JEPIN

Vol. 5

No. 1

April 2019

*(Jurnal Edukasi dan Penelitian Informatika)*

ISSN(e): 2548-9364 / ISSN(p) : 2460-0741

Klasterisasi Menggunakan Algoritma *K-Means Clustering* untuk Memprediksi Waktu Kelulusan Mahasiswa

Hendro Priyatman#1, Fahmi Sajid\*2, Dannis Haldivany\*3

*#Jurusan Teknik Elektro, Universitas Tanjungpura Jl. Prof. Hadary Nawawi, Pontianak*

[1hendro.priyatman@ee.untan.ac.id](mailto:1hendro.priyatman@ee.untan.ac.id)

*\*Jurusan Informatika, Universitas Tanjungpura Jl. Prof. Hadary Nawawi, Pontianak*

[2](mailto:2fahmisajid@student.untan.ac.id)[fahmisajid@student.untan.ac.id](mailto:fahmisajid@student.untan.ac.id) [3](mailto:3dpriyatman@student.untan.ac.id)[dpriyatman@student.untan.ac.id](mailto:dpriyatman@student.untan.ac.id)

***Abstrak*— Prediksi waktu kelulusan mahasiswa bertujuan memberikan sarana untuk mengetahui perkiraan waktu lulus mahasiswa dengan melihat mahasiswa-mahasiswa mana saja yang masuk kedalam suatu *cluster* tertentu berdasarkan parameter IPK dan kehadiran. Penelitian ini menerapkan algoritma *K-Means*.Algoritma *K-Means* merupakan model *cendroid*. Model *cendroid* adalah model yang menggunakan *cendroid* untuk membuat *cluster*. *K- Means* merupakan salah satu metode *cluster*ing non hirarki yang berusaha mempartisi data yang ada ke dalam bentuk satu atau lebih *cluster*. Metode ini mempartisi data ke dalam *cluster* sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu *cluster* yang sama dan data yang mempunyai karateristik yang berbeda dikelompokan ke dalam *cluster* yang lain. Penelitian ini diharapkan dapat membantu pihak kampus maupun mahasiswa untuk memprediksi tingkat kelulusan tepat waktu dan untuk meningkatkan reputasi bagi pihak kampus itu sendiri dan kelulusan tepat waktu bagi mahasiswa supaya kelulusan mereka tidak terlambat, selain itu pihak kampus bisa melakukan hal-hal yang perlu dilakukan apabila mereka di prediksi lulus tidak tepat waktu seperti dengan melakukan bimbingan dan hal lainnya.**

***Kata kunci***— **prediksi waktu, kelulusan mahasiswa, *cluster*, Algoritma K-Means**

1. PENDAHULUAN

Kelulusan mahasiswa adalah satu *moment* yang sangat dinanti-nanti oleh setiap mahasiswa dan lulus tepat pada waktunya adalah suatu hal yang diharapkan.Selain itu setiap kampus juga pasti menginginkan mahasiswa nya untuk lulus tepat waktu, karena hal ini bermanfaat bagi pihak kampus untuk meningkatkan reputasi dan akreditasinya. Tapi pada kenyataannya para mahasiswa tidak bisa memprediksinya, mereka hanya mengikuti alur

perkuliahan, sehingga akhirnya waktu kelulusan mereka menjadi terlambat. hal ini dapat diketahui dari tidak seimbangnya presentase kelulusan mahasiswa dari tahun ke tahun ataupun perbandingan jumlah yang sudah lulus dan belum lulus pada beberapa angkatan.

Salah satu metode yang digunakan untuk memprediksi waktu kelulusan mahasiswa ini adalah dengan menggunakan *metode K-Means Clustering,* yaitu dengan mengelompokan ‗n‘ buah objek kedalam ‗k‘ kelas berdasarkan jarak nya dengan pusat *cluster* [1]. Analisis *cluster* merupakan teknik multivariat yang mempunyai tujuan utama untuk mengelompokkan objek-objek berdasarkan karakteristik yang dimilikinya. Analisis *cluster* mengklasifikasi objek sehingga setiap objek yang paling dekat kesamaannya dengan objek lain berada dalam *cluster* yang sama. Analisis *cluster* merupakan salah satu alat analisis yang berguna sebagai peringkas data. Dalam meringkas data ini dapat dilakukan dengan cara mengelompokkan objek-objek berdasarkan kesamaan karakteristik tertentu diantara objek-objek yang hendak diteliti.

Beberapa penelitian sebelumnya, telah berhasil melakukan pengelompokan dengan menggunakan algoritma K-Means, diantaranya adalah : Fauzan dkk. [2] yang mengimplementasikan algoritma K-Means untuk menklaster kinerja akademik mahasiswa, dimana hasil clustering menunjukkan bahwa pendapatan orang tua tidak mempengaruhi tingkat kinerja akademik mahasiswa. Sulistyowati dkk. [3] menganalisis segmentasi konsumen menggunakan algoritma *k-means clustering*, dari hasil perbandingan aplikasi dengan menggunakan software SPSS diperoleh hasil selisih dari masing-masing *cluster* dengan jumlah rata-rata 7%. Hal ini disebabkan karena centroid awal yang dipilih secara acak. Johan [4]

menggunakan algoritma Kmeans untuk menentukan strategi marketing president university dengan cara mengolah data-data yang telah didapatkan dari mahasiswa yang telah lulus seperti nama mahasiswa, kota asal, jurusan yang diambil dan yang terkahir adalah nilai IPK, kemudian dikelompokan kedalam beberapa cluster berdasarkan kemiripan dari data-data tersebut. Bastian [5] mengkaji bagaimana penggunaan Algoritma *K-Means Cluster Analysis* dalam studi kasus penyakit menular manusia pada suatu objek berdasarkan set variabel yang dibentuk per kecamatan tiap puskesmas yang jumlahnya ada 32 Kantor Puskesmas di Kabupaten Majalengka. Rendy dkk. [6] menganalisis perbandingan metode clustering menggunakan metode single Linkage dan K- Means pada pengelompokan dokumen, dari hasil perbandingan dapat disimpulkan metode single likage memiliki perfomansi yang lebih baik dibandingkan dengan metode K-means. Khotimah [7] menggunakan algoritma Kmeans untuk mengelompokkan surat dalam Al-Qur'an berdasarkan jumlah ayat, jumlah ruku‘ dan lama waktu membaca surat. Dari percobaan yang dilakukan, nilai optimal diperoleh pada iterasi ke-14 dengan nilai k sebanyak 4 cluster. Irfiani dan Rani[8] menggunakan algoritma Kmeans untuk membantu para kader Posyandu dan orang tua balita dalam penanganan dini kondisi gizi balita dengan kategori obesitas, gizi lebih, gizi baik, gizi kurang dan gizi buruk. Pengklasteran mahasiswa untuk kepentingan akademis juga pernah dilakukan oleh peneliti-peneliti sebelumnya [9]-[12].

Penelitian ini fokus kepada penggunaan algoritma K- Means untuk memprediski kelulusan mahasiswa. Diharapkan dengan adanya prediksi waktu kelulusan mahasiswa ini, apabila mereka diprediksi lulus tidak tepat pada waktunya, pihak kampus dapat melakukan kegiatan- kegiatan yang dapat meminimisasi jumlah mahasiswa yang diprediksi lulus tidak tepat waktu, seperti dengan mengadakan bimbingan belajar, mengadakan semester pendek ataupun kegiatan-kegiatan bermanfaat yang dapat membantu meningkatkan prestasi belajar mahasiswa.

Pada dasarnya penelitian ini mempunyai tujuan untuk :

(1) mengklasterisasi mahasiswa dengan algoritma *K- Means*, (2) mengklaster data-data mahasiswa berdasarkan parameter yang ada yaitu IPK dan kehadiran, dan (3) mengetahui kemampuan algoritma *K-Means Clustering* dalam memprediksi waktu kelulusan mahasiswa serta memberikan sarana untuk mengetahui perkiraan waktu lulus.

1. METODOLOGI
2. *K-Means Clustering*

Pada artikel yang diterbitkan Elsevier ―Data clustering: 50 years beyond K-means‖ [13] diungkapkan bahwa pengorganisasian data ke dalam suatu kluster merupakan suatu model yang paling mendasar untuk pemahaman dan pembelajaran. Analisis kluster adalah studi formal untuk mengelompokkan, atau *clustering* benda-benda sesuai dengan karakteristik yang diukur

berdasarkan kemiripan satu sama lain. *Clustering* adalah pengelompokan menggunakan teknik *unsupervised learning* dimana tidak diperlukan *fase learning* serta tidak menggunakan pelabelan pada setiap kelompok.

Metode *clustering* mempartisi data ke dalam kelompok sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu *cluster* yang sama (Refaat, 2007). Tujuan dari *clustering* ini adalah untuk meminimalisasi fungsi tujuan yang ditetapkan dalam proses *clustering*, yang umumnya berusaha meminimalisasi variasi dalam suatu *cluster* dan memaksimalisasi variasi antar *cluster* [14].

*Clustering* atau analisis *cluster* adalah proses pembentukan kelompok data (*cluster*) dari himpunan data yang tidak diketahui kelompok-kelompok atau kelas- kelasnya dan proses menentukan data-data termasuk ke dalam *cluster* yang mana. *Clustering* merupakan proses untuk mengetahui kelas-kelas taksonomi atau batryologi, atau analisis topologi dari data-data yang ada. Dilihat dari kacamata data mining, clustering bukanlah proses klarifikasi. Karena dalam proses klarifikasi, data dikelompokkan ke dalam kelas-kelas yang telah diketahui sebelumnya.

1. *Analisis Kebutuhan dan Perancangan*

## Analisis Kebutuhan

Pada analisis kebutuhan ini membahas beberapa kebutuhan atau persyaratan yang terkait dalam penganalisisan predikisi kelulusan mahasiswa. Berdasarkan data yang diperoleh hasil analisis kebutuhan tersebut adalah sistem yang akan dibangun membutuhkan input diantaranya NIM, nama, jenis kelamin, jurusan, IPK dan kehadiran sebagai data akademik atau data mahasiswa. Dan untuk kebutuhan proses, proses yang digunakan untuk mengolah data input adalah teknik *clustering* menggunakan algoritma *k-means.* Yang terakhir adalah kebutuhan *output* yang diharapkan adalah analisis dari hasil *clustering* sehingga dapat diketahui informasi yang dibutuhkan. Berikut adalah istilah-istilah didalam algoritma *k-means clustering :*

1. Cluster : Cluster adalah kelompok atau grup
2. Cendroid : Cendroid adalah titik tengah atau pusat
3. Iterasi : Iterasi adalah pengulangan

## Analisis Data

Pada tahap ini dilakukan analisis data dari hasil pengclusteran menggunakan algoritma *k-means*. Data-data yang dibutuhkan untuk membangun aplikasi ini berupa parameter yang kita ambil sebagai acuan prediksi kelulusan mahasiswa yaitu IPK dan kehadiran. Dari dua parameter tersebut data diubah menjadi tipe data yang memudahkan untuk diproses. Tingkat kelulusan diukur dari IPK dan kehadiran. Misalkan IPK dikategorikan berdasarkan predikat kelulusan yang diatur dalam kaidah akademik yang berbunyi ―predikat kelulusan program sarjana dan program diploma‖ adalah sebagi berikut :

1. IPK memuaskan dengan IPK 2.00 – 2.75

* Masukkan data login
* Masukan data mahasiswa

- Data random, Pusat clus-ter

* Data pengujian ( prediksi kelulusan )

Aplikasi Data Mining Memprediksi Kelulusan Mahasiswa

Pengguna

* Informasi data hasil prediksi
* Informasi data tingkat keberhasilan

-

- Informasi data hasil kelompok cluster

1. IPK sangat memuaskan dengan IPK 2.76 – 3.35
2. IPK dengan pujian ( cumlaude ) dengan IPK 3.51 – 4.00
3. HASIL DAN PEMBAHASAN

Tabel I memperlihatkan pengkategorian data kelulusan berdasarkan kehadiran :

TABEL I KATEGORI KEHADIRAN

|  |  |  |
| --- | --- | --- |
| **Keterangan Nilai Kehadiran** | | |
| 86-100 % | A (5) | Rajin Sekali |
| 71-85 % | B (4) | Rajin |
| 56-70 % | C (3) | Cukup Rajin |
| 41-55 % | D (2) | Malas |
| <40 % | E (1) | Sangat Malas |

1. *Prosedur Kerja Sistem*

Mulai

Banyaknya Cluster

Menentukan centroid atau pusat

Menghitung jarak objek ke centroid atau pusat

Ya

Centroid berubah ?

Mengelompokkan data atau objek berdasarkan jarak terdekat

Tidak

Selesai

Gambar 1. Flowchart Algoritma Metode K-Means

Gambar 1 memperlihatkan flowchar dari algoritma K- Means, sedangkan Gambar 2 adalah diagram konteks dari aplikasi data mining dalam memprediksi kelulusan mahasiswa menggunakan *algortma k-means clustering*.

- Login valid

Gambar 2 Diagram Konteks

1. *Langkah-Langkah Algoritma Metode K-Means*
   1. Menentukan jumlah *cluster*

Dari data yang diperoleh, dapat ditentukan bahwa objek-objek tersebut memiliki 2 atribut (IPK dan kehadiran), dimana tiap-tiap mahasiswa mewakili 1 titik dengan 2 atribut (X, Y).

* 1. pengesetan nilai awal titik tengah

Untuk menentukan nilai awal *cendroid* dilakukan secara acak. Untuk penentuan awal diasumsikan :

1. Sebagai pusat *cluster* pertama ( C1 ) : 3.13 ; 5
2. Sebagai pusat *cluster* kedua ( C2 ) : 3.18 ; 4
3. Sebagai pusat *cluster* ketiga ( C3 ) : 3.38 ; 3
4. Sebagai pusat *cluster* keempat ( C4 ) : 2.92 ; 2.
5. menghitung jarak obyek ke centroid dengan menggunakan rumus jarak euclid [15]

De = √ (xi - si)2 + (yi - ti)2 dengan,

De = Enclidean Distance

i = adalah banyaknya objek

( x, y ) merupakaan koordinat objek

( s, t ) merupakan koordinat *cendroid*

Perhitungan jarak dari data ke 1 terhadap pusat cluster adalah sebagai berikut :

C1 = De =



=

C2 = De =



=

C3 = De = 

=

C4 = De =



= .

Perhitungan jarak dari data ke 2 terhadap pusat cluster adalah sebagai berikut :

C1 = De = 

=

C2 = De = 

=

C4 =







C3 = De =

=

Perhitungan jarak dari data ke 3 terhadap pusat cluster adalah sebagai berikut :



C4 = De

=

C1 = De =



=

C2 = De =



=

C3 = D



|  |  |  |
| --- | --- | --- |
| **Cluster ke** | **IPK** | **Kehadiran** |
| C1 |  | 5 |
| C2 |  | 4 |
| C3 |  | 3 |
| C4 |  | 2 |

C4 = ( , )

Pusat *cluster* yang baru pada iterasi ke 1 dapat dilihat pada Tabel II.

TABEL II

PUSAT *CLUSTER* PADA ITERASI KE 1

e

=

C4 = De = 

=

Perhitungan seterusnya sampai jarak dari data ke 100 terhadap pusat cluster.

1. Menghitung pusat *cluster* baru

Pusat *cluster* yang baru ditentukan berdasarkan pengelompokan anggota masing-masing *cluster.* Berdasarkan tabel 3.3, C1 untuk parameter IPK dan kehadiran mahasiswa memiliki 13 anggota, maka perhitungan *cluster* menjadi :

C1 = ,



C1 = ( , 5 )

Kemudian, karena *cluster* ke 2 mempunyai 32 anggota maka *cluster* baru menjadi :

C2 =



=

C2 = ( , )

Kemudian, karena *cluster* ke 3 mempunyai 17 anggota maka *cluster* baru menjadi :

C3 =



# =

C3 = ( , )

Kemudian, karena *cluster* ke 3 mempunyai 6 anggota maka *cluster* baru menjadi :

1. Pengulangan langkah ke 2 hingga posisi data tidak mengalami perubahan.

Perhitungan jarak dari data ke 1 terhadap pusat cluster, dilakukan seperti perhitungan jarak pada iterasi 1 dengan pusat cluster ipk dan kehadiran sesuai dengan hasil pusat cluster pada iterasi 1. Dan hasil dari penghitungan pusat *cluster* pada iterasi ke 2 dapat dilihat pada tabel III.

TABEL III

PUSAT *CLUSTER* PADA ITERASI KE 2

|  |  |  |
| --- | --- | --- |
| **Cluster ke** | **IPK** | **Kehadiran** |
| C1 |  | 5 |
| C2 |  | 4 |
| C3 |  | 3 |
| C4 |  | 2 |

1. KESIMPULAN DAN SARAN
   1. *Kesimpulan*

Setelah dilakukan pengujian dan analisis program, maka dapat diperoleh kesimpulan bahwa pada kasus ini implementasi algoritma k-means dalam data mining sudah berhasil, dan bisa menampilkan informasi prediksi kelulusan mahasiswa, namun untuk melihat tingkat kemampuan *real k-means clustering* dalam memprediksi waktu kelulusan tergantung pada mahasiswa itu sendiri

* 1. *Saran*

Dari beberapa kesimpulan yang telah diambil, maka dapat dikemukakan saran-saran yang akan sangat membantu untuk pengembangan aplikasi ini selanjutnya

1. Untuk para peneliti yang akan melakukan penelitian serupa dengan ini, akan sangat menarik dan lebih baik apabila parameter yang digunakan ditambah

lebih banyak lagi guna mendapatkan hasil yang lebih optimal dan akurat.

1. Untuk mengetahui pengembangan analisis dan aplikasi data mining yang lainnya serta untuk mendapat keputusan yang beragam, kita juga bisa menggunakan algoritma lain dalam data mining didalam melakukan penelitian seperti C4.5, Algoritma Apriori, *Naive Bayes* dan algoritma lainnya untuk mengetahui perbedaan dan cara kerja dari setiap masing-masing algoritma dalam data mining tersebut.

REFERENSI

1. T. Rismawan S. Kusumadewi, K-Means Untuk Pengelompokan Mahasiswa Berdasarkan Nilai Body Mass Index (BMI) dan Ukuran Kerangka: Yogyakarta, 2008.
2. F.N.R. Fauzan, J. Aziz, B.D. Setiawan, dan I. Arwani,

―Implementasi Algoritma K-Means untuk Klasterisasi Kinerja Akademik Mahasiswa‖, *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 2. No. 6, 2018.

1. S. Sulistyowati, B.E. Ketherin, A.A. Arifiyanti, dan A, Sodik. Analisa Segmentasi Konsumen Menggunakan Algoritma K- Means Clustering, [Prosiding Seminar Nasional Sains dan](https://ejurnal.itats.ac.id/sntekpan/issue/view/24) [Teknologi Terapan VI (SNTEKPAN VI), 2018](https://ejurnal.itats.ac.id/sntekpan/issue/view/24)
2. Ong J.O, "Implementasi Algoritma K-Means Clustering Untuk Menentukan Strategi Marketing President University*", Jurnal Ilmiah Teknik Industri*, Vol. 12, No. 1, pp 10-20, Juni. 2013.
3. A. Bastian, ―Penerapan Algoritma K-Means Clustering Analysis Pada Penyakit Menular Manusia (Studi Kasus Kabupaten Majalengka)‖, *Jurnal Sistem Informasi*, vol. 14, no. 1, pp. 28-34, Apr. 2018.
4. Handoyo .R , M. Rumani .R, dan Nasution S.M, "Perbandingan Metode Clustering Menggunakan Metode Single Linkage Dan K - Means Pada Pengelompokan Dokumen", *Jurnal SIFO Mikroskil*, Vol 15, No. 1, pp 73-82, Oktober. 2014.
5. Khotimah T, "Pengelompokan Surat Dalam Al Qur‘an Menggunakan Algoritma K-Means", *Jurnal Simetris*, Vol 5, No. 1, pp 83-88, April. 2014.
6. E. Irfiani, S.S. Rani, ―Algoritma K-Means Clustering untuk Menentukan Nilai Gizi Balita‖, *Jurnal Sistem dan Teknologi Informasi (JUSTIN)*, Vol. 6, No. 4, 2018.
7. Asroni, R. Adrian, "Penerapan Metode K-Means untuk Clustering Mahasiswa Berdasarkan Nilai Akademik dengan Weka Interface Studi Kasus Pada Jurusan Teknik Informatika UMM Magelang", *Jurnal Ilmiah Semesta Teknika ,* Vol. 18, No. 1, pp. 76-82, 2015.
8. Widodo dan D.Wahyuni, "Implementasi Algoritma K-Means Clustering Untuk Mengetahui Bidang Skripsi Mahasiswa Multimedia Pendidikan Teknik Informatika Dan Komputer Universitas Negeri Jakarta", *JURNAL Pendidikan Teknik Informatika dan Komputer*, Vol 1 No 2, 2017.
9. B. Poerwanto, R.Y. Fa‘rifah, ―Analisis Cluster K-Means Dalam Pengelompokan Kemampuan Mahasiswa‖, *Indonesian Journal of Fundamental Sciences (IJFS)*, Vol. 2, No. 2, 2016.
10. Rosmini, A. Fadlil, dan Sunardi, Implementasi Metode K-Means Dalam Pemetaan Kelompok Mahasiswa Melalui Data Aktivitas Kuliah, IT Journal Research and Development, Vol.3, No.1,

Agustus 2018

1. Jain. Data clustering: 50 years beyond K-means. 2009.
2. E. Muningsih1 dan S. Kiswati. ―Penerapan Metode K-Means Untuk Clustering Produk Online Shop Dalam Penentuan Stok Barang‖, *Jurnal Bianglala Informatika*, Vol. 3, No. 1, 2015.
3. M. Anggara, H. Sujaini, H. Nasution, "Pemilihan Distance Measure Pada K-Means Clustering untuk Pengelompokkan Member DiAlvaro Fitness", Jurnal Sistem dan Teknologi Informasi (JUSTIN), Vol. 4, No. 1, 2016.