**DANA BANTUAN**

**UNIVERSITAS NASIONAL**

**PROPOSAL STIMULUS PENELITIAN**



**PERBANDINGAN ALGORITMA K-MEANS DAN K-MEDOIDS UNTUK KLASTERISASI DATA MAHASISWA BERPRESTASI**

Peneliti Pengusul:

Frenda Farahdinna, S.Kom., M.Kom/ 0305119402

|  |  |
| --- | --- |
| 1. Moch Rafindra Aditias | / 227006516059 |
| 1. Khansa Zakiyah | / 227006516074 |

**FAKULTAS TEKNOLOGI KOMUNIKASI DAN INFORMATIKA**

**PROGRAM STUDI INFORMATIKA**

**UNIVERSITAS NASIONAL**

**AGUSTUS 2024**

# **HALAMAN PENGESAHAN**

# **RINGKASAN**

Pendidikan tinggi saat ini dihadapkan pada tantangan besar dalam mengelola data mahasiswa yang semakin melimpah, terutama dengan adanya penerimaan mahasiswa baru setiap semester dan penilaian prestasi akademis. Dalam konteks ini, pengolahan data menjadi krusial untuk mendapatkan informasi penting yang dapat memandu institusi pendidikan dalam meningkatkan kualitas pendidikan. Penelitian ini membahas perbandingan antara algoritma K-Means dan K-Medoids dalam mengklasterisasi data mahasiswa berprestasi. Kedua algoritma tersebut digunakan untuk mengelompokkan mahasiswa berdasarkan berbagai indikator prestasi seperti nilai akademik, partisipasi dalam kegiatan ekstrakurikuler, dan penghargaan yang diperoleh. K-Means merupakan algoritma yang populer karena kesederhanaan dan kecepatannya, di mana ia menggunakan centroid sebagai pusat kluster. Namun, K-Means memiliki kelemahan dalam menangani outlier yang dapat mempengaruhi hasil klasterisasi secara signifikan. Sebagai alternatif, K-Medoids menggunakan medoid, yaitu titik data aktual yang paling representatif dalam kluster, sehingga lebih tahan terhadap outlier dan memberikan hasil yang lebih stabil. Meskipun K-Medoids cenderung lebih lambat dalam proses komputasi dibandingkan K-Means, algoritma ini lebih cocok untuk dataset yang mengandung nilai ekstrem atau distribusi data yang tidak teratur. Penelitian ini menunjukkan bahwa pemilihan algoritma klasterisasi harus disesuaikan dengan karakteristik data dan tujuan analisis. K-Means lebih efisien untuk dataset yang besar dan homogen, sedangkan K-Medoids lebih unggul dalam menghasilkan kluster yang akurat dan robust terhadap outlier, terutama dalam konteks klasterisasi mahasiswa berprestasi.

**Kata kunci:** Machine Learning, Klasterisasi, K-Means, K-Medoids, Perbandingan

# **DAFTAR ISI**

[HALAMAN PENGESAHAN ii](#_Toc175052465)

[RINGKASAN iii](#_Toc175052466)

[DAFTAR ISI iv](#_Toc175052467)

[A. Latar Belakang 1](#_Toc175052468)

[B. Permasalahan 2](#_Toc175052469)

[C. Urgensi (Keutamaan) Penelitian 2](#_Toc175052470)

[D. RoadMap Penelitian 3](#_Toc175052471)

[E. Tujuan Penelitian 3](#_Toc175052472)

[F. Hipotesis 3](#_Toc175052473)

[BAB II KAJIAN PUSTAKA 4](#_Toc175052474)

[A. *Data Mining* 4](#_Toc175052475)

[B. *Clustering* 4](#_Toc175052476)

[C. CRISP DM 4](#_Toc175052477)

[D. *Machine Learning* 5](#_Toc175052478)

[E. K-Means 6](#_Toc175052479)

[F. K-Medoids 6](#_Toc175052480)

[G. *Mean Absolute Percentage Error* (MAPE) 6](#_Toc175052481)

[BAB III METODE PENELITIAN 7](#_Toc175052482)

[A. Waktu dan Lokasi Penelitian 7](#_Toc175052483)

[B. Bahan dan Alat 7](#_Toc175052484)

[C. Desain Penelitian 7](#_Toc175052485)

[D. Analisis Data 8](#_Toc175052486)

[BAB IV JADWAL DAN PEMBIAYAAN PENELITIAN 9](#_Toc175052487)

[A. Jadwal Penelitian 9](#_Toc175052488)

[B. Pembiayaan Penelitian 9](#_Toc175052489)

[DAFTAR PUSTAKA 10](#_Toc175052490)

**BAB I  
PENDAHULUAN**

## **Latar Belakang**

Dalam era digital dan perkembangan teknologi informasi, pengelolaan data menjadi krusial, terutama dalam konteks pendidikan tinggi. Penerimaan mahasiswa baru setiap semester dan prestasi akademis menghasilkan data yang berlimpah, berdasarkan data tersebut terdapat informasi yang dapat di ketahui dengan cara melakukan pengolahan data. Pengolahan data mahasiswa perlu dilakukan untuk mengetahui informasi penting berupa pengetahuan baru. Tugas besar bagi institusi pendidikan untuk mampu mengenali kemampuan akademis mahasiswanya dan mengarahkannya untuk mendapatkan mutu yang baik.

Nilai merupakan salah satu bagian yang terpenting dalam dunia akademisi, karena dengan nilai tersebut mahasiswa dapat dinyatakan lulus atau tidaknya. Selain itu, prestasi akademis mahasiswa juga termasuk faktor kunci dalam mengevaluasi efektivitas pendidikan tinggi. Identifikasi pola dan karakteristik kelompok mahasiswa berprestasi tinggi dapat memberikan wawasan berharga bagi lembaga pendidikan untuk meningkatkan strategi pembelajaran dan memberikan dukungan lebih baik kepada mahasiswa. Klasterisasi data mahasiswa berprestasi menggunakan metode k-means dapat membantu dalam pemetaan kelompok mahasiswa dengan karakteristik serupa. Hasil pengolahan informasi dapat di gunakan sebagai penunjang dalam pengambilan keputusan bagi institusi. Masalah yang dihadapi dalam klasterisasi adalah karena adanya data yang cukup besar sehingga membutuhkan waktu yang lama untuk langsung mengambil keputusan tanpa adanya pengolahan data.

Klasterisasi merupakan salah satu metode penting dalam analisis data yang bertujuan untuk mengelompokkan data ke dalam beberapa kelompok (kluster) berdasarkan kemiripan atau kedekatan karakteristiknya. Dalam konteks pendidikan, klasterisasi dapat digunakan untuk mengelompokkan nilai mahasiswa. Algoritma K-Means dan K-Medoids merupakan dua metode yang sering digunakan dalam klasterisasi. Keduanya memiliki prinsip dasar yang sama, yaitu membagi data ke dalam sejumlah kluster yang telah ditentukan sebelumnya. Namun, terdapat perbedaan mendasar dalam cara kedua algoritma ini memilih representasi pusat kluster, yang pada akhirnya mempengaruhi hasil klasterisasi. Dengan demikian, latar belakang ini menekankan pentingnya membandingkan metode klasterisasi data mahasiswa berprestasi menggunakan k-means yang berfokus pada pembagian objek ke dalam kelompok (klaster) berdasarkan kesamaan atribut tertentu. Hasil dari klasterisasi ini dapat membantu perguruan tinggi untuk mengidentifikasi tren dan pola dalam data mahasiswa berprestasi, yang pada gilirannya dapat digunakan untuk merancang program pendukung atau pengembangan kurikulum yang lebih efektif.

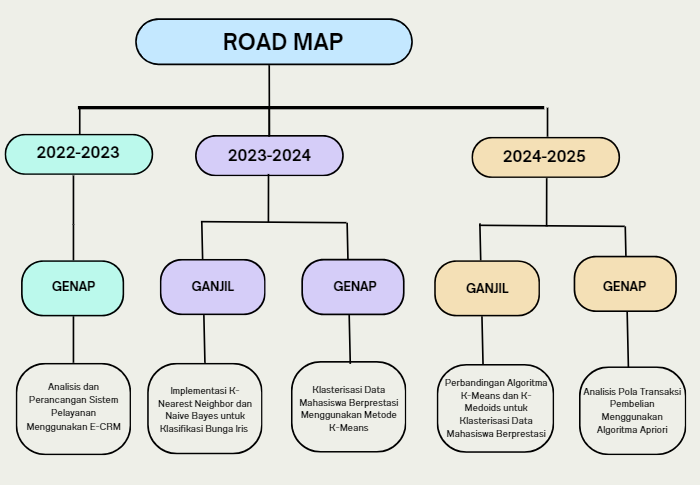
## **Permasalahan**

Berdasarkan latar belakang, maka didapat permasalahan yang ada yaitu kesulitan dalam menentukan algoritma yang tepat untuk klasterisasi mahasiswa berprestasi.

## **Urgensi (Keutamaan) Penelitian**

Pada penelitian ini terdapat urgensi atau keutamaan dibuatnya penelitian dengan tema klasterisasi data mahasiswa berprestasi adalah menentukan algoritma terbaik untuk membuat pemodelan yang dapat mengklasterisasi mahasiswa berprestasi berdasarkan nilai.

## **RoadMap Penelitian**



**Gambar 1.3** Roadmap Penelitian

## **Tujuan Penelitian**

Berdasarkan uraian tersebut maka tujuan penelitian ini adalah membuat, merancang dan menentukan algoritma terbaik untuk membuat pemodelan klasterisasi mahasiswa berprestasi berdasarkan nilai yang nantinya dapat digunakan sebagai bahan rekomendasi penerimaan beasiswa.

## **Hipotesis**

Hipotesis dari penelitian ini yaitu diduga dengan membuat dan merancang model dengan membandingkan algoritma K-Means dan K-Medoids untuk mengklasterisasi mahasiswa berprestasi.

# **BAB II KAJIAN PUSTAKA**

## ***Data Mining***

Data mining merupakan serangkaian proses untuk menggali nilai tambah berupa informasi yang selama ini tidak diketahui secara manual dari suatu basis data. Data mining mulai ada sejak 1990-an sebagai cara yang benar dan tepat untuk mengambil pola dan informasi yang digunakan untuk menemukan hubungan antara data untuk melakukan pengelompokkan ke dalam satu atau lebih cluster sehingga objek-objek yang berada dalam satu cluster akan mempunyai kesamaan yang tinggi antara satu dengan lainnya. Data mining merupakan bagian dari proses penemuan pengetahuan dari basis data Knowledge Discovery in Databases (Alkhairi, *et al* 2019).

Menurut Larose dalam Hasugian (2018), data mining dibagi menjadi beberapa kelompok berdasarkan tugas yang dapat dilakukan, yaitu sebagai berikut:

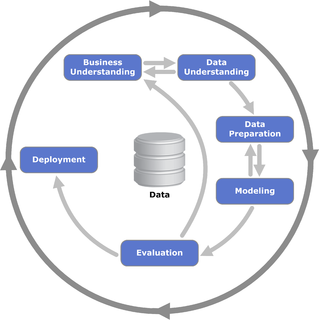
1. Deksripsi (*Description*)
2. Klasifikasi (*Classification*)
3. Estimasi (*Estimation*)
4. Peramalan (*Forecasting*)
5. Pengklusteran (*Clustering*)
6. Asosiasi (*Association*)

## ***Clustering***

*Clustering* adalah proses pembagian data ke dalam kelas atau cluster berdasarkan tingkat kesamaannya. Dalam clustering, data yang memiliki kesamaan dimasukkan ke dalam cluster yang sama, sedangkan data yang tidak memiliki kesamaan dimasukkan dalam cluster yang berbeda (Lestari, 2019).

## **CRISP DM**

Daimler Chrysler (Daimler-Benz), SPSS (ISL), NCR adalah 3 (tiga) penggagas data mining market yang menyusun CRISP-DM (*Cross Industry Standart Process for Data* *Mining*) sebagai suatu standarisasi data mining yang memiliki 6 (enam) fase seperti yang tertera pada Gambar 2.1. (Yuniarti, 2018).



Gambar 2.1 Proses Model CRISP-DM (Yuniarti, 2018)

Berdasarkan CRISP-DM, sebuah proyek data mining merupakan sebuah siklus hidup yang terdiri atas 6 (enam) tahap:

1. Pemahaman Bisnis (*Business Understanding*)
2. Pemahaman Data (*Data Understanding*)
3. Pengolahan Data (*Data Preparation*)
4. Pemodelan (*Modelling*)
5. Evaluasi (*Evaluation*)
6. Penerapan (*Deployment*)

## ***Machine Learning***

Menurut IBM, metode machine learning merupakan cabang dari kecerdasan buatan (Artificial Intelligence) dan ilmu komputer yang berfokus pada penggunaan data dan algoritma untuk meniru cara manusia belajar dan secara bertahap dapat meningkatkan akurasinya. Semakin bagus algoritma machine learning yang digunakan maka akan semakin baik pula keputusan yang keluar

Menurut Somvanshi dalam kutipan (Roihan et al., 2020) machine learning terbagi menjadi tiga kategori: *Supervised Learning*, *Unsupervised Learning*, dan *Reinforcement Learning*. Dalam metode *Supervised Learning* melakukan pemberian label pada dataset yang digunakan untuk mengklasifikasikan kelas yang tidak dikenal. Sedangkan *Unsupervised Learning* tidak ada kebutuhan untuk memberikan label dalam dataset dan hasilnya tidak mengidentifikasi contoh di kelas yang ditentukan. *Reinforcement Learning* berada diantara *Supervised Learning* dan *Unsupervised* *Learning* yang mana konsepnya harus menyelesaikan tujuan tanpa adanya pemberitahuan dari komputer jika tujuan telah tercapai.

## **K-Means**

K-means clustering meruapakan metode dalam machine learning yang tanpa pengawasan (unsupervised learning) yang secara efektif digunakan untuk memecah data set yang dikelompokan ke dalam grup k atau kluster k, di mana k mewakili jumlah klaster. Dalam pengelompokan k-means, setiap klaster diwakili oleh titik pusat (centroid) atau rata-rata yang dan dihitung sebagai nilai rata-rata data dalam klaster tersebut (Aggarwal, 2019)

Menurut Yahya K-means adalah metode analisa datapada Data Mining yang mana proses permodelannya dilakukan tanpa supervisi dan merupakansalah satu metode pengelompokan data secara partisi, Metode ini menimalisir perbedaan antar data di dalam satu kelompok dan memaksimalkan perbedaan dengan sebuah kelompok dengankelompok yang lain (Irfiani, 2018)

## **K-Medoids**

K-Medoids adalah teknik pengelompokan partisi untuk mengelompokkan kumpulan n objek ke dalam k cluster dengan memanfaatkan objek pada kumpulan objek guna mewakilkan sebuah cluster yang dinamakan medoid (Purba, Saifullah, & Dewi, 2019).

## ***Mean Absolute Percentage Error* (MAPE)**

*Mean Absolute Percentage Error* (MAPE) dihitung dengan menggunakan kesalahan absolut pada tiap periode dibagi dengan nilai observasi yang nyata untuk periode itu. Kemudian, merata-ratakan kesalahan persentase absolut tersebut Azmi, et al (2020).

# **BAB III METODE PENELITIAN**

## **Waktu dan Lokasi Penelitian**

Penelitian ini dilakukan paling lambat 6 bulan (semester ganjil 2024/2025) dengan perincian melakukan identifikasi masalah, studi literatur, pengumpulan data, pengolahan data, membuat model, pengujian, implementasi dan membuat laporan. Lokasi penelitian di kampus Universitas Nasional.

## **Bahan dan Alat**

1. Instrumen perangkat keras (*hardware*) berupa laptop Acer Swift 3 dengan spesifikasi sebagai berikut:

a) Processor: Core i5

b) Harddisk: 500 GB

c) RAM: 4 GB

d) Sistem Operasi: Windows 10 Pro 64 Bit

1. Instrumen perangkat lunak (*software*) yaitu:

a) Microsoft Office

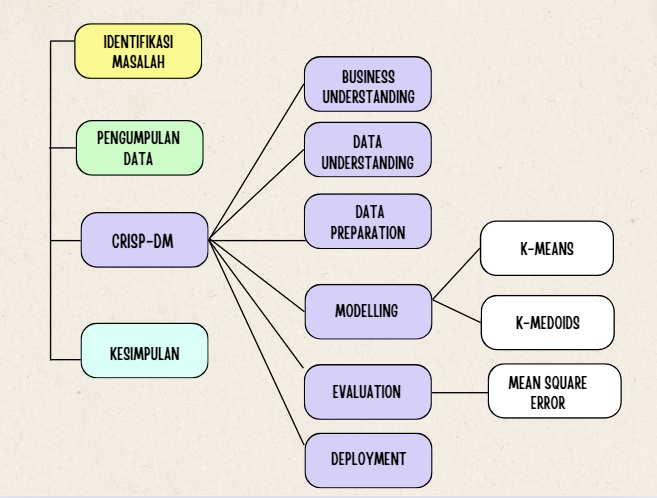
b) Python

c) Streamlit

d) Google Colab

## **Desain Penelitian**

Metodologi Penelitian yang digunakan dalam penelitian ini adalah adalah penelitian deskriptif dengan pendekatan kuantitatif dan membuat model klasterisasi mahasiswa berprestasi dengan perbandingan metode K-Means dan K-Medoids. Teknik perancangan model mengadopsi metode *Cross Industry Standart Process for Data Mining* (CRISP-DM) yang terdiri dari *Business Understanding*, *Data Understanding*, *Data Preparation*, *Modelling*, *Evaluation* dan *Deployment*. Hasil penelitian ini diharapkan mendapatkan nilai akurasi terbaik. Gambar 3.1 merupakan desain penelitiannya.



**Gambar 3.1** Desain Penelitian

## **Analisis Data**

Analisis data dilakukan pada dataset nilai mahasiswa. Penulis melakukan pengumpulan data yang berkaitan dengan pemodelan K-Means dan K-Medoids dan melakukan proses pengujian dari hasil klastrisasi untuk mengetahui tingkat akurasi pemodelan pada masing-masing algoritma.

# **BAB IV JADWAL DAN PEMBIAYAAN PENELITIAN**

## **Jadwal Penelitian**

**Tabel 4.1** Jadwal Penelitian

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Kegiatan** | **Bulan ke-1** | **Bulan ke-2** | **Bulan ke-3** | **Bulan ke-4** | **Bulan ke-5** | **Bulan ke-6** |
| Menyiapkan studi literatur |  |  |  |  |  |  |
| Pengumpulan data |  |  |  |  |  |  |
| Pemahaman dan pengolahan data |  |  |  |  |  |  |
| Pengembangan sistem |  |  |  |  |  |  |
| Pengujian sistem |  |  |  |  |  |  |
| Pembuatan laporan |  |  |  |  |  |  |

# **DAFTAR PUSTAKA**

Alkhairi, P., & Windarto, A. P. (2019). Penerapan K-Means Cluster pada Daerah Potensi Pertanian Karet Produktif di Sumatera Utara. Seminar Nasional Teknologi Komputer & Sains, 762–767.

D. Aggarwal and D. Sharma. (2019). Application Of Clustering For Student Result Analysis,” International Journal of Recent Technology and Engineering, vol. 7, no. 6, pp. 50–53

Hasugian, P. S. (2018) ‘Penerapan Data Mining Untuk Klasifikasi Produk Menggunakan Algortima K-Means (Studi Kasus: Toko Usaha Maju Barabai)’, Jurnal Mantik Penusa, 2(2), pp. 191–198.

Irfiani and S. S. Rani, (2018). “Algoritma K-Means Clustering untuk Menentukan Nilai Gizi Balita,” J. Sist. dan Teknol. Inf., vol. 6, no. 4, p. 161, doi: 10.26418/justin.v6i4.29024.

Lestari, Widhi (2019). Clustering Data Mahasiswa Menggunakan Algoritma K-Means Untuk Menunjang Strategi Promosi (Studi Kasus: STMIK Bina Bangsa Kendari). SIMKOM, Vol. 4, No. 2

Nahdliyah, Milla, et al. (2019). Metode K-Medoids Clustering Dengan Validasi Silhouette Index dan C-Index.Jurnal Gaussian.Vol 8 No.2

Purba, L., Saifullah, & Dewi, R. (2019). Pengelompokan Kasus Penyakit Aids Berdasarkan Provinsi Dengan Data Mining K-Medoids Clustering. KOMIK (Konferensi Nasional Teknologi Informasi Dan Komputer), 3(1), 687–694. <https://doi.org/10.30865/komik.v3i1.1679>

Roihan, A., Sunarya, P. A., & Rafika, A. S. (2020). Pemanfaatan Machine Learning dalam Berbagai Bidang : Review paper. 5(April), 75–82.

Yuniarti, A. S. (2018) Perbandingan Metode Weighted Moving Average Dengan Weighted Moving Average Berbasis Particle Swarm Optimization Untuk Prediksi Penjualan: Studi Kasus PT. Citra Mitra Nusantara.