

## Implementasi Data Mining Klasifikasi Kelulusan Mahasiswa di Perguruan Tinggi Menggunakan *K-Nearest Neighbors*

Legi Octa Sofyan Firmandala <sup>1</sup>, Zaehol Fatah<sup>2</sup>

<sup>1,2</sup>. Universitas Ibrahimy Situbondo, Indonesia

### 1. PENDAHULUAN

Di era digital saat ini, banyak institusi pendidikan tinggi menghadapi tantangan dalam menganalisis data akademik mahasiswa untuk meningkatkan proses pembelajaran dan memperbaiki tingkat kelulusan[1]. Perguruan Tinggi merupakan institusi pendidikan yang menyediakan layanan pembelajaran di tingkat lanjut, yang menjadi tahap akhir opsional dalam pendidikan formal. Lembaga pendidikan tinggi ini biasanya berbentuk Universitas, Akademi, Institut, atau Sekolah Tinggi[2].

Namun, dalam beberapa tahun terakhir, perguruan tinggi ini menghadapi beberapa tantangan signifikan terkait kelulusan mahasiswa[3]. Salah satu masalah krusial yang dihadapi adalah rendahnya tingkat kelulusan tepat waktu. Banyak mahasiswa yang gagal lulus sesuai jadwal yang ditentukan, yang berdampak pada efisiensi akademik, baik bagi mahasiswa itu sendiri maupun bagi perguruan tinggi secara keseluruhan. Selain itu, mekanisme evaluasi kelulusan yang dilakukan secara manual sering kali memerlukan waktu yang lama dan memiliki potensi bias dalam menilai kinerja mahasiswa[4].

Beberapa penelitian sebelumnya telah menunjukkan efektivitas penerapan algoritma *K-Nearest Neighbors* (KNN) dalam klasifikasi dan prediksi kelulusan mahasiswa. Sebagai contoh, penelitian yang dilakukan oleh Aldi Solihin[5] pada tahun 2024 yang mengkaji penerapan algoritma *K-Nearest Neighbors* (KNN) untuk analisis prestasi akademik mahasiswa. Dengan menggunakan dataset 250 mahasiswa, penelitian ini menunjukkan bahwa algoritma KNN menghasilkan tingkat akurasi prediksi kelulusan sebesar 91.5% melalui metode *Cross Validation 10-fold*. Hasil ini lebih tinggi dibandingkan dengan metode klasifikasi tradisional, yang menunjukkan efektivitas KNN dalam memberikan prediksi akurat terkait kelulusan mahasiswa. Penelitian menurut Supri Arrohman dan Zaehol Fatah[6] algoritma *K-Nearest*

*Neighbors* (K-NN) mampu memberikan akurasi yang cukup baik dalam melakukan klasifikasi prediksi diabetes pada perempuan suku Indian Pima. Dengan tingkat akurasi sebesar 70,13%, dapat dikatakan bahwa model ini memiliki potensi untuk digunakan dalam mengidentifikasi pasien yang berisiko menderita diabetes

Permasalahan menunjukkan bahwa Perguruan Tinggi perlu mengembangkan sistem yang lebih efisien dan akurat dalam memprediksi kelulusan mahasiswa. Data mining, yang merupakan teknik untuk menggali informasi dari kumpulan data besar, bisa menjadi solusi untuk mengatasi permasalahan ini. Algoritma *K-Nearest Neighbors* (KNN), salah satu teknik data mining yang populer, dapat digunakan untuk melakukan klasifikasi terhadap data mahasiswa[7], sehingga memudahkan dalam memprediksi kelulusan mereka berdasarkan pola historis.

Penerapan KNN dalam data mining memungkinkan Perguruan Tinggi untuk mengklasifikasikan mahasiswa berdasarkan faktor-faktor yang mempengaruhi kelulusan, serta faktor sosial-ekonomi. Dengan demikian, sistem ini dapat memberikan prediksi yang lebih akurat mengenai kelulusan, yang nantinya dapat digunakan sebagai acuan dalam menentukan langkah intervensi yang diperlukan untuk meningkatkan tingkat kelulusan tepat waktu di Perguruan Tinggi.

Oleh karena itu, penelitian ini bertujuan untuk mengimplementasikan algoritma KNN dalam data mining guna mengoptimalkan proses klasifikasi kelulusan mahasiswa di Perguruan Tinggi. Dengan penerapan teknologi ini, diharapkan dapat diperoleh sistem prediksi kelulusan yang lebih akurat dan efisien, serta mampu membantu pihak Perguruan Tinggi dalam mengambil kebijakan yang tepat terkait peningkatan kualitas pendidikan dan kelulusan mahasiswa.

## 2. METODE

### 2.1. Metode Penelitian

Penelitian ini bersifat deskriptif kuantitatif[8], di mana data yang diperoleh akan dianalisis secara statistik untuk menggambarkan kondisi kelulusan mahasiswa berdasarkan faktor-faktor tertentu. Penelitian deskriptif kuantitatif merupakan pendekatan yang bertujuan untuk mendeskripsikan, menganalisis, dan menjelaskan fenomena yang diteliti sebagaimana adanya. Penelitian kuantitatif adalah penyelidikan terstruktur mengenai suatu fenomena yang melibatkan pengumpulan data yang dapat diukur melalui teknik statistik[9], matematika, atau komputasi. Metode ini sering digunakan dalam bidang ilmu alam dan ilmu fisika.[10].

### 2.2. Pengumpulan Data

Teknik pengumpulan data adalah metode yang digunakan untuk memperoleh informasi atau fakta-fakta yang terdapat di lapangan[11]. Data yang diperlukan untuk analisis dapat dikumpulkan dari sistem informasi akademik Perguruan Tinggi, yang mencakup data mahasiswa seperti nama, jenis kelamin, pekerjaan, status pernikahan, Indeks Prestasi Semester dan Indeks Prestasi Kumulatif. Teknik pengumpulan data dapat dilakukan melalui survei, observasi, dan akses ke database akademik.

### 2.3. Pemilihan Atribut

Pemilihan atribut dilakukan dengan secara heuristik memilih kriteria pemisahan yang paling efektif untuk membagi partisi data yang ada[12]. Mengidentifikasi atribut yang berpengaruh terhadap kelulusan, seperti Indeks Prestasi Semester(IPS), pekerjaan dan status pernikahan, Indeks Prestasi Kumulatif(IPK). Atribut ini akan digunakan sebagai input untuk model KNN

**2.4. Implementasi Algoritma *K-Nearest Neighbors* (KNN)**

Algoritma *K-Nearest Neighbor* (KNN) merupakan metode klasifikasi yang menentukan kelas sebuah objek berdasarkan data pembelajaran yang memiliki jarak terdekat dengan objek tersebut[13]. Algoritma *K-Nearest Neighbor* (KNN) adalah metode klasifikasi yang menentukan kategori atau label suatu objek dengan melihat kategori mayoritas dari k tetangga terdekatnya dalam dataset[14]. Dalam proses ini, objek baru akan dikategorikan berdasarkan kedekatannya dengan data lain yang sudah memiliki label[15]. Algoritma ini akan menghitung jarak antara titik data mahasiswa dengan data kelulusan sebelumnya dan menentukan kelas (lulus atau tidak lulus) berdasarkan kedekatannya dengan titik data lainnya. Tahapan menggunakan algoritma KNN sebagai berikut:

1. Tentukan terlebih dahulu jumlah tetangga K (sebaiknya dalam angka ganjil).
2. Hitung jarak antara data untuk dibandingkan dengan dataset dengan menggunakan rumus jarak Euclidean.

$$dist(p, q) = \sqrt{\sum_{i=1}^n (p_i - q_i)^2}$$

**Gambar 1.** Rumus jarak Euclidean

$p_i$  : nilai indeks ke-i dari data p  
 $q_i$  : nilai indeks ke-i dari data q  
 $d(p,q)$  : jarak antara titik data p dan titik data q  
 di mana n adalah jumlah atribut dalam dataset.

3. Urutkan semua titik data pelatihan berdasarkan jarak yang telah dihitung dan pilih K tetangga terdekat yang memiliki jarak terkecil.
4. Tentukan kelas untuk data yang sedang dianalisis berdasarkan mayoritas kelas dari K tetangga terdekat.
5. Uji hasil klasifikasi dengan menggunakan data uji dan hitung metrik evaluasi seperti akurasi, presisi, *recall*, dan *F1-score* untuk menilai efektivitas model.
6. Analisis hasil klasifikasi untuk memahami faktor-faktor yang berkontribusi pada prediksi, serta mengevaluasi pola yang muncul dari data.

**3. HASIL DAN PEMBAHASAN**

**3.1. Data preparation**

Pada tahap ini adalah proses mengumpulkan, membersihkan, dan mengatur data sebelum digunakan untuk analisis atau model machine learning[16]. Tujuan dari proses ini adalah untuk memastikan bahwa data yang digunakan dalam penelitian memiliki kualitas yang optimal dan siap untuk dianalisis menggunakan metode *Naive Bayes* dan *K-Nearest Neighbors* (KNN)[17].

1	NAMA	JENIS KELAMIN	STATUS MAHASISWA	UMUR	STATUS NIKAH	IPS 1	IPS 2	IPS 3	IPS 4	IPS 5	IPS 6	IPS 7	IPS 8	IPK	STATUS KELULUSAN
2	ANIK WIDAYANTI	PEREMPUAN	BEKERJA	28	BELUM MENIKAH	2,76	2,8	3,2	3,17	2,98	3	3,03	0	3,07	TERLAMBAT
3	DWI HESTYNA PRIHASTI	PEREMPUAN	MAHASISWA	32	BELUM MENIKAH	3	3,3	3,14	3,14	2,84	3,13	3,25	0	3,17	TERLAMBAT
4	MURYA ARIEF BASUKI	PEREMPUAN	BEKERJA	29	BELUM MENIKAH	3,5	3,3	3,7	3,29	3,53	3,72	3,73	0	3,54	TERLAMBAT
5	NANIK SUSANTI	PEREMPUAN	MAHASISWA	27	BELUM MENIKAH	3,17	3,41	3,61	3,36	3,48	3,63	3,46	0	3,41	TERLAMBAT
6	RIFKA ISTIOFARINA	PEREMPUAN	BEKERJA	29	BELUM MENIKAH	2,9	2,89	3,3	2,85	2,98	3	3,08	0	3,09	TERLAMBAT
7	SUHARYONO	LAKI - LAKI	BEKERJA	27	BELUM MENIKAH	2,95	2,82	3,09	3,1	2,78	3,16	3,23	0	3,07	TERLAMBAT
8	FARIKHATUN NAZLI	PEREMPUAN	MAHASISWA	26	BELUM MENIKAH	2,76	3,14	2,6	2,95	3,23	3,33	3,3	3,3	3,06	TEPAT
9	FIFI SUNALISA	PEREMPUAN	MAHASISWA	27	BELUM MENIKAH	2,62	2,89	2,32	2,5	2,5	2,86	3,05	2,5	2,91	TEPAT
10	HENDRIK MULIYANTO	PEREMPUAN	BEKERJA	25	MENIKAH	3,6	3,54	3,52	3,39	3,52	3,68	3,15	0	3,4	TERLAMBAT
11	IMAM AGUNG RIBOWO	PEREMPUAN	BEKERJA	28	BELUM MENIKAH	2,71	2,55	1,77	2,11	1,93	2,13	1,78	0,2	2,2	TERLAMBAT
12	IMAM SANTOSA	PEREMPUAN	BEKERJA	27	BELUM MENIKAH	3,14	3,46	3,4	3,43	3,27	3,15	3,81	4	3,44	TERLAMBAT
13	IRFAN EKO WAHYUDI	PEREMPUAN	BEKERJA	32	BELUM MENIKAH	2,67	2,3	1,57	1,44	1,58	1,68	1,13	0,94	2,4	TERLAMBAT
14	IWAN HAMBALI	PEREMPUAN	BEKERJA	26	BELUM MENIKAH	2,57	2,82	2,2	2,45	2,1	2,42	1	1,42	2,45	TERLAMBAT
15	M SYAIFULLAH	PEREMPUAN	BEKERJA	31	BELUM MENIKAH	2,71	3	2,65	2,27	2,13	3,34	2,5	0	2,57	TERLAMBAT
16	DIANA LAILY FITHRI	PEREMPUAN	MAHASISWA	26	BELUM MENIKAH	3,24	3,38	3,44	3,3	3,56	3,45	3,35	3,35	3,45	TERLAMBAT
17	DONNY PRABOWO	LAKI - LAKI	MAHASISWA	27	BELUM MENIKAH	2,86	2,86	2,45	1,86	3,19	3,14	2,98	2	3,04	TERLAMBAT
18	EDI JATMIKO	PEREMPUAN	MAHASISWA	27	BELUM MENIKAH	2,71	3,27	2,54	3,36	3,23	3,28	3,15	2	3,19	TERLAMBAT
19	ADI SUBEKTI	LAKI - LAKI	BEKERJA	29	BELUM MENIKAH	2,67	2,2	1,45	3	3,17	3,08	3,33	0	2,95	TERLAMBAT
20	AHMAD IBROZI	LAKI - LAKI	MAHASISWA	27	BELUM MENIKAH	2,67	2,68	1,95	1,61	2,36	2,21	0,78	0,92	2,11	TERLAMBAT
21	ANDI HARDIANTO	LAKI - LAKI	MAHASISWA	27	BELUM MENIKAH	3,1	3,71	2,96	3,4	3,4	3,61	3,73		3,38	TEPAT
22	ANDI SUNARWAN	LAKI - LAKI	BEKERJA	28	BELUM MENIKAH	3	3,33	2,88	2,8	3,36	3,07	3,43	4	3,29	TERLAMBAT
23	ANNY AZIZAH	LAKI - LAKI	BEKERJA	26	BELUM MENIKAH	3,12	3,23	2,96	3	3,31	3	3,08	3	3,19	TERLAMBAT
24	ARIEF PRAM WAHYUDI	LAKI - LAKI	MAHASISWA	26	BELUM MENIKAH	2,9	3,32	2,83	2,91	3,3	2,83	3,27	4	3,11	TERLAMBAT

**Gambar 2.** Tabel Data Preparation

### 3.2. Perhitungan manual *K-Nearest Neighbor*

Perhitungan Manual *K-Nearest Neighbor* (KNN) adalah suatu metode yang digunakan untuk menentukan klasifikasi dari data baru dengan menganalisis kedekatannya terhadap data yang sudah ada dalam dataset[18]. Proses ini melibatkan beberapa langkah sistematis untuk mencapai hasil yang akurat dan informatif. Langkah-langkah dalam perhitungan untuk memprediksi kelulusan:

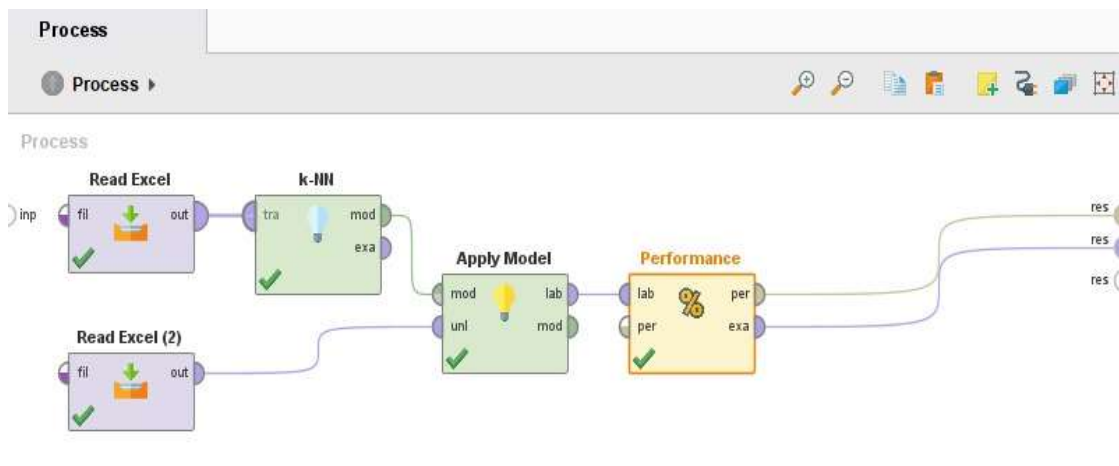
1. Nilai K yang digunakan dalam penelitian ini adalah K=3
2. Kumpulkan data penting tentang siswa, seperti nilai akademik dan kehadiran, serta faktor lain yang memengaruhi kelulusan.
3. Gunakan rumus sebagai berikut:

**Tabel 1.** Perhitungan KNN

No		D	Klasifikasi
1	$\sqrt{((2,76 - 3,94)^2 + (2,8 - 3,8)^2 + (3,2 - 3,8)^2 + (2,98 - 3,76)^2 + (3,03 - 3,8)^2 + (0 - 3,58)^2 + (2,62 - 3,75)^2)}$	4,33	Terlambat
2	$\sqrt{((3 - 3,94)^2 + (3,3 - 3,8)^2 + (3,14 - 3,8)^2 + (2,84 - 3,59)^2 + (3,13 - 3,25 - 3,8)^2 + (0 - 3,58)^2 + (2,73 - 3,75)^2)}$	8,21	Terlambat
3	$\sqrt{((2.76 - 3.94)^2 + (3.14 - 3.8)^2 + (2.6 - 3.8)^2 + (2.95 - 3.8)^2 + (3.23 - 3.59)^2 + (3.83 - 3.76)^2 + (3.3 - 3.8)^2 + (3.3 - 3.58)^2 + (3.08 - 3.75)^2)}$	9,26	Tepat
....	.....	.....	.....

### 3.3. Permodelan *K-Nearest Neighbor* dengan *RapidMiner*

*RapidMiner* adalah perangkat lunak dengan antarmuka grafis (GUI) yang dibuat khusus untuk analisis data, pembelajaran mesin (*machine learning*), serta penambangan data (*data mining*)[19]. Dikembangkan oleh *RapidMiner GmbH*[20], platform ini menyediakan solusi menyeluruh untuk berbagai tahap dalam proses *extract, transform, load* (ETL), analisis prediktif, dan pembuatan model AI. Alat ini digunakan oleh berbagai kalangan profesional mulai dari akademisi hingga pebisnis dan praktisi industri untuk mempermudah pengelolaan data dan pengembangan model analitik tanpa memerlukan banyak kode pemrograman.



**Gambar 3.** Permodelan dengan *RapidMiner*

Permodelan ini adalah proses penerapan algoritma *K-Nearest Neighbor* (KNN) menggunakan platform AI Studio di *RapidMiner* untuk tujuan klasifikasi data.

accuracy: 88.39%

	true TERLAMBAT	true TEPAT	class precision
pred. TERLAMBAT	140	21	86.96%
pred. TEPAT	23	195	89.45%
class recall	85.89%	90.28%	

**Gambar 4.** Pemodelan klasifikasi

Dalam pemodelan klasifikasi ini, hasil akurasi model *K-Nearest Neighbor* (KNN) menunjukkan performa yang signifikan dalam memprediksi status tepat waktu dan status terlambat. Berdasarkan hasil evaluasi, beberapa metrik utama yang dianalisis meliputi prediksi terlambat, prediksi tepat, *true* terlambat, *true* tepat, *class recall*, dan *class precision* dengan hasil akurasi 88,39%.

#### 4. KESIMPULAN DAN SARAN

Penerapan algoritma *K-Nearest Neighbor* (KNN) efektif dalam memprediksi kelulusan mahasiswa, baik tepat waktu maupun terlambat, dengan tingkat akurasi yang tinggi. Algoritma KNN mampu mengidentifikasi pola akademik yang relevan dan membantu perguruan tinggi dalam mengambil langkah intervensi yang lebih tepat guna meningkatkan tingkat kelulusan. Penerapan KNN pada data mahasiswa menghasilkan akurasi 88,39%.

Namun, penelitian ini memiliki keterbatasan dalam hal jumlah data dan atribut yang digunakan. Oleh karena itu, disarankan agar penelitian selanjutnya menggunakan dataset yang lebih luas dan atribut tambahan, serta mengeksplorasi algoritma lain seperti *Decision Tree* atau *Random Forest* untuk mendapatkan perbandingan terhadap metode yang dipilih.

#### 5. DAFTAR PUSTAKA

- [1] J. Santoso, "Mengatasi Tantangan Keterlibatan Mahasiswa: Strategi Efektif untuk Menciptakan Lingkungan Belajar yang Menarik," *J. Ilm. Kanderang Tingang*, vol. 14, no. 2, pp. 469–478, 2023.
- [2] A. F. Rodli, S. E. Nikma Yucha, and M. SM, *Manajemen Kinerja Institusi Perguruan Tinggi*. CV Rey Media Grafika, 2022.
- [3] M. Rohman, W. Kurniawan, M. L. Nawawi, and H. H. Yana, "Pelatihan Penyusunan Artikel Ilmiah Bagi Mahasiswa Perguruan Tinggi Keagamaan Islam," *J. Inov. Penelit. dan Pengabd. Masy.*, vol. 4, no. 2, pp. 185–198, 2024.
- [4] M. Syawal, M. Damapolii, and M. Hasan, "PERAN SISTEM INFORMASI MANAJEMEN KEPEGAWAIAN (SIMPEG) DALAM PENGELOLAAN AKREDITASI DI MAN 1 BONE," *Cendikia J. Pendidik. dan Pengajaran*, vol. 2, no. 7, pp. 181–191, 2024.
- [5] A. S. Fauzan, A. I. P. Sari, and I. Ali, "ANALISIS PERBANDINGAN ALGORITMA DECISION TREE DAN NAÏVE UNTUK MENGEVALUASI PRESTASI BELAJAR SISWA," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 8, no. 1, pp. 741–747, 2024.
- [6] S. Arrohman and Z. Fatah, "Prediksi Diabetes Menggunakan Algoritma Klasifikasi K-Nearest Neighbors (K-NN) pada Perempuan Indian Pima," *Gudang J. Multidisiplin Ilmu*, vol. 2, no. 10, pp. 220–226, 2024.
- [7] 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.

- [8] M. Waruwu, “Pendekatan penelitian pendidikan: metode penelitian kualitatif, metode penelitian kuantitatif dan metode penelitian kombinasi (Mixed Method),” *J. Pendidik. Tambusai*, vol. 7, no. 1, pp. 2896–2910, 2023.
- [9] M. S. Jailani, “Teknik pengumpulan data dan instrumen penelitian ilmiah pendidikan pada pendekatan kualitatif dan kuantitatif,” *IHSAN J. Pendidik. Islam*, vol. 1, no. 2, pp. 1–9, 2023.
- [10] M. Ramdhan, *Metode penelitian*. Cipta Media Nusantara, 2021.
- [11] M. Hasan *et al.*, “Metode penelitian kualitatif,” *Penerbit Tahta Media*, 2023.
- [12] Z. Setiawan *et al.*, *Buku Ajar Data Mining*. PT. Sonpedia Publishing Indonesia, 2023.
- [13] J. Indriyanto, *Algoritma K-Nearest Neighbor untuk prediksi nasabah asuransi*. Penerbit NEM, 2021.
- [14] Shudiq, W. J. F., As, A. H., & Rahman, M. F. (2020). Penentuan Metode Terbaik Dalam Menentukan Jenis Pohon Pisang Menurut Tekstur Daun (Metode K-NN dan SVM). *Jurnal Teknologi dan Manajemen Informatika*, 6(2), 128-136.
- [15] F. M. N. Akbar, “Metode KNN (K-Nearest Neighbor) untuk Menentukan Kualitas Air,” *J. Tekno Kompak*, vol. 18, no. 1, pp. 28–40, 2024.
- [16] R. Aprilian, R. Habibi, and M. Y. H. Setyawan, *Algoritma KNN dalam memprediksi cuaca untuk menentukan tanaman yang cocok sesuai musim*. Kreatif, 2020.
- [17] Permatasari, U. O. R., Shudiq, W. J. F., & Jasri, M. (2024). Prediksi Kelayakan Mahasiswa sebagai Penerima Beasiswa Bank Indonesia pada Tahap Seleksi Administrasi di Universitas Nurul Jadid Menggunakan Algoritma K Nearest Neighbor. *Journal of Electrical Engineering and Computer (JEECOM)*, 6(1), 252-260.
- [18] D. Pratmanto and F. F. D. Imaniawan, “Analisis Sentimen Terhadap Aplikasi Canva Menggunakan Algoritma Naive Bayes Dan K-Nearest Neighbors,” *Comput. Sci.*, vol. 3, no. 2, pp. 110–117, 2023.
- [19] N. Ahmad, S. Hafizh, and R. Sulthanah, “Prediksi Kelulusan Mata Kuliah Mahasiswa Teknologi Informasi Menggunakan Algoritma K-Nearest Neighbor,” *J. Manaj. Inform.*, vol. 14, no. 2, pp. 135–149, 2024.
- [20] F. Fitrianiingsih and B. Zuraeni, “Analisis Ramalan Cuaca di Sekupang, Kota Batam Menggunakan Algoritma Decision Tree dan Confusion Matrix,” *Ekosph. J. Ekon. Pembang. dan Manaj.*, vol. 1, no. 3, pp. 15–26, 2024.
- [21] Y. Resti, E. S. Kresnawati, N. Eliyati, and I. Yani, “STATISTIKA DAN TEORI PROBABILITAS UNTUK PENGUATAN LITERASI SAINS DAN NUMERASI MELALUI SOFTWARE RAPIDMINER PADA MATA KULIAH STATISTIKA PEMBELAJARAN MESIN BAGI SISWA SEKOLAH MENENGAH DAN MAHASISWA MATEMATIKA,” *Appl. Innov. Eng. Sci. Res.*, vol. 15, no. 1, pp. 400–403, 2023.