

# PENENTUAN STRATEGI PROMOSI PROGRAM STUDI PENDIDIKAN TEKNOLOGI INFORMASI MENGGUNAKAN ALGORITMA K-MEANS CLUSTERING

Satrio Agung Wicaksono<sup>\*1)</sup>, Satrio Hadi Wijoyo<sup>2)</sup>, Himawat Aryadita<sup>3)</sup>

1. Pendidikan Teknologi Informasi, Fakultas Ilmu Komputer, Universitas Brawijaya, Indonesia
2. Pendidikan Teknologi Informasi, Fakultas Ilmu Komputer, Universitas Brawijaya, Indonesia
3. Sistem Informasi, Fakultas Ilmu Komputer, Universitas Brawijaya, Indonesia

## Article Info

**Kata Kunci:** Penambangan Data; Strategi Promosi; Klasterisasi K-Means

**Keywords:** Data Mining; Promotion Strategy; K-Means Clustering

## Article history:

Received 27 November 2025

Revised 22 December 2025

Accepted 31 December 2025

Available online 31 December 2025

## DOI :

<https://doi.org/10.29100/jipi.v10i4.9576>

\* Corresponding author.

Corresponding Author

E-mail address:

[satriohadi@ub.ac.id](mailto:satriohadi@ub.ac.id)

## ABSTRAK

Program Studi Pendidikan Teknologi Informasi (PS PTI) merupakan salah satu Program Studi pada Fakultas Ilmu Komputer Universitas Brawijaya di Kota Malang, dimana tidak terlepas dari kegiatan promosi pada setiap tahunnya. Kegiatan promosi yang dilakukan oleh Program Studi menggunakan biaya yang cukup besar sehingga promosi yang dilakukan diharapkan juga mendapatkan hasil maksimal. Kegiatan promosi dilakukan lewat himpunan mahasiswa yang meminta mahasiswa untuk promosi ke sekolah masing-masing. Kegiatan promosi juga melalui kunjungan sekolah ke Fakultas Ilmu Komputer. Suksesnya promosi yang dilakukan dapat dilihat dari banyaknya jumlah mahasiswa/i yang mendaftar pada tahun ajaran baru. Beberapa tahun penerimaan mahasiswa baru dari PS PTI tidak memenuhi daya tampung yang sudah ditentukan. Promosi yang sukses dapat dilihat dari strategi promosi yang digunakan, penentuan strategi promosi bukanlah pekerjaan yang mudah dikarenakan harus mempertimbangkan beberapa aspek. Data mining digunakan penentuan strategi promosi pada PS PTI, dimana dengan data mining dapat menemukan sebuah pola baru dari kumpulan data yang tersimpan. Salah teknik yang digunakan pada data mining adalah teknik klusterisasi dengan algoritma yang digunakan adalah K-Means. Nilai SSE tertinggi adalah 600 pada  $k=2$  sedangkan nilai SSE terendah sebesar 406 pada  $k=10$ .

## ABSTRACT

The Information Technology Education Study Program (PS PTI) is one of the Study Programs at the Faculty of Computer Science, Brawijaya University in Malang City, which is inseparable from promotional activities every year. Promotional activities carried out by the Study Program use quite large costs so that the promotion carried out is expected to also get maximum results. Promotional activities are carried out through student associations that ask students to promote to their respective schools. Promotional activities are also through school visits to the Faculty of Computer Science. The successful promotion carried out can be seen from the large number of students who register in the new academic year. Several years of new student admissions from PS PTI did not meet the specified capacity. Successful promotion can be seen from the promotional strategy used, determining a promotional strategy is not an easy job because it must consider several aspects. Data mining is used to determine promotional strategies at PS PTI, where data mining can find a new pattern from a collection of stored data. One of the techniques used in data mining is the clustering technique with the algorithm used is K-Means. The highest SSE value is 600 at  $k = 2$  while the lowest SSE value is 406 at  $k = 10$ .

## I. PENDAHULUAN

**D**ata Mining sudah banyak diterapkan pada berbagai bidang, misalnya bidang analisis pemasaran, kedokteran, pendidikan, dan lain-lain. *Data mining* dapat diterapkan pada bidang lembaga atau institusi Pendidikan dan sering disebut juga dengan *Educational Data Mining* (EDM) yaitu

sebuah pengembangan metode dalam mengeksplorasi jenis tipe data Pendidikan yang bersifat unik yang bertujuan untuk mempelajari dalam memahami kinerja siswa dan pengaturan lingkungan di tempat siswa belajar [1]. EDM digunakan untuk mengklasifikasikan, menganalisis, dan memprediksi kinerja akademik siswa, dan tingkat putus sekolah siswa, serta kinerja pengajar untuk meningkatkan proses pengajaran-pembelajaran. Teknik *data mining* diterapkan pada data pendidikan ini untuk mengekstrak pola yang bermakna dan bermanfaat yang selanjutnya dapat dipertimbangkan untuk meningkatkan proses pengajaran-pembelajaran. Berbagai teknik penambangan data seperti klasifikasi, pengelompokan, regresi, jaringan saraf, algoritma genetik, algoritma aturan asosiasi, dll. digunakan untuk menemukan pengetahuan dari dataset pendidikan ini. Dalam hal ini, berbagai model telah diusulkan untuk menganalisis dan memprediksi kinerja siswa atau mahasiswa dalam pembelajaran akademik maupun daring [2]. Prediksi kegagalan sekolah/ perguruan tinggi/ putus sekolah merupakan permasalahan utama yang dihadapi lembaga dalam bidang Pendidikan [3],[4],[5],[6],[7]. Prediksi karir alumni perguruan tinggi merupakan permasalahan yang dihadapi yang berdampak pada akreditasi program studi tau perguruan tinggi [8]. Pencarian pola jawaban mahasiswa dalam mengerjakan soal atau ujian [9]. Data teks di bidang Pendidikan merupakan permasalahan sering dilakukan Analisa untuk menggali informasi yang tersembunyi [10],[11].

Program Studi Pendidikan Teknologi Informasi (PS PTI) merupakan salah satu Program Studi pada Fakultas Ilmu Komputer Universitas Brawijaya. Fakultas Ilmu Komputer (FILKOM) dibentuk sejak tanggal 27 Oktober 2011 yang bernama Program Teknologi Informasi dan Ilmu Komputer (PTI IK). PTI IK merupakan gabungan dari dua Program Studi Teknik Perangkat Lunak dari Fakultas Teknik dan Program Studi Ilmu Komputer dari Fakultas MIPA. Kemudian pada tanggal 10 Desember 2014, PTI IK resmi menjadi Fakultas Ilmu Komputer (FILKOM). Saat ini, FILKOM telah memiliki dua jurusan (saat ini berganti nama menjadi Departemen) dan masing-masing jurusan (Departemen) menawarkan tiga program studi (prodi) sebagai berikut: Jurusan Teknik Informatika (JTIF) saat ini berganti nama menjadi Departemen Teknik Informatika (DTIF) terdiri atas tiga prodi: Program Studi (S-2) Magister Ilmu Komputer, Program Studi (S-1) Teknik Informatika, dan Program Studi (S-1) Teknik Komputer. Jurusan Sistem Informasi (JSI) saat ini berganti nama menjadi Departemen Sistem Informasi (DSI) terdiri atas tiga prodi: Program Studi (S-1) Sistem Informasi, Program Studi (S-1) Teknologi Informasi, dan Program Studi (S-1) Pendidikan Teknologi Informasi.

Program Studi Pendidikan Teknologi Informasi, Fakultas Ilmu Komputer, Universitas Brawijaya, didirikan pada tahun 2014, dengan Surat Keputusan Menteri Pendidikan dan Kebudayaan Republik Indonesia nomor 595/E/O/2014 tentang izin penyelenggaraan program-program studi pada Universitas Brawijaya di Malang. Program Studi Pendidikan Teknologi Informasi Universitas Brawijaya mulai menerima mahasiswa pada tahun 2015 dan berakreditasi A pada tahun 2020 berdasarkan Keputusan BAN-PT No. 4793/SK/BAN-PT/Akred/S/VIII/2020. Program studi ini merupakan Program Sarjana Pendidikan yang lulusannya akan dianugerahi gelar Sarjana Pendidikan (S.Pd.) dan pengelolaannya sudah disesuaikan dengan Standar Pendidikan Guru pada Peraturan Menteri Riset, Teknologi, dan Pendidikan Tinggi nomor 55 tahun 2017. Keilmuan yang dikembangkan di dalam program studi ini adalah keilmuan terkait didaktik-metodik, integrasi, pemodelan data, dan rekayasa teknologi informasi. Lulusan program studi Pendidikan Teknologi Informasi memiliki kompetensi sebagai (1) pendidik bidang teknologi informasi; (2) pengembang dan pengelola layanan teknologi informasi; (3) pengembang konten dan media pembelajaran interaktif; (4) saintis data bidang pendidikan; dan (5) pelaku wirausaha bidang teknologi informasi (techno-edupreneur). Melalui pengembangan keilmuan tersebut, diharapkan program studi ini mampu berkontribusi untuk menghasilkan lulusan yang mampu berperan sebagai pendidik bidang teknologi informasi, pengembang serta pengelola sistem dan teknologi informasi, pengembang konten dan media pembelajaran interaktif, saintis bidang pendidikan, ataupun pelaku wirausaha bidang teknologi informasi (techno-edupreneur).

Jumlah mahasiswa PS PTI yang aktif saat ini sebanyak 250 lebih dan jumlah alumni sebanyak 148 lebih. PS PTI tidak terlepas dari kegiatan promosi pada setiap tahunnya, dimana kegiatan promosi dilakukan guna menarik minat bagi mahasiswa/i baru. Didalam kegiatan promosinya PS PTI harus bersaing dengan Program Studi pada PTN atau PTS lainnya guna menyaring mahasiswa/i baru untuk mendaftar pada PS PTI FILKOM UB. Kegiatan promosi yang dilakukan membutuhkan dana yang cukup besar menggunakan media promosi atau bahkan turun langsung ke masyarakat baik didalam kota, luar kota bahkan luar provinsi. Dengan kegiatan promosi yang membutuhkan dana yang cukup besar tersebut diharapkan sesuai dengan hasil yang didapatkan dimana dapat dilihat dari banyaknya mahasiswa/i yang mendaftar pada tahun ajaran baru. Namun didalam menentukan strategi promosi khususnya menentukan lokasi dan tujuan promosi yang tepat bukanlah hal yang mudah, ada banyak faktor yang harus diperhatikan sehingga membutuhkan waktu untuk menentukan strategi promosi tersebut. Jumlah mahasiswa/i yang mendaftar berasal dari berbagai daerah dan berasal dari latar belakang lulusan pendidikan yang berbeda baik yang Negeri ataupun Swasta. Beberapa tahun sejak angkatan pertama pada tahun 2015, PS PTI sering tidak memenuhi daya tampung mahasiswa baru. Hal tersebutlah yang menjadi tolak ukur ataupun pertimbangan untuk menentukan

strategi promosi pada asal daerah ataupun sekolah tertentu agar hasil yang didapatkan maksimal bertambah dari pada tahun – tahun sebelumnya.

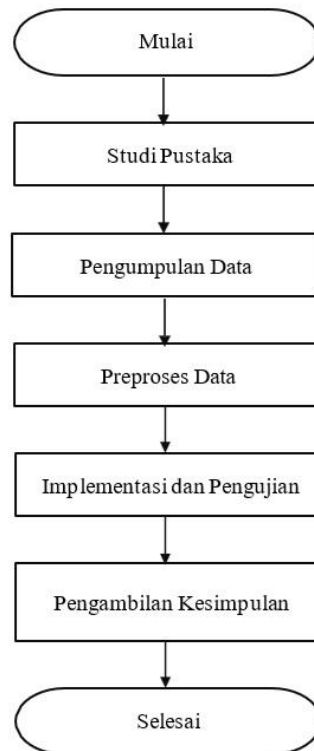
Penentuan strategi promosi yang dapat menggunakan data mining. Data mining merupakan sebuah cara yang tepat, dimana data mining mengolah data – data yang telah tersedia sebelumnya khususnya data – data mahasiswa yang ada pada PS PTI FILKOM UB. Data Mining adalah salah satu cara untuk mencari pengetahuan dan pola dari data dalam jumlah besar. Data mining adalah suatu proses ekstraksi atau penggalian data yang belum diketahui sebelumnya, namun dapat dipahami dan berguna dari database yang besar serta digunakan untuk membuat suatu keputusan bisnis yang sangat penting. Teknik – teknik data mining telah digunakan untuk menemukan pola yang tersembunyi dan memprediksi tren masa depan [12]. Pada data mining terdapat banyak teknik – teknik didalamnya salah satunya adalah Teknik Clustering dan Algoritma K-Means merupakan bagian dari pada Teknik Clustering. Algoritma Clustering K-Means merupakan salah satu algoritma dalam data mining yang berbasis jarak terpendek yang membagi data kedalam beberapa cluster berdasarkan kriteria yang berbasis numerik [13],[14],[15],[16],[17],[18].

Pada penelitian terdahulu yang dilakukan oleh Tria Titiani Chasanah dan Widiyono mengatakan bahwa Clustering K-Means membagi nilai variabel yang berbeda – beda. Secara keseluruhan cluster diperoleh berdasarkan dari kota asal dan juga asal jurusan. Penentuan strategi promosi yang dapat ditentukan dengan pendekatan nilai hasil clustering K-Means, pendekatan yang didapat memungkinkan memiliki peluang yang besar untuk meningkatkan jumlah mahasiswa/i baru [19]. Penelitian terdahulu menggunakan algoritma K-Means untuk membantu pihak pengambil keputusan pada Universitas Budi Darma guna menentukan strategi promosi kampus. Agar hasil promosi yang didapatkan juga maksimal sesuai dengan biaya yang digunakan. Variabel-variabel yang digunakan asal daerah, asal sekolah, dan pekerjaan orang tua. Hasil clustering didapatkan bahwa untuk daerah asal yang menjadi target promosi adalah Kota Medan dan latar belakang pendidikan sekolah SMA Swasta dengan penjurusan IPA dan IPS [20]. Penelitian terdahulu memiliki tujuan untuk menganalisis data calon mahasiswa baru dengan menggunakan algoritma K-Means pada Universitas Katolik Widya Mandala (UKWMS) Kampus Kota Madiun. Atribut yang digunakan dalam penelitian tersebut adalah data mahasiswa baru tahun akademik 2019/2020 dan sumber data berasal dari bagian Penerimaan Mahasiswa Baru (PMB). Beberapa atribut yang digunakan yaitu Asal Kota, Program Studi, Asal Sekolah, Nilai Matematika, dan Nilai Bahasa Inggris. Adanya pertimbangan dalam menentukan tempat promosi berdasarkan kota asal calon mahasiswa dan program studi apa yang menjadi pilihan calon mahasiswa selain itu jenis sekolah yang dijadikan target promosi, sehingga pilihan asal kota, program studi dan jenis sekolah tidak salah sasaran dalam pengambilan strategi promosi yang dilakukan tim promosi UKWMS Kampus Kota Madiun [21]. Penelitian terdahulu bertujuan untuk menentukan strategi promosi berdasarkan daerah asal dari profil mahasiswa baru yang mendaftar. Metode yang digunakan adalah K-Means Clustering. Proses implementasi dalam penelitian tersebut menggunakan software WEKA versi 3.8.5. Hasil penelitiannya terdapat tiga cluster yaitu cluster 1 berjumlah 189 data dengan presentase 42%, cluster 2 berjumlah 186 data dengan presentase 41% dan cluster 3 berjumlah 78 data dengan presentase 17%. [22].

Berdasarkan uraian permasalahan tersebut, tujuan penelitian ini mengusulkan penentuan strategi promosi Program Studi Pendidikan Teknologi Informasi menggunakan algoritma K-Means Clustering. Penelitian sebelumnya lebih pada karakteristik umum universitas bukan program studi kependidikan teknologi informasi secara spesifik. Belum ada penelitian yang secara khusus mengkaji strategi promosi berbasis data mining pada Program Studi Pendidikan Teknologi Informasi, khususnya di lingkungan Fakultas Ilmu Komputer universitas negeri besar seperti Universitas Brawijaya. Penelitian ini inline untuk menopang arah kebijakan pembangunan iptek Nasional sesuai dengan RPJPN 2005-2025 yang menyatakan secara eksplisit bahwa pembangunan iptek diarahkan salah satunya untuk menciptakan dan memanfaatkan teknologi informasi dan komunikasi, khususnya dalam bidang pendidikan. Dengan penentuan strategi promosi Program Studi Pendidikan Teknologi Informasi menggunakan algoritma K-Means Clustering, sehingga penelitian dapat membantu pihak pengambil keputusan.

## II. METODE PENELITIAN

Pada bab ini dibahas metode penelitian yang digunakan dan langkah-langkah yang dilakukan dalam penelitian mengenai penentuan strategi promosi Program Studi Pendidikan Teknologi Informasi menggunakan algoritma K-Means Clustering. Diagram alir metode penelitian sesuai dengan tahapan-tahapan penelitian yang dilakukan dapat dilihat pada Gambar 1. berikut :



Gambar 1. Alur Metode Penelitian

#### A. Studi Pustaka

Studi Literatur merupakan langkah awal dalam proses pengumpulan informasi sebagai acuan, studi literatur merupakan metode pengumpulan informasi melalui dokumen-dokumen baik berupa buku, jurnal maupun artikel dari internet. Dalam penelitian ini literatur yang dipelajari meliputi:

##### 1. Strategi Promosi

Strategi adalah konsep yang mengacu pada suatu jaringan yang kompleks dari pemikiran, ide-ide, pengertian yang mendalam, pengalaman, sasaran, keahlian, memori, persepsi, dan harapan yang membimbing untuk menyusun suatu kerangka pemikiran umum agar kita dapat memutuskan tindakan-tindakan yang spesifik bagi tercapainya tujuan. Keberadaan strategi tidak terlepas dari tujuan yang dicapai.

Hal ini ditunjukkan oleh suatu jaringan kerja yang membimbing tindakan yang akan dilakukan dan pada saat yang sama, strategi akan mempengaruhi tindakan tersebut. Ini berarti bahwa prasyarat yang diperlukan untuk merumuskan strategi adalah meningkatkan pemahaman tentang tujuan. Artinya setelah kita bersama-sama memahami hakikat dan makna suatu tujuan, maka kita menentukan strategi untuk mencapai tujuan. Tanpa tujuan, maka tindakan yang dibuat semata-mata sekadar suatu taktik yang dapat meningkatkan cepat namun sebaliknya dapat merosot ke dalam satu masalah lain [13].

Strategi merupakan simpulan taktik dalam keperluan bagaimana tujuan yang di inginkan dapat diperoleh atau didapat. Oleh sebab itu, strategi biasanya terdiri atas dua atau lebih taktik dengan anggapan yang satu lebih bagus dari yang lain. Dengan demikian strategi merupakan kumpulan taktik dengan maksud mencapai tujuan dan sasaran dari perusahaan, institusi atau badan. Bila strateginya sudah benar, maka pertempuran sudah separuh dimenangkan. Sebaliknya, bila pelaksanaan kurang baik, pertempurannya lebih dari separuh dinyatakan kalah, seperti menurut Sun Tzu dalam *The Art War*. Guna memudahkan pengertian strategi dalam pemasaran, maka parapelaku pasar harus mengartikan strategi minimal sebagai sebuah gambaran besar, pandangan jarak jauh (jangka panjang), cara mencapai tujuan, ringkasan taktik, pedoman taktik, platform dan lain-lain [13].

Promosi merupakan salah satu variabel dalam bauran pemasaran yang sangat penting dilaksanakan oleh perusahaan dalam memasarkan produk jasa. Kegiatan promosi bukan saja berfungsi sebagai alat komunikasi antara perusahaan dengan konsumen, melainkan juga sebagai alat untuk mempengaruhi konsumen dalam kegiatan pembelian atau penggunaan jasa sesuai dengan keinginan dan kebutuhannya [20]. Promosi adalah “kegiatan yang ditujukan untuk mempengaruhi konsumen agar mereka dapat menjadi kenal akan produk yang ditawarkan oleh perusahaan kepada mereka dan kemudian mereka menjadi senang lalu membeli produk tersebut”.

Promosi digital (optimasi website, media sosial, dan iklan digital), promosi tatap muka (roadshow ke

sekolah, seminar, dan kerja sama dengan bimbingan belajar), serta strategi peningkatan nilai tambah (meraih akreditasi, menawarkan beasiswa, dan menjalin kerja sama internasional). Kunci utamanya adalah memahami target calon mahasiswa, baik melalui analisis data demografi maupun perilaku, lalu menyusun pesan yang menarik dan konsisten.

## 2. K-Mean Clustering

K-Means merupakan metode clustering non hierarki yang mempartisi data dalam satu atau lebih cluster. Data yang memiliki karakteristik sama akan dikelompokkan ke dalam satu cluster dan data yang mempunyai karakteristik berbeda akan dikelompokkan ke dalam cluster lain [13],[23],[25]. K-Means menggunakan suatu persamaan untuk mengukur jarak paling terdekat pada tiap data yaitu Distance Space atau Euclidean Distance Space. Distance Space adalah suatu komponen dari K-Means Clustering yang digunakan untuk menghitung jarak antar data [24].

Secara umum Langkah - langkah melakukan clustering dengan menggunakan algoritma K-Means Clustering adalah sebagai berikut:

1. Tentukan jumlah cluster K.
  - a. Memberikan nilai secara random.
  - b. Mengambil sampel awal data.
  - c. Menentukan nilai awal hasil cluster hirarki dengan jumlah cluster yang sesuai dengan penentuan awal.
2. Inisialisasi k pusat klaster.
3. Tempatkan data ke cluster terdekat.
  - a. Berdasarkan kedekatannya dengan pusat cluster.
  - b. Dihitung jarak tiap data ke tiap pusat cluster.
  - c. Jarak paling dekat tersebut akan menjadi penentu masuk ke cluster mana.
4. Hitung kembali pusat cluster.
  - a. Hitung nilai rata - rata data dalam cluster tertentu
  - b. Pada kasus tertentu menggunakan median, karena median tidak sensitif terhadap outlier data.
5. Ulangi langkah ketiga dengan pusat cluster baru. Jika nilai pusat cluster tidak mengalami perubahan, maka proses clustering dihentikan.

## 3. Evaluasi Algoritma

Fungsi kinerja yang sering digunakan untuk Algoritma K-Means Clustering adalah Mean Square Error (MSE). Fungsi ini akan mengambil rata-rata kuadrat error yang terjadi antara output jaringan dan target. Fungsi lain yang dapat digunakan adalah Sum Square Error (SSE) seperti Persamaan (1) berikut ini.

$$SSE = \sum_{i=1}^K e(i)^2 = \sum_{i=1}^K (t(i) - y(i))^2 \quad (1)$$

### B. Metode Pengumpulan Data

Pengumpulan data dilakukan dengan cara melakukan wawancara kepada pihak Program Studi Pendidikan Teknologi Informasi. Pada penelitian ini juga menggunakan data sekunder. Data yang dibutuhkan pada penelitian ini diperoleh dari database program studi Pendidikan Teknologi Informasi Fakultas Ilmu Komputer Universitas Brawijaya melalui bagian akademik dan PSIK Fakultas Ilmu Komputer.

### C. Preproses Data

Pada bagian ini menjelaskan tahap preproses data yang digunakan dalam proses mining. Preproses data ini terdiri dari pembersihan data dan transformasi data. Pada proses pembersihan data dilakukan untuk membersihkan noise, data tidak konsisten, dan tidak relevan. Contohnya ada beberapa records data yang atribut seperti nilai mata kuliah yang kosong atau bernilai 0. Records yang demikian tersebut perlu dihilangkan karena berpotensi membuat hasil tidak valid. Proses transformasi dilakukan pada data yang akan digunakan dalam proses mining. Tujuan dari transformasi data adalah untuk meningkatkan efisiensi dari model pembelajaran data mining pada algoritma. Beberapa atribut yang menyimpan nilai rentang seperti Gaji Orang Tua.

### D. Implementasi dan Pengujian

Proses implementasi sistem merupakan proses pengimplementasian ke dalam program komputer dari hasil preproses data untuk menentukan strategi promosi Program Studi Pendidikan Teknologi Informasi menggunakan algoritma K-Means Clustering telah dibuat. Implementasi dilakukan dengan pembuatan program menggunakan software Weka.

Pengujian dilakukan terhadap sistem apakah sistem dapat melakukan proses perhitungan dari algoritme yang telah diimplementasi untuk penelitian ini. Pengujian dilakukan dengan beberapa skenario untuk mengetahui seberapa baik hasil optimasi yang dapat dilakukan oleh sistem.

### III. HASIL DAN PEMBAHASAN

#### A. Hasil Pengumpulan Data

Data yang dibutuhkan pada penelitian ini diperoleh dari database program studi Pendidikan Teknologi Informasi Fakultas Ilmu Komputer Universitas Brawijaya melalui bagian akademik dan PSIK Fakultas Ilmu Komputer. Penelitian ini hanya menggunakan data mahasiswa program studi Pendidikan Teknologi Informasi hanya angkatan 2020 dan 2021 karena Angkatan menggunakan kurikulum baru atau kurikulum 2020. Data yang digunakan sebanyak 150 data mahasiswa. Faktor eksternal dari mahasiswa yang digunakan seperti Pekerjaan Ayah, Pekerjaan Ibu, Pendidikan Ayah, Pendidikan Ibu, Gaji Orang Tua, Asal Sekolah, Propinsi, Jurusan di Sekolah, dan Kategori UKT. Beberapa faktor yang dipilih merujuk dari penelitian sebelumnya [22]. Faktor-faktor yang digunakan sesuai dengan kebutuhan pihak program studi Pendidikan teknologi informasi. Atribut data yang digunakan dapat dilihat pada Tabel 1.

TABEL I.  
 ATRIBUT DATA

| No. | Nama Atribut | Keterangan                |
|-----|--------------|---------------------------|
| 1.  | Kerja Ayah   | Pekerjaan Ayah            |
| 2.  | Kerja Ibu    | Pekerjaan Ibu             |
| 3.  | Didik Ayah   | Pendidikan Ayah           |
| 4.  | Didik Ibu    | Pendidikan Ibu            |
| 5.  | Gaji Ortu    | Gaji Orang Tua            |
| 6.  | Seleksi      | Seleksi Masuk Universitas |
| 7.  | Sekolah      | Asal Sekolah              |
| 8.  | Propinsi     | Propinsi Asal             |
| 9.  | Jurusan      | Jurusan saat sekolah      |
| 10. | KATEGORIUKT  | Kategori UKT              |

#### B. Hasil Preproses Data

Pada bagian ini menjelaskan tahap preproses data yang digunakan dalam proses mining. Preproses data ini terdiri dari pembersihan data, seleksi data, dan transformasi data. Pada proses pembersihan data dilakukan untuk membersihkan noise, data tidak konsisten, dan tidak relevan. Pembersihan data lainnya adalah data mahasiswa yang masuk melalui jalur SAP karena atribut dan matakuliah pada semester 1 dan 2 berbeda dengan matakuliah mahasiswa S1 yang masuk melalui jalur regular, sbmptn, maupun mandiri. Dari 150 records data mahasiswa angkatan tahun 2020 – 2021, 8 data mahasiswa angkatan tidak digunakan karena terdapat atribut yang kosong. Hasil akhir dari cleaning data berjumlah 142 data.

Pada proses seleksi ini dilakukan berdasarkan batasan atribut yang digunakan untuk clustering hanya berasal dari faktor eksternal mahasiswa. Data akademik yang diberikan memiliki 26 atribut antara lain ID Mhs, Nama, NIM, Jenis kelamin, Jalur Masuk, dan Angkatan, Tahun, dan lain-lainnya. Penelitian hanya menggunakan 10 atribut yaitu : Pekerjaan Ayah, Pekerjaan Ibu, Pendidikan Ayah, Pendidikan Ibu, Gaji Orang Tua, Asal Sekolah, Propinsi, Jurusan di Sekolah, dan Kategori UKT. Atribut data yang digunakan dapat dilihat pada Tabel 1.

Proses transformasi dilakukan pada data yang akan digunakan dalam proses mining. Tujuan dari transformasi data adalah untuk meningkatkan efisiensi dari model pembelajaran data mining pada algoritma. Atribut yang menyimpan nilai rentang seperti Gaji Orang Tua berupa nilai numerik ditransformasikan pada nilai interval nominal seperti tabel 2.

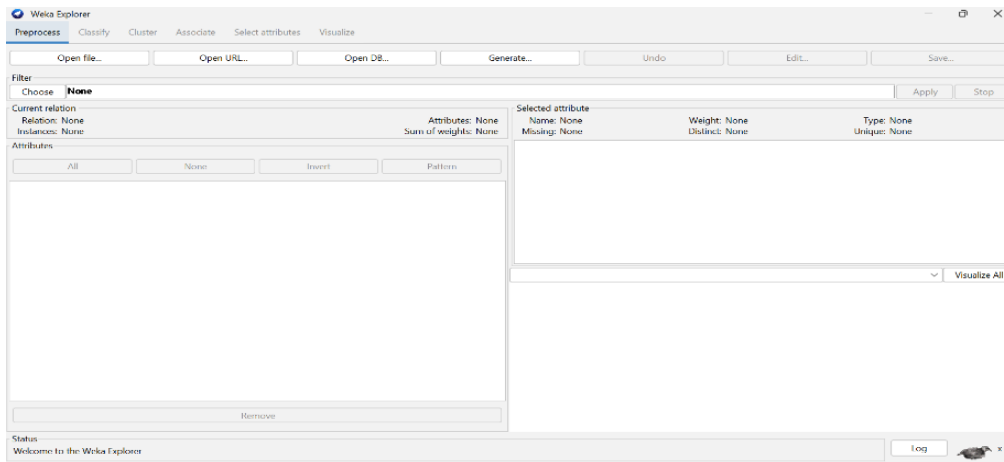
TABEL II.  
 TRANSFORMASI GAJI ORANG TUA

| No. | Interval             |
|-----|----------------------|
| 1.  | kurang dari 1,8 juta |
| 2.  | 1,8 juta - 3,5 juta  |
| 3.  | 3,5 juta - 6,9 juta  |
| 4.  | lebih dari 6,9 juta  |

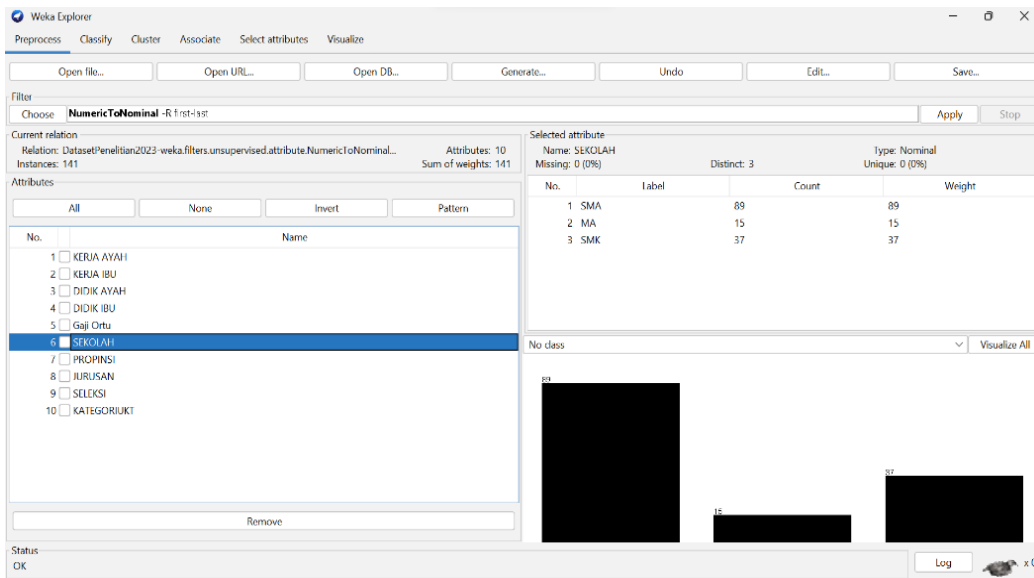
#### C. Hasil Implementasi Algoritma

Implementasi algoritma K-means Clustering dengan tools WEKA menggunakan data sebanyak 142 data dengan beberapa skenario. Gambar 2. menunjukkan tampilan utama dari software WEKA. Tampilan utama terdapat beberapa menu, seperti Praprocess, Classify, Cluster, Associate, Select attributes, dan Visualize. Pada penelitian ini menggunakan menu Cluster seperti Gambar 3. Pada menu Cluster menggunakan algoritma

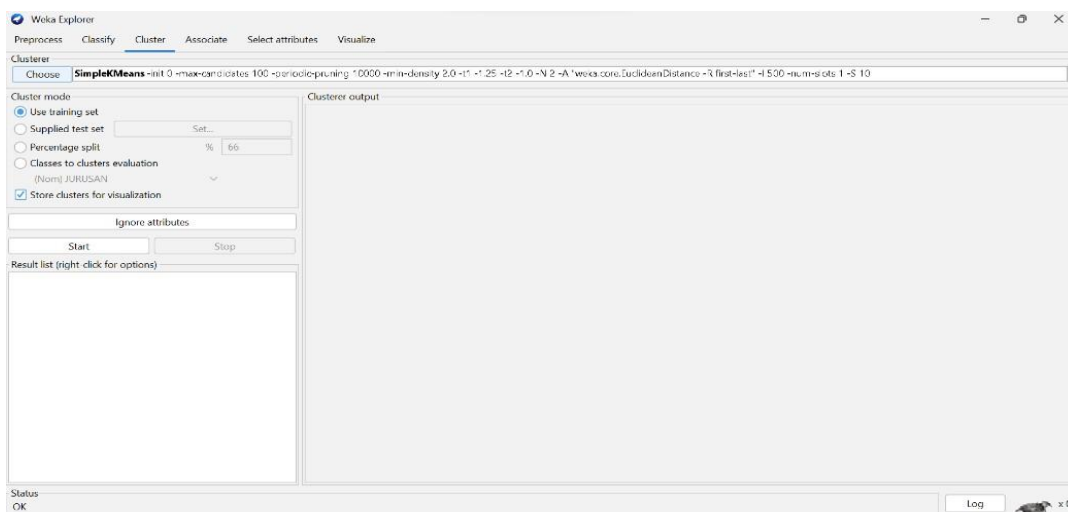
SimpleKMeans. Preprocess dilakukan sebelum Cluster dengan menggunakan button Open File seperti Gambar 4.



Gambar 2. Tampilan Utama Software WEKA



Gambar 3. Tampilan setelah Open File Data



Gambar 4. Tampilan pada Menu Cluster

#### D. Hasil Pengujian dan Analisis

Pengujian dibagi dalam 3 skenario untuk melihat performa dataset yang diolah dan algoritma yang digunakan. Pengujian pertama yaitu membandingkan persentase kluster yang terbentuk dari Algoritma SimpleKMeans (SK) dengan menggunakan distance function yang berbeda-beda. Distance function yang digunakan adalah Euclidean dan Manhattan. Parameter awal kluster  $k=2$ ,  $\text{maxIterations} = 500$ , dan nilai seed  $s=10$ . Tabel 4.3 menunjukkan tidak ada perbedaan nilai persentase kluster dari Euclidean dan Manhattan. Sehingga skenario kedua menggunakan Distance function yang adalah Euclidean.

TABEL IV.  
 PERSENTASE PEMBENTUKAN KLUSTER DENGAN DISTANCE FUNCTION YANG BERBEDA

| Distance Function | Kluster 0 | Kluster 1 |
|-------------------|-----------|-----------|
| Euclidean         | 43%       | 43%       |
| Manhattan         | 57%       | 57%       |

Pengujian kedua yaitu menggunakan fitur classes to cluster evaluation yang bertujuan untuk melihat seberapa baik dataset yang diuji dalam melakukan klustering tanpa diberikan class antar data. Untuk menggunakan fitur ini, semua dataset dalam bentuk Numeric harus diubah dalam bentuk nominal agar dapat diproses. Algoritma SimpleKMeans (SK) menggunakan parameter  $\text{maxIterations} = 500$ , nilai seed  $s=10$ , Class attribute: KATEGORIUKT, dan distance function = Euclidean, sedangkan parameter kluster  $k=2, 3, 4, 5, 6, 7, 8, 9, 10$ . Class attribute: KATEGORIUKT memiliki 7 label atau class. Tabel 5. menjelaskan nilai SSE tertinggi adalah 521 pada  $k=2$  sedangkan nilai SSE terendah sebesar 365 pada  $k=10$ . Jika pengujian menggunakan  $k=7$  seperti banyaknya 7 Class menunjukkan nilai SSE sebesar 393. Nilai Correctly clustered instances tertinggi adalah 43.9716% pada  $k=10$  sedangkan Nilai Correctly clustered instances terendah adalah 39.0071% pada  $k=3$ .

TABEL V.  
 HASIL KINERJA K-MEANS DARI FITUR CLASSES TO CLUSTER EVALUATION

| Jumlah k | Incorrectly clustered instances | Correctly clustered instances | Nilai SSE |
|----------|---------------------------------|-------------------------------|-----------|
| 2        | 82.0 (58.156 %)                 | 41.844%                       | 521.0     |
| 3        | 86.0 (60.9929 %)                | 39.0071%                      | 469.0     |
| 4        | 84.0 (59.5745 %)                | 40.4255%                      | 436.0     |
| 5        | 76.0 (53.9007 %)                | 46.0993%                      | 404.0     |
| 6        | 80.0 (56.7376 %)                | 43.2624%                      | 394.0     |
| 7        | 82.0 (58.156 %)                 | 41.844%                       | 393.0     |
| 8        | 85.0 (60.2837 %)                | 39.7163%                      | 379.0     |
| 9        | 84.0 (59.5745 %)                | 40.4255%                      | 372.0     |
| 10       | 79.0 (56.0284 %)                | 43.9716%                      | 365.0     |

Pengujian ketiga yaitu menggunakan tanpa fitur classes to cluster evaluation yang bertujuan untuk melakukan klustering dataset dan bentuk hasil cluster. Algoritma SimpleKMeans (SK) menggunakan parameter  $\text{maxIterations} = 500$ , nilai seed  $s=10$ , dan distance function = Euclidean, sedangkan parameter kluster  $k=2, 3, 4, 5, 6, 7, 8, 9, 10$ . Nilai SSE tertinggi adalah 600 pada  $k=2$  sedangkan nilai SSE terendah sebesar 406 pada  $k=10$ . Tabel 6. menjelaskan hasil akhir dari centroid cluster pada  $k=2$ .

TABEL VI.  
 HASIL KINERJA ALGORITMA K-MEANS CLUSTERING PADA PENGUJIAN KETIGA

| Jumlah k | Nilai SSE |
|----------|-----------|
| 2        | 600.0     |
| 3        | 549.0     |
| 4        | 504.0     |
| 5        | 485.0     |
| 6        | 463.0     |
| 7        | 459.0     |
| 8        | 438.0     |
| 9        | 431.0     |
| 10       | 405.0     |

Tabel 7 menjelaskan Cluster 0 cenderung punya ayah dengan pekerjaan mandiri (wirausaha), sedangkan Cluster 1 ayahnya bekerja sebagai pegawai. Ini dapat menunjukkan perbedaan stabilitas pendapatan dan pola ekonomi

keluarga. Cluster 1 memiliki tingkat pendidikan ayah yang lebih tinggi daripada Cluster 0. Ini berpotensi berkorelasi dengan tingkat pendapatan dan orientasi akademik anak. Perbedaan paling signifikan ada pada tingkat ekonomi keluarga. Cluster 1 berasal dari keluarga dengan ekonomi jelas lebih tinggi dibanding Cluster 0. Ini dapat berpengaruh terhadap pola pilihan sekolah, seleksi kuliah, hingga kategori UKT yang diperoleh. Cluster 0 memiliki latar belakang vokasional (SMK), sedangkan Cluster 1 berasal dari jalur akademik (SMA). Cluster 0 cenderung diterima melalui jalur nilai rapor/prestasi sekolah, sedangkan Cluster 1 melalui tes tertulis. Sangat selaras dengan perbedaan pendapatan orang tua. Cluster 1 berada pada kategori UKT lebih tinggi karena pendapatan keluarga jauh lebih besar. Ringkasan pola utama dapat di lihat pada Tabel 8.

Strategi Promosi untuk Cluster 0 merupakan mahasiswa Ekonomi Rendah, Lulusan SMK, Jalur SNMPTN, dan UKT 3. Pesan promosi harus menekankan aksesibilitas dan keberlanjutan studi, bukan kemewahan fasilitas. Media promosi dengan cara sosialisasi langsung ke sekolah (low-cost, high impact), Kerja sama guru BK dan wali kelas, Poster dan booklet cetak sederhana, WhatsApp Group sekolah dan orang tua, dan Instagram Reels/TikTok edukatif (tanpa kesan elit). Konten materi yang ditonjolkan Skema UKT rendah (UKT 3), Beasiswa KIP-K dan bantuan pendidikan lain, Prospek kerja lulusan PTI (guru TIK, IT support, edupreneur), dan Testimoni mahasiswa PTI dari SMK dan ekonomi menengah ke Bawah.

Strategi Promosi untuk Cluster 1 merupakan mahasiswa Ekonomi Tinggi, Lulusan SMA, Jalur SBMPTN, dan UKT 6. Pesan promosi menekankan kualitas akademik, reputasi, dan keunggulan kompetitif lulusan. Media promosi dengan cara webinar akademik dan open house daring, Instagram, YouTube, website resmi, Brosur digital professional, dan Iklan berbayar tersegmentasi (Instagram Ads). Konten materi yang ditonjolkan Akreditasi A, Dosen bergelar doktor dan riset aktif, Kurikulum berbasis teknologi dan data, Koneksi industri dan sekolah, dan Prestasi mahasiswa dan alumni.

Penelitian ini sejalan dengan tren internasional yang menggunakan data mining dan cluster analysis seperti K-Means untuk memahami calon mahasiswa dan menyusun strategi pemasaran pendidikan tinggi. Namun, penelitian ini juga membawa kontribusi baru dengan memadukan variabel-variabel sosial, ekonomi, dan jalur seleksi yang jarang dianalisis secara bersamaan dalam literatur internasional yang fokus pada pemasaran umum. Hal ini menunjukkan bahwa pendekatan data-driven yang digunakan tidak hanya relevan secara lokal, tetapi juga memberikan insight yang berkontribusi pada pengetahuan global terkait segmentasi calon mahasiswa dan desain strategi rekrutmen.

TABEL VII.  
FINAL CLUSTER CENTROIDS UNTUK K=2

| Attribute   | Cluster 0            | Cluster 1           |
|-------------|----------------------|---------------------|
| KERJA AYAH  | Wiraswasta           | Bekerja             |
| KERJA IBU   | Tidak bekerja        | Tidak bekerja       |
| DIDIK AYAH  | Tamat SLTA           | Sarjana S1          |
| DIDIK IBU   | Tamat SLTA           | Tamat SLTA          |
| Gaji Ortu   | kurang dari 1,8 juta | lebih dari 6,9 juta |
| SEKOLAH     | SMK                  | SMA                 |
| PROPINSI    | Jawa Timur           | Jawa Timur          |
| JURUSAN     | IPA / MIA            | IPA / MIA           |
| SELEKSI     | SNMPTN               | SBMPTN              |
| KATEGORIUKT | 3                    | 6                   |

TABEL VIII.  
RINGKASAN POLA UTAMA

| Aspek           | Cluster 0 | Cluster 1 | Kesimpulan                    |
|-----------------|-----------|-----------|-------------------------------|
| Ekonomi         | Rendah    | Tinggi    | Paling membedakan             |
| Pendidikan Ayah | SLTA      | Sarjana   | Cluster 1 lebih berpendidikan |
| Sekolah Asal    | SMK       | SMA       | Fokus pendidikan berbeda      |
| Seleksi         | SNMPTN    | SBMPTN    | Cara masuk berbeda            |
| UKT             | 3         | 6         | Konsisten dengan ekonomi      |

#### IV. KESIMPULAN

Berdasarkan pembahasan penelitian yang telah dilakukan, maka penelitian disimpulkan bahwa weka merupakan perangkat data mining yang dapat digunakan untuk melakukan tugas klustering dari dataset yang tersedia karena menyediakan banyak algoritma klusterer yang mendukung untuk pengujian. Pengujian pertama yaitu membandingkan persentase kluster yang terbentuk dari Algoritma SimpleKMeans (SK) dengan menggunakan distance function adalah Euclidean dan Manhattan. Hasil pengujian menunjukkan tidak ada perbedaan nilai persentase kluster dari Euclidean dan Manhattan.

Hasil pengujian kedua, nilai SSE tertinggi adalah 521 pada k=2 sedangkan nilai SSE terendah sebesar 365 pada k=10. Jika pengujian menggunakan k=7 seperti banyaknya 7 Class menunjukkan nilai SSE sebesar 393. Nilai Correctly clustered instances tertinggi adalah 43.9716% pada k=10 sedangkan Nilai Correctly clustered instances

terendah adalah 39.0071% pada  $k=3$ . Hasil pengujian ketiga, Clustered Instances yang menunjukkan presentase yang hampir seimbang dalam setiap kluster adalah  $k=2$ . Cluster 0 sebesar 43% dan Cluster 1 sebesar 57%. Nilai SSE tertinggi adalah 600 pada  $k=2$  sedangkan nilai SSE terendah sebesar 406 pada  $k=10$ .

Kelemahan klustering terletak pada penggunaan jumlah kluster  $k$  untuk melakukan inisiasi centroid. Diperlukan lebih banyak informasi atribut dari dataset sehingga tingkat kesalahan dalam pengelompokan instance lebih kecil. Penelitian selanjutnya melakukan optimasi untuk mencari nilai kluster  $k$  dan centroid awal yang terbaik.

Keterbatasan dari penelitian Adalah hanya mahasiswa program studi Pendidikan Teknologi Informasi hanya angkatan 2020 dan 2021. Penelitian selanjutnya menggunakan data seluruh mahasiswa Fakultas Ilmu Komputer sebanyak lima program studi. Penelitian ini hanya menggunakan data akademik dan lingkungan keluarga. Penelitian selanjutnya dapat menambahkan data non akademik atau prestasi.

#### DAFTAR PUSTAKA

- [1] ROMERO,C., & VENTURA, S. 2010. Educational Data Mining: A Review of the State of the Art. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 40(6), 601–618. <https://doi.org/10.1109/TSMCC.2010.2053532>
- [2] Sunita M. Dol a, & Pradip M. Jawandhiya. Juni 2023. Classification Technique and its Combination with Clustering and Association Rule Mining in Educational Data Mining — A survey. Engineering Applications of Artificial Intelligence. Volume 122. <https://doi.org/10.1016/j.engappai.2023.106071>
- [3] Satrio Hadi Wijoyo, Satrio Agung Wicaksono, dan Admaja Dwi Herlambang. April 2023. ANALISIS SELEKSI ATRIBUT DALAM MEMPREDIKSI KEGAGALAN SKRIPSI MAHASISWA MENGGUNAKAN ALGORITMA NAÏVE BAYES PADA JURUSAN SISTEM INFORMASI UNIVERSITAS BRAWIJAYA. Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK). Vol. 10, No. 2. 379-384.
- [4] Imam Riadi, Rusydi Umar, dan Rio Anggara. April 2024. PREDIKSI KELULUSAN TEPAT WAKTU BERDASARKAN RIWAYAT AKADEMIK MENGGUNAKAN METODE K-NEAREST NEIGHBOR. Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK). Vol. 11, No. 2. 249-256.
- [5] Ermatita & M. Hafyz Sytar. Januari 2025. Prediksi Kelulusan Mahasiswa Menggunakan Metode K-Means dan Random Forest. Jurnal Pendidikan dan Teknologi Indonesia (JPTI). Vol. 5, No. 1. <https://doi.org/10.52436/1.jpti.577>
- [6] Stefanus Alvian Setiono1 & Eko Purwanto. 2025. Prediksi Kelulusan Mahasiswa Menggunakan Algoritma Decision Tree. Seminar Nasional Teknologi Informasi dan Bisnis (SENATIB) 2025. ISSN: 2962-1968
- [7] Elyzia Janara Khansa, Satrio Hadi Wijoyo, dan Djoko Pramono. Juli 2025. PREDIKSI PRESTASI AKADEMIK SISWA SMPN 2 WUNGU KABUPATEN MADIUN MENGGUNAKAN DECISION TREE. Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer e-Vol. 9, No. 7, Juli 2025.
- [8] Willy Aidil Basrean Hasibuan, Satrio Hadi Wijoyo, dan Faizatul Amalia. Juli 2025. Prediksi Karier Mahasiswa Pendidikan Teknologi Informasi Berdasarkan IP Semester dan SKS Menggunakan Algoritma K-Nearest Neighbor. Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer. Vol. 9, No. 7.
- [9] Sandi Fajar Rodyansyah & Ardi Mardiana. Agustus 2017. Ekstraksi Pola Kesalahan Jawaban Siswa Menggunakan Algoritma Apriori. Jurnal Infotel Informatika - Telekomunikasi - Elektronika. Vol. 9 No. 3
- [10] Admaja Dwi Herlambang & Satrio Hadi Wijoyo. Agustus 2019. Algoritma Naive Bayes untuk Klasifikasi Sumber Belajar Berbasis Teks pada Mata Pelajaran Produktif di SMK Rumpun Teknologi Informasi dan Komunikasi. Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK). Vol 6 No 4. 431-436. <https://doi.org/10.25126/jtiik.2019641323>
- [11] Alif Rizal Maulana, Satrio Hadi Wijoyo, dan Yusi Tyroni Mursityo. Juni 2023. Analisis Sentimen Kebijakan Penerapan Kurikulum Merdeka Sekolah Dasar dan Sekolah Menengah pada Media Sosial Twitter dengan Menggunakan Metode Word Embedding dan Long Short Term Memory Networks (LSTM). Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK). Vol 10 No 3. 523-530. <https://doi.org/10.25126/jtiik.2023106977>
- [12] E. H. Pujiarini. 2019. "ANALISIS ASOSIASI UNTUK MENENTUKAN STRATEGI PROMOSI PERGURUAN TINGGI DENGAN ALGORITMA APRIORI," Jurnal Informatika dan Komputer (JIKO), vol. 4, no. 1, pp. 45-51.
- [13] Darmi, Y.D. & Setiawan, A., 2017. Penerapan Metode Clustering K-Means Dalam Pengelompokan Penjualan Produk. Jurnal Media Infotama, 12(2), pp.148–157.
- [14] Widhi Lestari. Juli 2019. Clustering Data Mahasiswa Menggunakan Algoritma K-Means Untuk Menunjang Strategi Promosi (Studi Kasus : STMIK Bina Bangsa Kendari). Jurnal Sistem Informasi dan Sistem Komputer. Vol. 4. No. 2.
- [15] Rafli Danu Kusuma Putra, Kinanthi Sekar Palupi, dan Nur Wakhidah. Oktober 2025. Pengelompokan Data Nilai Mahasiswa Menggunakan Metode K-Means. Jurnal Algoritma, Vol. 6. No. 1. 88-99
- [16] Totok Suprawoto. Februari 2016. KLASIFIKASI DATA MAHASISWA MENGGUNAKAN METODE K-MEANS UNTUK MENUNJANG PEMILIHAN STRATEGI PEMASARAN. Jurnal Informatika dan Komputer (JIKO) – Vol. 1. No. 1.
- [17] Yogi Bachtiar, Rayong Wulan, dan Aswin Fitriansyah. Agustus 2025. K-MEANS CLUSTERING DATA NIK MAHASISWA BARU MENGGUNAKAN ALTAIR AI-STUDIO. STRING (Satuan Tulisan Riset dan Inovasi Teknologi). Vol. 10 No. 1
- [18] Achmad Maulidin, Rudiman, Ahmad Ridhani Mubaraq, Mohamad Febriansyah Al Akbar, dan Muhammad Farid Azis. 2024. IMPLEMENTASI K-MEANS UNTUK CLUSTERING KEPUASAN MAHASISWA TEKNIK INFORMATIKA TERHADAP LAYANAN AKADEMIK. Journal Genta Mulia. Vol. 15, No. 2.
- [19] T. T. Chasanah dan Widiyono. 2017. "PENENTUAN STRATEGI PROMOSI PENERIMAAN MAHASISWA BARU DENGAN ALGORITMA CLUSTERING K-MEANS," IC-Tech, vol. XII, no. 1, pp. 39-44.
- [20] E. Buulolo dan R. Syahputra. 2019. "Implementasi Algoritma Clustering K-Means Untuk Mengelompokkan Mahasiswa Baru Yang Berpotensi (Studi Kasus: Stmik Budi Darma)," dalam Seminar Nasional Riset Information Science (SENARIS), Siantar.
- [21] Setiyo, L. A. dan Ign. F. Bayu Andoro. 2021. Penerapan Algoritma K-Means Untuk Menentukan Strategi Promosi (Studi Kasus : Universitas Katolik Widya Mandala Kampus Kota Madiun). Seminar dan Konferensi Nasional IDEC. Hal 26-27, bulan Juli. Silalahi, N. 2020. "Penentuan Strategi Promosi Universitas Budi Darma Menggunakan Algoritma K-Means Clustering". TIN: Terapan Informatika Nusantara. Vol. 1, no. 1, hal 40-46.
- [22] Rahmalinda, N. A. dan Arief Jananto. 2022. Penerapan Metode K-Means Clustering Dalam Menentukan Strategi Promosi Berdasarkan Data Penerimaan Mahasiswa Baru. Jurnal Tekno kompak. Vol. 16, No. 2, Hal. 163-175.
- [23] Putri Margaretta, Betha Nurina Sari, dan Azhari Ali Ridha. November 2024. Clustering Data Penjualan Toko XYZ Menggunakan Metode K-Means. Jurnal Ilmiah Wahana Pendidikan. Vol. 10, No. 22. 1092-1101
- [24] Satriawan, M.A., Andreswari, R. & Pratiwi, O.N., 2021. Segmentasi Pelanggan Telkomsel Menggunakan Metode Clustering Dengan RFM Model Dan Algoritma K-Means. e-Proceeding of Engineering, 8(2), pp.2876–2883.
- [25] Prianto, C. & Bunyamin, S., 2020. Pembuatan aplikasi clustering gangguan jaringan menggunakan metode K-Means clustering. Bandung. Kreatif Industri Nusantara.