JELTec (Journal of Learning Technology)
JELTec, Volume 04, Number 01, March 2026
E ISSN: 2986-3538; P ISSN: 2986-4364, pp. 1-12
https://journal.yhmm.or.id/index.php/JELTec
DOI: 10.56327/**JELTec**.v13i2.1297

Educational Data Mining for Student Graduation Pattern Analysis Using K-Means Clustering: A Case Study of STIT Pringsewu

Muhamad Muslihudin[1], Fitria Cahyani[2], Miswan Gumanti[3]
[1,2,3]Prodi Sistem Informasi, Institut Bakti Nusantara, Lampung
Jalan Wisma Rini. No 09 Pringsewu, Lampung
E-Mail: mmuslihudin415@gmail.com*

Article history:
Received: December 22, 2025
Revised: January 8, 2026
Accepted: February 3, 2025

Corresponding authors*
mmuslihudin415@gmail.com

Keywords:
Educational;
Data Mining;
K-Means Clustering;
Learning Analytics;
Student Graduation Patterns.

Abstract
The increasing adoption of information technology in higher education has generated large volumes of academic data that can be utilized to support evidence-based institutional decision-making. However, many Islamic higher education institutions still face challenges in identifying student graduation patterns, particularly in distinguishing students who graduate on time from those who experience academic delays. This study aims to analyze student graduation patterns using the K-Means clustering algorithm as an Educational Data Mining approach. The study utilized academic records of 634 undergraduate students from an Islamic higher education institution in Lampung, Indonesia, with 127 records (20% of the population) selected through representative sampling for clustering analysis. Data preprocessing included cleaning, normalization, and attribute selection before clustering was performed using the Orange Data Mining platform. The resulting clusters were evaluated through visual exploration using scatter plots to identify similarities and differences in graduation characteristics. The analysis revealed distinct student clusters representing different graduation profiles, including timely graduates and students with delayed completion. These findings provide valuable insights for academic management by enabling early identification of at-risk students and supporting data-driven academic interventions, strategic planning, and quality assurance initiatives. The proposed approach demonstrates that K-Means clustering is an effective method for discovering hidden patterns in academic data and can contribute to improving student success management in Islamic higher education institutions.

I. PENDAHULUAN

Perguruan tinggi memiliki peran strategis dalam mencetak sumber daya manusia berkualitas, di mana tingkat kelulusan tepat waktu menjadi indikator keberhasilan utamanya. STIT Pringsewu, sebagai salah satu PTKI di Lampung, menghadapi tantangan berupa keberagaman latar belakang mahasiswa yang memengaruhi pola studi mereka.

Dalam beberapa tahun terakhir, dinamika jumlah mahasiswa di Perguruan Tinggi Keagamaan Islam (PTKI) wilayah Lampung, khususnya di STIT Pringsewu, menunjukkan peningkatan yang cukup signifikan. pada tahun akademik 2022/2023 jumlah mahasiswa masuk sebanyak 180 orang, meningkat menjadi 200 orang pada 2023/2024, dan mencapai 220 orang pada 2024/2025. Peningkatan ini menunjukkan daya tarik institusi yang semakin kuat di masyarakat. Peningkatan kuantitas mahasiswa ini membawa tantangan besar dalam hal retensi dan ketepatan waktu lulus. Menurut data yang sama, angka mahasiswa yang keluar juga mengalami kenaikan dari 30 orang hingga mencapai 40 orang dalam tiga tahun terakhir. Kondisi ini mempertegas bahwa keberhasilan sebuah perguruan tinggi tidak hanya diukur dari jumlah pendaftar, tetapi juga dari kemampuannya meluluskan mahasiswa secara tepat waktu, karena hal ini merupakan indikator kualitas akreditasi (Rahman et al., 2025).

Permasalahan yang dihadapi oleh pengelola institusi adalah kesulitan dalam mengambil keputusan strategis terkait penanganan mahasiswa yang berisiko mengalami keterlambatan studi. Kendala utamanya adalah beban kerja manual dalam menganalisis data akademik yang sangat besar. Mengumpulkan dan menganalisis variabel seperti IPK, jumlah SKS, dan faktor sosial ekonomi secara manual memerlukan waktu dan tenaga yang besar[1]. Hal ini menghambat pengambilan keputusan yang cepat dan efisien bagi pihak manajemen kampus. Dengan banyaknya data mahasiswa yang tersedia di sistem informasi akademik, diperlukan metode otomatis yang dapat membantu memahami pola kelulusan secara cepat. Penggunaan metode seperti *K-Means Clustering* dapat menjadi solusi untuk mengelompokkan mahasiswa berdasarkan karakteristik kemiripan data mereka[2].

Transformasi digital di bidang pendidikan tinggi telah menghasilkan akumulasi data akademik yang semakin besar dan kompleks, sehingga memerlukan pendekatan analitis yang mampu menghasilkan informasi bernilai untuk mendukung pengambilan keputusan. Salah satu pendekatan yang berkembang pesat adalah *Educational Data Mining* (EDM) melalui penerapan algoritma *clustering* seperti K-Means, yang mampu mengidentifikasi pola tersembunyi dari data akademik mahasiswa. Berbagai penelitian terdahulu menunjukkan bahwa algoritma K-Means efektif dalam mengelompokkan mahasiswa berdasarkan karakteristik akademik maupun pola kelulusan. [3] memanfaatkan K-Means untuk mengidentifikasi kelompok mahasiswa berdasarkan lama studi, sedangkan [4] berhasil memetakan tingkat kelulusan mahasiswa menggunakan pendekatan klasterisasi. Penelitian lain oleh [5] [6] juga menunjukkan bahwa hasil pengelompokan mampu memberikan informasi yang bermanfaat bagi pengelola perguruan tinggi dalam menyusun strategi peningkatan mutu akademik. Meskipun demikian, sebagian besar penelitian tersebut hanya berfokus pada variabel akademik, seperti IPK dan jumlah SKS, tanpa mengintegrasikan faktor sosial ekonomi mahasiswa sebagai bagian dari analisis pola kelulusan.

Kondisi tersebut menunjukkan adanya kesenjangan penelitian, khususnya pada perguruan tinggi keagamaan Islam di Indonesia. Penelitian yang mengkaji pola kelulusan mahasiswa di Sekolah Tinggi Ilmu Tarbiyah (STIT) Pringsewu menggunakan pendekatan *Educational Data Mining* masih sangat terbatas. Padahal, karakteristik mahasiswa pada perguruan tinggi keagamaan memiliki latar belakang akademik, sosial, dan ekonomi yang beragam sehingga berpotensi menghasilkan pola kelulusan yang berbeda dibandingkan perguruan tinggi umum. Selain itu, proses evaluasi akademik di STIT Pringsewu masih didominasi oleh analisis deskriptif sehingga belum mampu mengungkap hubungan multidimensi antara capaian akademik, beban studi, dan kondisi ekonomi keluarga mahasiswa. Kebaruan penelitian ini terletak pada integrasi atribut akademik (IPK dan jumlah SKS) dengan atribut sosial ekonomi (biaya pendidikan, pekerjaan, dan penghasilan orang tua) dalam proses klasterisasi menggunakan algoritma K-Means untuk menghasilkan profil kelulusan mahasiswa yang lebih komprehensif. Pendekatan ini diharapkan mampu

memberikan informasi yang lebih akurat mengenai karakteristik mahasiswa yang berpotensi lulus tepat waktu maupun mengalami keterlambatan studi.

Berdasarkan permasalahan tersebut, penelitian ini bertujuan untuk menganalisis dan mengelompokkan pola kelulusan mahasiswa STIT Pringsewu menggunakan algoritma K-Means sebagai pendekatan *Educational Data Mining*. Secara khusus, penelitian ini bertujuan mengidentifikasi atribut-atribut yang berkontribusi terhadap pembentukan klaster kelulusan, menginterpretasikan karakteristik setiap klaster yang dihasilkan, serta menghasilkan informasi berbasis data yang dapat dimanfaatkan sebagai dasar penyusunan kebijakan akademik. Hasil penelitian diharapkan mampu mendukung pengembangan *data-driven academic decision making*, seperti implementasi *early warning system*, penyusunan program pendampingan akademik, penyaluran bantuan pendidikan yang lebih tepat sasaran, serta peningkatan efektivitas sistem penjaminan mutu internal. Dengan demikian, penelitian ini tidak hanya memberikan kontribusi terhadap pengembangan kajian *Educational Data Mining* pada perguruan tinggi keagamaan Islam, tetapi juga menyediakan model analisis yang dapat direplikasi pada institusi pendidikan tinggi lainnya dalam upaya meningkatkan ketepatan waktu kelulusan mahasiswa.

II. METODE PENELITIAN

Teknik Pengumpulan Data

Metode pengumpulan data adalah cara yang dipakai peneliti untuk mengumpulkan informasi yang nantinya akan dianalisis. Data harus dikumpulkan dengan tepat, teratur, dan hati-hati supaya hasilnya akurat dan sesuai dengan kondisi yang sebenarnya. Beberapa cara yang biasa digunakan antara lain:

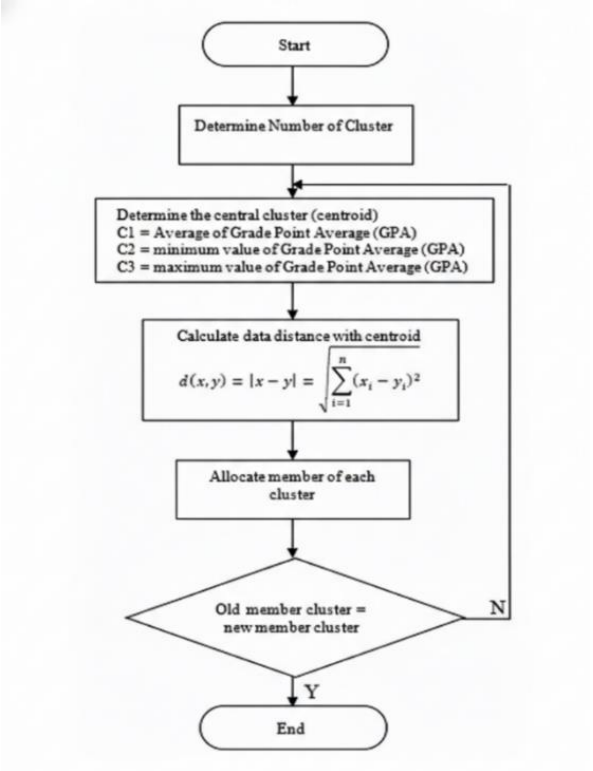
1. Observasi
   Observasi merupakan teknik pengambilan data yang dilakukan dengan mengamati guru dan siswa secara sistematis untuk memperoleh informasi relevan terhadap praktik pembelajaran di kelas. Data diambil langsung dari dokumen atau database resmi yang memuat informasi akademik mahasiswa, seperti IPK, lama studi, jumlah SKS, dan status kelulusan. Tujuannya supaya peneliti bisa melihat kondisi akademik mahasiswa secara langsung dan mendapatkan gambaran yang nyata.
2. Wawancara
   Wawancara merupakan teknik pengumpulan data yang sering digunakan dalam penelitian kualitatif untuk memperoleh informasi mendalam melalui proses tanya jawab antara peneliti dan narasumber. Wawancara dilakukan dengan ketua STIT Pringsewu ibu Iis Maysaroh dan bagian BAAK untuk mengetahui bagaimana proses kelulusan mahasiswa dan cara pengelolaan data akademik. Dari hasil temuan awal, diketahui bahwa data seperti IPK, jumlah SKS, dan lama studi sudah tersimpan dengan baik di sistem, tetapi penggunaannya masih sebatas untuk administrasi dan laporan. Data tersebut belum dianalisis lebih lanjut untuk melihat pola kelulusan atau mengetahui mahasiswa yang berisiko terlambat lulus. Pihak akademik juga menyampaikan bahwa belum ada metode khusus untuk mengelompokkan mahasiswa berdasarkan kondisi akademiknya. Dari hasil ini dapat disimpulkan bahwa perlu adanya analisis berbasis data mining agar data yang ada bisa dimanfaatkan lebih maksimal dan membantu kampus dalam mengambil keputusan yang lebih tepat.
3. Studi Pustaka
   Beberapa penelitian menunjukkan bahwa K-Means mampu membagi mahasiswa menjadi beberapa kelompok berdasarkan prestasi akademik, sehingga memudahkan identifikasi pola kelulusan dan memberikan klasifikasi objektif terhadap pencapaian mahasiswa. Studi pustaka merupakan landasan krusial bagi

peneliti untuk mencari tempat berpijak yang kokoh melalui pengkajian berbagai sumber bacaan terkait variabel masalah dan tindakan dalam penelitian.

Metode K-Means

Metode K-Means merupakan algoritma yang paling populer dalam teknik klasterisasi data mining karena kemampuannya dalam membagi data ke dalam beberapa cluster yang bermakna. Algoritma K-Means adalah salah satu metode dalam data mining yang digunakan untuk mengelompokkan data ke dalam beberapa kelompok berdasarkan kemiripan datanya. Cara kerjanya dengan menentukan titik tengah kelompok, lalu data yang mirip akan dimasukkan ke kelompok yang sama. Proses ini dilakukan berulang sampai kelompok yang terbentuk sudah stabil. Metode K-Means cukup sering digunakan karena langkah-langkahnya sederhana dan mudah dipahami. Selain itu, metode ini juga bisa mengolah data dalam jumlah banyak dengan cukup cepat[7], [8].

Dalam analisis data akademik, algoritma K-Means digunakan untuk mengelompokkan mahasiswa berdasarkan kondisi akademiknya. Menurut[8], [9], [10] metode K-Means merupakan algoritma unsupervised learning yang efektif untuk mengelompokkan data berdasarkan kesamaan karakteristik. Data yang digunakan biasanya berupa IPK, lama studi, jumlah SKS yang telah ditempuh, dan status kelulusan. Dengan pengelompokan ini, mahasiswa yang memiliki kondisi akademik mirip akan berada dalam satu kelompok. Dari hasil tersebut, bisa terlihat pola kelulusan yang berbeda di setiap kelompok. Hasil analisis ini dapat dimanfaatkan sebagai bahan evaluasi dan perencanaan kebijakan akademik.



Gambar 1. Flowchart Algoritma K-Means
Sumber : muslihudin [11]

Proses dimulai dengan menentukan jumlah klaster yang diinginkan, kemudian menetapkan titik pusat awal (*centroid*) untuk masing-masing klaster berdasarkan nilai rata-rata, minimum, dan maksimum dari data IPK (*Grade Point Average*). Selanjutnya, algoritma menghitung jarak antara setiap data dengan ketiga centroid tersebut menggunakan rumus *Euclidean Distance* untuk menentukan kedekatan data terhadap pusat kelompok.

Berdasarkan perhitungan jarak terdekat, setiap data kemudian dialokasikan ke dalam klaster yang sesuai. Langkah ini diulangi terus menerus (iterasi) hingga keanggotaan klaster stabil, yaitu ketika tidak ada lagi data yang berpindah klaster, dan proses berakhir setelah mendapatkan hasil pengelompokan yang optimal.

Tahapan Metode K-Means

1. **Penentuan Centroid Awal**
   Centroid awal merupakan titik pusat klaster yang ditentukan secara acak pada tahap awal proses clustering. Centroid ini berfungsi sebagai representasi awal dari setiap klaster yang akan dibentuk.
   $\boldsymbol{C_k = (x_1, x_2, \ldots, x_n)}$ **……(3)**
   Keterangan:
   a. C_k adalah centroid klaster ke-k
   b. $x_1, x_2, \ldots, x_n$ adalah nilai atribut pada centroid
   c. n adalah jumlah atribut data
2. **Perhitungan Jarak Data ke Centroid**
   Untuk mengetahui kedekatan antara suatu data dengan centroid klaster, digunakan rumus jarak Euclidean. Data akan dimasukkan ke dalam klaster yang memiliki jarak paling dekat dengan centroid.
   $\boldsymbol{d(x_i, c_j) = \sqrt{\sum_{k=1}^{n} (x_{ik} - c_{jk})^2}}$ **……(4)**
   Keterangan:
   a. $d(x_i, c_j)$ adalah jarak data ke-i terhadap centroid klaster ke-j
   b. x_{ik} adalah nilai atribut ke-k pada data ke-i
   c. c_{jk} adalah nilai atribut ke-k pada centroid klaster ke-j
   d. n adalah jumlah atribut

3. **Penentuan Keanggotaan Klaster**
   Tahap penentuan keanggotaan klaster adalah proses mengelompokkan data berdasarkan tingkat kemiripan. Setiap data akan dihitung jaraknya ke pusat klaster, lalu dimasukkan ke klaster dengan jarak paling dekat. Dengan cara ini, data yang memiliki karakteristik serupa akan berada dalam kelompok yang sama.
4. **Pembaruan Centroid**
   Setelah seluruh data dikelompokkan, centroid baru dihitung berdasarkan nilai rata-rata dari seluruh data yang berada dalam satu klaster.
   $c_j = \frac{1}{m} \sum_{i=1}^{m} x_i$ ……(5)
   Keterangan:
   a. c_j adalah centroid klaster ke-j yang baru
   b. m adalah jumlah data dalam klaster
   c. x_i adalah data ke-i dalam klaster
5. **Iterasi dan Konvergensi**
   Proses perhitungan jarak dan pembaruan centroid dilakukan secara berulang hingga tidak terjadi perubahan klaster atau perubahan centroid sangat kecil.
6. **Fungsi Objektif (Sum of Squared Error / SSE)**
   Untuk mengukur kualitas hasil clustering, digunakan fungsi SSE yang menghitung total kuadrat jarak antara data dan centroid klasternya.
   $SSE = \sum_{j=1}^{k} \sum_{i=1}^{n} d(x_i, c_j)^2$ ……(6)
   Keterangan:
   a. k adalah jumlah klaster

b. n adalah jumlah data

c. $d(x_i, c_j)$ adalah jarak data ke centroid klaster

Nilai SSE yang semakin kecil menunjukkan bahwa data dalam klaster semakin homogen dan hasil clustering semakin baik [12], [13], [14].

Data dan Variabel

Penelitian ini menggunakan data kuantitatif yang mencakup mahasiswa yang masih menjalani studi pada periode 2022–2025. Pemilihan data kuantitatif bertujuan agar analisis dapat dilakukan secara terukur dan sistematis, sehingga pola akademik mahasiswa dapat teridentifikasi dengan lebih mudah. Seluruh data diperoleh dari dokumen resmi akademik STIT Pringsewu, sehingga informasi yang digunakan dapat dipastikan valid dan mencerminkan kondisi sebenarnya mahasiswa. Variabel yang digunakan dalam penelitian ini adalah data akademik mahasiswa, yaitu:

Tabel 1. Variabel Penelitian.

No	Peneliti	Variabel
1	[15][16]	Motivasi Belajar, Jalur Masuk, Kualititas Pengajaran.
2	[17] [14]	IPK, SKS yg ditempuh, Masa Studi, Prestasi Akademik.
3	[6], [18]	Usia Mahasiswa, Jenis Kelamin, Faktor Kelulusan, Ekonomi, Model Prediksi Kelulusan
4	[4], [19], [20]	Status Sosial Ekonomi, Pendapatan Orang Tua, Biaya Pendidikan, Minat Melanjutan Perguruan Tinggi

Berdasarkan data tersebut, variabel-variabel yang diteliti mencakup berbagai faktor penting dalam dunia pendidikan, mulai dari kondisi internal mahasiswa hingga lingkungan keluarganya. Penelitian difokuskan pada indikator kinerja akademik, seperti IPK, jumlah SKS, dan lama masa studi, serta faktor pendidikan dan pembelajaran, termasuk motivasi belajar dan kualitas pengajaran. Selain itu, penelitian juga menekankan faktor sosial ekonomi, seperti pendapatan orang tua, biaya pendidikan, dan status sosial, yang memengaruhi minat mahasiswa untuk melanjutkan pendidikan. Secara keseluruhan, hubungan antara data demografis, jalur masuk, dan model prediksi kelulusan menunjukkan upaya untuk memahami faktor-faktor yang memengaruhi keberhasilan studi mahasiswa secara menyeluruh dan komprehensif.

Tabel 2. Hasil Variabel

Variabel X	IPK, SKS, Pekerjaan, Biaya Pendidikan
Variabel Y	Tepat Waktu, Lulus Terlambat, Drop Out (DO)

Penelitian ini bertujuan untuk mengetahui faktor-faktor yang memengaruhi keberhasilan studi mahasiswa. Untuk itu, digunakan dua kategori variabel utama yang membantu menganalisis hubungan antara kondisi mahasiswa dan hasil studi.

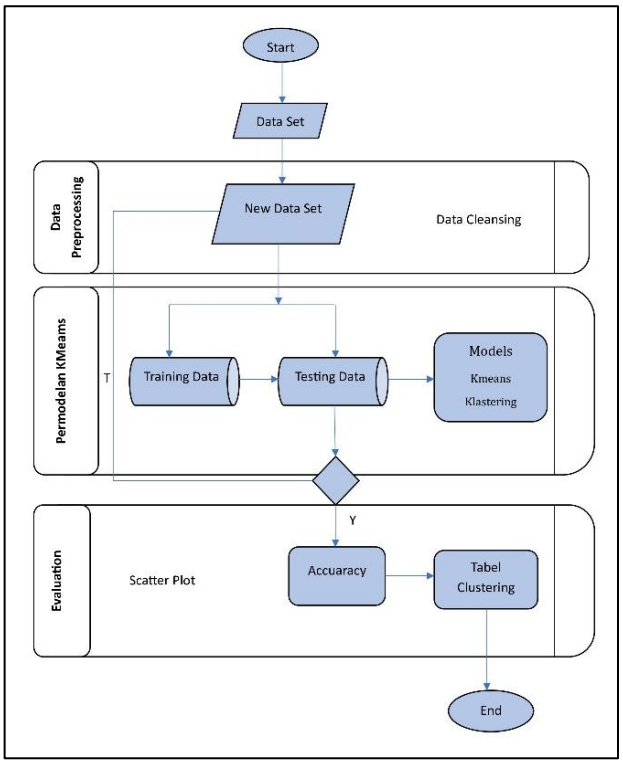
1. Variabel Independen (Variabel X) merupakan faktor-faktor yang diduga memengaruhi hasil studi mahasiswa. Variabel ini meliputi Indeks Prestasi Kumulatif (IPK) sebagai indikator pencapaian akademik mahasiswa, jumlah SKS yang ditempuh selama masa perkuliahan, serta faktor sosial ekonomi seperti pekerjaan mahasiswa dan biaya pendidikan yang dikeluarkan selama menempuh pendidikan. Variabel-variabel tersebut digunakan untuk melihat keterkaitan dengan pola kelulusan mahasiswa.

2. Variabel Dependen (Variabel Y) merupakan hasil yang diukur, yaitu status kelulusan mahasiswa. Variabel ini dibagi menjadi tiga kategori: lulus tepat waktu dan tidak tepat waktu.

Dengan demikian, penelitian ini memberikan gambaran menyeluruh mengenai bagaimana berbagai faktor, baik yang berasal dari kemampuan akademik mahasiswa maupun kondisi sosial ekonomi, berperan dalam menentukan keberhasilan mereka untuk menyelesaikan studi tepat waktu.

Populasi dan Sampel

Menurut [21], [22] apabila jumlah populasi kurang dari 100 maka seluruh populasi dijadikan sampel, sedangkan jika jumlah populasi lebih dari 100 maka sampel dapat diambil antara 10%–25% dari jumlah populasi. Sejalan dengan hal tersebut, [9], [23] menyatakan bahwa jika jumlah subjek kurang dari 100, disarankan untuk memasukkan semua subjek sehingga penelitian dikategorikan sebagai penelitian populasi. Sebaliknya, jika jumlah subjek melebihi 100, peneliti dapat memilih sampel mulai dari 10% hingga 15% atau bahkan 20% hingga 25%. Dalam penelitian ini, karena jumlah populasi melebihi 100 orang, maka sampel yang diambil adalah sebesar 20% dari total 634 data kelulusan mahasiswa, sehingga menghasilkan total 127 data kelulusan mahasiswa yang akan digunakan sebagai objek analisis



Gambar 2. *Flowchart Algoritma Kmeans*

Penjelasan:

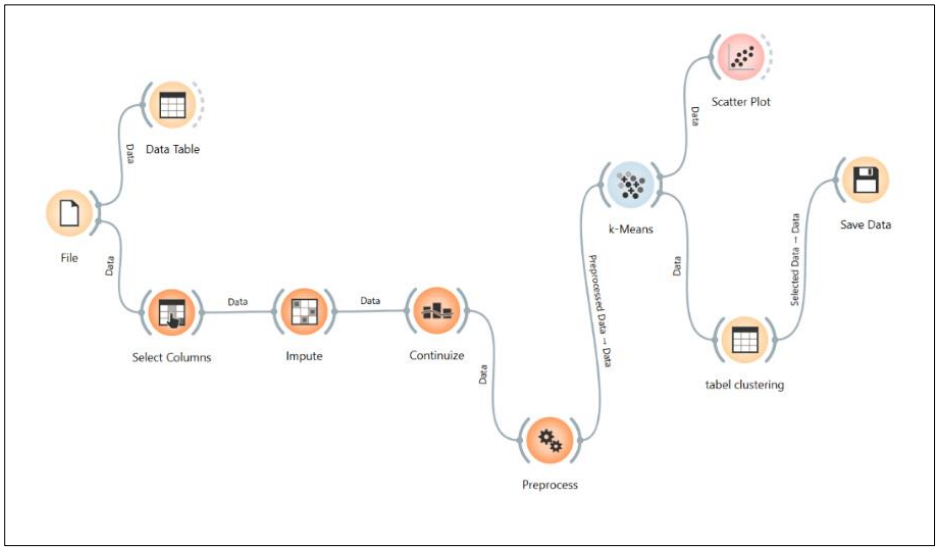
1. Penelitian dimulai dengan perancangan konsep dan identifikasi masalah.
2. Data Set: Pengumpulan data mentah yang akan digunakan sebagai objek penelitian (misalnya data kelulusan mahasiswa).
3. Tahap ini bertujuan untuk menyiapkan data mentah melalui proses seleksi menjadi *New Data Set* agar hanya data relevan yang digunakan. Selanjutnya, dilakukan *Data Cleansing* untuk membersihkan data dari nilai yang kosong, data duplikat, dan ketidakteraturan format. Proses ini memastikan data dalam kondisi bersih dan konsisten sehingga siap diolah secara optimal oleh algoritma K-Means.

4. Tahap ini dimulai dengan pembagian data menjadi *training data* untuk pembentukan pola dan *testing data* sebagai data uji. Algoritma K-Means kemudian mengelompokkan data berdasarkan kemiripan karakteristik melalui perhitungan jarak terhadap titik pusat (*centroid*). Proses ini dilakukan secara iteratif; jika hasil klaster belum mencapai kriteria optimal, alur akan kembali ke tahap pemrosesan data, namun jika sudah terpenuhi, proses akan langsung berlanjut ke tahap evaluasi.
5. Tahap ini menggunakan *Scatter Plot* untuk memvisualisasikan persebaran dan pemisahan antar klaster secara jelas. Selanjutnya, dilakukan uji *Accuracy* dan tabel *clustering* menggunakan metrik tertentu guna mengukur kualitas dan kekuatan pengelompokan yang dihasilkan. Proses ini memastikan hasil *clustering* sudah optimal sebelum ditarik kesimpulan akhir.
6. End: Setelah hasil evaluasi dinyatakan valid dan menjawab tujuan penelitian, alur kerja selesai dengan penarikan kesimpulan dari hasil pengelompokan tersebut.

III. HASIL DAN PEMBAHASAN

Pengolahan Data Menggunakan *Tools Orange*

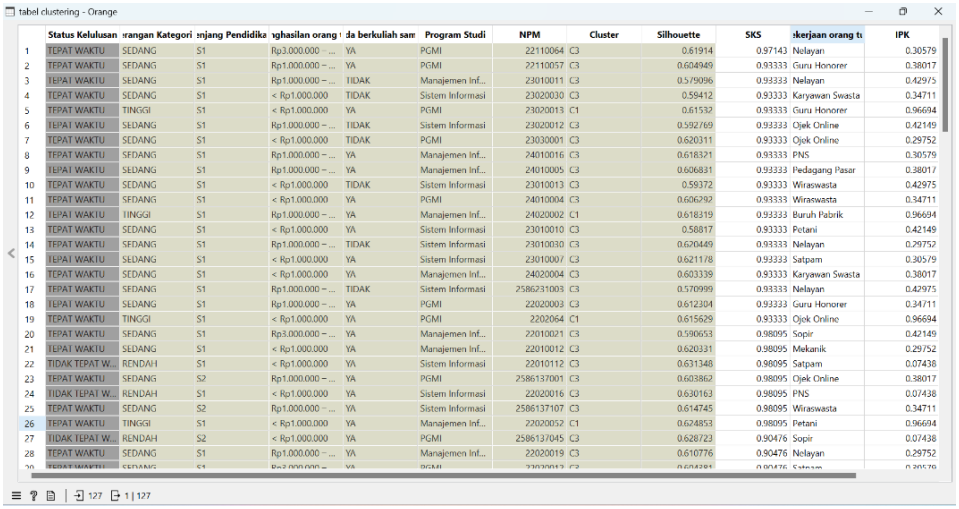
Perancangan sistem *data mining* pada skenario penelitian yang akan dilakukan ditunjukkan dengan menggunakan bantuan *software* Orange Data Mining, di mana tahapan pengolahan data diawali dengan langkah pertama File untuk memuat dataset akademik mahasiswa STIT Pringsewu, dilanjutkan dengan Select Columns untuk memilih variabel relevan seperti IPK dan SKS, serta Data Sampler untuk mengambil sampel data yang representatif. Selanjutnya, diterapkan metode K-Means untuk melakukan pengelompokan data secara otomatis yang kemudian divalidasi kekuatannya melalui Silhouette Plot serta dievaluasi akurasinya menggunakan Test and Score. Tahap akhir dari perancangan ini mencakup visualisasi hasil melalui Scatter Plot untuk melihat persebaran data, Data Table untuk menyajikan rincian klasifikasi mahasiswa secara individu, dan Tabel Clustering untuk menampilkan ringkasan statistik serta pembagian jumlah mahasiswa pada setiap kelompok kelulusan guna mempermudah penarikan kesimpulan penelitian.



Gambar 3. Desain Sistem Penelitian

Tabel clustering menyajikan hasil pengelompokan mahasiswa berdasarkan kemiripan karakteristik akademik yang telah diproses secara sistematis menggunakan algoritma K-Means. Data yang ditampilkan dalam tabel ini memberikan gambaran yang terstruktur mengenai distribusi mahasiswa pada masing-masing kelompok, sehingga tidak lagi berupa

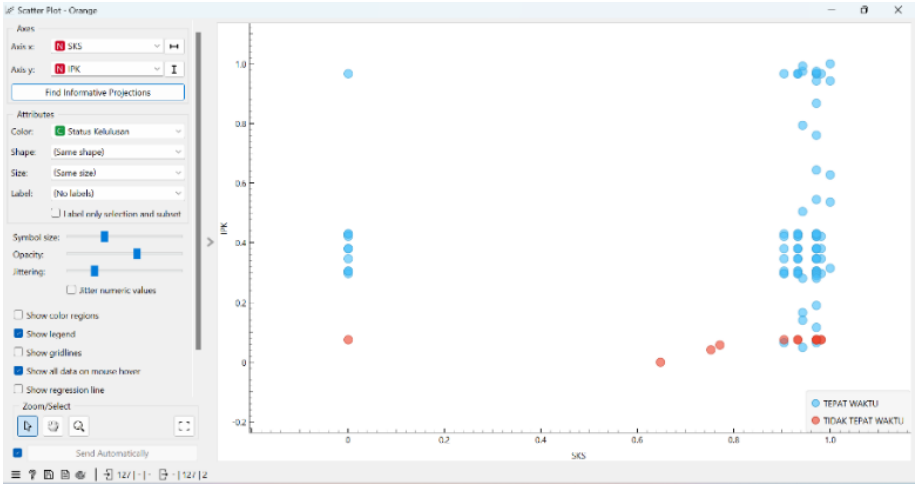
data mentah yang sulit diinterpretasikan. Melalui pengorganisasian data ini, pihak institusi dapat melihat perbandingan profil mahasiswa secara akurat, mulai dari perolehan IPK hingga akumulasi SKS, guna menentukan langkah-langkah strategis dalam meningkatkan kualitas serta ketepatan waktu kelulusan.



Gambar 4. Tabel *Clustering*

Dalam tabel clustering yang dipaparkan, terlihat pembagian jumlah mahasiswa ke dalam beberapa klaster utama menunjukkan dominasi sebanyak 88 mahasiswa yang dikategorikan lulus tepat waktu karena performa akademik yang stabil. Sementara itu, Cluster terbawah mengidentifikasi adanya 15 mahasiswa yang berada dalam kategori berisiko atau tidak tepat waktu, yang ditandai dengan nilai rata-rata variabel yang lebih rendah dibandingkan kelompok lainnya. Analisis pada pusat klaster (centroid) ini memberikan dasar informasi yang kuat bagi pimpinan STIT Pringsewu untuk melakukan intervensi dini dan mengevaluasi kebijakan akademik guna menekan angka keterlambatan studi mahasiswa secara efektif.

Scatter plot yang ditampilkan dalam gambar merupakan visualisasi data yang menunjukkan hubungan antara variabel akademik mahasiswa untuk mengidentifikasi pola kelulusan. Dalam hal ini, sumbu x dan sumbu y menggambarkan variabel kunci seperti Indeks Prestasi Kumulatif (IPK) dan jumlah SKS yang telah ditempuh.



Gambar 5. *Scatter Plot*

Pada Scatter Plot ini, terlihat dua kelompok utama titik data yang ditampilkan dalam

dua warna berbeda. Warna biru menunjukkan kelompok mahasiswa yang lulus tepat waktu sedangkan warna merah menunjukkan kelompok mahasiswa yang tidak tepat waktu. Pemisahan warna ini mempermudah identifikasi pola bahwa mahasiswa dengan IPK dan SKS tinggi cenderung mengelompok pada kategori lulus tepat waktu.

Berdasarkan hasil analisis terhadap 127 data sampel mahasiswa, diperoleh sebanyak 112 mahasiswa lulus tepat waktu dan 15 mahasiswa tidak tepat waktu, sehingga dapat disimpulkan bahwa mayoritas mahasiswa berada pada kategori lulus tepat waktu. Hasil ini memberikan gambaran yang jelas mengenai pola kelulusan mahasiswa dan dapat dimanfaatkan oleh pihak perguruan tinggi sebagai dasar dalam pengambilan kebijakan untuk meningkatkan kualitas akademik serta mengurangi keterlambatan kelulusan. Berdasarkan hasil tersebut, mahasiswa yang tidak lulus tepat waktu perlu mendapatkan perhatian melalui beberapa upaya yang mendukung kelancaran studi. Salah satu langkah yang dapat dilakukan adalah pemberian keringanan biaya pendidikan bagi mahasiswa yang memiliki keterbatasan ekonomi agar tidak terkendala secara finansial. Selain itu, pihak kampus dapat meningkatkan bimbingan akademik secara rutin untuk memantau perkembangan studi dan memberikan arahan yang lebih terstruktur. Mahasiswa juga diharapkan lebih disiplin dalam mengatur waktu dan aktif mengikuti perkuliahan. Dengan adanya dukungan dari kampus serta kesadaran mahasiswa, keterlambatan kelulusan dapat diminimalkan.

Penelitian mengenai perkiraan kelulusan tepat waktu mahasiswa penting dilakukan untuk membantu perguruan tinggi dalam meningkatkan kualitas akademik serta mengurangi keterlambatan masa studi. Kelulusan mahasiswa dipengaruhi oleh beberapa faktor utama, seperti nilai Indeks Prestasi Kumulatif (IPK), jumlah Satuan Kredit Semester (SKS), serta kondisi sosial ekonomi yang tercermin dari biaya pendidikan dan pekerjaan orang tua. Mahasiswa dengan IPK tinggi dan jumlah SKS yang memenuhi standar kelulusan umumnya memiliki peluang lebih besar untuk lulus tepat waktu. Metode K-Means digunakan dalam penelitian ini karena mampu mengelompokkan data berdasarkan kemiripan karakteristik tanpa memerlukan label awal. Algoritma ini bekerja dengan membagi data ke dalam beberapa cluster berdasarkan kedekatan jarak terhadap pusat cluster. Penggunaan metode ini dinilai tepat karena data yang digunakan memiliki banyak variabel dan belum memiliki kategori kelulusan yang pasti, sehingga dapat membantu menemukan pola dan mengelompokkan mahasiswa secara lebih objektif.

Penggunaan K-Means juga didukung oleh penelitian terdahulu. Alalawi et al. (2023) menyatakan bahwa K-Means efektif dalam mengelompokkan data performa mahasiswa, sedangkan Talib et al. (2023) menunjukkan bahwa metode ini dapat digunakan untuk menganalisis pola keberhasilan akademik. Selain itu, Khodijah et al. (2023) menjelaskan bahwa K-Means mampu mengelompokkan mahasiswa ke dalam beberapa kategori berdasarkan performa akademik. Dengan demikian, metode ini memiliki dasar yang kuat untuk digunakan dalam penelitian ini. Berdasarkan uraian tersebut, dapat disimpulkan bahwa perkiraan kelulusan tepat waktu mahasiswa dipengaruhi oleh faktor IPK, jumlah SKS, serta kondisi sosial ekonomi. Penggunaan metode K-Means dinilai tepat karena mampu mengelompokkan data secara objektif berdasarkan kemiripan karakteristik tanpa memerlukan label awal. Selain itu, metode ini didukung oleh penelitian terdahulu yang menunjukkan efektivitasnya dalam menganalisis performa akademik mahasiswa. Dengan demikian, hasil penelitian ini dapat digunakan sebagai dasar dalam menentukan kategori kelulusan mahasiswa secara lebih akurat.

IV. KESIMPULAN

Penelitian ini menunjukkan bahwa penerapan algoritma K-Means sebagai pendekatan Educational Data Mining mampu mengidentifikasi dan mengelompokkan pola kelulusan mahasiswa berdasarkan karakteristik akademik dan sosial ekonomi. Proses clustering

menghasilkan kelompok mahasiswa dengan karakteristik yang berbeda, sehingga memberikan gambaran yang lebih komprehensif mengenai kecenderungan mahasiswa untuk lulus tepat waktu maupun mengalami keterlambatan studi. Temuan ini membuktikan bahwa teknik unsupervised learning dapat dimanfaatkan sebagai instrumen analitik untuk mengekstraksi pola tersembunyi dari data akademik yang sulit diidentifikasi melalui analisis konvensional. Hasil analisis menunjukkan bahwa atribut Indeks Prestasi Kumulatif (IPK), jumlah SKS yang ditempuh, biaya pendidikan, serta pekerjaan dan penghasilan orang tua merupakan variabel yang berkontribusi dalam pembentukan klaster. Berdasarkan analisis terhadap 127 data sampel yang merepresentasikan 634 data mahasiswa, diperoleh dua kelompok utama, yaitu 112 mahasiswa yang memiliki karakteristik kelulusan tepat waktu dan 15 mahasiswa yang menunjukkan kecenderungan mengalami keterlambatan kelulusan. Pengelompokan tersebut memberikan informasi yang relevan mengenai profil mahasiswa berdasarkan kombinasi faktor akademik dan kondisi sosial ekonomi. Hasil penelitian ini dapat dimanfaatkan oleh perguruan tinggi, khususnya Sekolah Tinggi Agama Islam, sebagai dasar dalam merancang kebijakan akademik berbasis data (*data-driven decision making*). Informasi hasil clustering dapat digunakan untuk mengembangkan sistem peringatan dini (*early warning system*), menetapkan program pendampingan akademik bagi mahasiswa yang berisiko terlambat lulus, serta menjadi dasar dalam penyaluran bantuan finansial yang lebih tepat sasaran. Dengan demikian, penerapan metode K-Means tidak hanya memberikan kontribusi pada analisis pola kelulusan mahasiswa, tetapi juga mendukung peningkatan kualitas layanan akademik dan efektivitas pengelolaan pendidikan tinggi.

DAFTAR PUSTAKA

[1] C. Yusuf Bayu Wicaksono, Juliane, "Pemanfaatan Manajemen Pengetahuan Untuk Membantu Persiapan Data Pada Proses Data Mining," *J. Tek. Inform. dan Sist. Inf.*, vol. 10, no. 1, pp. 298–308, 2023, doi: https://doi.org/10.35957/jatisi.v10i1.2424.

[2] S. Mochamad Wahyudi, Masitha, Risna Saragih, *Data Mining: Penerapan Algoritma K-Means Clustering dan K-Medoids Clustering*. 2020.

[3] S. F. & H. D. Priyatman H., "Klasterisasi Menggunakan Algoritma K-Means Clustering untuk Memprediksi Waktu Kelulusan," *JEPIN J. Edukasi Dan Penelit. Inform. 5(1), 62–66*, 2025, [Online]. Available: https://doi.org/10.26418/jp.v5i1.29611

[4] M. P. A. Ariawan, I. B. A. Peling, and G. B. Subiksa, "Prediksi Nilai Akhir Matakuliah Mahasiswa Menggunakan Metode K-Means Clustering (Studi Kasus : Matakuliah Pemrograman Dasar)," *J. Nas. Teknol. dan Sist. Inf.*, vol. 9, no. 2, pp. 122–131, Aug. 2023, doi: 10.25077/TEKNOSI.v9i2.2023.122-131.

[5] A. Anas, A. J. Zebua, and Akhmadi, "Clustering Nilai Mahasiswa Menggunakan Algoritma K-Means," *J. Ilm. Media Sisfo*, vol. 19, no. 1, pp. 7–14, Apr. 2025, doi: 10.33998/mediasisfo.2025.19.1.2200.

[6] D. Dwi Aulia and N. Nurahman, "Comparison Performance of K-Medoids and K-Means Algorithms In Clustering Community Education Levels," *J. Nas. Pendidik. Tek. Inform.*, vol. 12, no. 2, pp. 273–282, Jul. 2023, doi: 10.23887/janapati.v12i2.59789.

[7] S. I. Rulin Swastika, Siti Mukodimah, Ferry Susanto, Muhamad Muslihudin, *Implementasi Data Mining (Clastering, Association, Prediction, Estimation, Classification)*. Indramayu: Penerbit Adab (CV. Adanu Abimata), 2023.

[8] F. S. Fauzi, Rita Irviani, Andino Maseleno, "Revolutionizing Education through Technology : Big Data and Online Learning," in *CICCSE*, 2017, p. 44. doi: 10.1166/asl.2011.1261.

[9] M. Muslihudin, "Decision Support System for Lecturers Achieving Using Algorithm C.45 (Study: Stit Pringsewu Lampung)," *Technol. Accept. Model*, vol. 11, pp. 118–124,

2020.
[10] D. A. C, D. A. Baskoro, L. Ambarwati, and I. W. S. Wicaksana, *Belajar Data Mining dengan RapidMiner*. 2013.
[11] M. Muslihudin, "Analisis Prediksi Mahasiswa Tidak Tepat Waktu Menyelesaikan Studi Dengan Menggunakan Metode Algoritma C 4.5 (Studi Kasus: STMIK Pringsewu)," IBI Darmajaya, 2015.
[12] R. Swastika, S. Mukodimah, F. Susanto, M. Muslihudin, and S. Ipnuwati, *Implementasi Data Mining (Clastering, Association, Estimation, Classification)*. Penerbit Adab, 2023.
[13] A. M. Muhammad Muslihudin, Rita Irviani, Prayugo Khoir, "Decision Support System Level Economic Classification Of Citizens Using Fuzzy Multiple Attribute Decision Makin," in *ICCSE*, 2017, pp. 1–75.
[14] S. Mukodimah, M. Muslihudin, D. R. Mustofa, and D. Susianto, "Naive Bayes Classifier Method Analysis and Support Vector Machine (SVM) Student Graduation Prediction," *NEUROQUANTOLOGY*, vol. 20, no. 12, pp. 3522–3533, 2022, doi: 10.14704/NQ.2022.20.12.NQ77360.
[15] S. L. & A. W. Putra A. R., "The Implementation of Data Mining Techniques for Predicting Student Study Period Using the C4," *5 Algorithm*, 2023.
[16] & M. S. Fatah Z., "Analisis Data Mining dengan Algoritma K-Means Clustering untuk Menentukan Siswa Berprestasi di MTS Miftahul Ulum Bengkak," *JAMASTIKA, 4(2), 133–140*, 2025, [Online]. Available: https://doi.org/10.35473/jamastika.v4i2.4527
[17] N. W. Pratiwi, S. Hendriani, and R. Roesi, "Faktor internal dan eksternal yang mempengaruhi lama masa studi mahasiswa sarjana manajemen Universitas Riau," *J. Bisnis Mhs.*, vol. 5, no. 5, pp. 2755–2768, Oct. 2025, doi: 10.60036/jbm.698.
[18] V. S. Davina Rizky Efendi, Deci Irmayani, "Klasifikasi Tingkat Kelulusan Mahasiswa Menggunakan Algoritma K-Nearest Neighbor (K-NN) Pada Data Akademik Perguruan Tinggi," *J. Comput. Sci. Inf. Syst.*, vol. 6, no. 3, pp. 409–421, 2025, doi: https://doi.org/10.36987/jcoins.v6i3.8041.
[19] T. H. Hasibuan and D. Mahdiana, "Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Algoritma C4.5 Pada Uin Syarif Hidayatullah Jakarta," *SKANIKA*, vol. 6, no. 1, pp. 61–74, Jan. 2023, doi: 10.36080/skanika.v6i1.2976.
[20] K. A. A. F. A. & Z. N. A. P. Fatkhudin A., "Implementasi Algoritma Clustering K-Means Dalam Pengelompokan Mahasiswa Studi Kasus (Prodi Manajemen Informatika)," *J. Minfo Polgan, 12(2), 777–783*, 2023, [Online]. Available: https://doi.org/10.33395/jmp.v12i2.12494
[21] A. Suharsimi, *Prosedur Penelitian Suatu Pensekatan Praktek*. Rineka Cipta, 1998.
[22] S. Arikunto, *Prosedur Penelitian: Suatu Pendekatan Praktik*. Jakarta: Rineka Cipta, 2010.
[23] S. Mukodimah and M. Muslihudin, "The Naïve Bayes Method as A Measurement Model Effectiveness of Online Learning," *Technol. Accept. Model. J. TAM*, vol. 13, no. 2, pp. 131–137, 2022.