
PENGELOMPOKAN POTENSI KELULUSAN MAHASISWA FAKULTAS TEKNIK UNIB MENGUNAKAN ALGORITME K-MEANS (STUDI KASUS: FAKULTAS TEKNIK UNIVERSITAS BENGKULU)

Viryandra Virtusena¹, Asahar Johar², Andang Wijanarko³

¹Program Studi Infomatika, Fakultas Teknik, Universitas Bengkulu.

^{2,3}Program Studi Sistem Informasi, Fakultas Teknik, Universitas Bengkulu.

Jl. WR. Supratman Kandang Limun Bengkulu 38371A INDONESIA

(tel: 0736-341022; fax: 0736-341022)

¹viryandravirtusena03@gmail.com

²asahar.johar@unib.ac.id

³andang@unib.ac.id

Abstrak: Ketepatan waktu kelulusan merupakan permasalahan umum bagi pihak akademik maupun mahasiswa, karena kedua pihak tersebut sama-sama tidak dapat mengetahui ketepatan waktu kelulusan mahasiswa. Dengan adanya masalah ini perlu menciptakan suatu sistem yang dapat mengetahui ketepatan waktu kelulusan mahasiswa. Teknik *clustering* dapat memecahkan masalah ini dengan menggunakan metode algoritma *K-Means*. Metode algoritma *K-Means* merupakan metode yang mempartisipasi data ke dalam *cluster* sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam *cluster* yang sama dan data yang mempunyai karakteristik yang berbeda dikelompokkan ke dalam *cluster* yang lain dalam penelitian ini membagi menjadi dua kelompok *cluster* yaitu lulus tepat waktu dan tidak tepat waktu. Hasil dari proses pengelompokan *cluster* ditampilkan dalam bentuk *website* dengan metode pengembangan sistem yang digunakan adalah *waterfall* dengan bahasa pemrograman PHP yang menggunakan *Frame work Code Ignitor*.

Kata Kunci : Kelulusan, *Clustering*, Algoritma *K-Means*

Abstract: The timeliness of graduation is a common problem for academics and students, cause both parties are equally unable to know the timeliness of student graduation. With this problem, it is necessary to create a system that can know the timeliness of student graduation. Clustering techniques can solve this problem by using the K-Means algorithm method. The K-Means algorithm method is a method that articulates data into clusters so that has the same characteristics are grouped into the same cluster and data that has different characteristics are grouped into other clusters in this study divided into two cluster groups that are passed on time and not on time. The result of the cluster grouping process is displayed in the form of a website with the system development method used is a waterfall with PHP programming language that uses Frame work Code Ignitor.

Keyword : Graduation, Clustering, Algorithm K-Means

I. PENDAHULUAN

Universitas adalah bentuk lembaga pendidikan lanjutan yang dinamakan Perguruan Tinggi dan memiliki fakultas-fakultas yang mana didalam fakultas tersebut mempunyai program studi yang beragam. Universitas Bengkulu merupakan salah satu perguruan tinggi negeri yang berada di Bengkulu dan berdiri pada 24 April 1982 berdasarkan keputusan presiden RI nomor 17 tahun 1982 yang diresmikan oleh Menteri Pendidikan Kebudayaan Republik Indonesia. Di kalangan akademik, wisuda merupakan penanda kelulusan mahasiswa yang telah menempuh masa studi pada suatu perguruan tinggi. Pada umumnya masa studi perkuliahan di perguruan tinggi dapat ditempuh

selama 4 tahun atau sebanyak 8 semester. Kelulusan dengan waktu yang tepat merupakan salah satu keinginan mahasiswa karna untuk mahasiswa yang dapat lulus dengan tepat waktu semakin cepat mereka lulus maka semakin banyak kesempatan mereka untuk mengikuti berbagai seleksi dalam mencari pekerjaan.

Hingga saat ini waktu kelulusan mahasiswa tidak selalu dapat dideteksi secara dini untuk itu pihak perguruan tinggi dapat menggunakan sejumlah data mahasiswa yang tersimpan *database*. Perguruan tinggi adalah salah satu institusi yang sudah pasti memiliki data yang tidak kecil volumenya. *Database* pada perguruan tinggi menyimpan berbagai data misalnya seperti data akademik dan data mahasiswa. Dari data tersebut apabila diolah maka dapat dikembangkan menjadi informasi untuk diterapkan pada perguruan tinggi diantaranya mengenai kelulusan mahasiswa. Salah satu teknik yang dapat digunakan untuk memberikan informasi dari data tersebut adalah teknik data *minning*. Data *minning* bertujuan untuk mendapatkan informasi yang berguna dari sekumpulan dokumen yang berskala besar bagi pihak-pihak tertentu yang membutuhkan informasi.

Penelitian yang penulis lakukan adalah di salah satu fakultas di Universitas Bengkulu yaitu pada Fakultas Teknik. Fakultas Teknik Universitas Bengkulu atau dikenal sebagai FT Unib merupakan salah satu fakultas yang terdapat di Universitas Bengkulu. Secara formal Fakultas Teknik Unib dibentuk pada tahun 2006 dengan izin pendirian Fakultas Teknik melalui surat keputusan direktur jenderal pendidikan tinggi no 3233/D/T/2006 tanggal 30 agustus 2006. Tercatat Sejak tahun 2012 hingga 2020 Di

fakultas tersebut menyimpan sebanyak 3.194 data mahasiswa di dalam *database*. Untuk itu peneliti tertarik melakukan penelitian untuk mengelompokan mahasiswa berdasarkan ketepatan waktu kelulusan mahasiswa berdasarkan jumlah data mahasiswa yang tersimpan dalam *database* di Fakultas Teknik. Karena jumlah data yang tersimpan cukup banyak maka penelitian ini dirasa cocok menggunakan data *minning* mengingat bahwa data *minning* bertujuan untuk memberikan informasi dari jumlah data yang berskala besar. Salah satu metode yang terdapat dalam data *minning* adalah metode *clustering*. *Clustering* adalah teknik yang digunakan dalam data *minning* yang cara kerjanya mencari dan mengelompokan data yang mempunyai kemiripan karakteristik antara data satu dengan data lainnya yang diperoleh (Wardhani, 2016).

Clustering dalam penelitian ini menggunakan algoritma *k-means*. Algoritma *k-means* merupakan salah satu metode *clustering* 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 dikelompokan ke dalam satu *cluster* yang sama dan data yang mempunyai karakteristk yang berbeda dikelompokan ke dalam *cluster* yang lain (Priyatman, Sajid, & Haldivany, 2019). Algoritma *k-means* memiliki ketelitian yang cukup tinggi terhadap ukuran objek sehingga algoritma *k-means* ini lebih terukur dan efisien untuk pengolahan objek dalam jumlah yang besar (Ediyanto, Mara, & Satyahadewi, Pengklasifikasian Karakteristik Dengan Metode K-Means Cluster Analysis, 2013). Dalam implementasi menyelesaikan masalah, algoritma

k-means sangat *simple* serta *fleksibel* artinya perhitungan komputasinya tidak terlalu rumit dan algoritma ini dapat diimplementasikan pada segala bidang termasuk dalam data *minning* (Priati, 2017).

Berdasarkan latar belakang di atas, maka dibuatlah penelitian yang berjudul “Pengelompokan Potensi Kelulusan Mahasiswa Fakultas Teknik UNIB Menggunakan Algoritma *K-Means*.” Alasan penggunaan algoritma *k-*

II. LANDASAN TEORI

A. *Clustering*

Clustering adalah teknik menemukan sekelompok data dari pemecahan atau pemisahan sekumpulan data menurut karakteristik tertentu yang telah ditentukan. Dalam pengelompokan tersebut nilai labelnya belum diketahui sehingga diharapkan setelah melakukan pengelompokan data dapat diketahui label dari data tersebut. Metode *clustering* juga sering disebut tahapan awal sebelum melakukan metode lain seperti klasifikasi.

Cluster analysis adalah mengelompokan data objek pada informasi yang mirip atau memiliki kesamaan antara satu dengan yang lainnya, tujuannya agar dapat menemukan kelompok yang berkualitas seperti kelompok yang merupakan objek-objek yang mirip atau memiliki hubungan satu sama lain dan sebaliknya yaitu kelompok yang tidak berhubungan dengan objek dalam kelompok yang lain.

Clustering cocok digunakan untuk menjelajahi data. Jika ada banyak kasus tapi tidak ada pengelompokan yang jelas, algoritma *clustering* dapat digunakan untuk mencari pengelompokan dari data tersebut. *Clustering*

dikarenakan dari data yang tersimpan dalam *database* fakultas teknik yang jumlahnya cukup besar yaitu berjumlah 3.194 data yang tercatat sejak tahun 2012 sampai 2020, *k-means* dirasa dapat membantu untuk mengelompokan data sesuai dengan *cluster* tertentu berdasarkan Indeks Prestasi Semester dan Satuan Kredit Semester mahasiswa, sehingga hasilnya nanti akan diketahui mahasiswa mana saja yang dapat lulus dengan tepat waktu dan tidak tepat waktu.

juga dapat berguna sebagai data-*preprocessing* yaitu langkah untuk mengidentifikasi kelompok-kelompok yang berhubungan dalam membangun model (Prasetyo, 2012).

B. Normalisasi

Normalisasi merupakan proses yang digunakan untuk standarisasi semua variabel dataset dan memberikan bobot yang sama (Shofiani, 2017). Normalisasi penting dilakukan untuk memberikan bobot yang sama terhadap nilai-nilai fitur yang berbeda dari hasil ekstrasi. Normalisasi dapat dilakukan dengan berbagai cara, salah satunya dapat menggunakan persamaan *Min-Max Normalization* (Mustaffa & Yusuf, 2011), yang ditunjukkan pada persamaan 1 berikut ini :

$$X_n = \frac{X_0 - X_{min}}{X_{max} - X_{min}} \quad (1)$$

Dimana,

X_n : adalah nilai normalisasi

X_0 : adalah nilai awal variabel

X_{min} : adalah nilai *minimum* variabel

X_{max} : adalah nilai *maximum* variabel

C. Pengenalan Algoritma *K-Means*

Algoritma *K-Means* masuk ke dalam penerapan data *minning clustering* (He & Tan, 2012). Tujuan dari *clustering* adalah mengelompokan data ke dalam beberapa *cluster*

berdasarkan tingkat kemiripan (Xiang, Zhu, Mha, Meng, & An, 2015), tingkat kemiripan dihitung berdasarkan jarak antar data (Zhang, Ouyang, & Ning, 2010). Langkah-langkah algoritma *k-means* adalah sebagai berikut (Suntoro, 2019) :

1. Siapkan *dataset*.
2. Tentukan jumlah *cluster* (*K*).
3. Pilih titik *centroid* secara acak.
4. Kelompokkan data sehingga terbentuk *K* buah *cluster* dengan titik *centroid* dari setiap *cluster* yang ditunjukkan pada persamaan 2 berikut :

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (2)$$

Dimana,

$d(x, y)$: adalah jarak antara data x ke data y

x_i : adalah data testing ke- i

y_i : adalah data training ke- i

5. Perbaharui nilai titik *centroid* yang ditunjukkan pada persamaan 3 berikut ini :

$$\mu_k = \frac{1}{N_k} \sum_{i=1}^{N_k} x^i \quad (3)$$

Dimana,

μ_k : adalah titik *centroid* dari *cluster* ke- K

N_k : adalah banyaknya data pada *cluster* ke- K

x_i : adalah data ke- i pada *cluster* ke- K

6. Ulangi Langkah 3 sampai 5 sampai nilai dari titik *centroid* tidak lagi berubah.

III. METODE PENELITIAN

A. Metode Pengumpulan Data

Berdasarkan cara memperoleh data, jenis data yang digunakan dalam penelitian ini adalah data sekunder. Data sekunder adalah data yang diperoleh atau dikumpulkan dan disatukan oleh studi-studi sebelumnya atau yang diterbitkan oleh berbagai instansi lain.

1. Studi Pustaka

Studi pustaka dilakukan dengan cara menelaah beberapa literatur, yaitu:

a. Buku referensi

Buku yang digunakan sebagai referensi adalah buku-buku yang membahas tentang Kelulusan Mahasiswa, *K-Means* dan *Web*.

b. Jurnal Ilmiah

Jurnal ilmiah yang digunakan diperoleh dengan cara mengunduhnya melalui internet. Informasi yang (2) adalah informasi yang membahas tentang Kelulusan Mahasiswa, *K-Means* dan *Web*.

2. Observasi

a. Studi Lapangan

Metode ini dilakukan dengan cara terjun langsung ke lapangan yaitu ke Fakultas Teknik Universitas Bengkulu untuk data Akademik Mahasiswa Fakultas Teknik Universitas Bengkulu. (3)

b. Wawancara

Wawancara dilakukan dengan Dosen Fakultas Teknik Universitas Bengkulu. Wawancara yang dilakukan diharapkan akan memberikan informasi detail data Akademik Mahasiswa Fakultas Teknik Universitas Bengkulu. Wawancara dilakukan bersama bapak Funny Farady Coastera, S.Kom., M.T selaku Dosen Fakultas Teknik Universitas Bengkulu dan juga selaku Kepala Pusat Sistem Informasi LPTIK Universitas Bengkulu.

B. Metode Pengembangan Sistem

Metode pengembangan sistem pada penelitian ini menggunakan model pengembangan *waterfall*. Disebut dengan *waterfall* karena tahap demi tahap yang dilalui harus menunggu selesainya tahap sebelumnya

dan berjalan berurutan. Alur yang digunakan yaitu dari analisis, desain, pengkodean, pengujian dan tahap pendukung (*Support*) (Pressman, 2012). Adapun tahapan-tahapan dalam metode *waterfall* adalah :

1. Analisis Kebutuhan

Sistem yang akan dibuat memerlukan masukan, keluaran, dan kebutuhan *interface*. Tujuan dari analisis kebutuhan adalah sebagai batasan dari sistem yang akan dibangun, menentukan kemampuan dan fungsi sistem sesuai dengan kebutuhan pengguna, dan fitur-fitur yang merupakan nilai tambah. Adapun analisis kebutuhan sistem yang akan dibangun adalah sebagai berikut :

a. Kebutuhan Data Masukan

Data-data masukan yang digunakan dalam penelitian ini adalah sebagai berikut :

- Keterangan Kelulusan Mahasiswa.
- Program Studi Mahasiswa.
- NPM Mahasiswa.
- Nama Mahasiswa.
- Jenis Kelamin Mahasiswa.
- Tanggal Lahir Mahasiswa.
- Asal Sekolah Mahasiswa
- Indeks Prestasi Semester (IPS) Mahasiswa.

b. Kebutuhan Data Keluaran

Adapun data keluaran yang dibutuhkan adalah hasil dari metode K - *Means* untuk mengetahui pengelompokan potensi kelulusan mahasiswa Fakultas Teknik Unib yang didapat dari masukan oleh *user*.

c. Kebutuhan *Interface*

Kebutuhan *interface* pada aplikasi adalah kemudahan dan kenyamanan pengguna saat mengakses aplikasi sesuai dengan permasalahan yang ada.

2. Desain

Setelah tahap analisa kebutuhan, maka tahapan kedua pada model pengembangan sistem *waterfall* ini adalah tahap desain. Desain perangkat lunak adalah proses multilangkah yang fokus pada desain pembuatan program perangkat lunak. Pada tahap ini, langkah yang dilakukan adalah penetapan alur program, dan alur desain *interface*.

a. Pembuatan Kode Program

Setelah tahap desain dilakukan maka selanjutnya ketahap pembuatan kode program. Tahap ini adalah mentranslasikan desain yang telah dibuat kedalam kode-kode program. Dalam hal ini, bahasa pemrograman yang digunakan adalah PHP dengan menggunakan *Frame work Code Ignitor*.

b. Pengujian

Pengujian program dilakukan menggunakan dua metode, yaitu *black box* dan *white box*. Pengujian ini dilakukan untuk melihat apakah program sudah berjalan sesuai rancangan atau belum.

c. Pendukung atau pemeliharaan

Tidak menutup kemungkinan sebuah perangkat lunak mengalami perubahan ketika sudah dikirimkan oleh *user*. Tahapan ini mengantisipasi jika ada ketidaksesuaian sistem setelah dilakukan pengujian. Langkah pada tahapan ini

mengulangi tahapan-tahapan sebelumnya. Dalam tahap perancangan desain sistem, tahap ini merupakan suatu tahap yang mengharuskan analisis dalam perancangan desain sistem untuk berusaha mengetahui hal-hal yang menjadi kebutuhan pengguna. Tahapan perancangan desain sitem yaitu :

- **Diagram Alir**
Diagram alir pada desain sistem yaitu diagram alir penelitian, diagram alir metode, diagram alir kerja sistem, diagram alir pengujian fungsional, dan diagram alir pengujian *black box*.
- **Diagram *Unified Modeling Language* (UML)**
Pada percancangan *Unified Modeling Language* menggunakan empat diagram yaitu *use case diagram*, *activity diagram*, *sequence diagram*, *class diagram*.
- ***Entity Relationship Diagram* (ERD)**
Pada perancangan diagram *Entity Relationship Diagram* (ERD) merupakan diagram relasi dari database yaitu relasi antar tabel *database*.
- **Desain Perancangan Antarmuka**
Desain sistem antarmuka menggunakan *software balsamiq mocup* yang menampilkan *form-form* yang akan dibangun nantinya.
- **Desain Pengujian**
Pada penelitian ini pengujian menggunakan pengujian fungsional *blackbox testing*.

d. Implementasi

Tahapan selanjutnya dalam pembuatan aplikasi ini secara nyata adalah tahap implementasi desain yang telah dibuat kedalam kode-kode program. Dalam hal ini, yaitu Pengelompokan Potensi Kelulusan Mahasiswa. Perancangan basis data yang digunakan yaitu menggunakan *database mysql* dengan bantuan *software xampp*. Aplikasi Pengelompokan Potensi Kelulusan Mahasiswa Fakultas Teknik UNIB menggunakan algoritma *K-Means* dengan bahasa pemrograman yang digunakan adalah bahasa pemrograman *Hypertext Preprocessor* (PHP) dengan bantuan *framework codeigniter* dengan menggunakan *software sublime text* dengan menggunakan bantuan *software google chrome* agar dapat dioperasikan.

e. Pengujian

Pengujian yang dilakukan pada penelitian ini fokus pada perangkat lunak dari segi fungsional pada aplikasi yang dibangun, apakah sesuai dengan tujuan dari penelitian ini, yaitu aplikasi berjalan dengan baik dan benar sehingga dapat Mengelompokan Potensi Kelulusan Mahasiswa Berdasarkan IP Semester 1 sampai dengan IP Semester 4. Pengujian fungsional adalah teknik pengujian yang digunakan untuk menguji fitur atau fungsi dari sistem. Pada penelitian ini pengujian fungsional menggunakan pengujian sistem *blackbox testing*.

f. Penggunaan dan pemeliharaan

Setelah sistem selesai maka pengguna akan menggunakan sistem. Jika terdapat pengembangan fungsional dari sistem

yang diinginkan pengguna, maka akan dilakukannya pemeliharaan.

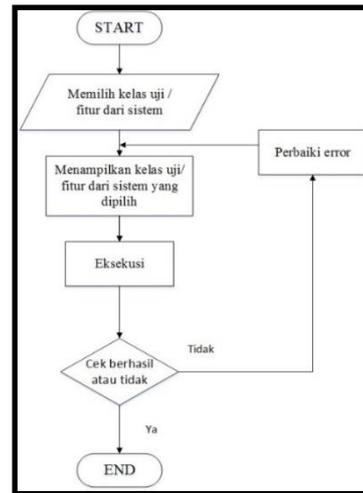
C. Metode Pengujian Sistem

Pada penelitian ini, pengujian sistem yang dilakukan menggunakan *blackbox testing*. Pengujian *black box* adalah suatu pengujian yang dilakukan untuk mengamati hasil eksekusi antarmuka (*interface*) melalui data uji dan memeriksa fungsional dari sistem yang dibuat. Pengujian *black box* merupakan metode perancangan data uji yang didasarkan pada spesifik perangkat lunak.

Data uji dibangkitkan, dieksekusi pada perangkat lunak dan kemudian keluaran dari perangkat lunak diuji apakah sudah sesuai dengan yang diharapkan atau belum. Pengujian *black box* berusaha menemukan kesalahan seperti berikut ini:

- a. Fungsi-fungsi yang tidak benar atau hilang.
- b. Kesalahan antarmuka (*interface*).
- c. Kesalahan dalam struktur data atau akses database eksternal.
- d. Kesalahan kinerja.
- e. Inisialisasi dan kesalahan terminasi.

Teknik pengujian *black box* yang dilakukan dalam Pengelompokan Potensi Kelulusan Mahasiswa Berdasarkan IP Semester dengan menggunakan algoritma *K-Means*. Berikut ini merupakan gambar dari diagram alir sistem yang ditunjukkan pada gambar 1 berikut ini :



Gambar 1 Diagram Pengujian Black Box

IV. ANALISIS DAN DESAIN

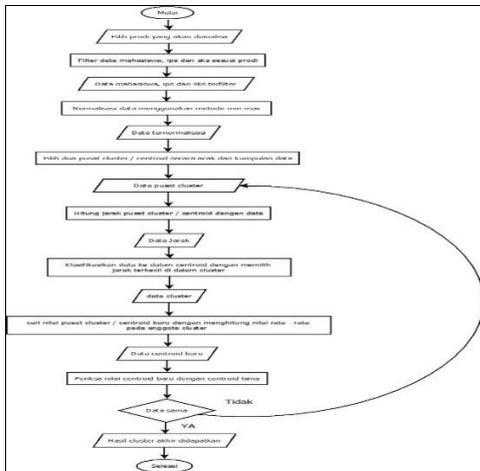
A. Analisis Sistem

Analisis sistem merupakan suatu proses untuk menentukan hal-hal yang akan dikerjakan oleh sistem yang diusulkan. Tahapan analisis sistem dilakukan untuk mengembangkan sistem yang sudah ada atau mengatasi masalah-masalah yang belum tertangani dengan demikian dapat disimpulkan bahwa analisis sistem adalah sebuah pembelajaran mengenai sistem yang sedang berjalan saat ini untuk dapat bisa merancang sistem baru untuk memberikan sistem yang terbaik bagi para user.

B. Analisis Alur Kerja Sistem

Alur sistem akan menjelaskan analisis tahapan kerja sistem yang akan dibangun. Analisis alur kerja sistem bertujuan untuk menjelaskan alur kerja dari sebuah sistem secara berurutan yang dimulai dari *user* memasukan sebuah masukan sampai dengan *user* menerima sebuah keluaran dari sistem yang keluaran tersebut diproses terlebih dahulu oleh sistem. Berikut ini adalah analisis alur kerja sistem dari

aplikasi yang akan dibuat yang ditunjukkan pada gambar 2 berikut ini :



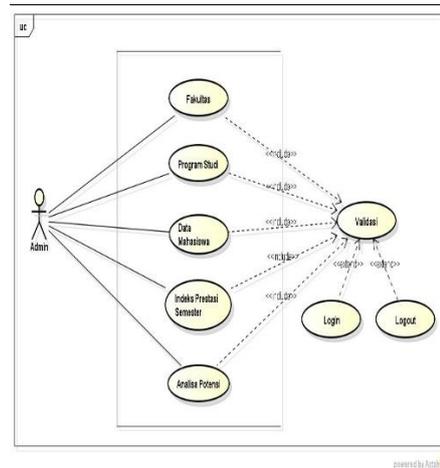
Gambar 2 Diagram Alir Sistem

C. Perancangan Sistem

Perancangan sistem adalah sebuah proses yang menentukan bagaimana suatu sistem akan menyelesaikan apa yang mesti diselesaikan. Tahap ini berupa penggambaran, perencanaan dan pembuatan sketsa atau pengaturan dari berbagai elemen yang terpisah ke dalam satu kesatuan yang utuh dan berfungsi. Perancangan yang dilakukan dalam sistem ini adalah perancangan *Unified Modelling Language* (UML), perancangan basis data (*database*) dan perancangan antarmuka (*interface*).

1. Use Case Diagram

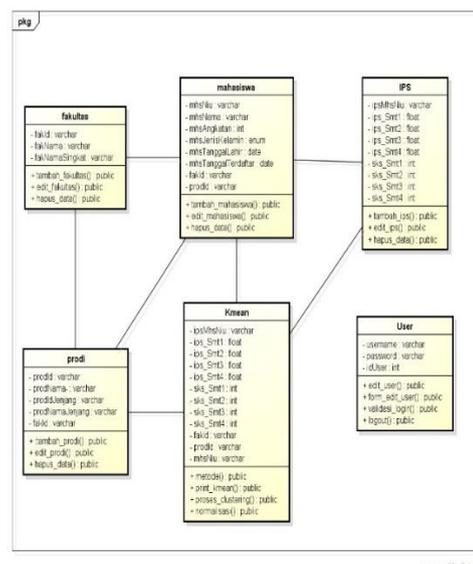
Use case diagram merupakan permodelan untuk menggambarkan kelakuan dari sistem yang dibuat dan mendeskripsikan sebuah interaksi antara satu atau lebih aktor dengan sistem yang dibuat. *Use case diagram* dari sistem Pengelompokan Potensi Kelulusan Mahasiswa Fakultas Teknik Unib dapat dilihat pada gambar 3 berikut ini :



Gambar 3 Use Case Diagram

2. Class Diagram

Class diagram adalah diagram yang menggambarkan relasi atau hubungan antar kelas-kelas dalam sistem yang akan dibangun. *Class diagram* memberikan pandangan secara luas dari suatu sistem dengan menunjukan kelas-kelasnya dan hubungan mereka. *Class diagram* pada sistem Pengelompokan Potensi Kelulusan mahasiswa Fakultas Teknik Unib dapat dilihat pada gambar 4 berikut ini :



Gambar 4 Class Diagram

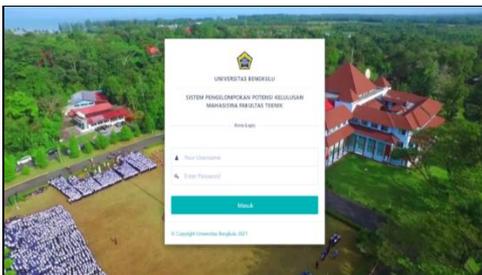
V. HASIL DAN PEMBAHASAN

A. Implementasi Antar Muka

Berikut ini adalah implementasi antarmuka dari sistem Pengelompokan Potensi Kelulusan Mahasiswa Fakultas Teknik Unib.

1. Halaman Log In

Halaman *log in* merupakan halaman dimana *admin* menginputkan *username* dan *password* untuk dapat masuk ke sistem. Berikut ini tampilan halaman *log in* dari sistem yang ditunjukkan pada gambar 5 berikut ini :



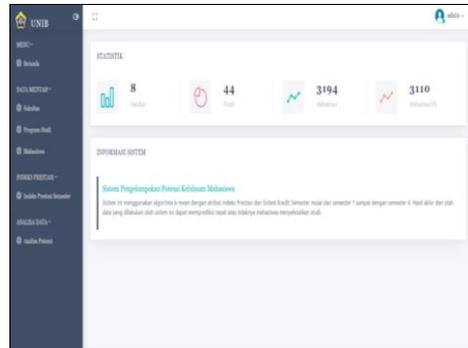
Gambar 5 Halaman Log In

Gambar 5 merupakan tampilan *log in* sistem Pengelompokan Potensi Kelulusan Mahasiswa Fakultas Teknik Unib. Pada halaman *log in* tersebut admin harus menginputkan *username* dan *password* pada *textbox* yang telah disediakan, apabila admin salah menginput *username* atau *password* maka sistem akan menampilkan notifikasi jika *username* atau *password* yang diinput salah. Sebaliknya apabila admin menginput *username* dan *password* yang benar maka sistem akan masuk ke halaman selanjutnya yaitu halaman beranda.

2. Halaman Beranda

Halaman beranda merupakan pertama yang tampil setelah proses *log in* berhasil. Gambar tampilan halaman dari sistem

Pengelompokan Potensi Kelulusan Mahasiswa Fakultas Teknik Unib ditunjukkan pada gambar 6 berikut ini :

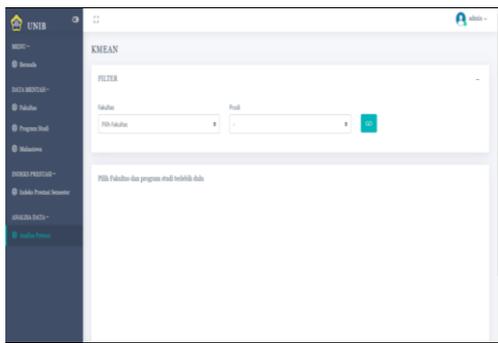


Gambar 6 Halaman Beranda

Gambar 6 merupakan tampilan halaman beranda dari sistem Pengelompokan Potensi Kelulusan Mahasiswa Fakultas Teknik Unib. Pada tampilan halaman beranda terdapat beberapa menu yaitu menu fakultas, prodi, mahasiswa, dan ips mahasiswa dan terdapat jumlah *record* masing-masing tabel yang terdapat disetiap menu. Lalu pada bagian sebelah kiri tampilan halaman beranda terdapat beberapa pilihan menu yang dapat diakses oleh admin.

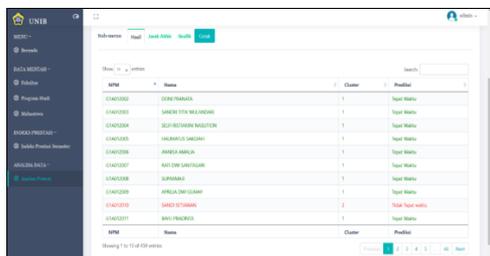
3. Halaman Analisa Potensi

Halaman analisa potensi merupakan proses untuk menjalankan algoritma *k-means*. Pada gambar 7 dibawah ini merupakan tampilan awal halaman analisa potensi dari sistem Pengelompokan Potensi Kelulusan Mahasiswa Fakultas Teknik Unib:



Gambar. 7 Tampilan Utama Halaman Analisa Potensi

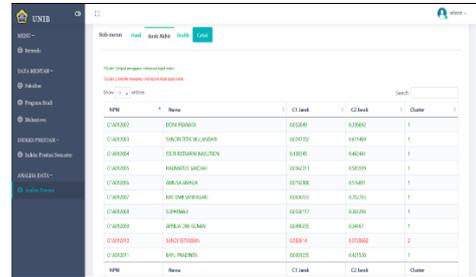
Pada gambar 7 merupakan tampilan halaman utama ini dimana admin harus memfilter terlebih dahulu data berdasarkan nama fakultas dan prodi. Setelah selesai memfilter berdasarkan fakultas dan prodi maka sistem akan menampilkan halaman sub-menu yang terdiri dari hasil, jarak akhir, dan grafik. Sub menu grafik pada halaman analisa potensi ditunjukkan pada gambar 8 berikut ini :



Gambar 8. Tampilan Sub-Menu Hasil Halaman Analisa Potensi

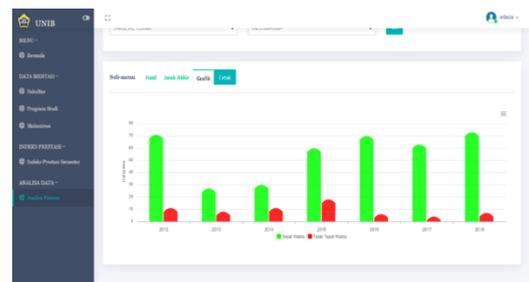
Pada gambar 8 merupakan Sub-menu hasil merupakan tampilan halaman yang menampilkan informasi hasil klasterisasi yang telah diberi keterangan prediksi ketepatan waktu kelulusan mahasiswa. Setelah itu admin dapat mengakses sub-menu yang lainnya yaitu sub-menu data jarak akhir halaman analisa potensi. Tampilan halaman sub-menu data jarak akhir halaman analisa

potensi yang ditunjukkan pada gambar 9 berikut :



Gambar 9 Tampilan Sub-Menu Jarak Akhir Analisa Potensi

Pada gambar 9 merupakan Sub-menu jarak akhir merupakan tampilan halaman yang menampilkan menu hasil cluster suatu data yang diambil dari jarak terdekat data tersebut terhadap suatu *cluster*. Selanjutnya admin dapat mengakses sub-menu grafik halaman analisa potensi. Tampilan sub-menu grafik pada halaman analisa potensi ditunjukkan pada gambar 10.



Gambar 10 Tampilan Sub-Menu Grafik Halaman Analisa Potensi

B. Normalisasi Data

Tabel. 1 Data Awal

Sebelum melakukan proses perhitungan algoritma *K-Means* diperlukan dahulu tahap normalisasi data. Proses normalisasi yang digunakan pada penelitian ini adalah dengan menggunakan metode *min-max*. Data awal

merupakan data sebelum dilakukan perhitungan normlisasi data. Berikut ini data awal yang ditunjukkan data pada tabel 1 berikut ini :

MhsNiu	Ips_1	Sks_1	Ips_2	Sks_2	Ips_3	Sks_3	Ips_4	Sks_4
G1A012002	3.05	19	3.22	23	2.74	23	2.71	21
G1A012003	3.68	19	3.48	23	3.36	22	3.43	21
G1A012004	2.95	19	2.8	20	3.15	20	2.75	24
G1A012005	2.89	19	3.25	20	3.09	23	3	24
G1A012006	3.26	19	3.61	23	3.26	23	3	21
G1A012007	3.79	19	3.74	23	3.36	22	3.57	21
G1A012008	3.11	19	3.09	23	3.09	23	2.57	21
G1A012010	2.37	19	2	20	2.4	15	2.5	18
G1A012011	3.16	18	3.22	23	3.36	22	2.86	21

Selanjutnya dari data di atas dilakukan proses normalisasi data. Adapun tahapan yang dilakukan untuk proses normalisasi adalah mencari nilai *minimum* dan *maximum* variabel

terlebih dahulu. Nilai *minimum* dan *maximum* dari data awal pada tabel 1 diatas ditunjukkan pada tabel 2 dibawah ini :

Tabel. 2 Nilai Minimum dan Maximum Data

MhsNiu	Ips_1	Sks_1	Ips_2	Sks_2	Ips_3	Sks_3	Ips_4	Sks_4
G1A012002	3.05	19	3.22	23	2.74	23	2.71	21
G1A012003	3.68	19	3.48	23	3.36	22	3.43	21
G1A012004	2.95	19	2.8	20	3.15	20	2.75	24
G1A012005	2.89	19	3.25	20	3.09	23	3	24
G1A012006	3.26	19	3.61	23	3.26	23	3	21
G1A012007	3.79	19	3.74	23	3.36	22	3.57	21
G1A012008	3.11	19	3.09	23	3.09	23	2.57	21
G1A012010	2.37	19	2	20	2.4	15	2.5	18
G1A012011	3.16	18	3.22	23	3.36	22	2.86	21
Min	2.37	18	2	20	2.4	15	2.5	18
Max	3.79	19	3.74	23	3.36	23	3.57	24

Berdasarkan tabel 2 didapatlah nilai *minimum* dan *maximum* dari setiap masing-masing variabel. Nilai *minimum* didapatkan dengan cara melihat nilai terkecil dari data dan nilai *maximum* didapatkan dengan cara mencari nilai terbesar data langkah selanjutnya adalah menghitung nilai normalisasi setelah didapatkan nilai

minimum dan *maximum* langkah selanjutnya adalah menormalisasikan nilai data. Nilai data setelah dinormalisasikan hasilnya ditunjukkan pada tabel 3 berikut :

Tabel. 3 Nilai Data Setelah Dinormalisasikan

MhsNiu	Ips_1	Sks_1	Ips_2	Sks_2	Ips_3	Sks_3	Ips_4	Sks_4
G1A012002	0.478873239	1	0.701149425	1	0.3541666667	1	0.19626	0.5
G1A012003	0.922535211	1	0.850574713	1	1	0.875	0.869158879	0.5
G1A012004	0.408450704	1	0.459770115	0	0.78125	0.625	0.23364486	1
G1A012005	0.366197183	1	0.718390805	0	0.71875	1	0.4678972	1
G1A012006	0.626760563	1	0.925287356	1	0.895833333	1	0.46728972	0.5
G1A012007	1	1	1	1	1	0.875	1	0.5
G1A012008	0.521126761	1	0.626436782	1	0.71875	1	0.065420561	0.5
G1A012010	0	1	0	0	0	0	0	0
G1A012011	0.556338028	0	0.701149425	1	1	0.875	0.336448598	0.5

Dari tabel 3 dapat dilihat hasil perhitungan nilai normalisasi data setelah menggunakan nilai *minmum* dan *maximum* variabel. Hasil perhitungan nilai normalisasi data didapat dari nilai lama variabel dikurangi dengan nilai terkecil variabel (*min*) kemudian dibagi dengan nilai terbesar variabel (*max*) lalu dikurangi nilai terkecil variabel (*min*) atau menggunakan persamaan (1). Berikut ini persamaan untuk menghitung nilai normalisasi data yang ditunjukkan pada persamaan 1 :

$$X_n = \frac{X_0 - X_{min}}{X_{max} - X_{min}}$$

Berikut ini contoh perhitungan nilai normalisasi untuk variabel IPS dan SKS menggunakan persamaan (1) :

- Nilai normalisasi variabel IPS :

Pada data pertama yaitu mahasiswa dengan nomor pokok G1A012002 diketahui untuk nilai awal variabelnya (X_0) adalah 3,05 nilai *minimum* variabelnya (X_{min}) adalah 2,37 dan nilai *maximum* variabelnya (X_{max}) adalah 3,79 sehingga :

$$X_n = \frac{X_0 - X_{min}}{X_{max} - X_{min}}$$

$$X_n = \frac{3,05 - 2,37}{3,79 - 2,37}$$

$$= 0.478873239$$

- Nilai normalisasi variabel SKS :

Pada data pertama mahasiswa dengan nomor pokok G1A012002 diketahui untuk nilai awal variabelnya (X_0) adalah 19 nilai *minimum* variabelnya (X_{min}) adalah 18 dan nilai *maximum* variabelnya (X_{max}) adalah 19 sehingga:

$$X_n = \frac{X_0 - X_{min}}{X_{max} - X_{min}}$$

$$X_n = \frac{19 - 18}{19 - 18}$$

$$= 1$$

C. Perhitungan Menggunakan Algoritma

K-Means

Setelah perhitungan normalisasi selesai langkah selanjutnya adalah melakukan perhitungan menggunakan metode algoritma *k-means*. Berikut ini langkah-langkah perhitungan menggunakan metode algoritma *k-means* :

1. Menyiapkan Dataset

Langkah yang pertama adalah menyiapkan data set. Dataset yang digunakan disini adalah data yang nilainya

sudah dinormalisasikan. Dalam hal ini menggunakan data pada tabel 3.

2. Menentukan Jumlah Cluster (K)

Langkah kedua adalah menentukan jumlah cluster. Jumlah cluster yang ditentukan adalah sebanyak 2 cluster yaitu cluster tepat waktu dan cluster tidak tepat waktu.

3. Memilih Titik Centroid

Tabel. 4 Titik Centroid Yang Dipilih Secara Acak

Data Ke	Ips_1	Sks_1	Ips_2	Sks_2	Ips_3	Sks_3	Ips_4	Sks_4
Data Ke-6	1	1	1	1	1	0.875	1	0.5
Data Ke-7	0.521126761	1	0.626436782	1	0.71875	1	0.065420561	0.5

Langkah ketiga memilih titik *centroid*. Titik *centroid* yang dipilih secara acak dari tabel 3 adalah data ke-6 dan data ke-7. Nilai titik *centroid* yang dipilih secara acak ditunjukkan pada tabel 4 berikut :

4. Kelompokan Data Sehingga Terbentuk K Buah Cluster Dengan Titik Centroid

Dalam melakukan perhitungan dilangkah menghitung jarak data terhadap suatu cluster lalu dikelompokan berdasarkan jarak terdekatnya menggunakan persamaan yang ada ditunjukkan pada persamaan 2 dibawah ini :

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

Berikut ini contoh perhitungan jarak data pertama terhadap masing-masing cluster menggunakan persamaan (2) :

- **Perhitungan jarak data pertama dengan**

C1 :

$$\begin{aligned} &\sqrt{(0,478873 - 0,849765258)^2 + (1 - 1)^2} \\ &\sqrt{(0,701149425 - 1)^2 + (1 - 1)^2} \\ &\sqrt{(0,354166667 - 1)^2 + (1 - 0,875)^2} \\ &\sqrt{(0,196261682 - 1)^2 + (0,5 - 0,5)^2} \\ &=1,439605744 \end{aligned}$$

- **Perhitungan jarak data pertama dengan**

C2 :

$$\sqrt{(0,478873239 - 0,521126761)^2 + (1 - 1)^2}$$

$$\begin{aligned} &\sqrt{(0,701149425 - 0,626436782)^2 + (1 - 1)^2} \\ &\sqrt{(0,354166667 - 0,71875)^2 + (1 - 1)^2} \\ &\sqrt{(0,196261682 - 0,65420561)^2 + (0,5 - 0,5)^2} \\ &=0,157407745 \end{aligned}$$

Lakukan yang sama untuk perhitungan data selanjutnya untuk mendapatkan jarak data pada masing-masing cluster. Berikut ini adalah hasil perhitungan jarak untuk setiap anggota cluster dari data ke 1 hingga ke 9 ditunjukkan pada tabel 5 berikut:

Tabel. 5 Hasil Perhitungan Jarak

MhsNiu	Jarak C1	Jarak C2	Jarak terdekat
G1A012002	1.439605744	0.157407745	0.157407745
G1A012003	0.045448109	0.952088402	0.045448109
G1A012004	2.589430661	1.463304336	1.463304336
G1A012005	2.109516555	1.443957537	1.443957537
G1A012006	0.455145593	0.293327494	0.293327494
G1A012007	0	1.337034348	0
G1A012008	1.337034348	0	0
G1A012010	6.015625	3.434877554	3.434877554
G1A012011	1.726448074	1.175004572	1.175004572

Setelah masing-masing data dihitung jaraknya untuk tiap *cluster*, langkah selanjutnya adalah mengelompokkan data sesuai *clusternya*. Kelompok *cluster* suatu data diambil dari jarak terdekat data tersebut terhadap suatu *cluster*. Contoh berdasarkan tabel 5 untuk data pertama memiliki jarak 1,439605744 terhadap *cluster* 1 dan memiliki jarak 0,157407745 terhadap *cluster* 2. Dari kedua *cluster* tersebut, data pertama memiliki jarak terdekat dengan *cluster* 2. Oleh karena itu data pertama masuk dalam *cluster* 2. Untuk hasil pengelompokkan setiap data berdasarkan jarak terdekatnya ditunjukkan pada tabel 6 berikut ini :

Tabel. 6 Pengelompokan Data Berdasarkan Jarak Terdekat

MhsNiu	C1	C2
G1A012002		1
G1A012003	1	
G1A012004		1
G1A012005		1
G1A012006		1
G1A012007	1	
G1A012008		1
G1A012010		1
G1A012011		1

Setelah setiap data dikelompokkan berdasarkan jarak terdekatnya maka dapat disimpulkan bahwa data ke-2 dan ke-6 berada dalam *cluster* 1 sedangkan ke-1, ke-3, ke-4, ke-5, ke-7, ke-8, dan ke-9 berada dalam *cluster* 2.

5. Perbaharui Nilai Titik *Centroid*

Setelah data dikelompokkan sesuai dengan *clusternya*, langkah selanjutnya adalah menghitung nilai *centroid* baru pada masing-masing *cluster* menggunakan persamaan yang ditunjukkan pada persar (3) 3 berikut ini

$$\mu_k = \frac{1}{N_k} \sum_{i=1}^{N_k} x^i$$

Berikut ini contoh perhitungan nilai *centroid* terbaru terhadap masing-masing *cluster* menggunakan persamaan (3) :

- Perhitungan Nilai *Centroid* Baru Untuk *Cluster* 1 :

$$C_1 = \frac{(0,922535211+1)}{2} = 0,961267606$$

- Perhitungan Nilai *Centroid* Baru Untuk *Cluster* 2 :

$$C_2 = \frac{(0,4788732239+0,408450704+0,366197183)}{2}$$

$$\frac{(0,626760563+0,626760563+0,366197183)}{2}$$

$$= 0,422535211$$

Untuk hasil perhitungan nilai *centroid* baru pada masing-masing *cluster* ditunjukkan pada tabel 7 berikut ini :

Tabel. 7 Hasil Perhitungan Nilai Centroid Baru

Data Ke	Ips_1	Sks_1	Ips_2	Sks_2	Ips_3	Sks_3	Ips_4	Sks_4
Data Ke-6	0,961267606	1	0,925287 356	1	1	0,875	0,93457943 9	0,5
Data Ke-7	0,422535211	0.8571 43	0,590311 987	0,5714 29	0,638393	0,785714	0,25233644 9	0,571429

Setelah nilai *centroid* telah diperbarui langkah selanjutnya adalah membandingkan nilai *centroid* yang baru dengan yang sebelumnya. Berdasarkan hasil perbandingan nilai *centroid* yang baru di tabel 7 dengan yang sebelumnya di tabel 4 belum memiliki nilai yang sama maka perlu melakukan proses iterasi kembali.

Perhitungan Iterasi Kedua :

6. Kelompokan Data Sehingga Terbentuk K Buah Cluster Dengan Titik Centroid

Berikut ini contoh perhitungan jarak data pertama terhadap masing-masing *cluster* menggunakan persamaan (2) pada iterasi kedua :

- Perhitungan jarak data pertama dengan C1 :

$$\begin{aligned} &\sqrt{(0,478873 - 0,961267606)^2 + (1 - 1)^2} \\ &\sqrt{(0,478873 - 0,961267606)^2 + (1 - 1)^2} \\ &\sqrt{(1 - 1)^2 + (0,354167 - 1)^2 + (0,196261682 - 0,93457439)^2} \\ &\sqrt{(0,5 - 0,5)^2} \\ &=1,260780941 \end{aligned}$$

- Perhitungan jarak data pertama dengan C2 :

$$\begin{aligned} &\sqrt{(0,478873 - 0,422535211)^2 + (1 - 0,857143)^2} \\ &\sqrt{(0,701149425 - 0,590311987)^2 + (1 - 0,571429)^2} \\ &\sqrt{(0,354167 - 0,638393)^2 + (1 - 0,785714)^2} \\ &\sqrt{(0,19626182 - 0,252336449)^2} \\ &\sqrt{(0,5 - 0,571429)^2} \end{aligned}$$

=0,354489859

Lakukan yang sama untuk perhitungan data selanjutnya untuk mendapatkan jarak data pada masing-masing *cluster*. Berikut ini adalah hasil perhitungan jarak untuk setiap anggota *cluster* dari data ke 1 hingga ke 9 pada iterasi kedua yang ditunjukkan pada tabel 8 berikut :

Tabel. 8 Hasil Perhitungan Jarak Iterasi

2

MhsNiu	Jarak C1	Jarak C2	Jarak terdekat
G1A012002	1.260780941	0.354489859	0.354489859
G1A012003	0.011362027	1.046121934	0.011362027
G1A012004	2.373973676	0.594438419	0.594438419
G1A012005	1.960001235	0.648770948	0.648770948
G1A012006	0.356730338	0.521499041	0.356730338
G1A012007	0.011362027	1.408225974	0.011362027
G1A012008	1.133199348	0.307522155	0.307522155
G1A012010	5.66925583	2.289039696	2.289039696
G1A012011	1.571966278	1.099464038	1.099464038

Untuk data pertama memiliki jarak 1,260780941 terhadap *cluster* 1 dan memiliki jarak 0,354489859 terhadap *cluster* 2. Dari kedua *cluster* tersebut, data pertama memiliki jarak terdekat dengan *cluster* 2. Oleh karena itu data pertama pada iterasi kedua ini masuk dalam *cluster* 2. Untuk hasil pengelompokan setiap data berdasarkan jarak terdekatnya ditunjukkan pada tabel 9 berikut:

Setelah setiap data dikelompokkan berdasarkan jarak terdekatnya maka dapat disimpulkan bahwa pada iterasi kedua ini terjadi perubahan posisi dimana data ke-2, ke-5, dan ke-6 berada dalam *cluster* 1 sedangkan untuk data ke-1, ke-3, ke-4, ke-7, ke-8, dan ke-9 berada dalam *cluster* 2.

7. Perbaharui Nilai *Centroid*

Berikut ini contoh perhitungan nilai *centroid* terbaru terhadap masing-masing *cluster* menggunakan persamaan (3) pada iterasi kedua :

- Perhitungan Nilai *Centroid* Baru Untuk *Cluster* 1
- $$C_1 = \frac{(0,922535211+0,626760563+1)}{3}$$
- $$= 0,849765258$$

MhsNiu	C1	C2
GIA012002		1
GIA012003	1	
GIA012004		1
GIA012005		1
GIA012006	1	
GIA012007	1	
GIA012008		1
GIA012010		1
GIA012011		1

- Perhitungan Nilai *Centroid* Baru Untuk *Cluster* 2

$$C_2 = \frac{(0,478873239+0,408450704+0,366197183)}{6}$$

$$= \frac{0,521126761+0+0,556338028}{6}$$

$$= 0,388497653$$

Untuk hasil perhitungan nilai *centroid* baru pada masing-masing *cluster* pada iterasi ke-2 ditunjukkan pada tabel 10 berikut :

Tabel. 10 Hasil Perhitungan Nilai *Centroid* Baru Iterasi 2

Data Ke	Ips_1	Sks_1	Ips_2	Sks_2	Ips_3	Sks_3	Ips_4	Sks_4
Data Ke-6	0,849765258	1	0,92528735 6	1	0,965278	0,916667	0,778816199	0,5
Data Ke-7	0,388497653	0,833333	0,53448275 9	0,5	0,595486	0,75	0,216510903	0,583333

Berdasarkan hasil perbandingan nilai *centroid* terbaru pada tabel 10 dengan yang sebelumnya di tabel 7 ternyata masih belum memiliki nilai yang sama sehingga perlu melakukan proses iterasi kembali.

Perhitungan Iterasi Ketiga :

8. Kelompokan Data Sehingga Terbentuk K Buah *Cluster* Dengan Titik *Centroid*

Berikut ini contoh perhitungan jarak data pertama terhadap masing-masing *cluster* menggunakan persamaan (2) pada iterasi ketiga.

- Perhitungan jarak data pertama dengan **C1 :**

$$\sqrt{(0,478873 - 0,849765258)^2 + (1 - 1)^2}$$

$$\sqrt{(0,701149425 - 0,925287356)^2 + (1 - 1)^2}$$

$$\sqrt{(0,354166667 - 0,965278)^2 + (1 - 0,916666667)^2}$$

$$\sqrt{(0,196261682 - 0,778816199)^2 + (0,5 - 0,5)^2}$$

$$= 0,907569702$$

- Perhitungan jarak data pertama dengan **C2 :**

$$\sqrt{(0,478873 - 0,388497653)^2 + (1 - 0,833333)^2}$$

$$\sqrt{(0,701149425 - 0,534482759)^2 + (1 - 0,5)^2}$$

$$\sqrt{(0,354167 - 0,595486)^2 + (1 - 0,75)^2}$$

$$\sqrt{(0,196261682 - 0,216510903)^2 + (0,5 - 0,583333333)^2}$$

$$= 0,441812852$$

MhsNiu	C1	C2
G1A012002		1
G1A012003	1	
G1A012004		1
G1A012005		1
G1A012006	1	
G1A012007	1	
G1A012008		1
G1A012010		1
G1A012011		1

Lakukan langkah yang sama untuk perhitungan data selanjutnya untuk mendapatkan jarak data pada masing-masing *cluster*. Berikut ini adalah hasil perhitungan jarak untuk setiap anggota *cluster* dari data ke 1 hingga ke 9 pada iterasi ketiga yang ditunjukkan pada tabel 11 dibawah ini :

Tabel. 11 Hasil Perhitungan Jarak Iterasi

3

MhsNiu	Jarak C1	Jarak C2	Jarak terdekat
G1A012002	0.907569702	0.441812852	0.441812852
G1A012003	0.021980989	1.275038325	0.021980989
G1A012004	2.077612295	0.507795787	0.507795787
G1A012005	1.691413404	0.62629237	0.62629237
G1A012006	0.158546817	0.709818139	0.158546817
G1A012007	0.080016474	1.668475296	0.080016474
G1A012008	0.773968655	0.411290523	0.411290523

Untuk data pertama memiliki jarak 0,907569702 terhadap *cluster* 1 dan memiliki jarak 0,4418122852 terhadap *cluster* 2. Dari kedua *cluster* tersebut, data pertama mahasiswa memiliki jarak terdekat

dengan *cluster* 2. Oleh karena itu data pertama masuk dalam *cluster* 2.

Setelah setiap data dikelompokan berdasarkan jarak terdekatnya maka dapat disimpulkan bahwa pada iterasi ketiga ini tidak terjadi perubahan posisi dimana data ke-2, ke-5, dan ke-6 berada dalam *cluster* 1 sedangkan untuk data ke-1, ke-3, ke-4, ke-7, ke-8, dan ke-9 berada dalam *cluster* 2. Hasil dari pengelompokan data berdasarkan jarak terdekatnya pada iterasi ketiga ini memiliki hasil yang sama dengan iterasi kedua.

9. Perbaharui Nilai *Centroid*

Berikut ini contoh perhitungan nilai *centroid* terbaru terhadap masing-masing *cluster* menggunakan persamaan (3) pada iterasi ketiga.

• Perhitungan Nilai *Centroid* Baru Untuk *Cluster* 1

$$C_1 = \frac{(0,922535211+0,626760563+1)}{3}$$

$$= 0,849765$$

• Perhitungan Nilai *Centroid* Baru Untuk *Cluster* 2

$$C_2 = \frac{(0,478873239+0,408450704+0,366197183 + 0,521126761 + 0 + 0,556338028)}{6}$$

$$= 0,388497653$$

Untuk hasil perhitungan nilai *centroid* terbaru pada iterasi ketiga menggunakan persamaan (3) ditunjukkan pada tabel 13 berikut :

Tabel. 13 Hasil Perhitungan Nilai *Centroid* Baru Iterasi 3

Data Ke	Ips_1	Sks_1	Ips_2	Sks_2	Ips_3	Sks_3	Ips_4	Sks_4
Data Ke-6	0,849765258	1	0,925287356	1	0,965278	0,916667	0,778816199	0,5
Data Ke-7	0,388497653	0,833333	0,534482759	0,5	0,595486	0,75	0,216510903	0,5833333

Berdasarkan hasil perbandingan nilai *centroid* terbaru pada tabel 13 dengan nilai yang sebelumnya di tabel 10 ternyata hasilnya memiliki nilai yang sama sehingga proses iterasi dapat dihentikan.

D. Hasil Perhitungan Algoritma K-Means

Pada perhitungan ini menggunakan metode algoritma k-means dengan jumlah data yang digunakan berjumlah 9 data Pada

perhitungan ini iterasi terhenti pada iterasi ketiga karna tidak ada lagi anggota *cluster* yang berpindah. Dalam penelitian ini terdapat dua macam *cluster* yaitu tepat waktu dan tidak tepat waktu. Adapun hasil dari perhitungan menggunakan algoritma k-means ditunjukkan pada tabel 14 dibawah ini :

Tabel. 14 Tabel Hasil Perhitungan

No	MhsNiu	C1	C2	Keterangan
1	G1A012002		1	Tepat Waktu
2	G1A012003	1		Tepat Waktu
3	G1A012004		1	Tidak Tepat Waktu
4	G1A012005		1	Tidak Tepat Waktu
5	G1A012006	1		Tepat Waktu
6	G1A012007	1		Tepat Waktu
7	G1A012008		1	Tidak Tepat Waktu
8	G1A012010		1	Tidak Tepat Waktu
9	G1A012011		1	Tepat Waktu

Berdasarkan tabel 14 diatas dapat dilihat dari 9 data yang digunakan sebagai sampel untuk perhitungan Algoritma K-Means secara manual didapatkanlah 5 mahasiswa yang masuk kedalam *cluster* lulus dengan tepat waktu dan 4

mahasiswa yang masuk ke dalam *cluster* lulus dengan tidak tepat waktu.

E. Hasil Penelitian

Tujuan penelitian ini dilakukan penulis ialah untuk lama waktu penyelesaian masa studi oleh mahasiswa berdasarkan Indeks

Prestasi Semester dan Satuan Kredit Semester yang telah diselesaikan oleh mahasiswa. Algoritma yang digunakan penulis merupakan algoritma *K-Means* menggunakan dua buah kelompok atau *cluster*. Kelompok pertama mahasiswa digolongkan ke dalam mahasiswa yang berpotensi menyelesaikan studi tepat waktu dan kelompok kedua merupakan mahasiswa yang berpotensi terlambat menyelesaikan dalam menyelesaikan studi.

Untuk menganalisa apakah program yang telah penulis kembangkan berjalan dengan baik maka perlu diadakan perbandingan antara pengelompokan potensi yang dihasilkan program dengan data mahasiswa yang telah lulus sebelumnya. Dari data yang penulis dapatkan mahasiswa yang menyelesaikan studi kurang dari 4 tahun dikelompokkan mahasiswa tepat waktu dan mahasiswa yang menyelesaikan studi lebih dari 4 tahun digolongkan kedalam mahasiswa tidak tepat waktu. Data pembandingan yang digunakan merupakan data mahasiswa di program studi teknik informatika mulai dari tahun 2012 sampai 2016 yang telah berhasil menyelesaikan masa studi datanya berjumlah sebanyak 190 data yang terdapat pada bagian halaman lampiran. Hasil perbandingan prediksi dengan data asli ditunjukkan pada tabel 15 berikut :

Tabel. 15 Hasil Perbandingan Data Asli dan Output Sistem

Total Data	190
Prediksi Benar	151
Prediksi Salah	39
Presentase Prediksi Benar	79.48%
Presentase Prediksi Salah	20.52%

Berdasarkan tabel 15 diatas, dari total 190 data yang dibandingkan, sistem yang menggunakan algoritma *k-means* memprediksi 151 data dengan benar dan 39 data mengalami kesalahan prediksi. Jika dilihat berdasarkan presentase, sistem memprediksi 79,48% data dengan benar dan 20.52% data salah diprediksi. Jadi, dapat disimpulkan program yang dibuat dengan algoritma *k-means* menggunakan data indeks prestasi semester dan sistem kredit semester dapat memprediksi data dengan ketepatan sebesar 79.48%.

VI. KESIMPULAN DAN SARAN

Pada penulisan ini menjelaskan mengenai kesimpulan dan saran dari hasil penelitian.

A. Kesimpulan

Dari penelitian yang telah dilakukan dapat diambil beberapa kesimpulan yaitu :

1. Algoritma *kmean* berhasil di implementasikan dalam mengelompokkan mahasiswa kedalam 2 kelompok mahasiswa yang menyelesaikan studi tepat waktu dan tidak tepat waktu dengan variabel indeks prestasi semester dan sistem kredit semester dengan tingkat keberhasilan prediksi 79.48%.
2. Sistem pengelompokan potensi kelulusan mahasiswa fakultas teknik unib menggunakan algoritma *k-means* berhasil dikembangkan menggunakan sstem pemrograman berbasis *web*.

B. Saran

Dari kesimpulan diatas maka dapat dikemukakan berupa saran yang akan sangat membantu untuk pengembangan sistem ini selanjutnya diantaranya :

1. Untuk mengetahui pengembangan aplikasi data *minning* yang lainnya serta untuk mendapatkan keputusan yang beragam maka juga bisa menggunakan algoritma lainnya untuk mengetahui perbedaan dan cara kerja dari setiap masing-masing algoritma dalam data *minning*.
2. Untuk peneliti yang akan melakukan penelitian serupa dengan ini akan sangat menarik dan lebih baik apabila parameter yang digunakan ditambah lebih banyak lagi guna mendapat hasil yang lebih optimal lagi.

REFERENSI

- [1] Ediyanto, Mara, M. N., & Satyahadewi, N. (2013). Pengklasifikasian Karakteristik Dengan Metode K-Means Cluster Analysis. *Buletin Ilmiah Mat. Stat. dan Terapannya (Bimaster)*, 133-136.
- [2] He, H., & Tan, Y. (2012). A Two-Stage Genetic Algorithm for Automatic Clustering. *Neurocomputing*, 81, 49-59.
- [3] Mustaffa, Z., & Yusuf, Y. (2011). A Comparison of Normalization Techiques in Predicting Dengue Outbreak. *International Conference on Business and Economics Reasearch*, 01, 345-349.
- [4] Prasetyo, E. (2012). *Data Mining Konsep dan Aplikasi Menggunakan Matlab*. Yogyakarta: Andy.
- [5] Pressman, R. S. (2012). *Rekayasa Perangkat Lunak : Pendekatan Praktisi Edisi 7*. Yogyakarta: Penerbit Andi.
- [6] Priati, A. F. (2017). Data Minning Dengan Teknik Culustering Menggunakan Algoritma K-Means Pada Data Transaksi Superstore. *Seminar Nasional Informatika dan Aplkasinya (SNIA)*, 15-19.
- [7] Priyatman, H., Sajid, F., & Haldivany, D. (2019). Klasterisasi Menggunakan Algoritma K-Means Clustering Untuk Memprediksi Waktu Kelulusan Mahasiswa. *Jurnal Edukasi dan Penelitian Informatika*, 5, 62-66.
- [8] Shofiani, N. (2017). *Segmentasi Supplier Menggunakan Metode K-Means Clustering (Studi Kasus: PTPN X PG Merritjan)*. Surabaya: Institut Teknologi Sepuluh Nopember.
- [9] Suntoro, J. (2019). *Data Mining Algoritma dan Implementasi Dengan Pemrograman PHP*. Jakarta: Elex Media Computindo.
- [10] Wardhani, A. K. (2016). Implementasi Algoritma K-Means Untuk Pengelompokan Pasien Pada Puskesmas Kajen Pekalongan. *Jurnal Transformatika*, 14, 30-36.
- [11] Xiang, W. L., Zhu, N., Mha, S. F., Meng, X. L., & An, M. Q. (2015). A Dynamic Shuffled Differential Evolution Algorithm For Data Clustering. *Neurocomputing*, 158, 144-154.
- [12] Zhang, C., Ouyang, D., & Ning, J. (2010). An Artificial bee Colony approach For Clustering. *Expert Systems with Applications*, 37(7), 4761-4767.