

Penerapan Data Mining Untuk Prediksi Perkiraan Hujan dengan Menggunakan Algoritma K-Nearest Neighbor

Nursobah*, Siti Lailiyah, Bartolomius Harpad, Muhammad Fahmi

STMIK Widya Cipta Dharma, Samarinda, Indonesia

Email: ^{1,*}nursobah@wicida.ac.id, ²sitilailiyah@wicida.ac.id, ³harpad@wicida.ac.id, ⁴fahmi@wicida.ac.id

Email Penulis Korespondensi: nursobah@wicida.ac.id

Submitted: 21/11/2022; Accepted: 24/12/2022; Published: 30/12/2022

Abstrak—Hujan merupakan sebuah kondisi dimana jatuhnya tetesan air dari awan ke bumi. Didalam kehidupan kehadiran hujan sangatlah dinantikan, hujan dapat membantu bagi masyarakat yang memiliki profesi sebagai petani. Hujan yang terjadi dalam skala yang besar akan benar – benar memberikan kendala bagi masyarakat, selain terhambatnya kegiatan ataupun aktivitas terkhususnya yang dilakukan pada outdoor hujan juga dapat menimbulkan musibah bagi masyarakat berupa banjir. Memperkirakan hujan bagi masyarakat sangatlah penting, dengan mengetahui apakah terjadi hujan atau tidak dapat memberikan kemudahan bagi masyarakat untuk melakukan antisipasi terhadap kemungkinan – kemungkinan yang terjadi dikarenakan hujan. Namun dalam proses penyampaian perkiraan sering terjadi tidak meratanya informasi dan juga terlambatnya untuk penyampaian informasi kepada masyarakat terkait dengan terjadinya hujan atau tidak Masyarakat seharusnya mampu secara mandiri untuk memperkirakan keadaan terjadinya hujan atau tidak. Proses pengolahan data seharusnya dilakukan dengan baik dan benar. Data mining merupakan sebuah cara yang dapat dilakukan untuk membantu dalam proses pengolahan data. Pada penelitian ini akan dilakukan proses penyelesaian dengan menggunakan algoritma K-Nearest Neighbor (K-NN). Hasil yang didapatkan bahwasannya keputusan data testing yaitu TIDAK. Dengan kata lain bahwasannya data mining dan algoritma K-Nearest Neighbor dapat membantuk untuk proses penyelesaian masalah.

Kata Kunci: Data Mining; Perkiraan; Hujan; Prediksi; K-Nearest Neighbor

Abstract—Rain is a condition where water droplets fall from clouds to the earth. In life, the presence of rain is highly anticipated, rain can help people who have a profession as farmers. Rain that occurs on a large scale will really provide obstacles for the community, in addition to hampering activities or activities especially those carried out on outdoor rain can also cause disaster for the community in the form of flooding. Estimating rain for the community is very important, knowing whether it will rain or not can make it easier for the community to anticipate the possibilities that may occur due to rain. However, in the process of delivering forecasts, there is often an uneven distribution of information and delays in conveying information to the public regarding whether or not rain will occur. The community should be able to independently predict whether or not rain will occur. Data processing should be done properly and correctly. Data mining is a way that can be done to assist in data processing. In this study, the settlement process will be carried out using the K-Nearest Neighbor (K-NN) algorithm. The results obtained show that the data testing decision is NO. In other words, data mining and the K-Nearest Neighbor algorithm can help the problem solving process.

Keywords: Data Mining; Estimation; Rain; Predictions; K-Nearest Neighbor

1. PENDAHULUAN

Hujan merupakan sebuah kondisi dimana jatuhnya tetesan air dari awan ke bumi. Hujan yang terjadi didasari dengan kondisi terjadinya kondensasi uap air pada atmosfer berubah menjadi butiran – butiran air yang terkumpul menjadi awan. Butiran – butiran air yang sudah melebihi kapasitas serta terjadinya penambahan terhadap pendinginan udara dan penambahan uap air menjadikan butiran – butiran air tersebut terpecah dan jatuh ke daratan [1], [2].

Didalam kehidupan kehadiran hujan sangatlah dinantikan, hujan dapat membantu bagi masyarakat yang memiliki profesi sebagai petani. Dengan hujan dapat membantu dalam proses penyiraman tanaman hingga tidak terjadi kekeringan air serta menyebabkan tanaman mati. Selain itu juga hujan bisa menjadi sebuah hambatan bagi kehidupan masyarakat, hal tersebut dikarenakan jika terjadinya hujan maka akan terbatasnya kegiatan ataupun aktivitas yang dilakukan oleh masyarakat [3], [4].

Hujan yang terjadi dalam skala yang besar akan benar – benar memberikan kendala bagi masyarakat, selain terhambatnya kegiatan ataupun aktivitas terkhususnya yang dilakukan pada outdoor hujan juga dapat menimbulkan musibah bagi masyarakat berupa banjir. Maka dari itu, sudah sepantasnya bagi masyarakat untuk dapat mengetahui kondisi – kondisi untuk dapat memperkirakan kapan terjadinya hujan atau tidak [5], [6].

Memperkirakan hujan bagi masyarakat sangatlah penting, dengan mengetahui apakah terjadi hujan atau tidak dapat memberikan kemudahan bagi masyarakat untuk melakukan antisipasi terhadap kemungkinan – kemungkinan yang terjadi dikarenakan hujan. Tetapi dalam proses yang dilakukan dalam melakukan perkiraan hujan masih didapati kendala tersendiri.

Pada umumnya proses perkiraan hujan disampaikan oleh BMKG, BMKG sebagai lembaga yang melakukan proses pengolahan data pada cuaca dan iklim. Namun dalam proses penyampaian perkiraan sering terjadi tidak meratanya informasi dan juga terlambatnya untuk penyampaian informasi kepada masyarakat terkait dengan terjadinya hujan atau tidak [7], [8].

Masyarakat seharusnya mampu secara mandiri untuk memperkirakan keadaan terjadinya hujan atau tidak. Dimana untuk melakukan perkiraan dapat menggunakan data – data terdahulu yang telah dikumpulkan pada BMKG.

Dengan berdasarkan data terkait terjadinya hujan seharusnya masyarakat dapat untuk memperkirakan terjadinya hujan dengan melakukan proses pengolahan data.

Proses pengolahan data seharusnya dilakukan dengan baik dan benar, kesalahan dalam proses pengolahan data akan mengakibatkan hasil dari pemroses tidaklah valid. Maka dari itu perlu kiranya dilakukan proses pengolahan berdasarkan dengan ketentuan ataupun aturan – aturan yang dapat dipergunakan untuk melakukan proses pengolahan data hingga mendapatkan hasil yang valid[9], [10].

Data mining merupakan sebuah cara yang dapat dilakukan untuk membantu dalam proses pengolahan data. Data mining merupakan proses pengolahan terhadap data yang telah tersimpan pada kumpulan data ataupun bisa disebut dengan data warehouse. Proses pengolahan data berdasarkan dengan data mining dapat dimanfaatkan pada berbagai macam bidang keilmuan, hal tersebut dikarenakan hasil yang diperoleh pada data mining berdasarkan dengan model hasil seperti apa[11]–[13].

Tetapi pada data mining sendiri proses pengolahan data yang dilakukan tidak dapat berdiri sendiri, pada data mining proses pengolahan data yang dilakukan dengan menggunakan metode - metode ataupun algoritma yang terbagi dalam berbagai macam teknik penyelesaian pada data mining. Berbagai contoh metode ataupun algoritma yang dapat dipergunakan untuk menyelesaikan permasalahan pada data mining seperti FP-Growth, Apriori, Naïve Bayes, SVM, Regresi Linear, K-Means, K-Medoids dan juga beberapa metode lainnya[14]–[19].

Pada penelitian ini akan dilakukan proses penyelesaian dengan menggunakan algoritma K-Nearest Neighbor (K-NN). K-Nearest Neighbor (K-NN) merupakan bagian dari teknik klasifikasi data mining. Proses penyelesaian dengan menggunakan K-Nearest Neighbor (K-NN) berdasarkan dengan perhitungan jarak terdekat antar tiap objek (euclidean distance). Proses perhitungan berdasarkan dengan jarak dari objek yang baru kemudian menghitung setiap jarak pada objek yang lama. Kemudian, dalam proses K-Nearest Neighbor (K-NN) setelah dilakukan proses perhitungan jarak antar tiap objek selanjutnya memilih nilai K. Penentuan nilai K nantinya sebagai dasar proses pengambilan keputusan[20]–[22].

Beberapa penelitian terdahulu yang dilakukan oleh Sri Widaningsih dan Sonia Yusuf pada tahun 2022 didapatkan hasil penelitian bahwa data mining dapat dipergunakan untuk memprediksi siswa berprestasi, proses prediksi berdasarkan dengan perhitungan jarak terdekat antar setiap objek[23]. Penelitian lainnya yang dilakukan oleh Silvilestari pada tahun 2021 dengan hasil penelitian bahwa penerapan algoritma k-nearest neighbor dapat mempermudah untuk melihat pola data dari kumpulan – kumpulan data dan hasil dari pengujian memiliki akurasi yang baik[24].

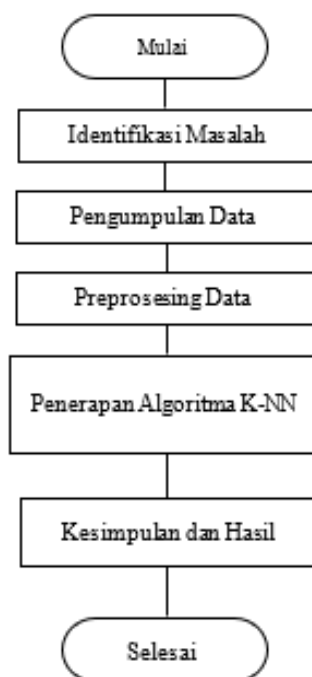
Penelitian lainnya yang dilakukan oleh Aluisius Dwiki Adhi Putra dan Safitri Juanita pada tahun 2021 dengan hasil penelitian didapatkan modelling dari proses pengujian algoritma didapatkan tingkat akurasi sebesar 91,91% [25]. Ditahun yang sama juga dilakukan penelitian oleh Nanda Satria Halim Pratama, dkk dengan hasil penelitian yaitu penerapan dengan data mining dan algoritma k-nearest neighbor dimana tingkat akurasi hasil pengujian didapatkan sebesar 71% [26].

Berdasarkan dengan hal tersebut maka pada penelitian ini akan dilakukan sebuah penelitian terkait dengan prediksi untuk perkiraan hujan. Melakukan prediksi merupakan sebuah hal penting terkhususnya bagi kehidupan masyarakat, hal tersebut dapat membantu masyarakat untuk mempersiapkan segala sesuatunya. Hujan bisa saja dibutuhkan oleh masyarakat tetapi bisa jadi sebagai pembatas kegiatan masyarakat, terkhususnya bagi yang melakukan kegiatan di luar ruangan. Maka dari itu, masyarakat seharusnya dapat untuk memperkirakan secara mandiri terhadap akan terjadi hujan atau tidak. Proses dapat dilakukan dengan menggunakan data cuaca sebelumnya. Proses penyelesaian dapat diselesaikan dengan menggunakan data mining dan memanfaatkan proses dari algoritma k-nearest neighbor.

2. METODE PENELITIAN

2.1 Kerangka Kerja Penelitian

Kerangka kerja penelitian merupakan sebuah kerangka dari tahapan proses yang dilakukan pada penelitian, dimana tahapan tersebut dari proses identifikasi hingga proses dokumentasi.



Gambar 1. Kerangka Kerja Penelitian

2.2 Data Mining

Data mining merupakan sebuah proses penggalian data. Penggalian data yang dilakukan dengan menggunakan data mining bertujuan untuk menemukan informasi baru yang tersimpan pada kumpulan data. Informasi yang dimaksud bisa berupa sebuah pola, prediksi, estimasi ataupun pengelompokan. Pada data mining proses penggalian data yang dilakukan juga berdasarkan dengan metode ataupun algoritma – algoritma tertentu. Berdasarkan hal tersebut data mining banyak digunakan untuk membantu proses penyelesaian masalah pada banyak rumpun ilmu yang berkaitan dengan pengolahan data [27], [28].

2.3 Algoritma K-Nearest Neighbor

Algoritma K-Nearest Neighbor merupakan algoritma pada data mining yang termasuk dalam klasifikasi data mining. Proses pada algoritma K-Nearest Neighbor dimulai dengan penentuan nilai K terlebih dahulu. Setelah penentuan nilai K maka selanjutnya proses yang dilakukan menghitung kedekatan nilai antar objek. Pada K-Nearest Neighbor objek yang terpilih untuk memenuhi nilai K dihitung berdasarkan dengan euclidean distance. Euclidean distance merupakan proses perhitungan jarak terdekat antara objek baru dengan objek lama yang tersimpan pada data. Adapun rumus yang digunakan untuk menyelesaikan algoritma K-Nearest Neighbor yaitu [20]–[22]:

$$Dq = \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2 + (a_n - b_n)^2} \quad (1)$$

3. HASIL DAN PEMBAHASAN

3.1 Analisa Masalah

Analisa masalah merupakan proses penjabaran terhadap permasalahan yang dihadapi pada penelitian. Pada penelitian permasalahan yang dihadapi terkait dengan proses prediksi perkiraan hujan. Didalam kehidupan hujan kadang kala kondisi yang dinantikan oleh masyarakat, tetapi hujan bisa juga sebagai kondisi untuk menimbulkan kendala terhadap aktivitas masyarakat. Proses perkiraan hujan pada umumnya dilakukan oleh badan terkait seperti BMKG, namun kadang kala informasi yang berikan terlambat untuk diterima oleh masyarakat. Maka dari itu masyarakat diharapkan mampu secara mandiri untuk dapat memperkirakan terjadinya hujan atau tidak. Proses yang dapat dilakukan dengan melakukan pengolahan data berdasarkan dengan data terjadinya hujan dimasa lampau, data tersebut dilakukan pengolahan hingga didapatkan informasi baru yang dapat dipergunakan ataupun proses ini disebut dengan data mining. Pada data mining proses penyelesaian juga membutuhkan peran dari pada algoritma ataupun metode, pada penelitian ini algoritma

3.1.1 Penerapan Algoritma K-Nearest Neighbor

Dalam proses penyelesaian algoritma K-Nearest Neighbor tahapan pertama yang dilakukan yaitu menentukan data yang akan digunakan sebagai dataset. Dataset ini nantinya akan dilakukan proses pengolahan data hingga didapatkan informasi baru. Sampel dataset dapat dilihat tabel berikut:

Tabel 1. Sampel Dataset

No	Kondisi	Suhu	Kelembapan Udara	Kecepatan Angin	Hasil
1	A1	23	65	8,5	Hujan
2	A2	27	70	9,0	Tidak
3	A3	28	80	8,8	Tidak
4	A4	24	70	10,2	Hujan
5	A5	25	75	6,3	Tidak
6	A6	26	83	4,6	Tidak
7	A7	23	68	8,6	Hujan

Setelah diketahui sampel dataset seperti tabel 1, selanjutnya yaitu mengetahui data testing. Data testing merupakan data yang akan dilakukan pengujian serta pengambilan keputusan. Pada data testing nantinya akan diketahui apakah terjadi hujan atau tidak. Adapun data testing dapat dilihat berikut:

Tabel 2. Data Testing

No	Kondisi	Suhu	Kelembapan Udara	Kecepatan Angin	Hasil
1	A1	25	77	5,8	???

Dari tabel diatas dapat dilihat bahwasannya data baru sebagai data testing yang akan dilakukan proses perhitungan dengan menggunakan algoritma K-Nearest Neighbor. Adapun proses perhitungan dapat dilihat berikut ini:

$$Dq = \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2 + (a_n - b_n)^2}$$

$$D1 = \sqrt{(25 - 23)^2 + (77 - 65)^2 + (5,8 - 8,5)^2}$$

$$= \sqrt{(2)^2 + (12)^2 + (-2,7)^2}$$

$$= \sqrt{4 + 144 + 7,29} = \sqrt{155,29} = 12,46$$

$$D2 = \sqrt{(25 - 27)^2 + (77 - 70)^2 + (5,8 - 9,0)^2}$$

$$= \sqrt{(-2)^2 + (7)^2 + (-3,2)^2}$$

$$= \sqrt{4 + 49 + 10,24} = \sqrt{63,24} = 7,95$$

$$D3 = \sqrt{(25 - 28)^2 + (77 - 80)^2 + (5,8 - 8,8)^2}$$

$$= \sqrt{(-3)^2 + (-3)^2 + (-3)^2}$$

$$= \sqrt{9 + 9 + 9} = \sqrt{27} = 5,20$$

$$D4 = \sqrt{(25 - 24)^2 + (77 - 70)^2 + (5,8 - 10,2)^2}$$

$$= \sqrt{(1)^2 + (7)^2 + (-4,4)^2}$$

$$= \sqrt{1 + 49 + 19,36} = \sqrt{69,36} = 8,33$$

$$D5 = \sqrt{(25 - 25)^2 + (77 - 75)^2 + (5,8 - 6,3)^2}$$

$$= \sqrt{(0)^2 + (2)^2 + (-0,5)^2}$$

$$= \sqrt{0 + 4 + 0,25} = \sqrt{4,25} = 2,06$$

$$D6 = \sqrt{(25 - 26)^2 + (77 - 83)^2 + (5,8 - 4,6)^2}$$

$$= \sqrt{(-1)^2 + (-6)^2 + (1,2)^2}$$

$$= \sqrt{1 + 36 + 1,44} = \sqrt{38,44} = 6,20$$

$$D7 = \sqrt{(25 - 23)^2 + (77 - 68)^2 + (5,8 - 8,6)^2}$$

$$= \sqrt{(2)^2 + (9)^2 + (-2,8)^2}$$

$$= \sqrt{4 + 81 + 7,84} = \sqrt{92,84} = 9,64$$

Setelah proses perhitungan jarak dari data testing dengan dataset, selanjutnya yaitu proses penyusunan data. Proses penyusunan data untuk melakukan proses perbandingan data dari jarak terdekat hingga jarak yang terjauh. Adapun hasil tersebut dapat dilihat pada tabel berikut:

Tabel 3. Hasil Perangkingan Data

No	Objek	Hasil Jarak	Ranking
1	D5	2,06	1
2	D3	5,20	2
3	D6	6,20	3
4	D2	7,95	4
5	D4	8,33	5
6	D7	9,64	6
7	D1	12,46	7

Setelah dilakukan proses pengurutan data, tahapan selanjutnya yang dilakukan yaitu menentukan nilai K. Pada penelitian ini nilai K yang diambil yaitu 3 (K=3) kemudian melakukan proses perangkingan kembali berdasarkan dengan data yang memiliki jarak terdekat. Data nilai K=3 dapat dilihat berikut:

Tabel 4. Data Nilai K=3

No	Objek	Hasil Jarak	Ranking	Hasil
1	D5	2,06	1	Tidak
2	D3	5,20	2	Tidak
3	D6	6,20	3	Tidak

Setelah proses perangkingan terhadap nilai K, maka proses selanjutnya yaitu pengambilan keputusan. Dari data nilai K dapat dilihat bahwasannya hasil yang didapatkan adalah TIDAK untuk keseluruhan. Maka dari itu dapat diambil keputusan bahwasannya dengan menggunakan data testing didapatkan hasil yaitu TIDAK

Setelah didapatkan data nilai K seperti tabel diatas, selanjutnya proses penentuan kelas berdasarkan dengan hasil pada nilai K. Pada data tersebut dapat dilihat bahwasannya untuk hasil terbanyak yaitu “YA” dengan 3 kali muncul. Maka dari itu hasil dari proses klasifikasi yaitu “YA”. Proses pengambilan keputusan berdasarkan dengan jumlah nilai kelas yang muncul dibandingkan dengan nilai lainnya. Maka hasil dari proses algoritma K-Nearest Neighbor yaitu

Tabel 5. Hasil Proses Algoritma K-NN

No	Kondisi	Suhu	Kelembapan Udara	Kecepatan Angin	Hasil
1	A1	25	77	5,8	Tidak

4. KESIMPULAN

Setelah dilakukan proses penelitian, tahapan akhir yang dilakukan yaitu penarikan kesimpulan. Berdasarkan dengan proses yang telah dilakukan maka dapat ditarik sebuah kesimpulan bahwasannya dengan data mining masyarakat dapat dengan sendiri untuk melakukan pengolahan data. Dalam hal ini masyarakat mampu secara mandiri untuk memperkirakan terjadinya hujan atau tidak. Hal tersebut dapat dilihat dari hasil yang didapatkan bahwasannya keputusan data testing yaitu TIDAK. Dengan kata lain bahwasannya data mining dan algoritma K-Nearest Neighbor dapat membantu untuk proses penyelesaian masalah.

REFERENCES

- [1] Syarifaturrahmah, A. Uperiati, and E. Suswaini, “PENERAPAN METODE K-NEAREST NEIGHBOR (KNN) UNTUK KLASIFIKASI CUACA DI WILAYAH TANJUNGPINANG,” *Student Online J.*, vol. 2, no. 1, pp. 250–256, 2022.
- [2] A. S. Utami, D. P. Rini, and E. Lestari, “Prediksi Cuaca di Kota Palembang Berbasis Supervised Learning Menggunakan Algoritma K-Nearest Neighbour,” *JUPITER (Jurnal Penelit. Ilmu ...)*, vol. 13, no. 1, pp. 9–18, 2021, [Online]. Available: <https://jurnal.polsri.ac.id/index.php/jupiter/article/view/3243%0Ahttps://jurnal.polsri.ac.id/index.php/jupiter/article/download/3243/1390>.
- [3] D. K. Pramudito *et al.*, “Jurnal Teknologi Pelita Bangsa DATA MINING IMPLEMENTATION ON JAVA NORTH COAST WEATHER FORECAST DATASET,” vol. 13, no. 3, 2022.
- [4] M. Y. R. Rangkuti, M. V. Alfansyuri, and W. Gunawan, “Penerapan Algoritma K-Nearest Neighbor (Knn) Dalam Memprediksi Dan Menghitung Tingkat Akurasi Data Cuaca Di Indonesia,” *Hexag. J. Tek. dan Sains*, vol. 2, no. 2, pp. 11–16, 2021, doi: 10.36761/hexagon.v2i2.1082.
- [5] A. Marbun and D. Nofriansyah, “ANALISA DATA MINING UNTUK MENGESTIMASI POTENSI CURAH HUJAN DENGAN MENGGUNAKAN METODE REGRESI,” vol. 4, no. 2, 2021.
- [6] M. Hudzaifah and A. A. Rismayadi, “Peramalan Arus Lalu Lintas Berdasarkan Waktu Tempuh Dan Cuaca Menggunakan Metode Time Series Decomposition,” *J. Responsif Ris. Sains dan Inform.*, vol. 3, no. 2, pp. 207–215, 2021, doi: 10.51977/jti.v3i2.559.
- [7] S. S. Purwandari, M. Bettiza, and A. Uperiati, “PENERAPAN ALGORITMA APRIORI UNTUK MENEMUKAN HUBUNGAN ANTARA FAKTOR CUACA DAN CURAH HUJAN (STUDI KASUS: KOTA TANJUNGPINANG),” *Student Online J.*, vol. 2, no. 1, pp. 178–183, 2022.
- [8] A. Purnamawati, M. N. Winnarto, and M. Mailasari, “Analisis Cart (Classification and Regression Trees) Untuk Prediksi

- Pengguna Sepeda Berdasarkan Cuaca,” *J. Teknoinfo*, vol. 16, no. 1, p. 14, 2022, doi: 10.33365/jti.v16i1.1478.
- [9] S. Khairunnisa and M. I. Jambak, “Pengelompokan Cuaca Kota Palembang Menggunakan Algoritma K- Means Clustering Untuk Mengetahui Pola Karakteristik Cuaca,” vol. 6, pp. 2352–2360, 2022, doi: 10.30865/mib.v6i4.4810.
 - [10] M. Andriyati, S. Ardiansyah, and M. G. L. Putra, “Penerapan Business Intelligence pada Data Cuaca Badan Meteorologi, Klimatologi, dan Geofisika (BMKG),” *SPECTA J. Technol.*, vol. 5, no. 1, pp. 13–21, 2021.
 - [11] R. R. Andarista and A. Jananto, “Penerapan Data Mining Algoritma C4 . 5 Untuk Klasifikasi Hasil Pengujian Kendaraan Bermotor,” vol. 16, no. 2, pp. 29–43.
 - [12] A. Damuri, U. Riyanto, H. Rusdianto, and M. Aminudin, “Implementasi Data Mining dengan Algoritma Naïve Bayes Untuk Klasifikasi Kelayakan Penerima Bantuan Sembako,” *J. Ris. Komput.*, vol. 8, no. 6, pp. 219–225, 2021, doi: 10.30865/jurikom.v8i6.3655.
 - [13] S. J. S. Tyas, M. Febianah, F. Solikhah, A. L. Kamil, and W. A. Arifin, “Analisis Perbandingan Algoritma Naive Bayes Dan C.45 Dalam Klasifikasi Data Mining Untuk Memprediksi Kelulusan,” *J. Teknol. Inf. Dan Komun.*, vol. 8, no. 1, pp. 86–99, 2021.
 - [14] Reza Gustrianda and D. I. Mulyana, “Penerapan Data Mining Dalam Pemilihan Produk Unggulan dengan Metode Algoritma K-Means Dan K-Medoids,” *J. Media Inform. Budidarma*, vol. 6, no. 1, pp. 27–34, 2022, doi: 10.30865/mib.v6i1.3294.
 - [15] R. A. Samosir, M. F. Rozy, and A. P. Windarto, “Penerapan Algoritma Regresi Linier Berganda dalam Mengestimasi Jumlah Perceraian di Pengadilan Agama Simalungun,” *TIN Terap. Inform. Nusantara*, vol. 2, no. 1, pp. 16–20, 2021, [Online]. Available: <https://ejurnal.seminar-id.com/index.php/tin>.
 - [16] I. Maryani, O. Revianti, H. M. Nur, and S. Sunanto, “Implementasi Data Mining Pada Penjualan Di Toko GOC Kosmetik Dengan Menggunakan Metode Algoritma Apriori,” *Indones. J. Softw. Eng.*, vol. 8, no. 1, pp. 92–98, 2022, doi: 10.31294/ijse.v8i1.13017.
 - [17] I. Romli, “Penerapan Data Mining Menggunakan Algoritma K-Means Untuk Klasifikasi Penyakit Ispa,” *Indones. J. Bus. Intell.*, vol. 4, no. 1, p. 10, 2021, doi: 10.21927/ijubi.v4i1.1727.
 - [18] E. Haryatmi and S. Pramita Hervianti, “Penerapan Algoritma Support Vector Machine Untuk Model Prediksi Kelulusan Mahasiswa Tepat Waktu,” *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 2, pp. 386–392, 2021, doi: 10.29207/resti.v5i2.3007.
 - [19] A. F. Boy, S. Yakub, I. Ishak, and Z. Azmi, “Implementasi Data Mining Pada Pengaturan Distribusi Barang Dengan Menggunakan Algoritma Fp-Growth,” *J. Sci. Soc. Res.*, vol. 5, no. 2, p. 431, 2022, doi: 10.54314/jssr.v5i2.947.
 - [20] A. E. Gumanti, Taslim, S. Handayani, and D. Toresa, “Penerapan Algoritma K-Nearest Neighbor Untuk Klasifikasi Topik Skripsi Mahasiswa di Fakultas Ilmu Komputer,” *JITaCS J. Inf. Technol. Comput. Sci.*, vol. 1, no. 2, pp. 44–50, 2022.
 - [21] S. Z. H. Rukmana, A. Aziz, and W. Harianto, “Optimasi Algoritma K-Nearest Neighbor (Knn) Dengan Normalisasi Dan Seleksi Fitur Untuk Klasifikasi Penyakit Liver,” *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 6, no. 2, pp. 439–445, 2022.
 - [22] H. W. Azizah, O. Nurdiawan, G. Dwilestari, and E. Tohidi, “Klasifikasi Pemberian Bantuan Umkm Cirebon Dengan Menggunakan Algoritma K-Nearest Neighbor,” vol. 3, no. 3, pp. 110–115, 2022, doi: 10.47065/josyc.v3i3.1392.
 - [23] S. Widaningsih and S. Yusuf, “Penerapan Data Mining untuk Memprediksi Siswa Berprestasi dengan Menggunakan Algoritma K Nearest Neighbor,” *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 9, no. 3, pp. 2598–2611, 2022, doi: 10.35957/jatisi.v9i3.859.
 - [24] S. Silvillestari, “Data Mining Menggunakan Algoritma K-Nearest Neighbor Dalam Menentukan Kredit Macet Barang Elektronik,” *J. Media Inform. Budidarma*, vol. 5, no. 3, p. 1063, 2021, doi: 10.30865/mib.v5i3.3100.
 - [25] A. D. Adhi Putra, “Analisis Sentimen pada Ulasan pengguna Aplikasi Bibit Dan Bareksa dengan Algoritma KNN,” *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 8, no. 2, pp. 636–646, 2021, doi: 10.35957/jatisi.v8i2.962.
 - [26] N. S. H. Pratama, D. T. Afandi, M. Mulyawan, I. Iin, and N. D. Nuris, “Menurunkan Presentase Kredit Macet Nasabah Dengan Menggunakan Algoritma K-Nearest Neighbor,” *Inf. Syst. Educ. Prof. J. Inf. Syst.*, vol. 5, no. 2, p. 131, 2021, doi: 10.51211/isbi.v5i2.1537.
 - [27] D. P. Utomo and Mesran, “Analisis Komparasi Metode Klasifikasi Data Mining dan Reduksi Atribut Pada Data Set Penyakit Jantung,” *J. Media Inform. Budidarma*, vol. 4, no. 2, pp. 437–444, 2020.
 - [28] D. P. Utomo, P. Sirait, and R. Yunis, “Reduksi Atribut Pada Dataset Penyakit Jantung dan Klasifikasi Menggunakan Algoritma C5. 0,” *J. Media Inform. Budidarma*, vol. 4, no. 4, pp. 994–1006, 2020, doi: 10.30865/mib.v4i4.2355.