

**PENERAPAN ALGORITMA K-MEANS UNTUK MENENTUKAN PENJURUSAN
MAHASISWA**

Fido Rizki, Asep Toyib Hidayat, Davit Irawan

PENERAPAN ALGORITMA K-MEANS UNTUK MENENTUKAN PENJURUSAN MAHASISWA

Fido Rizki¹, Asep Toyib Hidayat², Davit Irawan³
Informatika, Universitas Bina Insan, Lubuklinggau^{1,3}
Sistem Informasi, Universitas Bina Insan, Lubuklinggau¹
Email : fidorizki@univbinainsan.ac.id,
asep_toyib_hidayat@univbinainsan.ac.id,davit_irawan@univbinainsan.ac.id

Abstrak

Salah satu cara yang cukup efektif dapat digunakan untuk mendapatkan sebuah informasi baru dengan menggunakan klasifikasi data, klasifikasi data merupakan sebuah teknik yang dilakukan secara umum dengan memanfaatkan proses data mining, algoritma yang digunakan dalam penelitian ini adalah menggunakan Algoritma K-Means. K-Means merupakan algoritma clustering data berulang-ulang, dengan menetapkan nilai-nilai cluster (K) secara random, untuk sementara nilai tersebut menjadi pusat dari cluster atau biasa disebut dengan centroid, mean atau "means. Saat ini para mahasiswa di Universitas Bina Insan Lubuklinggau dihadapi pada penentuan penjurusan atau konsentrasi, penentuan penjurusan atau konsentrasi saat ini belum dilakukan klusterisasi berdasarkan riwayat mata kuliah yang telah diambil sebelumnya, berdasarkan masalah tersebut penulis akan melakukan pengelompokan dalam pemilihan penentuan penjurusan atau konsentrasi. Hasil dari penelitian ini didapatkan bahwa dengan menggunakan penerapan metode algoritma k-means dapat mengetahui pengelompokan dan kemampuan mahasiswa dalam penentuan penjurusan atau konsentrasi berdasarkan rata-rata nilai yang dominan dan IPK.

Kata Kunci : Algoritma; K-Means;; Mata Kuliah

Abstract

One quite effective way can be used to obtain new information by using data classification, data classification is a technique that is carried out in general by utilizing the data mining process, the algorithm used in this study is using the K-Means Algorithm. K-Means is an iterative data clustering algorithm, by assigning cluster values (K) randomly, temporarily the value becomes the center of the Penjurusan cluster or commonly referred to as centroid, mean or "means." Currently students at Bina Insan Lubuklinggau University are faced with determining majors or concentrations. Currently, clustering has not been carried out based on the history of courses taken before. Based on this problem, the author will group the selection of majors or concentrations. The results of this research found that using the application of the k-means algorithm method could find out the grouping and ability of students in determining majors or concentrations based on the dominant average value and GPA.

Keywords: Algorithm; K-Means; direction; Subject

1 Pendahuluan

Berbagai institusi, universitas, perusahaan atau organisasi pasti memiliki data yang tersimpan diserver masing-masing dan jumlahnya berlimpah. Berlimpahnya data ini merupakan akumulasi data transaksi yang terekam bertahun-tahun. Dengan data tersebut institusi, universitas, perusahaan atau organisasi bisa digunakan untuk mendapatkan informasi baru. Salah satu cara efektif dapat digunakan untuk mendapatkan informasi baru dengan menggunakan klasifikasi data, teknik yang dilakukan untuk klasifikasi data secara umum yaitu memanfaatkan proses data *mining* dengan menggunakan *Algoritma K-Means*. *K-Means* merupakan algoritma clustering data berulang-ulang, dengan menetapkan nilai-nilai cluster (K) secara random, untuk sementara nilai tersebut menjadi pisat dari cluster atau biasa disebut dengan *centroid, mean atau "means"*[1].

Universitas Bina Insan Lubuklinggau merupakan salah satu dari beberapa kampus yang ada di Kota Lubuklinggau, Universitas Bina Insan merupakan lembaga pendidikan yang didirikan oleh yayasan dwi tunggal, Universitas Bina Insan Lubuklinggau memiliki Sembilan program studi yaitu Magister Manajemen, Manajemen, Akuntansi, Hukum, Informatika, Sistem Informasi, Rekayasa Sistem Komputer, Agroteknologi dan Perikanan

Program studi informatika mempunyai konsentrasi keahlian seperti Software Engineering dan Sistem Inteligent yang sesuai dengan kurikulum program studi informatika di Universitas Bina Insan Lubuklinggau, proses penentuan Penjurusan saat ini belum dilakukan dengan alat bantu khusus yang dapat melakukan pemilihan penjurusan konsentrasi pada prodi informatika, Sistem penjurusan secara langsung dapat ditentukan berdasarkan minat mahasiswa yang telah memilih mata kuliah konsentersasi keahlian pada halaman Kartu Rencana Studi yang terdapat di website Sisfo Universitas Bina Insan. Hasil dari keseluruhan Kartu Rencana Studi yang telah dipilih setiap mahasiswa untuk menentukan

penjurusan konsentersasi, akan dapat diketahui Penjurusan yang paling banyak diminati oleh mahasiswa. Proses penentuan Penjurusan konsentersasi seperti ini tidak efektif karena belum tentu mahasiswa yang memilih konsentrasi Software Engineering mempunyai nilai mata kuliah pendukung konsentrasi Software Engineering yang baik dan juga sebaliknya mahasiswa yang memilih konsentrasi Sistem Inteligent belum tentu mempunyai nilai mata kuliah pendukung konsentrasi Sistem Inteligent yang baik juga.

Berdasarkan permasalahan yang ada, maka penulis akan menerapkan algoritma k-means untuk dapat membantu mengelompokkan pemilihan Penjurusan berdasarkan sampel pada nilai Kartu Hasil Studi mahasiswa semester empat dan lima pada angkatan 2020 dan minat setiap mahasiswa. Beberapa kelebihan dari k-means yang akan dapat memaksimalkan populasi penentuan Penjurusan yaitu dalam prosesnya cepat dan nilai cukup efisien untuk pengelompokan dalam jumlah data mahasiswa yang banyak.

2 Tinjauan Pustaka

2.1 Penerapan

Penerapan (Implementasi) adalah bermuara pada aktivitas, aksi, tindakan, atau adanya mekanisme suatu sistem. Implementasi bukan sekedar aktivitas, tetapi suatu kegiatan yang terencana dan untuk mencapai tujuan kegiatan [2].

Menurut definisi lain, penerapan adalah proses, cara, perbuatan menerapkan. Dapat disimpulkan bahwa penerapan adalah suatu perbuatan mempraktekkan teori, metode, dan hal lain untuk mencapai tujuan tertentu dan untuk suatu kepentingan yang diinginkan oleh kelompok atau golongan yang telah terencana dan tersusun sebelumnya[3].

2.2 Data Mining

Data Mining merupakan proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk

mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terakit dari berbagai database besar/Data Warehouse[1]

Data Mining mendefinisikan *data mining* adalah serangkaian proses untuk menggali nilai tambah berupa informasi yang selama ini tidak diketahui secara manual dari suatu basisdata. Informasi yang dihasilkan diperoleh dengan cara mengekstraksi dan mengenali pola yang menarik dari suatu data[4].

Data mining merupakan bidang dari beberapa bidang keilmuan yang menyatukan teknik dari pembelajaran mesin, pengenalan pola, statistik, database, dan visualisasi untuk pengenalan permasalahan pengambilan informasi dari database yang besar.

Beberapa tahapan proses *data mining* antara lain yaitu :

a. *Data Cleaning*

Pada proses *data cleaning* perlu dilakukan karena terdapat beberapa data yang mempunyai isian tidak sempurna seperti ada sel yang tidak punya nilai, data mengandung *error*, ada *outlier* serta ada inkonsistensi pada isian sel. Sel yang tidak mempunyai nilai bisa diabaikan, diisi dengan rata-rata data, diisi dengan rata-rata kelompok, dan estimasi (perkiraan).

b. Penentuan Sampel

Dalam penelitian penentuan sampel merupakan kebutuhan utama untuk memperoleh data yang menggambarkan realitas yang akan dijadikan fokus dalam penelitian dengan menggunakan rumus *slovin* sebagai berikut :

$$n = \frac{N}{1+Ne^2}$$

.....
.....

(1)

Keterangan :

n = Jumlah *sample*

N = Jumlah populasi

*e*² = Toleransi ketidaktelitian

3 Metodologi Penelitian

3.1 Metode Pengumpulan Data

Metode penelitian dibutuhkan untuk mengetahui kinerja sistem sehingga sistem perangkat lunak menjadi lebih baik dan memiliki data yang akurat. Metode-metode penelitian tersebut adalah sebagai berikut :

a. Observasi

Penulis melakukan Observasi pengamatan langsung ke Universitas Bina Insan Lubuklinggau pada jurusan Informatika untuk memperoleh informasi yang diperlukan Menentukan konsentrasi atau penjurusan. Hasil dari observasi ini penulis mendapatkan data mahasiswa, data mata kuliah, dan data Kartu Hasil Studi(KHS)

b. Wawancara

Penulis melakukan wawancara, merupakan teknik pengumpulan data dengan cara mengadakan tanya jawab kepada Kaprodi Informatika Universitas Bina Insan Lubuklinggau.

c. Literatur

Penulis mencari dan mengumpulkan materi-materi yang berkaitan dengan penyusunan penelitian ini, seperti buku-buku, jurnal, dan skripsi yang telah selesai dikerjakan. Materi-materi yang tersedia dapat digunakan sebagai bahan pembelajaran dan refrensi bagi pembuatan penelitian ini.

2.3 Algoritma K-Means

K-Means merupakan algoritma clustering data berulang-ulang, dengan menetapkan nilai-nilai cluster (*K*) secara random, untuk sementara nilai tersebut menjadi pisat dari cluster atau biasa disebut dengan *centroid*, *mean* atau "*means*" [1]

Berikut langkah-langkah dari penerapan algoritma *K-Means* Untuk dapat pemilihan rekomendasi tema tugas akhir pada Mahasiswa prodi Informatika Universitas Bina Insan Lubuklinggau :

a. Menentukan banyak *k-cluster* yang ingin dibentuk.

Dalam menentukan jumlah *cluster* berasal dari banyak datajurusan pada Prodi Teknik Informatika yang ada akan dibuat menjadi 2 *cluster* yaitu 1. Software

Engineering dan 2. Sistem Intelegnt, yang akan diidentifikasi sebagai *Cluster A*, *Cluster B* agar lebih mudah dalam proses perhitungan.

- b. Membangkitkan nilai random untuk pusat *cluster* awal(*centroid*) sebanyak *k-cluster*. Dalam menghitung nilai *centroid* awal akan digunakan nilai *mean* dari seluruh *Cluster*. Nilai mata kuliah pada masing-masing kolom konsentrasi diambil sampel 10 *dataset*.

- c. Menghitung *dataset* konsentrasi ke 1
Menghitung jarak setiap data *input* terhadap masing-masing *centroid* kosentrasi menggunakan rumus jarak (*Eucledian Distance*) hingga akan ditemukan jarak yang paling dekat dari setiap data dengan *centroid* konsentrasi. Berikut persamaan *Eucledian Distance*:

$$d(x_i, \mu_i) = \sqrt{(x_i - \mu_i)^2}$$

.....
.....
(2)

- d. Mengklasifikasikan setiap data konsentrasi berdasarkan kedekatannya dengan *centroid* konsentrasikan diidentifikasi sebagai jarak terkecil.

- e. Melakukan update nilai *centroid* konsentrasi. Nilai *centroid* konsentrasi baru diperoleh dari rata-rata *cluster* konsentrasi dengan menggunakan rumus:

$$C_k = \frac{1}{n_k} \sum d_i$$

.....
.....
(3)

Keterangan :

n_k = jumlah data dalam setiap *cluster* konsentrasi

d_i = jumlah nilai jarak masuk dalam setiap *cluster* konsentrasi

4 Analisa dan Perancangan

4.1 Analisa Data

Analisis sistem pada penelitian ini berperan sebagai penguraian dari sistem informasi utuh kedalam bagian-bagian komponennya dengan maksud untuk mengidentifikasi dan mengevaluasi

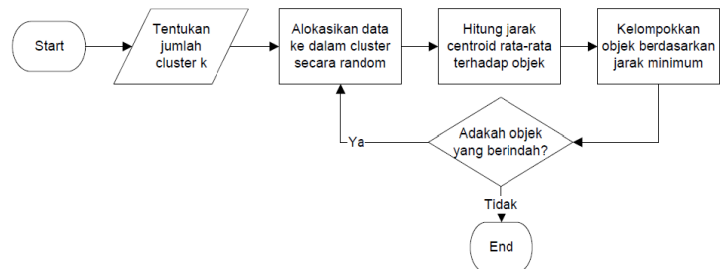
permasalahan-permasalahan, kesempatan-kesempatan, hambatan-hambatan yang terjadi dan kebutuhan yang diharapkan dapat terpenuhi.

Dalam analisis ini, penulis mengumpulkan data-data yang diperlukan dalam penyusunan penelitian ini, seperti observasi ke Universitas Bina Insan Lubuklinggau pada jurusan Informatika, melakukan wawancara dengan Kaprodi Informatika pada ke Universitas Bina Insan Lubuklinggau, serta studi pustaka terhadap judul penulis.

1. Menentukan Variabel, disini penulis menggunakan nilai rata-rata Mata Kuliah pada semester I dan V sesuai dengan hubungannya pada tema untuk dijadikan variabel.
2. Menentukan Sampel, berdasarkan jumlah mahasiswa reguler jurusan Informatika pada angkatan 2020.
3. Inisialisasi Data Nilai, data yang sudah diubah dikelompokan berdasarkan mata kuliah yang dominan kemasing-masing konsentrasi dan dicari rata-rata nilainya pada setiap mata kuliah.

4.2 Metode Clustering K-Means

Menentukan metode untuk penelitian penulis menggunakan metode *clustering k-means*. Beberapa kelebihan dari *k-means* yaitu dalam prosesnya cepat dan dinilai cukup efisien untuk proses pengelompokan dalam jumlah data yang banyak. Selain itu, dalam iterasinya, algoritma ini akan berhenti dalam kondisi *optimum local*[1]. Secara umum urutan proses *clustering* dengan algoritma *k-means* ditunjukkan pada gambar 4.1



Gambar 1 Flowchart Proses K-Means

Beberapa metode yang digunakan untuk menghitung selisih jarak adalah *Euclidean*

distance. Adapun persamaanya adalah sebagai berikut:

$$d(x,y) = \frac{||x - y||^2}{\sqrt{\sum_{i=1}^n (xi - yi)^2}}$$

Keterangan :

$d(x,y)$ = Jarak objek antara Xi dan Yi

n = Dimensi data

Xi = Koordinat dari objek Xi pada dimensi i

Yi = Koordinat dari objek Yi pada dimensi i

4.3 Proses Clustering

Tahap ini akan diterapkan metode *K-Means* untuk mengelompokkan data. Hasil pengelompokkan ini kemudian akan digunakan untuk pertimbangan Menentukan Penjurusan konsentrasi Mahasiswa Program Studi Informatika. Adapun algoritma *K-Means Clustering* pada Penjurusan konsentrasi Mahasiswa Program Studi Informatika adalah sebagai berikut:

1. Jumlah *cluster* yang dibentuk sebagai nilai k adalah dua ($k = 2$), yaitu Sistem Intelegent dan Software Engineering.
2. Membangkitkan nilai *random* untuk pusat *cluster* awal (*centroid*) sebanyak 2 dari data sample. *Centroid* kriteria 1 adalah rata-rata nilai mata kuliah yang dominan ke penjurusan konsentrasi *Sistem Intelegent* dan *centroid* kriteria 2 adalah rata-rata nilai mata kuliah yang dominan ke penjurusan konsentrasi *Software Engineering* dan *centroid* kriteria 3 adalah *IPK Semester V*.
3. Dalam menentukan penjurusan konsentrasi disini penulis menentukan paramater diantaranya :
 - Sistem Intelegent rata- rata = 3,00 – 4,00
 - Softwate Engineering rata - rata = 2,00 – 2,90

4.4 Hasil Pengujian

Selanjutnya akan digunakan algoritma klasifikasi *K-Means* untuk mengelompokkan data yang ada. Data yang ada akan dikelompokkan menjadi 4 kelompok. Adapun langkah dari pengelompokkan data adalah sebagai berikut :

1. Tentukan pusat cluster secara acak, misalkan kita tentukan

Tabel 1. Nilai Cluster Awal

	C1	C2
IPK	3,00	3,50
Rata-rata nilai dominan Sistem Intelegant	2,50	2,80
Rata-rata nilai dominan Software Engineering	2,80	3,00

2. Menghitung jarak setiap data yang ada terhadap setiap pusat cluster menggunakan rumus jarak *Eucledian Distance* hingga ditemukan jarak paling dekat dari setiap data dengan *centroid*. Berikut adalah persamaan *Eucledian Distance*:

$$d(x_i, \mu_j) = \sqrt{(x_{ia} - \mu_{ja})^2 + (x_{ib} - \mu_{jb})^2}$$

Keterangan :

x_i : data mahasiswa

μ_j : *centroid* pada *cluster* ke- j

x_{ia} : rata-rata nilai penjurusan *Sistem Intelegant*

x_{ib} : rata-rata nilai penjurusan *Software Engineering*

μ_{ja} : nilai kriteria 1 dari *centroid cluster* ke- j adalah data rata-rata nilai penjurusan *Sistem Intelegant*

μ_{jb} : nilai kriteria 2 dari *centroid cluster* ke- j adalah nilai

random data rata-rata nilai *Software Engineering* jarak data mahasiswa pertama dengan pusat cluster pertama adalah :

$$D1 = \sqrt{(3,22 - 3,00)^2 + (3,00 - 2,50)^2 + (3,29 - 2,80)^2} = 0,73$$

jarak data mahasiswa pertama dengan pusat cluster kedua adalah :

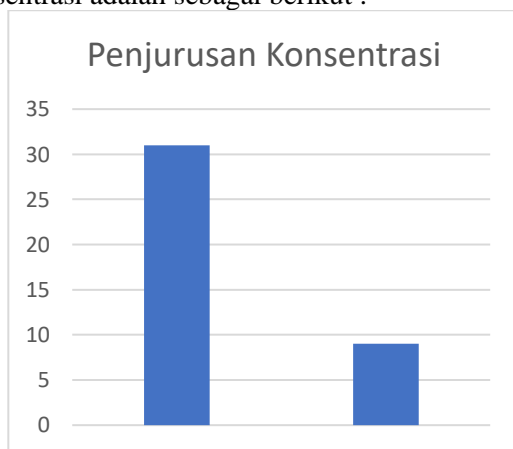
$$D2 = \frac{\sqrt{(3,22 - 3,50)^2 + (3,00 - 2,80)^2}}{(3,29 - 3,00)^2} = 0,45$$

4.5 Klasifikasi Hasil Clustering

Setelah proses *clustering*, tahap selanjutnya adalah proses klasifikasi. Di sini akan ditentukan *cluster* Jurusan Mahasiswa Program Studi Informatika apakah *Sistem Inteligent* dan *Software Engineering*, dan hasil akhir yang diperoleh adalah 2 cluster :

- Cluster pertama memiliki pusat (2,92 ; 2,06 ; 2,32) yang dapat diartikan bahwa mahasiswa tersebut dapat memilih konsentrasi Sistem intelgenet.
- Cluster kedua memiliki pusat (3,34 ; 3,02 ; 3,23) yang dapat diartikan bahwa mahasiswa tersebut dapat memilih konsentrasi software engineering.

Adapun grafik Penentuan Jurusan Konsentrasi adalah sebagai berikut :



Gambar 2 Grafik Penjurusan Konsentrasi

5 Kesimpulan

Berdasarkan hasil penelitian ini, dapat disimpulkan beberapa hal mengenai penerapan algoritma k-means untuk menentukan pemilihan penjurusan atau konsentrasi mahasiswa program studi informatika Universitas Bina Insan Lubuklinggau. Dengan menggunakan penerapan metode algoritma k-means dapat mengetahui

minat dan kemampuan mahasiswa berdasarkan rata-rata nilai yang dominan terhadap masing-masing konsentrasi penjurusan.

1. Setelah dilakukannya penerapan metode algoritma k-means ini dapat membantu mahasiswa dan program studi informatika dalam penentuan konsentrasi jurusan
2. Dari data yang diolah dapat dilihat dari 40 mahasiswa 31 orang layak di Software Engineering dan 9 orang di Sistem Inteligent.

6 Saran

Berdasarkan kesimpulan penelitian, maka penulis merekomendasikan berupa saran-saran sebagai berikut :

1. Pada penelitian ini dapat dikembangkan menjadi aplikasi maupun dibuat menggunakan bahasa pemrograman lainnya.
2. Untuk pengembangan penelitian ini dapat dikembangkan menggunakan perbandingan algoritma lainnya.

7 Daftar Pustaka

- [1] R. T. Vlandari, *Data Mining Teori dan Aplikasi Rapidminer*. Yogyakarta: Gava Media, 2017.
- [2] Y. P. Ario, "Implementasi Peraturan Presiden Nomor 26 Tahun 2009 Tentang Penerapan E-KTP Berbasis Nomor Induk Kependudukan Di Kecamatan Samarinda Utara Kota Samarinda." 2014.
- [3] Kbbi.web.id, "Definisi Penerapan," 2017. [Online]. Available: <https://kbbi.web.id/terap-2>.
- [4] W. R. Danang Aditya Nugraha, "Sistem Pendukung Keputusan Penjurusan Di Sma Menggunakan Metode Neural Network

- Backpropagation,” pp. 1–6, 2014.
- [5] M. N. Desta Putu Wikarta, “Penerapan Konseling Trait And Factor Pada Siswa Yang Mengalami Kesulitan Memilih Program Penjurusan Bahasa,” pp. 1–10, 2013.
- [6] S. Arikunto, *Prosedur Penelitian Suatu Pendekatan Praktik*. Jakarta: Rineka Cipta, 2010.
- [7] Sugiyono, *Statistika untuk Penelitan*. Bandung: Alfabeta, 2013.