

KLASTERISASI DATA HASIL STUDI PELACAKAN TENTANG KARIR DAN PEKERJAAN LULUSAN PERGURUAN TINGGI MENGGUNAKAN ALGORITMA K-MEANS

Joko Sutrisno¹, Arief Wibowo², Bayu Satria Pratama³

^{1,2,3} Universitas Budi Luhur, Jl. Ciledug Raya Petukangan Utara, Jakarta Selatan, Indonesia 12260

²Email: arief.wibowo@budiluhur.ac.id

ABSTRAK

Perguruan tinggi memiliki tanggung jawab untuk menghasilkan lulusan yang berkualitas. Salah satu indikator kualitas lulusan adalah status mendapatkan pekerjaan, kondisi kesesuaian bidang pekerjaan dengan program pendidikan yang ditempuh dan kondisi masa tunggu mendapatkan pekerjaan tersebut. Hal yang dilakukan untuk mengetahui kondisi-kondisi tersebut adalah dengan melakukan studi pelacakan (*tracking study*) bagi lulusan. Penelitian ini menganalisis data hasil studi pelacakan lulusan perguruan tinggi tentang karir dan pekerjaan menggunakan algoritma klasterisasi data mining yaitu *K-Means*. Hasil penelitian menunjukkan bahwa analisis data studi pelacakan mampu membentuk beberapa kluster lulusan dengan nilai evaluasi Davies-Bouldin *Index* (DBI) mencapai 0,287 pada uji coba pertama dan 0,291 pada uji coba kedua. Kluster-kluster yang terbentuk terdiri dari kelompok lulusan dengan status belum bekerja, atau sedang bekerja. Profil lulusan setiap kluster dapat diketahui berupa masa tunggu yang relatif singkat kurang dari enam bulan untuk mendapatkan pekerjaan pertama, atau masa tunggu yang relatif lambat di atas satu tahun. Spesifikasi kluster lain yang terbentuk adalah tentang profil lulusan dengan tingkat kesesuaian antara pendidikan yang ditempuh dengan bidang pekerjaan yang dilakukan. Hasil penelitian ini menjadi umpan balik bagi pengelola program studi untuk mengukur kualitas lulusan serta perbaikan proses pendidikan yang perlu dilakukan.

Kata kunci: klasterisasi, studi pelacakan, kesesuaian kerja, masa tunggu bekerja

ABSTRACT

Higher education has a responsibility to produce quality graduates. One indicator of the quality of graduates is the status of getting a job, the condition of the suitability of the field of work with the educational program pursued, and the waiting period to get the job. What is being done to find out these conditions is to conduct a tracer study for graduates. This study analyzes data from a college graduate tracking study about careers and jobs using a data mining clustering algorithm, namely K-Means. The results showed that the analysis of the tracking study data formed several graduate clusters with an evaluation value of the Davies-Bouldin Index (DBI) reaching 0.287 in the first trial and 0.291 in the second trial. The clusters formed consist of groups of graduates with status still needing to be working or currently working. The profile of graduates from each cluster can be identified in the form of a relatively short waiting period of less than six months to get a first job or a relatively slow waiting period of more than one year. Another cluster specification that is formed is about the profile of graduates with the level of compatibility between the education attained and the field of work carried out. The results of this study serve as feedback for study program managers to measure the quality of graduates and the improvements in the educational process that need to be made.

Keywords: *clustering, tracking studies, work suitability, job waiting periods*

1. PENDAHULUAN

Perguruan tinggi merupakan tempat dilaksanakannya pendidikan tinggi atau jenjang pendidikan yang disediakan bagi lulusan Sekolah Lanjutan Tingkat Atas. Perguruan tinggi mengelola program-program studi yang ditawarkan kepada masyarakat. Salah satu indikator peminatan dari masyarakat terhadap sebuah program studi adalah kualitas lulusan. Lulusan perguruan tinggi yang diharapkan para orang tua atau mahasiswa adalah keadaan cepat mendapatkan kerja, jenis pekerjaan yang memiliki kesesuaian dengan bidang ilmu serta waktu tunggu yang relatif cepat dalam mendapatkan pekerjaan pertama kali.

Program Studi merupakan satuan pengelola pendidikan yang memiliki kewenangan memberikan pengajaran, ujian dan menyatakan kelulusan bagi mahasiswa. Salah satu kebutuhan informasi bagi program studi adalah mengetahui hasil studi pelacakan (*tracer study*) berupa profil karir atau pekerjaan lulusan. Umumnya ada beberapa indikator yang ditanyakan dalam kegiatan studi pelacakan lulusan, meliputi berapa lama waktu tunggu untuk mendapatkan pekerjaan pertama, kesesuaian bidang kerja dengan program pendidikan yang ditempuh, serta penilaian dari pengguna lulusan di dunia kerja.

Telah banyak dilakukan berbagai penelitian yang menganalisis data lulusan atau perguruan tinggi dengan metode data mining klusterisasi. Peneliti terdahulu melakukan analisis data akademik menggunakan algoritma *K-Means* yang hasilnya berupa model klusterisasi kelompok mahasiswa yang berprestasi, berpotensi prestasi, berpotensi masalah, dan kelompok mahasiswa yang bermasalah [1]. Peneliti lain melakukan klusterisasi dengan algoritma yang sama, untuk menentukan predikat kelulusan mahasiswa sebagai data menganalisis kualitas lulusan [2].

Peneliti lainnya melakukan pengelompokan tingkat kepuasan pengguna lulusan perguruan tinggi berupa pengelompokan nilai kepuasan sangat puas, puas dan cukup puas. Pemodelan diselesaikan dengan berbantuan aplikasi WEKA menggunakan algoritma *K-Means* [3]. Penelitian tentang data institusi pendidikan juga dilakukan peneliti lain yang melakukan klusterisasi sekolah dalam rangka pendampingan oleh dinas terkait, menggunakan algoritma *K-Means* berdasarkan fasilitas, pendidik, dan tenaga pendidik. Studi tersebut menggunakan Davies-Bouldin *Index* untuk mengevaluasi model kluster yang dihasilkan [4]. Penelitian yang menggunakan algoritma klusterisasi juga dilakukan untuk memprediksi waktu tunggu lulusan mahasiswa [5] atau mengklusterisasi tingkat pendidikan masyarakat di wilayah tertentu [6].

Penelitian ini bertujuan untuk membentuk model klusterisasi data lulusan berbasis hasil studi pelacakan tentang karir dan pekerjaan. Dengan tujuan tersebut maka akan diketahui model klusterisasi lulusan berbasis data hasil studi pelacakan. Studi ini memberikan pengetahuan dari hasil penambangan data (*data mining*) tentang profil kluster lulusan perguruan tinggi, berdasarkan analisis terhadap data lulusan menggunakan algoritma *K-Means*.

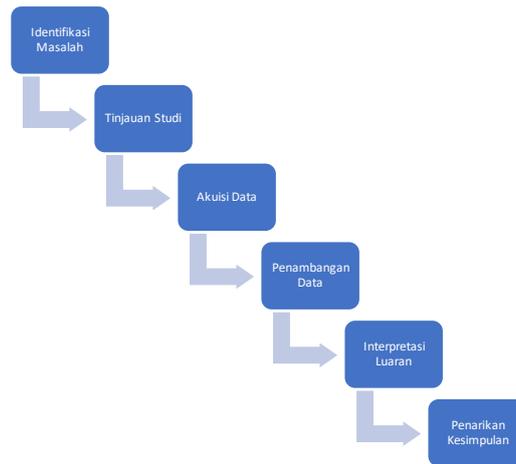
Hasil dari penelitian ini adalah pengelompokan atau klusterisasi lulusan berdasarkan profil kondisi karir dan pekerjaan dari hasil studi pelacakan. Kluster yang terbentuk akan dideskripsikan sesuai dengan kondisi *raw data* yang membentuknya. Hasil pemodelan klusterisasi dievaluasi menggunakan Davies-Bouldin *Index*, agar diketahui bentuk pembagian kluster terbaik dari berbagai uji coba klusterisasi yang dilakukan. Hasil penelitian dapat digunakan untuk mengevaluasi mutu lulusan dari suatu program studi atau fakultas.

2. MATERI DAN METODE

Studi yang dilakukan saat ini menggunakan pendekatan *Cross Industry Standard Process for Data Mining* (CRISP-DM), yaitu suatu model dalam proses penambangan data [7], dengan tahap-tahap sebagai berikut:

1. Pemahaman Bisnis
Merupakan tahap menjelaskan proses bisnis yang terjadi bahwa data diperoleh dari rangkaian kegiatan bisnis sehingga memiliki dampak berupa terciptanya data pada setiap tahap bisnis yang ada.
2. Pemahaman Data
Merupakan tahap yang menjelaskan ketersediaan data, periode data, dan bentuk data yang akan digunakan dalam proses penambangan.
3. Penyiapan Data
Merupakan tahap yang berisi rangkaian pekerjaan yang dilakukan agar data yang tersedia menjadi siap untuk dianalisis dalam proses pemodelan.
4. Pemodelan
Merupakan tahap pemilihan algoritma data mining yang sesuai dengan tujuan yang ingin dicapai.
5. Evaluasi
Merupakan tahap mengukur kinerja atau model yang terbentuk, sehingga pengetahuan yang ada dari model menjadi valid untuk digunakan.
6. Penyebaran
Merupakan tahap penyampaian pengetahuan (*knowledge*) yang berasal dari hasil penambangan data secara keseluruhan.

Untuk menyelesaikan rumusan masalah penelitian, maka studi ini dibagi menjadi beberapa tahap seperti terlihat pada Gambar 1.



Gambar 1. Tahap Penelitian

Terlihat pada Gambar 1 bahwa tahap penelitian saat ini dimulai dari proses mengidentifikasi masalah yang bersumber dari wawancara kepada pihak pengelola program studi, dilanjutkan dengan tahap tinjauan studi untuk mengetahui alternatif solusi yang bisa digunakan termasuk algoritma data mining yang sesuai. Tahap selanjutnya adalah akuisisi dan persiapan data, untuk selanjutnya dilakukan pemodelan berupa penerapan algoritma *K-Means* dalam proses klusterisasi yang dirancang disertai evaluasi menggunakan *Davies-Bouldin Index (DBI)*. Tahap berikutnya adalah interpretasi hasil atau luaran dan penarikan kesimpulan yang akan digunakan sebagai umpan balik pemilik data maupun peneliti.

Metode *K-Means* dipilih untuk memecahkan rumusan masalah pada studi ini berdasarkan studi literatur yang direferensi. Algoritma *K-Means* adalah algoritma yang sederhana namun memiliki metode kerja yang baik. Metode dasar analisis algoritma *K-Means* [8], [9] adalah sebagai berikut:

1. Tentukan jumlah kluster (k), tetapkan pusat kluster secara acak.
2. Hitung jarak setiap data ke pusat kluster dengan *Euclidian Distance*
3. Kelompokkan data ke dalam kluster dengan jarak yang paling pendek.
4. Hitung pusat *centroid* atau kluster baru.
5. Ulangi langkah 2 (dua) sampai 4 (empat) hingga sudah tidak ada lagi data yang berpindah ke kluster yang lain.

Proses klusterisasi dimulai dengan mengidentifikasi data yang akan dikluster, X_{ij} ($i = 1, \dots, n; j = 1, \dots, m$) dengan n adalah jumlah data yang akan dikluster dan m adalah jumlah variabel. Pada awal iterasi, pusat setiap kluster ditetapkan secara acak. Tahap selanjutnya adalah menghitung jarak setiap data ke pusat kluster dengan *Euclidian Distance*. *Euclidean Distance* adalah metrika yang paling sering digunakan untuk menghitung kesamaan dua vektor. Rumus *Euclidean Distance* adalah akar dari kuadrat perbedaan 2 vektor seperti terlihat pada Persamaan 1.

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

Keterangan:

d = jarak antara x dan y

x = data pusat kluster

y = data pada atribut

i = setiap data

n = jumlah data

x_y = data pada pusat kluster ke- i

x_{y_i} = nilai pada setiap data ke- i

Untuk menghitung *centroid* baru, dapat menggunakan Persamaan 2.

$$C = \frac{\sum m}{n} \quad (2)$$

Keterangan:

C = centroid data

m = anggota data yang termasuk ke dalam centroid tertentu

n = jumlah data yang menjadi anggota centroid tertentu

Pengukuran validitas dengan Davies-Bouldin Index menghitung rata-rata nilai setiap titik pada himpunan data. Perhitungan nilai setiap titik adalah jumlah nilai compactness yang dibagi dengan jarak antara kedua titik pusat kluster sebagai separation [10]. Pendekatan ini adalah untuk memaksimalkan jarak inter kluster dan meminimalkan jarak intra kluster yang dapat dihitung dengan Persamaan 3.

$$S_i = \frac{1}{|c_i|} \sum_{x \in c_i} \{|x - z_i|\} \quad (3)$$

dimana c_i sebagai banyaknya titik yang masuk ke dalam kluster i , x adalah data, dan z_i centroid dari kluster i . Sedangkan jarak antara kluster didefinisikan pada Persamaan 4.

$$d_{ij} = |z_i - z_j| \quad (4)$$

dimana z_i centroid dari kluster i dan z_j centroid dari kluster j . Perhitungan jarak d_{ij} dapat menggunakan euclidean. Selanjutnya akan mendefinisikan $R_{i,qt}$ untuk kluster c_i pada Persamaan 5.

$$R_{i,qt} = \max_{j, j \neq i} \left\{ \frac{S_{i,q} + S_{j,q}}{d_{ij,t}} \right\} \quad (5)$$

Selanjutnya Davies-Bouldin Index didefinisikan pada Persamaan 6.

$$DB = \frac{1}{k} \sum_{i=1}^k R_{i,qt} \quad (6)$$

Dari persamaan tersebut, k merupakan jumlah kluster yang digunakan. Semakin kecil nilai DBI yang diperoleh (non-negatif ≥ 0), maka semakin baik kluster yang diperoleh dari pengelompokan *K-means* yang digunakan [11].

3. HASIL DAN PEMBAHASAN

Berdasarkan pendekatan CRISP-DM maka, dapat diketahui dengan jelas fase-fase atau tahap penelitian ini terdiri dari enam tahap, sebagai berikut:

1. Pemahaman Bisnis

Pada tahap pemahaman bisnis dihasilkan suatu deskripsi proses bisnis yang berjalan, bahwa kegiatan studi pelacakan (*tracer study*) merupakan kegiatan rutin yang dilakukan pengelola program studi terhadap lulusannya. Kegiatan studi pelacakan dilakukan secara sistematis, terstruktur dengan metode dan pedoman serta petunjuk teknis (juknis) yang dipahami oleh Ketua Program Studi. Kegiatan studi pelacakan ini dilakukan berdasarkan data wisudawan atau lulusan dari sebuah program studi.

Kegiatan studi pelacakan menghasilkan data profil lulusan meliputi, status bekerja, belum bekerja, atau berwirausaha. Selain itu juga diketahui masa tunggu lulusan perguruan tinggi sewaktu mendapatkan pekerjaan pertama setelah diwisuda, dan kondisi kesesuaian antara bidang pekerjaan dengan program pendidikan yang telah ditempuh. Studi pelacakan ini dilakukan secara periodik biasanya pada saat menjelang wisuda, dan paling cepat enam bulan sejak diwisuda atau paling lambat satu tahun dari pelaksanaan wisuda.

Pada studi kali ini, data diperoleh dari hasil studi pelacakan yang dilakukan pada masa pandemi COVID-19, terbagi dalam empat periode waktu wisuda, diambil dari data lulusan yang diwisuda selama tahun 2020 hingga 2022. Studi ini menggunakan metode *purposive sampling* karena ingin mengetahui model klusterisasi untuk mengetahui profil lulusan yang diwisuda pada masa pandemi COVID-19 dan untuk mengetahui profil lulusan yang bekerja pada masa pandemi tersebut.

2. Pemahaman Data

Pada fase ini, diketahui bahwa data untuk kegiatan penelitian didapatkan dari pihak pengelola yaitu Ketua Program Studi (KPS) dan Direktorat Kemahasiswaan, Karir Dan Alumni. Data yang bersumber dari Ketua Program Studi meliputi data lulusan yang terdiri dari waktu wisuda, meliputi NIM, Nama, IPK per semester. Data yang diperoleh dari Direktorat Kemahasiswaan diperoleh data berupa hasil studi pelacakan yang terdiri dari atribut NIM, nama, status saat ini yang terbagi menjadi lima (bekerja,

belum memungkinkan bekerja, wirausaha, melanjutkan pendidikan, sedang tidak bekerja), kesesuaian bidang kerja dengan lima nilai (sangat erat, erat, cukup erat, kurang erat dan tidak sama sekali) dan juga data waktu tunggu untuk mendapatkan pekerjaan.

3. Penyiapan Data

Pada tahap ini dilakukan pengintegrasian data yang diperoleh dari Program Studi maupun data dari Direktorat Kemahasiswaan, sehingga menjadi data yang siap diolah pada proses pemodelan. Selain itu, pada dataset juga dilakukan kegiatan pra-pemrosesan berupa pembersihan (*cleaning*) pada hasil studi pelacakan lulusan, yaitu pada data yang tidak lengkap, *invalid*, atau duplikat.

Dari tahap pemahaman data sebelumnya diketahui bahwa jumlah lulusan yang diwisuda di masa pandemi dari perguruan tinggi yang menjadi obyek penelitian (populasi), ada sebanyak 3.098 orang. Dari jumlah lulusan tersebut dilakukan studi pelacakan sebanyak dua kali pada masa pandemi, dengan perolehan berupa 705 data atau 22,75% dari jumlah populasi yang menjadi dataset untuk pemodelan. Komposisi dataset terdiri dari lulusan Fakultas Teknologi Informasi (FTI) sebanyak 230 orang (32,6%), Fakultas Teknik (FT) sebanyak 18 orang (2,5%), Fakultas Ilmu Komunikasi dan Desain Kreatif (FKDK) sebanyak 182 orang (25,8%), Fakultas Ilmu Sosial dan Studi Global (FISSIG) sebanyak 47 orang (6,6%) serta Fakultas Ekonomi dan Bisnis (FEB) sebanyak 227 orang (32,2%).

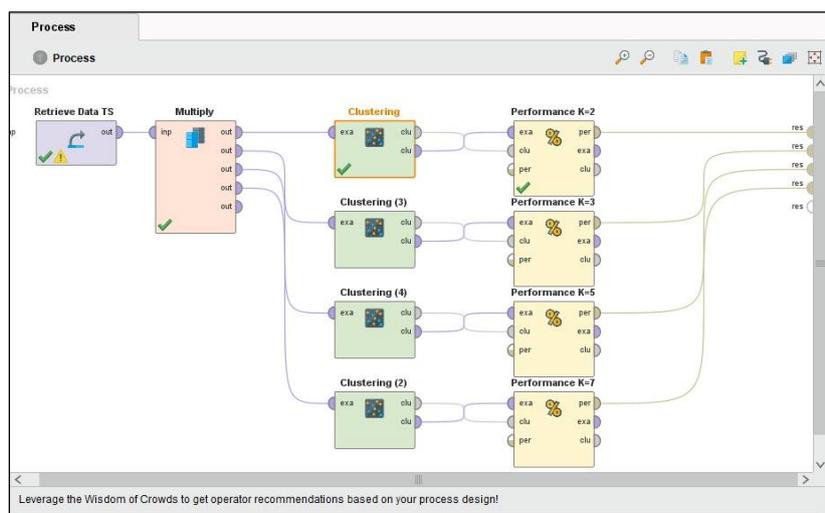
Pada tahap penyiapan data, studi ini menggunakan empat atribut yang dilibatkan dalam proses pemodelan klasterisasi, yaitu data nim/nama, status pekerjaan yang berisi lima jenis *data value* yaitu nilai satu untuk lulusan bekerja, nilai dua untuk lulusan yang belum memungkinkan bekerja, nilai tiga lulusan yang berwirausaha, nilai empat untuk lulusan yang sedang melanjutkan pendidikan, dan nilai lima bagi yang sedang mencari kerja.

Selain itu juga dilakukan transformasi pada atribut Kesesuaian Bidang Pekerjaan, dari semula lima nilai (sangat sesuai, sesuai, cukup sesuai, kurang sesuai dan tidak sesuai) menjadi tiga kategori yaitu nilai dua untuk sesuai, nilai satu untuk kurang sesuai dan nilai nol untuk tidak sesuai. Tujuan transformasi ini adalah untuk mengurangi varian data sehingga klaster yang terbentuk diharapkan lebih fit dan ideal. Data lain yang disiapkan adalah data waktu tunggu mendapatkan pekerjaan dalam satuan hitung berupa satuan bulan (per tiga puluh hari) dalam angka absolut.

4. Pemodelan

Pada tahap ini dilakukan pemilihan algoritma untuk pemecahan masalah penelitian yang ingin menghasilkan model klaster dari data lulusan berbasis hasil studi pelacakan. Berdasarkan studi literatur yang dijadikan referensi, algoritma *K-Means* merupakan metode yang paling sesuai untuk ketersediaan data penelitian saat ini. Pemodelan pada studi ini diselesaikan dengan berbantuan perangkat lunak RapidMiner Studio yang berjalan pada perangkat sistem operasi Mac (macOS).

Pada tahap pemodelan, data hasil pra-pemrosesan disimpan dalam format Excel untuk selanjutnya diimpor ke dalam jendela desain proses pada RapidMiner seperti terlihat pada Gambar 2.



Gambar 2. Desain Pemodelan Klasterisasi pada RapidMiner Studio

Terlihat pada Gambar 2 bahwa pemodelan dilakukan secara serentak untuk berbagai skenario (*multiply*), yaitu klusterisasi dengan $K=2$, $K=3$, $K=5$ dan $K=7$ terhadap data sebanyak 705 data lulusan. Hasil pemodelan terlihat pada Gambar 3.

Row No.	NIM	cluster	Status Peka...	Kesesuaian ...	Masa Tunggu
1	1732510191	cluster_1	1	2	72
2	1731510200	cluster_1	1	1	60
3	1731501001	cluster_1	1	2	60
4	1712530060	cluster_1	1	2	60
5	1731510036	cluster_1	1	1	60
6	1731510267	cluster_1	1	1	60
7	1611520188	cluster_1	1	1	56
8	1771500228	cluster_0	1	2	50
9	1771510151	cluster_0	5	1	50
10	1732520356	cluster_0	1	1	48
11	1731510747	cluster_0	1	2	48
12	1732520273	cluster_0	1	1	48
13	1731510655	cluster_0	1	2	48
14	1712530151	cluster_0	1	1	48
15	1732520034	cluster_0	1	1	48

Gambar 3. Hasil Pemodelan Klusterisasi

Terlihat pada Gambar 3 bahwa hasil proses *running* pemodelan pada aplikasi RapidMiner Studio menghasilkan pembagian klaster dari 705 baris data sesuai parameter nilai K yang dimasukkan. Pemodelan multi skenario ini juga mendapatkan nilai DBI terbaik dari empat model yang diuji coba serentak menggunakan algoritma *K-Means*. Hasil evaluasi model dari ujicoba serentak tahap pertama menunjukkan hasil seperti terlihat pada Tabel 1.

Tabel 1. Evaluasi Model pada Ujicoba Tahap Pertama

Indikator	Nilai DBI
$K=2$	0,287
$K=3$	0,483
$K=5$	0,405
$K=7$	0,346

Dari Tabel 1 diketahui bahwa ujicoba serentak yang dilakukan untuk masing-masing nilai K , memiliki kinerja yang berbeda-beda sesuai nilai DBI yang dihasilkan, namun demikian model klusterisasi terbaik diperoleh pada nilai $K=2$ dengan evaluasi DBI sebesar 0,287. Dari 705 baris data yang dikelompokkan menjadi dua klaster ($K=2$) pada aplikasi RapidMiner Studio diketahui bahwa klaster 0 memiliki anggota sebanyak 547 data lulusan, sementara klaster 1 memiliki anggota sebanyak 158 baris.

Klaster 0 merupakan kelompok lulusan yang jumlah anggotanya relatif banyak. Sebanyak 547 lulusan (77,5%) pada klaster 0 memiliki kesamaan ciri yaitu anggota kelompoknya merupakan lulusan yang memiliki masa tunggu relatif singkat untuk mendapatkan pekerjaan pertamanya, yaitu pada rentang 0-1 tahun. Dari 547 lulusan di klaster 0, mayoritas terdiri dari FEB sebanyak 177, FISSIG sebanyak 31, FKDK sebanyak 146, FT sebanyak 13 dan FTI sebanyak 180 lulusan.

Sementara klaster 1 yang anggotanya terdiri dari 158 lulusan dapat dideskripsikan sebagai lulusan yang memiliki masa tunggu relatif lama dalam mendapatkan pekerjaan pertamanya, yaitu satu tahun ke atas. Dari 158 lulusan di klaster 1, mayoritas terdiri dari FTI sebanyak 51 lulusan, FEB sebanyak 50 lulusan, FKDK sebanyak 36 lulusan, FISSIG sebanyak 16 lulusan, dan FT sebanyak 5 lulusan.

Untuk mendapatkan gambaran yang lebih spesifik, maka dilakukan ujicoba klusterisasi tahap kedua yang diterapkan khusus pada anggota klaster 0 dengan jumlah anggota sebanyak 547 anggota. Dengan ujicoba tahap kedua ini akan diketahui lebih spesifik kondisi profil anggota pada klaster yang terbentuk. Pemodelan serentak tahap kedua terhadap 547 data lulusan dilakukan menggunakan algoritma *K-Means* kembali pada nilai $K=2$, $K=3$, $K=5$ dan $K=7$. Hasil evaluasi model pada uji coba tahap kedua terlihat pada Tabel 2.

Tabel 2. Evaluasi Model pada Ujicoba Tahap Kedua

Indikator	Nilai DBI
K=2	0,291
K=3	0,457
K=5	0,450
K=7	0,398

Dari Tabel 2 diketahui bahwa ujicoba serentak tahap kedua yang dilakukan untuk masing-masing nilai K, memiliki kinerja yang berbeda-beda sesuai nilai DBI yang dihasilkan, namun demikian model klusterisasi terbaik diperoleh pada nilai K=2 dengan evaluasi DBI sebesar 0,291. Dari 547 baris data yang dikelompokkan menjadi dua klaster (K=2) pada aplikasi RapidMiner Studio diketahui bahwa klaster 01 memiliki anggota sebanyak 517 baris, sementara klaster 02 memiliki anggota sebanyak 30 data lulusan.

Klaster 01 merupakan kelompok lulusan yang jumlah anggotanya relatif banyak. Terdapat sebanyak 517 lulusan (94,5%) yang menjadi anggota pada klaster 01 dengan profil anggota kelompoknya merupakan lulusan yang cepat bekerja dengan masa tunggu pekerjaan seluruhnya kurang dari satu tahun, berstatus sedang bekerja pada saat studi pelacakan sedang dilakukan (82,4%), namun demikian ada sekitar 43 lulusan (8%) yang bekerja tidak sesuai bidang pendidikan yang ditempuhnya, sisanya 242 lulusan cukup sesuai bidang pekerjaannya dan 232 orang sangat sesuai pendidikan yang ditempuh dengan pekerjaannya.

Sementara itu, untuk klaster 02 yang memiliki anggota sebanyak 30 lulusan memiliki profil sebagai lulusan yang waktu tunggu mendapatkan pekerjaannya relatif sedang (antara 6 bulan hingga satu tahun), namun mayoritas berstatus sedang bekerja pada saat studi pelacakan sedang dilakukan (96%).

Dengan melakukan dua kali uji coba maka dapat diketahui bahwa klaster terakhir yang terbentuk adalah klaster 03 yang memiliki 158 anggota dengan ciri lulusan yang memiliki waktu tunggu mendapatkan pekerjaan relatif lambat (di atas satu tahun). Namun demikian, dari anggota klaster ini diketahui hanya ada 11 orang yang bidang kerjanya tidak sesuai, sementara 68 orang cukup sesuai bidang pekerjaannya dan 79 orang sangat sesuai pendidikan yang ditempuh dengan pekerjaannya.

5. Evaluasi

Berdasarkan nilai-nilai DBI yang diperoleh maka, seluruh pembagian klaster terbaik diperoleh pada K=2 yaitu pada ujicoba tahap pertama diperoleh angka Davies-Bouldin *Index* sebesar 0,287. Sementara dengan melakukan klusterisasi ulang pada salah satu klaster yang terbentuk, juga diperoleh nilai DBI terbaik pada ujicoba tahap kedua dengan nilai Davies-Bouldin *Index* sebesar 0,291 pada nilai K=2. Dengan perbedaan yang relatif tidak terlalu signifikan, maka pembagian klaster dengan K=2 adalah yang paling ideal dari berbagai rangkaian uji coba.

6. Penyebaran

Dengan seluruh hasil klusterisasi, maka dapat diperoleh pengetahuan yang cukup bermanfaat bagi pengelola program studi, yaitu diketahuinya pola pengelompokan lulusan dari perguruan tinggi yang menjadi obyek penelitian. Klaster utama adalah para lulusan yang memiliki masa tunggu relatif singkat untuk mendapatkan pekerjaan pertamanya, yaitu dalam waktu antara nol hingga enam bulan, yang mencapai 94,5% dari jumlah data latih. Namun demikian, tetap harus ada upaya perbaikan dalam peningkatan mutu lulusan karena pada analisis terhadap hasil klaster masih terdapat sekitar 22 persen lulusan yang bidang pekerjaannya tidak sesuai dengan bidang pendidikan yang ditempuh dan diselesaikan di program studinya.

4. KESIMPULAN DAN SARAN

Dari seluruh tahap penelitian yang sudah diselesaikan maka dapat disimpulkan bahwa dataset lulusan berbasis studi pelacakan dapat dianalisis menggunakan pendekatan data mining khususnya metode klusterisasi dengan algoritma *K-Means*. Hasil analisis mendapatkan pola pengelompokan kondisi lulusan perguruan tinggi dengan atribut seperti status lulusan dalam bekerja, berwirausaha atau sedang tidak bekerja, masa tunggu dalam mendapatkan pekerjaan pertama, dan tingkat kesesuaian pekerjaan dengan bidang pendidikan. Selain itu, pemodelan klusterisasi yang dilakukan mampu menghasilkan pengelompokan lulusan melalui ujicoba model klusterisasi tahap pertama dengan nilai evaluasi yang relatif baik, pada nilai K=2. Namun demikian karena salah satu klaster yang terbentuk masih memiliki jumlah anggota yang cukup besar maka dilakukan ujicoba model klusterisasi tahap kedua. Hasil pemodelan tetap menunjukkan bahwa pada K=2 diperoleh nilai evaluasi terbaik menurut standar Davies-Bouldin *Index* (DBI).

Pada studi mendatang, dapat dilakukan dengan pengembangan dengan menggunakan algoritma pembandingan seperti K-Medoids pada jumlah data yang lebih besar, dengan harapan mencapai nilai evaluasi yang lebih baik. Selain itu, penelitian ini masih dapat dikembangkan dengan pembuatan prototipe aplikasi untuk menyelesaikan proses pemodelan klasterisasi. Hasil penelitian berupa prototipe dapat diusulkan untuk pengakuan hak cipta program komputer dari Direktorat Jenderal Hak Cipta Kemenkumham RI.

DAFTAR PUSTAKA

- [1] Y. E. Fadrial, "Klasterisasi Hasil Evaluasi Akademik Menggunakan Metode K-Means (Studi Kasus Fakultas Ilmu Komputer UNILAK)," *Semin. Nas. Teknol. Inf. Ilmu Komput.*, vol. 1, no. 1, pp. 53–65, 2020, doi: [10.31849/semester.v1i1.5492](https://doi.org/10.31849/semester.v1i1.5492).
- [2] V. Novita Sari, Y. Yudianti, and D. Maharani, "Penerapan Metode K-Means Clustering Dalam Menentukan Predikat Kelulusan Mahasiswa Untuk Menganalisa Kualitas Lulusan," *Jurteksi*, vol. 4, no. 2, pp. 133–140, 2018, doi: [10.33330/jurteksi.v4i2.53](https://doi.org/10.33330/jurteksi.v4i2.53).
- [3] D. Praseptian M, A. Fadlil, and H. Herman, "Penerapan Clustering K-Means untuk Pengelompokan Tingkat Kepuasan Pengguna Lulusan Perguruan Tinggi," *J. Media Inform. Budidarma*, vol. 6, no. 3, p. 1693, 2022, doi: [10.30865/mib.v6i3.4191](https://doi.org/10.30865/mib.v6i3.4191).
- [4] N. Nurahman, A. Purwanto, and S. Mulyanto, "Klasterisasi Sekolah Menggunakan Algoritma K-Means berdasarkan Fasilitas, Pendidik, dan Tenaga Pendidik," *MATRIK J. Manajemen, Tek. Inform. dan Rekayasa Komput.*, vol. 21, no. 2, pp. 337–350, 2022, doi: [10.30812/matrik.v21i2.1411](https://doi.org/10.30812/matrik.v21i2.1411).
- [5] H. Priyatman, F. Sajid, and D. Haldivany, "Klasterisasi Menggunakan Algoritma K-Means Clustering untuk Memprediksi Waktu Kelulusan Mahasiswa," *J. Edukasi dan Penelit. Inform.*, vol. 5, no. 1, p. 62, 2019, doi: [10.26418/jp.v5i1.29611](https://doi.org/10.26418/jp.v5i1.29611).
- [6] R. Kurniawan, M. M. M. Mukarrob, and M. Mahradianur, "Klasterisasi Tingkat Pendidikan Di Dki Jakarta Pada Tingkat Kecamatan Menggunakan Algoritma K-Means," *Technol. J. Ilm.*, vol. 12, no. 4, p. 234, 2021, doi: [10.31602/tji.v12i4.5633](https://doi.org/10.31602/tji.v12i4.5633).
- [7] N. Mirantika, A. Tsamratul'Ain, and F. D. Agnia, "Penerapan Algoritma K-Means Clustering Untuk Pengelompokan Penyebaran Covid-19 di Provinsi Jawa Barat," *Nuansa Inform.*, vol. 15, no. 2, pp. 92–98, 2021, doi: [10.25134/nuansa.v15i2.4321](https://doi.org/10.25134/nuansa.v15i2.4321).
- [8] A. K. Wardhani, "Implementasi Algoritma K-Means Untuk Pengelompokan Penyakit Pasien Pada Puskesmas Kajen Pekalongan," *J. Transform.*, vol. 14, pp. 30–37, 2016.
- [9] M. Hariyanto and R. T. Shita, "Clustering Pada Data Mining Untuk Mengetahui Potensi Penyebaran Penyakit DBD Menggunakan Metode Algoritma K-Means dan Metode Perhitungan Jarak Euclidean Distance," *SKANIKA*, vol. 1, no. 1, pp. 117–122, 2018, doi: [10.36080/skanika.v13i2.1456](https://doi.org/10.36080/skanika.v13i2.1456).
- [10] A. F. Khairati, A. A. Adlina, G. F. Hertono, and B. D. Handari, "Kajian Indeks Validitas pada Algoritma K-Means Enhanced dan K-Means MMCA," in *Prosiding Seminar Nasional Matematika*, vol. 2, pp. 161–170, 2019.
- [11] R. D. Ramadhani and D. J. AK, "Evaluasi K-Means dan K-Medoids pada Dataset Kecil," *Semin. Nas. Inform. dan Apl.*, vol. 3, pp. 20–24, 2019.