

Metode K-Means untuk Mengelompokkan Alumni Berdasarkan Waktu Mencari Pekerjaan

(K-Means Method for Grouping Alumni Based on Time of Finding a Job)

Ridho Muktiadi¹, Abid Yanuar Badharudin²

^{1,2}*Teknik Informatika, Fakultas Teknik dan Sains, Universitas Muhammadiyah Purwokerto, Jalan Raya Dukuwaluh PO.BOX 202 Purwokerto, Jawa Tengah 53182*

¹ridho.muktiadi@gmail.com

²zaidsoft.indonesia@gmail.com

ABSTRAK

Alumni Universitas Muhammadiyah Purwokerto (UMP) sampai dengan tahun 2018 mencapai 31.178 alumni. Melalui alumni perguruan tinggi dapat melakukan perbaikan dan penyesuaian kurikulum sesuai dengan kebutuhan dunia kerja. *Tracer study* menjadi upaya UMP untuk memperoleh data alumni dari semua tahun lulusan yang sudah disajikan dalam bentuk sistem informasi *tracer study* (tracer.ump.ac.id). Hasil dari *tracer study* dikelompokkan berdasarkan waktu mencari kerja dengan menggunakan metode K-Means. Terdapat 3 kelompok yaitu: Sangat Baik, kelompok 2 yaitu Baik, Cukup dan Kurang dan kelompok 3 : Sangat Kurang. Sistem dibuat menggunakan bahasa pemrograman PHP dan MySQL sebagai sarana penyimpanan data. Pengujian sistem dilakukan dengan menguji sistem secara fungsionalitas dan pengujian penerapan metode K-Means. Dari hasil pengujian sistem yang dilakukan menunjukkan sistem dapat mengelompokkan data *tracer study* sebanyak 2.716 ke dalam 3 kelompok yang sudah ditentukan.

Kata Kunci: K-Means, Waktu Mencari Kerja, Tracer Study, Alumni

ABSTRACT

Alumni of the Universitas Muhammadiyah Purwokerto (UMP) up to 2018 reached 31,178 alumni. Through college alumni can make improvements and adjust curriculum according to the needs of the workforce. Tracer study is an effort of the UMP to obtain alumni data from all graduate years which has been presented in the form of a tracer study information system (tracer.ump.ac.id). The results of the tracer study are grouped by the time they look for work using the K-Means method. There are 3 groups: Very Good, group 2: Good, Enough and Poor and group 3: Very Poor. The system is made using PHP and MySQL programming languages as a means of data storage. System testing is done by testing the system's functionality and testing the application of the K-Means method. From the results of system testing conducted shows the system can group 2,716 tracer study data into 3 predefined groups.

Keywords: K-Means, Time Seeking Work, Tracer Study, Alumni

PENDAHULUAN

Alumni adalah orang-orang yang telah mengikuti atau tamat dari suatu sekolah atau perguruan tinggi (Kemendikbud). Alumni Universitas Muhammadiyah Purwokerto (UMP) sampai dengan tahun 2018 mencapai 31.178 alumni. Tracer study menjadi upaya

UMP untuk memperoleh data alumni dari semua tahun lulusan yang sudah disajikan dalam bentuk sistem informasi tracer study (tracer.ump.ac.id). Waktu mencari pekerjaan menjadi sebuah informasi bagaimana alumni mampu bersaing dalam memperoleh pekerjaan, serta sebagai dasar UMP meningkatkan kualitas dari lulusan. *Clustering* adalah proses untuk mengorganisasikan sekelompok data kedalam kelompok-kelompok sedemikian rupa sehingga obyek-obyek yang serupa akan menjadi satu *cluster* sedangkan obyek-obyek yang tidak serupa menjadi anggota *cluster* yang lain. Dalam setiap *cluster* akan berisi data yang semirip mungkin. Ukuran kemiripan biasanya dihitung dengan jarak. Jarak dalam satu *cluster* dibuat sedekat mungkin jadi dalam satu *cluster* harus semirip mungkin dan dengan *cluster* yang lain harus berbeda (Muflikhah, 2018). Metode K-Means adalah metode yang termasuk dalam algoritma *clustering* berbasis jarak yang membagi data ke dalam sejumlah *cluster* dan algoritma ini hanya bekerja pada atribut numerik (Witten, 2012). K-means mempartisi objek yang ada kedalam satu atau lebih *cluster* atau kelompok berdasarkan karakteristiknya, sehingga objek akan dikelompokkan berdasarkan karakteristik yang sama pada sebuah *cluster* dan objek yang mempunyai karakteristik yang berbeda dikelompokkan kedalam *cluster* yang lain. Pengelompokan data dengan karakteristik yang mirip pada setiap kelompok menggunakan metode ini diharapkan mampu mempermudah mengetahui keadaan data alumni berdasarkan waktu mencari pekerjaan sehingga bisa digunakan sebagai acuan peningkatan kualitas lulusan.

METODE

Data diperoleh dengan cara langsung memintakan data tersebut ke pihak pengelola sistem informasi *tracer study* di UMP. Sebelum mengelompokkan data, terlebih dahulu perlu dibuat *cluster* sebagai wadah pengelompokan data tersebut. Mengacu pada Buku VI Matriks Penilaian Instrumen Akreditasi Program Studi Sarjana tahun 2008 (Badan Akreditasi Nasional Perguruan Tinggi, 2008) maka akan ada tiga *cluster* pada sistem ini; kelompok 1 : Sangat Baik (waktu tunggu ≤ 3 bulan), kelompok 2 : Baik, Cukup dan Kurang ($3 \text{ bulan} < \text{waktu tunggu} < 18 \text{ bulan}$) dan kelompok 3 : Sangat Kurang (waktu tunggu ≥ 18 bulan). Kemudian tentukan pusat *cluster* dengan menggunakan persamaan 1 (Prasetyo, 2012).

$$C_i = \frac{1}{M} \sum_{j=1}^M x_j \quad (1)$$

Langkah selanjutnya, hitung jarak setiap data terhadap pusat *cluster* dengan menggunakan persamaan 2.

$$d = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \quad (2)$$

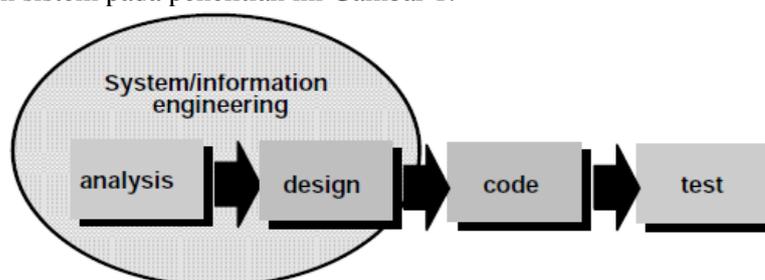
Kemudian kelompokkan data ke dalam *cluster* dengan jarak terpendek menggunakan persamaan 3.

$$a_{i1} = \begin{cases} 1 & d = \min \{D(x_i, c_1)\} \\ 0 & \text{lainnya} \end{cases} \quad (3)$$

Tahap terakhir, ulangi langkah 3, apabila masih ada data yang berpindah kelompok atau apabila ada perubahan nilai centroid.

HASIL DAN PEMBAHASAN

Data yang digunakan adalah data alumni yang sudah mengisi tracer study, data diperoleh dari Alumni Center UMP. Alumni yang sudah mengisi *tracer study* berjumlah 2716 alumni. Jumlah *cluster* ditentukan berdasarkan Buku VI Matriks Penilaian Instrumen Akreditasi Program Studi Sarjana tahun 2008, maka akan ada tiga *cluster* pada sistem ini; kelompok 1 : Sangat Baik (waktu tunggu ≤ 3 bulan), kelompok 2 : Baik, Cukup dan Kurang (3 bulan < waktu tunggu < 18 bulan) dan kelompok 3 : Sangat Kurang (waktu tunggu ≥ 18 bulan). Berdasarkan hal tersebut maka terdapat 3 *cluster* yang akan digunakan, yaitu: *cluster* Sangat Baik, *cluster* Baik, Cukup dan Kurang dan *cluster* Sangat Kurang. Sistem dibuat menggunakan bahasa pemrograman PHP dan database MySQL sebagai sarana penyimpanan data. Model *Waterfall* digunakan sebagai model pengembangan sistem pada penelitian ini Gambar 1.



Gambar 1. Tahapan Model *Waterfall* (Pressman, 2002)

Sesuai dengan Gambar 1 maka akan ada 4 tahapan dalam pengembangan sistem yang harus dilakukan. Hasil dari pengembangan sistem yang telah dibuat, sebagai berikut:

1. *Analysis*

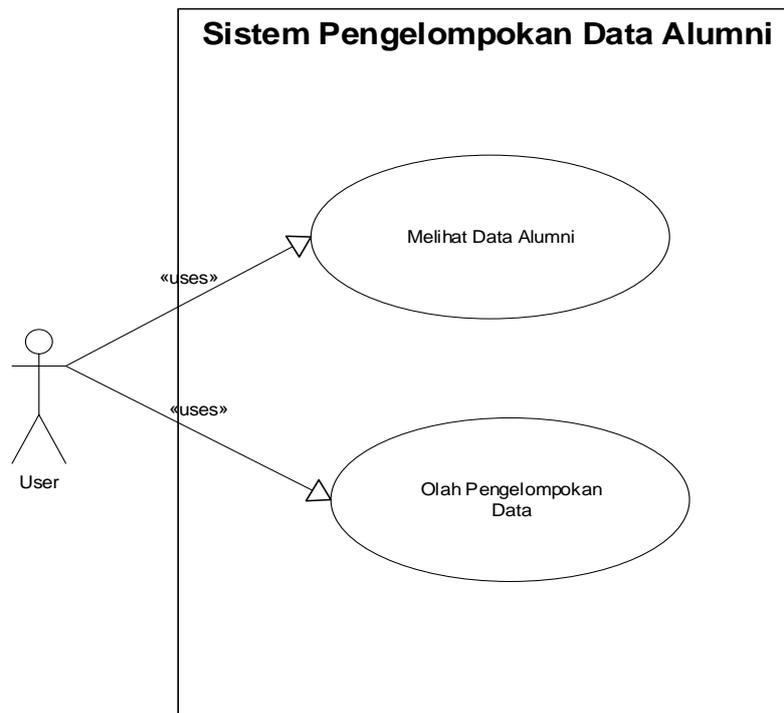
Tahap ini sudah terpenuhi berdasarkan tahap pengumpulan data, dimana data yang akan digunakan yaitu data alumni UMP yang sudah mengisi *tracer study*. Berdasarkan data tersebut dapat diketahui kebutuhan terkait data apa saja yang akan digunakan, kebutuhan tersebut meliputi: NIM, Nama Mahasiswa, Prodi, Fakultas, Tahun Lulus dan Jawaban pertanyaan waktu yang digunakan untuk mencari pekerjaan.

2. *Design*

Design yang dibuat meliputi: *Use Case Diagram* dan *Activity Diagram*.

a. *Use Case Diagram*

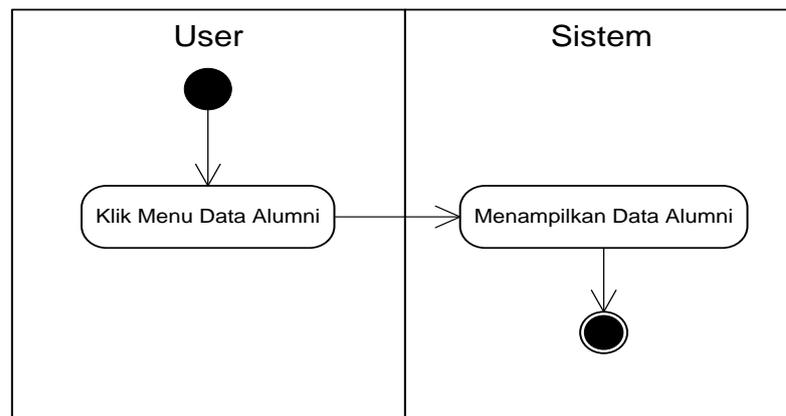
Use Case Diagram pada Gambar 2 menunjukkan *user* dapat melakukan 2 proses utama yaitu melihat data alumni dan olah pengelompokan data. Proses lihat data alumni *user* dapat mengetahui jumlah data yang akan diproses beserta isi data alumni dalam bentuk tabel. Pada proses pengelompokan data, *user* dapat mengetahui hasil dari pengelompokan data terhadap data alumni.



Gambar 2. Use Case Diagram Sistem Pengelompokan Data Alumni

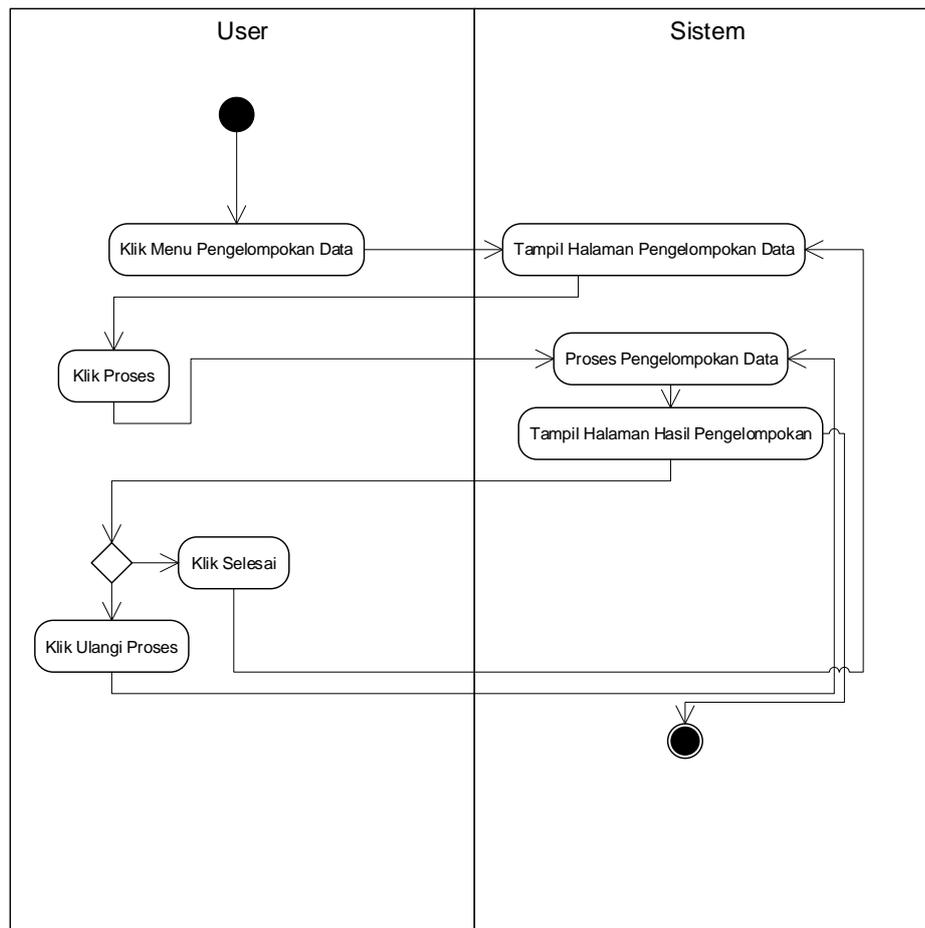
b. *Activity Diagram*

Berdasarkan *Use Case Diagram* pada Gambar 2 terdapat 2 proses utama, sehingga pada tahap perancangan ini dibuat 2 *activity diagram* yang ditunjukkan pada Gambar 5 dan Gambar 3.



Gambar 3. Activity Diagram Lihat Data Alumni

Activity Diagram pada Gambar 3 memperlihatkan bahwa *user* dapat melihat data alumni dengan terlebih dahulu mengklik menu lihat data alumni yang terdapat pada sistem.



Gambar 4. Activity Diagram Pengelompokan Data Alumni

Activity Diagram pada Gambar 4 memperlihatkan bahwa *user* dapat melakukan proses pengelompokan data dimulai saat *user* meng-klik tombol Proses, maka data yang berupa data alumni UMP akan diproses dan ditentukan pengelompokan dari masing-masing data.

3. Code

Tahap ini merupakan tahap implementasi dari tahap *design* dan penerapan metode K-Means *dengan* menggunakan bahasa pemrograman PHP. Berikut hasil implementasi menggunakan bahasa pemrograman PHP.



Gambar 5. Halaman Pengelompokan Data

Tombol Mulai Proses pada Gambar 5 digunakan untuk mengelompokkan data alumni berdasarkan waktu yang digunakan untuk memperoleh pekerjaan pertama. Berikut akan dijelaskan masing-masing tahap dari sistem saat menampilkan dan memproses data.

a. Alokasi Data Secara Acak dan Penentuan *Centroid* kelompok

Pada Gambar 6 diperlihatkan hasil dari pengalokasian data secara acak ke kelompok yang tersedia. Alokasi ini bersifat acak, jadi semua data dibagi 3 dan masing-masing bagian akan dimasukkan ke tiga kelompok yang sudah dibuat.

No	Nim	Nama	Prodi	Fakultas	Tahun Lulus	Waktu Sebelum	Waktu Sesudah	Normalisasi Waktu Sebelum	Normalisasi Waktu Sesudah	Kelompok
1	1420104021	CANGGIH PUTRA KHARISMA			2017	0	2	0.0000	0.0408	Kelompok 1
2	1420104014	RAZAK AKBAR TRIADI			2017	12	0	0.1111	0.0000	Kelompok 1
3	1420104027	GHUFRONI			2017	12	0	0.1111	0.0000	Kelompok 1
4	1220104016	MUFLIHAH			2017	0	0	0.0000	0.0000	Kelompok 1
5	1420104025	JOKO MULYANTO			2018	0	0	0.0000	0.0000	Kelompok 1

Gambar 6. Alokasi Data Acak Perkelompok

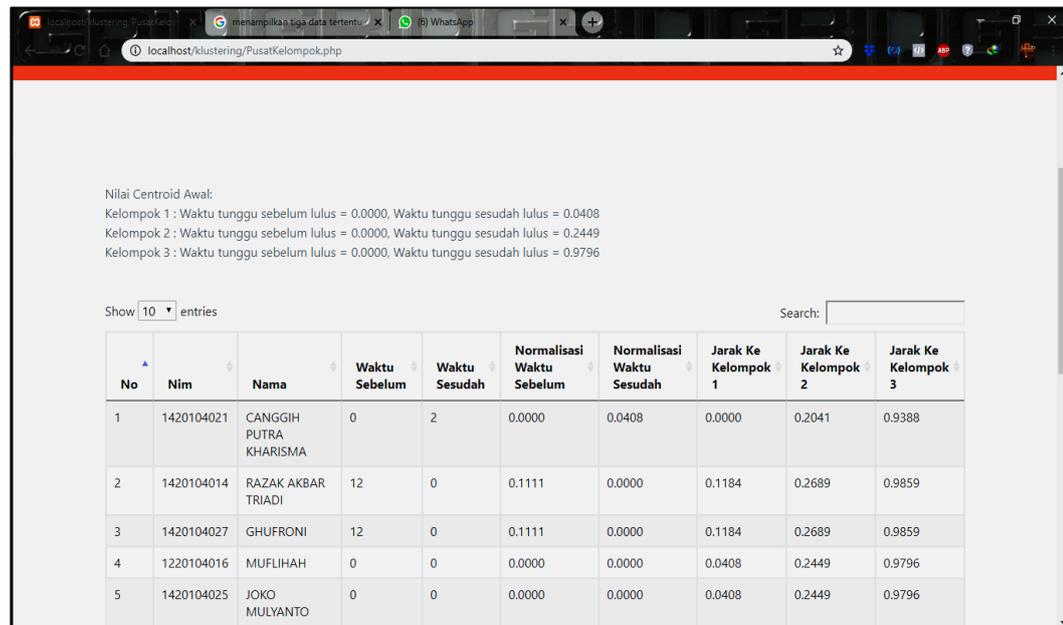
Pada halaman yang sama juga diperlihatkan nilai *centroid* awal Gambar 7, yang diambil secara acak.

Kelompok	NIM	Nama	Centroid Waktu Sebelum	Centroid Waktu Sesudah
Kelompok 1	1420104021	CANGGIH PUTRA KHARISMA	0.0000	0.0408
Kelompok 2	1404020039	FANI EKA KUSWANTI	0.0000	0.2449
Kelompok 3	1001100198	FATIH NURIJEKI	0.0000	0.9796

Gambar 7. Alokasi Data Acak Perkelompok

b. Pengelompokan data ke *centroid* terdekat

Pada tahap ini sistem akan menghitung jarak ke masing-masing *centroid*. Sistem akan menampilkan semua jarak dari masing-masing *centroid* kelompok Gambar 8.



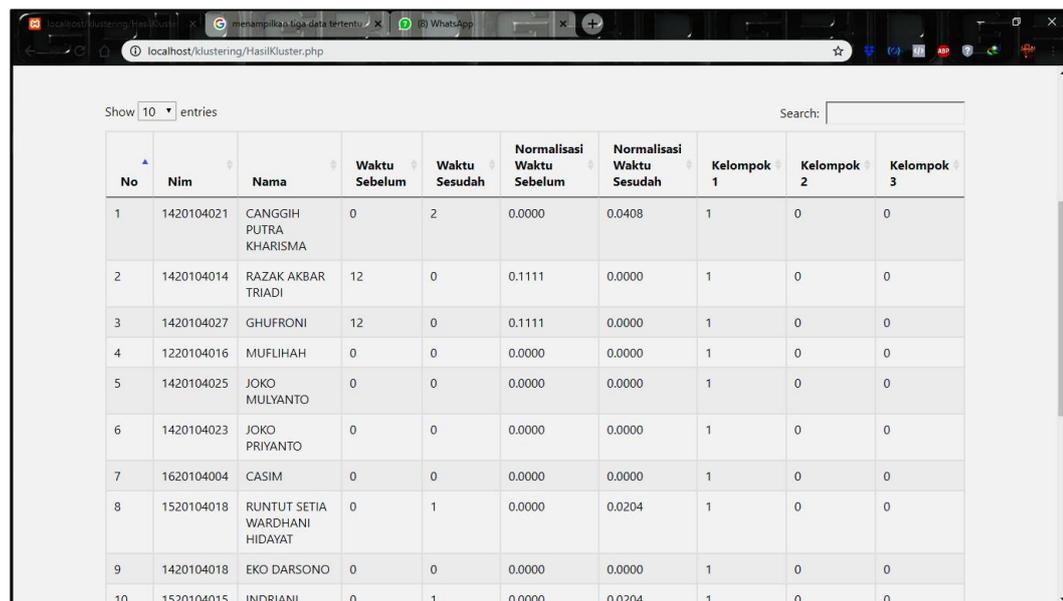
Nilai Centroid Awal:
 Kelompok 1 : Waktu tunggu sebelum lulus = 0.0000, Waktu tunggu sesudah lulus = 0.0408
 Kelompok 2 : Waktu tunggu sebelum lulus = 0.0000, Waktu tunggu sesudah lulus = 0.2449
 Kelompok 3 : Waktu tunggu sebelum lulus = 0.0000, Waktu tunggu sesudah lulus = 0.9796

Show 10 entries

No	Nim	Nama	Waktu Sebelum	Waktu Sesudah	Normalisasi Waktu Sebelum	Normalisasi Waktu Sesudah	Jarak Ke Kelompok 1	Jarak Ke Kelompok 2	Jarak Ke Kelompok 3
1	1420104021	CANGGIH PUTRA KHARISMA	0	2	0.0000	0.0408	0.0000	0.2041	0.9388
2	1420104014	RAZAK AKBAR TRIADI	12	0	0.1111	0.0000	0.1184	0.2689	0.9859
3	1420104027	GHUFRONI	12	0	0.1111	0.0000	0.1184	0.2689	0.9859
4	1220104016	MUFLIHAH	0	0	0.0000	0.0000	0.0408	0.2449	0.9796
5	1420104025	JOKO MULYANTO	0	0	0.0000	0.0000	0.0408	0.2449	0.9796

Gambar 8. Penentuan Jarak Kelompok

Setelah jarak masing-masing kelompok ditentukan, selanjutnya sistem akan menampilkan hasil pengelompokan data Gambar 9.



Show 10 entries

No	Nim	Nama	Waktu Sebelum	Waktu Sesudah	Normalisasi Waktu Sebelum	Normalisasi Waktu Sesudah	Kelompok 1	Kelompok 2	Kelompok 3
1	1420104021	CANGGIH PUTRA KHARISMA	0	2	0.0000	0.0408	1	0	0
2	1420104014	RAZAK AKBAR TRIADI	12	0	0.1111	0.0000	1	0	0
3	1420104027	GHUFRONI	12	0	0.1111	0.0000	1	0	0
4	1220104016	MUFLIHAH	0	0	0.0000	0.0000	1	0	0
5	1420104025	JOKO MULYANTO	0	0	0.0000	0.0000	1	0	0
6	1420104023	JOKO PRIYANTO	0	0	0.0000	0.0000	1	0	0
7	1620104004	CASIM	0	0	0.0000	0.0000	1	0	0
8	1520104018	RUNTA SETIA WARDHANI HIDAYAT	0	1	0.0000	0.0204	1	0	0
9	1420104018	EKO DARSONO	0	0	0.0000	0.0000	1	0	0
10	1520104015	INDRIANI	0	1	0.0000	0.0204	1	0	0

Gambar 9. Hasil Pengelompokkan Data

Tahap akhir dari pengelompokkan data alumni, sistem akan menampilkan kesimpulan hasil pengelompokkan data dengan disertai keterangan kelompok terbaru beserta keterangan dari kelompok tersebut Gambar 10.

No	Nim	Nama	Waktu Sebelum	Waktu Sesudah	Normalisasi Waktu Sebelum	Normalisasi Waktu Sesudah	Kelompok	Keterangan
1	1420104021	CANGGIH PUTRA KHARISMA	0	2	0.0000	0.0408	Kelompok 1	Sangat Baik
2	1420104014	RAZAK AKBAR TRIADI	12	0	0.1111	0.0000	Kelompok 1	Sangat Baik
3	1420104027	GHUFRONI	12	0	0.1111	0.0000	Kelompok 1	Sangat Baik
4	1220104016	MUFLIAH	0	0	0.0000	0.0000	Kelompok 1	Sangat Baik
5	1420104025	JOKO MUYANTO	0	0	0.0000	0.0000	Kelompok 1	Sangat Baik
6	1420104023	JOKO PRIYANTO	0	0	0.0000	0.0000	Kelompok 1	Sangat Baik
7	1620104004	CASIM	0	0	0.0000	0.0000	Kelompok 1	Sangat Baik
8	1520104018	RUNTUT SETIA WARDHANI HIDAYAT	0	1	0.0000	0.0204	Kelompok 1	Sangat Baik
9	1420104018	EKO DARSONO	0	0	0.0000	0.0000	Kelompok 1	Sangat Baik
10	1520104015	INDRIANI WAHYUJATI	0	1	0.0000	0.0204	Kelompok 1	Sangat Baik

Gambar 10. Hasil Akhir Pengelompokkan Data Tracer Study Berdasarkan Lama Waktu Tunggu

Pengujian

Tahap terakhir dalam penelitian ini adalah melakukan pengujian terhadap sistem yang sudah dibuat. Pengujian yang dilakukan meliputi:

1. Pengujian Fungsionalitas Sistem

Tabel 1. Pengujian Fungsionalitas Sistem

No	Komponen	Fungsi	Status	Keterangan
1	http://localhost/klusteri ng/utama.php	Menampilkan halaman awal sistem	Sesuai	-
2	Menu Navigasi	Mengarahkan ke tampilan yang halaman sesuai dengan yang dipilih	Sesuai	-
3	Tombol Mulai Proses	Menampilkan Alokasi Data Secara Acak dan Penentuan Centroid Perkelompok	Sesuai	-
4	Tombol Alokasi Data Ke Centroid Terdekat	Menampilkan hasil perhitungan jarak data terhadap masing-masing <i>centroid</i> kelompok	Sesuai	-
5	Tombol Kembali Ke Awal	Menampilkan halaman utama	Sesuai	-
6	Tombol Hasil Pengelompokkan	Menampilkan hasil pengelompokan dengan menampilkan nilai 0 atau 1 perkelompok	Sesuai	-
7	Tombol Hasil dan Kesimpulan	Menampilkan hasil pengelompokkan baru	Sesuai	-
8	Tombol Ulangi Proses	Menampilkan Alokasi Data Secara Acak dan Penentuan Centroid Perkelompok	Sesuai	-

2. Pengujian metode K-Means

Langkah awal dalam pengujian ini adalah menentukan *centroid* dari masing-masing kelompok. Pusat *centroid* diambil secara acak yang menjadi perwakilan dari masing-masing kelompok. Hasil dari penentuan *centroid* dapat dilihat pada Tabel 2.

Tabel 2. Penentuan *Centroid* Kelompok

Kelompok	NIM	Nama	Centroid Waktu Sebelum	Centroid Waktu Sesudah
Kelompok 1	1420104021	CANGGIH PUTRA KHARISMA	0	0.0408
Kelompok 2	1404020039	FANI EKA KUSWANTI	0	0.2449
Kelompok 3	1001100198	FATIH NURIJEKI	0	0.9796

Setelah itu, pengujian dilanjutkan dengan menghitung jarak terdekat terhadap *centroid* menggunakan data: NIM = 1420104014 dan Nama = RAZAK AKBAR TRIADI dengan waktu sebelum = 0.1111 dan waktu sesudah 0.0000. Data ini digunakan untuk menghitung jarak ke *centroid* masing-masing kelompok.

- Menghitung jarak ke *centroid* kelompok 1

$$\sqrt{(0,1111 - 0)^2 + (0,0000 - 0,0408)^2}$$

$$= 0.1184$$

- Menghitung jarak ke *centroid* kelompok 2

$$\sqrt{(0,1111 - 0)^2 + (0,0000 - 0,2449)^2}$$

$$= 0.2689$$

- Menghitung jarak ke *centroid* kelompok 3

$$\sqrt{(0,1111 - 0)^2 + (0,0000 - 0,9796)^2}$$

$$= 0.9859$$

Dari perhitungan di atas diperoleh data dengan NIM = 1420104014 dengan *centroid* Kelompok 1 diperoleh hasil 0.1184, dengan *centroid* Kelompok 2 diperoleh hasil 0.2689 dan dengan *centroid* Kelompok 3 diperoleh hasil 0.9859. Berdasarkan hasil perhitungan di atas dapat disimpulkan bahwa, jarak yang paling dekat adalah *centroid* Kelompok 1. Sehingga data tersebut dimasukkan ke dalam Kelompok 1.

Berdasarkan hasil di atas dan dibandingkan dengan hasil pengolahan sistem, diperoleh kesamaan. Data dengan NIM = 1420104014 pada sistem juga berada pada kelompok 1 disertai dengan nilai *centroid* yang sama (Gambar 11). Sesuai hasil tersebut berarti sistem sudah sesuai dalam menerapkan metode K-Means dalam mengelompokkan data alumni berdasarkan waktu mencari kerja.

Nilai Centroid Awal:
 Kelompok 1 : Waktu tunggu sebelum lulus = 0.0000, Waktu tunggu sesudah lulus = 0.0408
 Kelompok 2 : Waktu tunggu sebelum lulus = 0.0000, Waktu tunggu sesudah lulus = 0.2449
 Kelompok 3 : Waktu tunggu sebelum lulus = 0.0000, Waktu tunggu sesudah lulus = 0.9796

Show 10 entries

No	Nim	Nama	Waktu Sebelum	Waktu Sesudah	Normalisasi Waktu Sebelum	Normalisasi Waktu Sesudah	Jarak Ke Kelompok 1	Jarak Ke Kelompok 2	Jarak Ke Kelompok 3
1	1420104021	CANGGIH PUTRA	0	2	0.0000	0.0408	0.0000	0.2041	0.9388
2	1420104014	RAZAK AKBAR	12	0	0.1111	0.0000	0.1184	0.2689	0.9859
3	1420104027	GHUFRONI	12	0	0.1111	0.0000	0.1184	0.2689	0.9859
4	1220104016	MUFLIHAH	0	0	0.0000	0.0000	0.0408	0.2449	0.9796
5	1420104025	JOKO MULYANTO	0	0	0.0000	0.0000	0.0408	0.2449	0.9796
6	1420104023	JOKO PRIYANTO	0	0	0.0000	0.0000	0.0408	0.2449	0.9796
7	1620104004	CASIM	0	0	0.0000	0.0000	0.0408	0.2449	0.9796

Gambar 11. Perbandingan Hasil Perhitungan Sistem

KESIMPULAN

Penelitian ini berhasil membangun sebuah sistem untuk dapat mengelompokkan data alumni menggunakan metode K-Means berdasarkan waktu alumni mencari pekerjaan. Terdapat 3 kelompok yang digunakan dalam penelitian ini, dan setelah melakukan pengujian, sistem yang dibuat bisa mengelompokkan data alumni yang sudah mengisi *tracer study* sesuai jarak terdekat dengan *centroid* masing-masing kelompok.

DAFTAR PUSTAKA

- Badan Akreditasi Nasional Perguruan Tinggi, 2008, *Buku VI Matriks Penilaian Instrumen Akreditasi Program Studi Sarjana*, Jakarta.
- Kemendikbud (Pusat Bahasa), Kamus Besar Bahasa Indonesia, <https://kbbi.web.id/alumni>.
- Muflikhah, L., Ratnawati, D.E., dan Putri, R.R.M. 2018. *Buku Ajar Data Mining*. UB Press. Malang.
- Prasetyo, E., 2012, *Data Mining: Konsep dan Aplikasi Menggunakan MATLAB*, Andi, Yogyakarta
- Pressman, R.S. 2002. *Rekayasa Perangkat Lunak: Pendekatan Praktisi (Buku Dua)*. Andi, Yogyakarta.
- Witten, et al., 2012, *Data Mining Practical Machine Learning Tools and Technique, 2nd edition*, Morgan Kaufmann, San Fransisco.