1)
11		FLAT (2)
		UP (3)
	Penyakit Jantung	0 (Tidak)
12		1 (Ya)

HASIL DAN PEMBAHASAN

1. Dataset

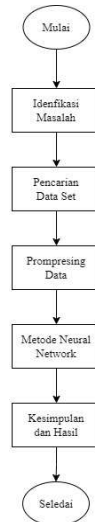
Dataset 300 pada data penyakit jantung. penelitian ini menggunakan metode *naïve bayes* dan knn untuk menghasilkan akurasi yang lebih baik dan akurat. *Dataset* terdiri dari 12 atribut diantaranya umur, jenis kelamin, tipe nyeri dada, tekanan darah istirahat, kolesterol, gula darah puasa, elektrokardiogram istirahat, denyut nadi maksimum, nyeri dada, pengukuran tes jantung, kemiringan elektrokardiogram penyakit jantung.

Tabel 1 DataSet Penyakit Jantung

No	Umur	Jenis Kelamin	Tipe Nyeri Dada	Tekanan Darah Istirahat	Kolesterol	Gula Darah Puasa	Elektrokardiogram Istirahat	Denyut Nadi Maksimum	Nyeri Dada	Pengukuran Tes Jantung	Kemiringan Elektrokardiogram	Penyakit Jantung
1	28	M	1	130	132	1	3	185	1	1	3	Tidak
2	29	M	1	120	243	1	1	160	1	1	3	Tidak
3	29	M	1	140	263	1	1	170	1	1	3	Tidak
4	30	F	4	170	237	1	2	170	1	1	3	Tidak
5	31	M	3	120	270	1	1	153	2	2,5	2	Ya
...	...	...	...	...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...	...	...	...
300	66	M	3	140	139	1	1	94	2	2	2	Ya

2. Metode Penelitian

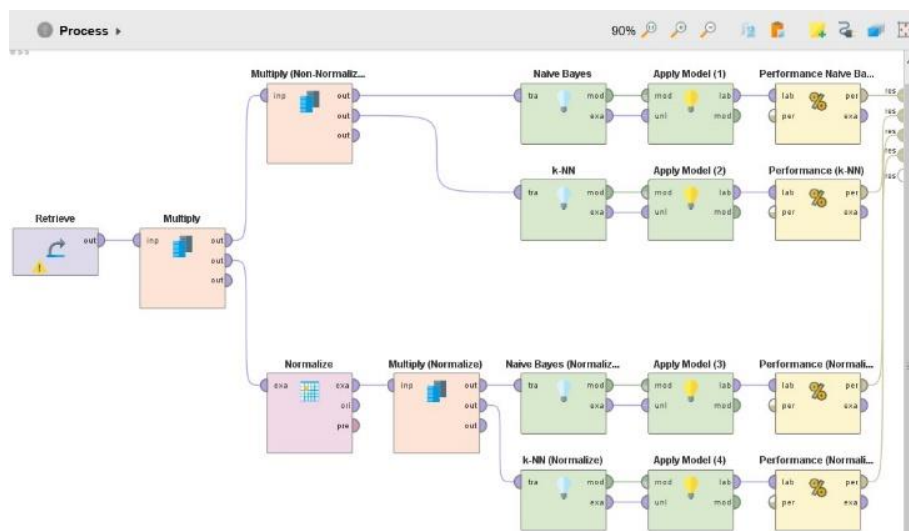
Metodologi penelitian merupakan tahapan-tahapan yang sistematis dilakukan pada penelitian ini sehingga penelitian ini terarah dengan baik. Berikut adalah metodologi penelitian yang dilakukan.



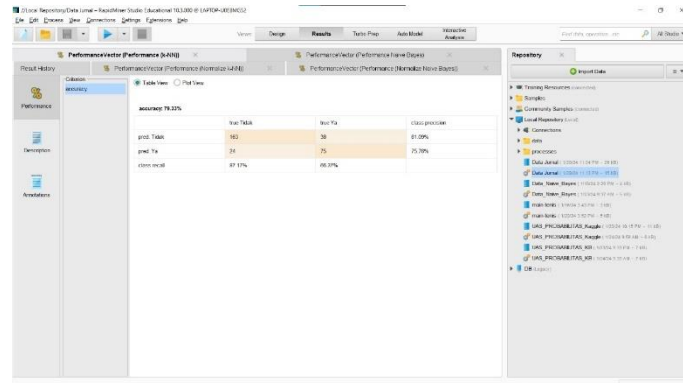
Gambar2 Medeologi Penelitian

3. Hasil Pengujian

Pengujian dilakukan untuk memutuskan apabila perhitungan yang telah dilakukan tepat untuk memutuskan status Penyakit Jantung dengan menggunakan *K-Nearest Neighbor* dan *Naïve Bayes*. Uji coba dilakukan dengan menentukan 12 data testing yang telah dipilih, Dimana 11 atribut berupa input dan 1 atribut berupa target. *Data testing* akan mencari nilai prediksinya menggunakan *RapidMiner*. Pada gambar 2 adalah proses data menggunakan *K-Nearest Neighbor* dan *Naïve Bayes*.

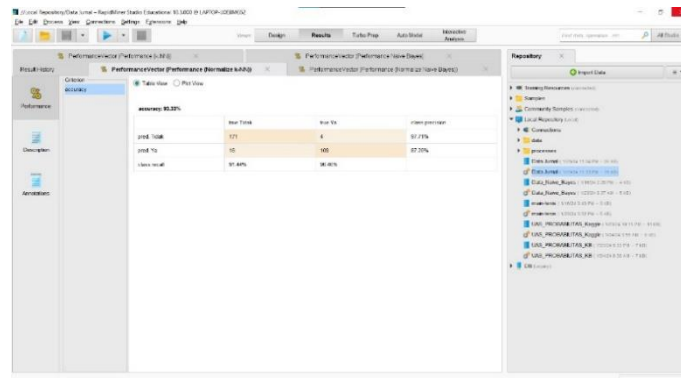


Gambar 3 Proses Data Penyakit Jantung



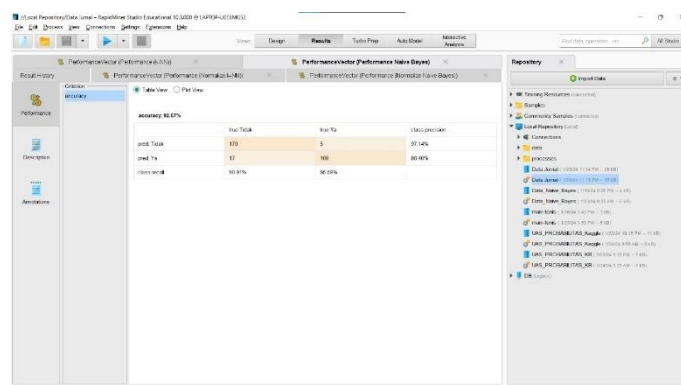
Gambar 4 K-Nearest Neighbor tanpa menggunakan normalisasi

Pada Gambar 4 adalah *accuracy* dari *K-Nearest Neighbor* tanpa menggunakan normalisasi hasilnya 79.33%. Hasil yang cukup baik.



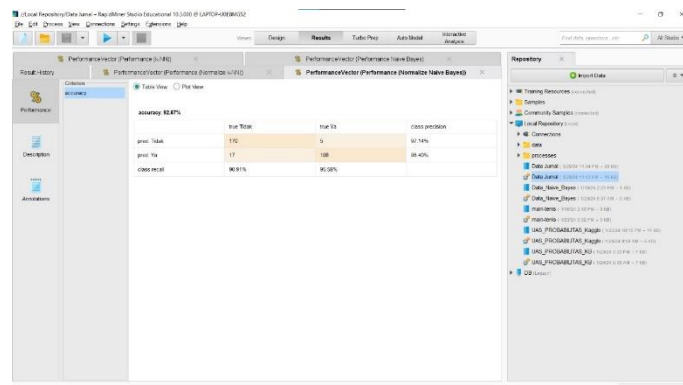
Gambar 5 K-Nearest Neighbor Menggunakan normalisasi

Pada Gambar 5 adalah *accuracy* dari *K-Nearest Neighbor* menggunakan normalisasi hasilnya 93.33%. Hasil yang sangat baik, karena keunggulan data *K-Nearest Neighbor* tidak terpengaruh data noise.



Gambar 6 Naïve Bayes tanpa menggunakan normalisasi

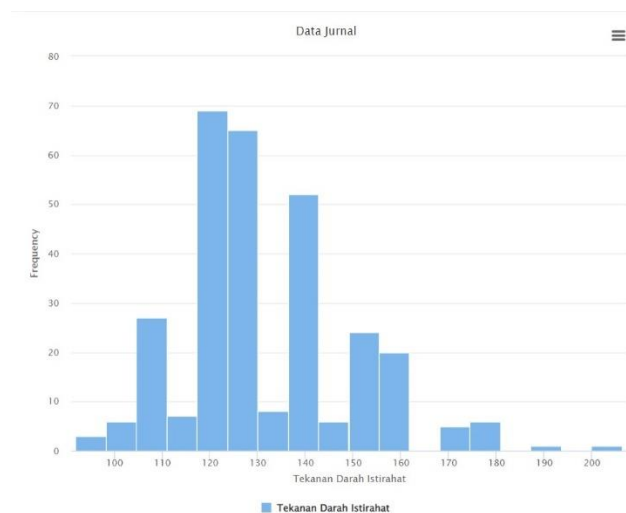
Pada Gambar 6 adalah *accuracy* dari *Naïve Bayes* tanpa menggunakan normalisasi hasilnya 92.67%. Hasil yang sangat baik, Karena keunggulan data *Naïve Bayes* adalah langsung menghitung dengan cepat hasil probabilitas.



Gambar 7 Naïve Bayes tanpa menggunakan normalisasi

Pada Gambar 7 adalah akurasi dari Naïve Bayes tanpa menggunakan normalisasi hasilnya 92.67%. Hasil yang sangat baik, Karena keunggulan data Naïve Bayes adalah langsung menghitung dengan cepat hasil probabilitas, Hasil *accuracy* Naïve Bayes tanpa menggunakan normalisasi dan menggunakan normalisasi signifikan tidak ada perubahan.

Berdasarkan 300 dataset yang telah diuji, didapatkan hasil perhitungan *accuracy* dari masing-masing *K-Nearest Neighbor* dan *Naïve Bayes*. Hasil pengukuran data *accuracy* yang diperoleh dari data *K-Nearest Neighbor* memiliki perbedaan hasil pengujian menggunakan normalisasi sebesar 93.33% sedangkan yang tidak menggunakan normalisasi sebesar 79.33%, untuk pengujian akurasi metode Naïve Bayes yang tidak menggunakan normalisasi maupun yang tidak menggunakan normalisasi memiliki hasil *accuracy* yang sama yaitu 92.67%.



Gambar 8 Tekanan darah istirahat

Pada gambar adalah salah satu data histogram batang yang memperngaruhi indikasi penyakit jantung. Dari uji coba untuk algoritma naive bayes *accuracy* yang dihasilkan cukup baik, hal ini karena keunggulan dari metode *naive bayes* sendiri yaitu sederhana dan cepat probabilitas dapat langsung dihitung dan karena data yang di normalisasi mempengaruhi hasil akurasi dari *K-Nearest Neighbor*.

KESIMPULAN

Berdasarkan hasil dari penelitian yang telah dilakukan terhadap *dataset* mengenai penyakit jantung menggunakan metode komparasi probabilitas memiliki kemampuan dalam meningkatkan akurasi untuk menentukan klasifikasi resiko penyakit jantung dengan menerapkan metode *Naive Bayes* dan *K-Nearest Neighbor*. Dari pengujian kinerja dari metode *K-Nearest Neighbor* menggunakan normalisasi lebih unggul dari metode *K-Nearest Neighbor*. Terbukti bahwa dari 300 data uji yang digunakan metode *K-Nearest Neighbor* mendapatkan hasil yang tidak signifikan tetapi saat menggunakan normalisasi *accuracy K-Nearest Neighbor* lebih tinggi karena keunggulan dari metode *K-Nearest Neighbor* tangguh terhadap noise. metode *Naive Bayes* tanpa normalisasi dan menggunakan normalisasi hasilnya signifikan. Oleh karena itu, disarankan jika ingin mendapatkan hasil *accuracy* yang baik, maka diperlukan metode *K-Nearest Neighbor*.

DAFTAR PUSTAKA

- [1] A. Ridwanmo, M. Fadillah, And T. H. Irfani, “Deteksi Dini Faktor Risiko Penyakit Jantung Dan Pembuluh Darah, Hubungan Antara Obesitas, Aktivitas Fisik Dan Kolesterol Total Di Kecamatan Kertapati, Kota Palembang,” *Jurnal Epidemiologi Kesehatan Komunitas*, Vol. 5, No. 2, 2020, Doi: 10.14710/Jekk.V5i2.6729.
- [2] K. Di Penyakit Jantung Koroner Pada Pasien Rsud Prof Dr W Z Johannes Kupang Ice J Johanis, I. A. Tedju Hinga, And A. B. Sir, “Media Kesehatan Masyarakat Faktor Risiko Hipertensi, Merokok Dan Usia Terhadap,” Vol. 2, No. 1, 2020, [Online]. Available: <https://ejurnal.undana.ac.id/mkm>
- [3] M. Saputri And S. R. Dewi, “Potensi Interaksi Polifarmasi Pasien Jantung Koroner (Pjk) Di Rumah Sakit I.A. Moeis Samarinda,” *Jurnal Sains Dan Kesehatan*, Vol. 5, No. 2, Pp. 109–114, Apr. 2023, Doi: 10.25026/Jsk.V5i2.1709.
- [4] A. Ridwanmo, M. Fadillah, Dan Tri Hari Irfani, And B. Ilmu Kesehatan Masyarakat Ilmu Kedokteran Komunitas Fakultas Kedokteran Universitas Sriwijaya, “Deteksi Dini Faktor Risiko Penyakit Jantung Dan Pembuluh Darah, Hubungan Antara Obesitas, Aktivitas Fisik Dan Kolesterol Total Di Kecamatan Kertapati, Kota Palembang.”
- [5] H. Sulastomo *Et Al.*, “Edukasi Pencegahan Dan Deteksi Dini Penyakit Kardiovaskular Di Masa Pandemi Covid-19,” 2023.
- [6] P. S. Ramadhan, “Penerapan Komparasi Teorema Bayes Dengan Euclidean Probability Dalam Pendiagnosaan Dermatic Bacterial,” *Infotekjar (Jurnal Nasional Informatika Dan Teknologi Jaringan)*, Vol. 4, No. 1, Pp. 1–7, Sep. 2019, Doi: 10.30743/Infotekjar.V4i1.1579.
- [7] K. Di Penyakit Jantung Koroner Pada Pasien Rsud Prof Dr W Z Johannes Kupang Ice J Johanis, I. A. Tedju Hinga, And A. B. Sir, “Media Kesehatan Masyarakat Faktor Risiko Hipertensi, Merokok Dan Usia Terhadap,” Vol. 2, No. 1, 2020, [Online]. Available: <https://ejurnal.undana.ac.id/mkm>

- [8] Hasran, “Indonesian Journal Of Data And Science Klasifikasi Penyakit Jantung Menggunakan Metode K-Nearest Neighbor,” Vol. 1, No. 1, Pp. 6–10, 2020, [Online]. Available: [Http://Bit.Ly/Datasetcardio](http://bit.ly/datasetcardio).
- [9] N. Noviyanto, “Penerapan Data Mining Dalam Mengelompokkan Jumlah Kematian Penderita Covid-19 Berdasarkan Negara Di Benua Asia,” *Paradigma - Jurnal Komputer Dan Informatika*, Vol. 22, No. 2, 2020, Doi: 10.31294/P.V22i2.8808.
- [10] J. Media And A. Issn, “Analisis Algoritma Clustering Dalam Kasus Penentuan Jenis Bunga Iris,” 2017.
- [11] R. N. Devita, H. W. Herwanto, And A. P. Wibawa, “Perbandingan Kinerja Metode Naive Bayes Dan K-Nearest Neighbor Untuk Klasifikasi Artikel Berbahasa Indonesia,” *Jurnal Teknologi Informasi Dan Ilmu Komputer*, Vol. 5, No. 4, Pp. 427–434, Oct. 2018, Doi: 10.25126/Jtiik.201854773.
- [12] I. A. Angreni, S. A. Adisasmita, M. I. Ramli, And S. Hamid, “Pengaruh Nilai K Pada Metode K-Nearest Neighbor (Knn) Terhadap Tingkat Akurasi Identifikasi Kerusakan Jalan,” *Rekayasa Sipil*, Vol. 7, No. 2, P. 63, Jan. 2019, Doi: 10.22441/Jrs.2018.V07.I2.01.
- [13] A. Nafalski And A. P. Wibawa, “Machine Translation With Javanese Speech Levels’ Classification,” *Informatics, Control, Measurement In Economy And Environment Protection*, Vol. 6, No. 1, Pp. 21–25, Feb. 2016, Doi: 10.5604/20830157.1194260.
- [14] A. Ashari Muin, “Metode Naive Bayes Untuk Prediksi Kelulusan (Studi Kasus: Data Mahasiswa Baru Perguruan Tinggi),” *Jurnal Ilmiah Ilmu Komputer*, Vol. 2, No. 1, 2016, [Online]. Available: [Http://Ejournal.Fikom-Unasman.Ac.Id](http://ejournal.fikom-unasman.ac.id)
- [15] A. A. Wulandari *Et Al.*, “Klasifikasi Data Mining Menggunakan Naïve Bayes Classifier Dengan Algoritma C5.0 (Classification Data Mining Using Naïve Bayes Classifier With C5.0 Algorithm).” [Online]. Available: [Https://Magestic.Unej.Ac.Id/](https://magestic.unej.ac.id/)

