

## KLASTERISASI LULUSAN PENDIDIKAN TINGGI VOKASI BERBASIS HASIL STUDI PELACAKAN MENGGUNAKAN ALGORITME K-MEANS

Dyah Retno Utari<sup>1</sup>, Fenti Sofiani<sup>2</sup>

Universitas Budi Luhur, Jakarta, Indonesia<sup>1</sup>

Akademi Sekretari Budi Luhur, Jakarta, Indonesia<sup>2</sup>

### ABSTRAK

Pendidikan vokasi kesekretarisan bersifat khas dan unik karena belum tersedianya program pendidikan dalam jenjang Strata Satu untuk program studi Sekretari di Indonesia. Kompetensi kerja dan profil lulusannya sangat spesifik serta mudah diidentifikasi di dunia kerja. Namun demikian, pada faktanya lulusan pendidikan kesekretarisan bekerja di berbagai sektor bidang kerja, karena lulusan dari program studi non-sekretari masih dapat menempati profesi sekretaris. Berdasarkan kondisi tersebut maka menjadi tantangan tersendiri bagi pengelola dan penyelenggara program studi vokasi kesekretarisan untuk mengetahui klusterisasi profil lulusannya berbasis hasil studi pelacakan (*tracer study*). Studi ini menganalisis data hasil pelacakan studi tentang lulusan pendidikan diploma tiga kesekretarisan untuk mengetahui kluster dan profil lulusan berdasarkan bidang kerja dan latar belakang akademiknya. Lingkup penelitian dibatasi pada data satu prodi Sekretari yang terakreditasi A di Indonesia untuk lulusan tiga tahun terakhir 2018-2020. Metode yang digunakan adalah pendekatan data mining klusterisasi dengan algoritme K-Means. Hasil pemodelan dievaluasi berdasarkan nilai Davies Bouldin Index (DBI) sebesar 0,221. Dari hasil pemodelan dapat disimpulkan bahwa model klusterisasi terbaik dapat dikembangkan menjadi prototipe aplikasi data lulusan ke kluster yang sesuai dengan saran pengembangan pada pembuatan sistem informasi klusterisasi berbasis web.

**Kata-kata Kunci:** Algoritme Klusterisasi; Studi Pelacakan; Data Mining; Klusterisasi Lulusan Program Pendidikan Vokasi Kesekretarisan.

## ***CLUSTERIZATION OF VOCATIONAL HIGHER EDUCATION GRADUATES BASED ON TRACKING STUDY RESULTS USING K-MEANS ALGORITHM***

### ABSTRACT

*Secretarial vocational education is unique as there is no educational program at the Strata One level for Secretarial study programs in Indonesia. Work competencies and graduate profiles are very specific and easy to identify in the world of work. However, in reality, secretarial education graduates work in various sectors of work because graduates from non-secretary study programs can still occupy the secretarial profession. Based on these conditions, it becomes a challenge for managers and administrators of secretarial vocational study programs to find out the clustering of graduates' profiles based on the results of tracer studies. This study analyzes data from tracking studies on graduates of three secretarial diploma education programs to identify clusters and profiles of graduates based on their field of work and academic background. The scope of the research is limited to data from one Secretarial study program accredited A in Indonesia for graduates of the last three years 2018-2020. The method used is a clustered data mining approach with the K-Means algorithm. The modeling results were evaluated based on the Davies Bouldin Index (DBI) value of 0.221. From the modeling results, it can be concluded that the best clustering model can be developed into a prototype application of graduate data to clusters in accordance with development suggestions for making a web-based clustering information system.*

**Keywords:** Clustering Algorithm; Tracer Study; Data Mining; Clustering of Secretarial Vocational Education Program Graduates.

**Korespondensi:** Dyah Retno Utari, M.Kom, Universitas Budi Luhur, Jl. Ciledug Raya, Petukangan Utara, Jakarta Selatan, 12260, Indonesia, *Email:* dyah.retnoutari@budiluhur.ac.id

**Submitted:** September 2021, **Accepted:** Oktober 2021, **Published:** October 2021

**OJS:** <https://journal.budiluhur.ac.id/index.php?journal=serasi>

## PENDAHULUAN

Pendidikan kesekretarisan merupakan salah satu jenjang pendidikan vokasi yang diselenggarakan oleh perguruan tinggi di Indonesia. Pendidikan vokasi kesekretarisan bersifat sangat unik dan klasik karena merupakan cabang ilmu administrasi yang diturunkan pada sub bidang ilmu administrasi perkantoran, yang bertujuan mempersiapkan lulusan yang akan berkarir sebagai sekretaris profesional.

Tantangan bagi pengelola program studi vokasi kesekretarisan, adalah bagaimana upaya menciptakan lulusan yang terampil, cepat mendapatkan pekerjaan, dan memiliki pekerjaan yang sesuai dengan bidang pembelajaran di perguruan tinggi, seperti ilmu administrasi, kearsipan, korespondensi, komunikasi dan teknologi informasi. Tantangan tersebut kerap kali menjadi lebih sulit karena profesi sekretaris di beberapa kantor dan dunia usaha, ditempati atau diisi oleh lulusan-lulusan program pendidikan non-kesekretarisan. Hal itu membuat semakin kompetitifnya lulusan program studi vokasi kesekretarisan untuk mendapatkan kerja sesuai bidang ilmu yang mereka pelajari.

Berdasarkan kondisi-kondisi tersebut maka pengelola program studi vokasi kesekretarisan memerlukan gambaran tentang kondisi lulusan program studi yang dikelolanya untuk mengevaluasi proses pendidikan yang diselenggarakan berdasarkan standar mutu pendidikan tinggi. Salah satu solusi untuk kebutuhan informasi tersebut, diperoleh melalui studi pelacakan lulusan atau yang dikenal dengan istilah *tracer study*. Data hasil studi pelacakan lulusan, diolah dan dianalisis sedemikian rupa untuk mendapatkan pengetahuan atau *knowledge* yang bermanfaat untuk umpan balik dalam hal pembaharuan strategi pembelajaran, model pendidikan maupun evaluasi akademik. Salah satu *knowledge* yang perlu diketahui untuk mendapatkan umpan balik adalah tentang pengelompokan atau klasterisasi para lulusan berdasarkan profil akademik maupun berbasis hasil studi pelacakan.

Studi terdahulu yang telah dilakukan antara lain, pembentukan klaster profil lulusan dengan algoritme K-Means yang dapat memberikan pengetahuan untuk menentukan strategi promosi di Politeknik LP3I Jakarta. Pengetahuan yang didapat dari hasil klasterisasi yaitu calon mahasiswa didominasi dari masyarakat ekonomi rendah dan menengah, dengan jumlah klaster yang terbentuk sebanyak empat (Setiawan, 2016). Studi ini hanya melakukan satu kali pemodelan dan belum mengeksplorasi kemungkinan pembentukan klaster alternatif. Studi lainnya mengklasterisasi data numerik yakni nilai Ujian Tengah Semester (UTS) dalam katgeori nilai UTS tinggi, nilai UTS sedang, dan nilai UTS rendah. Studi ini melakukan satu kali pemodelan menggunakan data numerik tanpa tahap pra-pemrosesan yang kompleks (Abdurrahman, 2016).

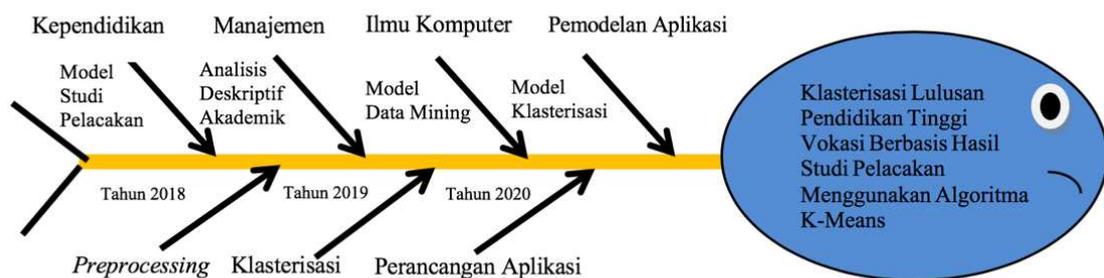
Peneliti lainnya melakukan analisis menggunakan algoritme K-Means untuk mengelompokkan mahasiswa berdasarkan IPK dan Prodynya (Novita Sari, Yupianti, & Maharani, 2018), selain itu juga dilakukan pengelompokan prestasi siswa dalam kategori tinggi, menengah dan cukup (Sibuea & Safta, 2017). Model-model klasterisasi tersebut belum dikembangkan ke dalam bentuk prototipe aplikasi.

Studi lain yang mendekati kemiripan dengan penelitian ini adalah pengelompokan mahasiswa yang lulus tepat waktu atau tidak tepat waktu, namun studi ini tidak menggunakan data hasil studi pelacakan (Rosmini, Fadlil, & Sunardi, 2018).

Berdasarkan studi literatur terdahulu, maka proses pengelompokan atau klasterisasi tersebut dapat diselesaikan menggunakan metode data mining klasterisasi, yaitu algoritme K-Means. Rumusan masalah pada penelitian ini adalah bagaimana model klasterisasi lulusan pendidikan tinggi vokasi berbasis hasil studi pelacakan menggunakan algoritme K-Means. Studi ini melakukan analisis data lulusan berbasis data akademik dan data hasil studi pelacakan yang ruang lingkupnya dibatasi pada mahasiswa program studi Sekretari yang sudah dinyatakan lulus studi. Pertanyaan penelitian yang dapat diajukan adalah apakah algoritme K-Means dapat menghasilkan model klasterisasi yang ideal dengan evaluasi berdasarkan nilai Davies Bouldin-Index . Model klasterisasi terbaik selanjutnya dikembangkan menjadi prototipe aplikasi klasterisasi lulusan yang diuji dengan metode *Focus Grup Discussion* (FGD) di antara para pemangku kepentingan pada obyek penelitian

## METODOLOGI PENELITIAN

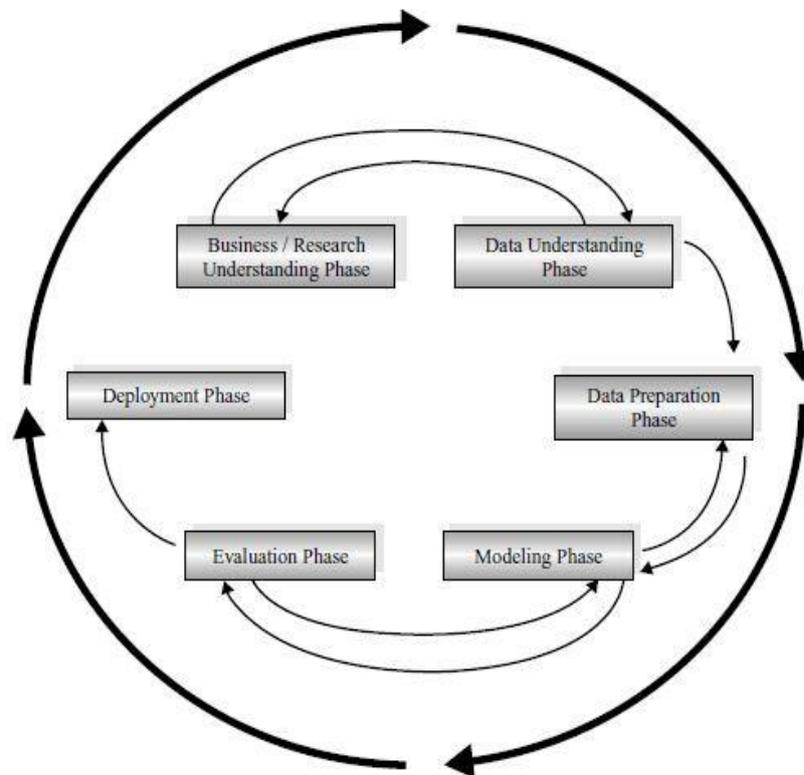
Roadmap atau peta jalan pada penelitian ini memiliki kerangka kerja yang menggabungkan dua topik bidang keilmuan yaitu ilmu komputer dan ilmu manajemen khususnya sub bidang manajemen pendidikan. Peta jalan penelitian ini digambarkan pada diagram fish-bone seperti terlihat pada Gambar 1.



**Gambar 1 Diagram Fish-bone Penelitian**

Pada penelitian ini, metode yang digunakan adalah pemodelan klasterisasi menggunakan algoritme K-Means. Eksperimen pemodelan dilakukan hingga mendapatkan klaster dengan indeks evaluasi terbaik, mengacu kepada studi terdahulu tentang data mining klasterisasi (Priyatman, Sajid, & Haldivany, 2019). Tahap penelitian ini terdiri dari lima langkah, mulai dari pengumpulan data, pengolahan data awal, metode usulan, eksperimen model dan pengujian, serta evaluasi dan validasi hasil.

Data sekunder pada penelitian merupakan kombinasi data akademik dan hasil studi pelacakan lulusan tahun 2016-2019 dari pendidikan vokasi kesekretarian terbaik di Indonesia yang memiliki peringkat Akreditasi A. Metodologi penelitian ini mengacu pada model CRISP-DM yang merupakan singkatan dari *Cross Industry Standard Process for Data Mining*. CRISP-DM merupakan standarisasi data mining yang disusun oleh tiga penggagas data mining market, menurut Larose sebagaimana dikutip dari studi terdahulu (Setiawan, 2016). Pada penelitian ini dilakukan pembagian siklus untuk proses data mining menjadi enam tahap, dimana ketergantungan antara setiap tahap digambarkan dengan panah. Gambar 2 menunjukkan metodologi penelitian yang diacu berdasarkan metode CRISP-DM.



**Gambar 2 Tahapan Metodologi CRISP-DM oleh Larose (Setiawan, 2016)**

Berdasarkan pada Gambar 2 di atas, lingkaran siklus paling luar menggambarkan bahwa data mining dengan metodologi CRISP-DM dapat mengambil pengalaman dari proyek masa lalu untuk menjadi masukan dalam proyek-proyek baru. Pada beberapa literatur digambarkan bahwa dari tahapan evaluasi dapat mengirim analisis kembali ke salah satu tahapan sebelumnya, namun untuk kesederhanaan maka pada literatur ini digambarkan proses yang paling umum dimana dari tahapan evaluasi dapat kembali ke tahap modeling.

Pada tahap *Business Understanding*, berfokus pada pemahaman mengenai tujuan dari proyek dan kebutuhan secara perspektif bisnis, kemudian mengubah hal tersebut menjadi sebuah permasalahan

data mining dan rencana awal untuk mencapai tujuan tersebut. Kegiatan yang dilakukan antara lain: menentukan tujuan dan persyaratan dengan jelas secara keseluruhan, menerjemahkan tujuan tersebut serta menentukan pembatasan dalam perumusan masalah data mining, dan selanjutnya mempersiapkan strategi awal untuk mencapai tujuan tersebut.

Pada tahap *Data Understanding*, dilakukan pengumpulan terhadap data. Data yang dikumpulkan merupakan data lulusan prodi Sekretari di Akademi Sekretari Budi Luhur untuk lulusan pada tahun 2018-2020. Prodi Sekretari merupakan satu-satunya prodi yang terakreditasi A pada saat penelitian ini dilakukan, sehingga dapat menjadi data penelitian yang representatif untuk model penelitian sejenis. Data berupa lulusan prodi Sekretari dengan atribut, tahun masuk, masa studi, tahun lulus, IPS semester 1 hingga semester 3, IPK, status asal SLTA, dan status mahasiswa reguler/non-reguler. Data tersebut lalu dipelajari dan dianalisis untuk identifikasi dan mengetahui kualitas data, serta mendeteksi subset yang menarik dari data yang dapat dijadikan hipotesa. Pada Tahap *Data Preparation* dilakukan persiapan mengenai data yang akan digunakan pada tahap berikutnya. Kegiatan yang dilakukan antara lain: memilih kasus dan parameter yang akan dianalisis (*Select Data*), melakukan transformasi terhadap parameter tertentu (*Transformation*), dan melakukan pembersihan data agar data siap untuk tahap modeling (*Cleaning*). Pada data dilakukan proses transformasi yang meliputi 183 baris data untuk perubahan nilai data (*data value*), yaitu mengetahui masa studi dari pengurangan antara tahun lulus dengan tahun masuk, selain itu juga dilakukan transformasi data dari asal SLTA yang dikategorikan menjadi data jabodetabek, non-jabodetabek pulau Jawa dan non-pulau Jawa. Transformasi ini dilakukan untuk mengurangi varian nilai data dari atribut status asal SLTA.

Tahap *Modeling* menentuka teknik data mining, alat bantu data mining, dan algoritme data mining yang tepat untuk diterapkan. Jika diperlukan penyesuaian data terhadap teknik data mining tertentu, dapat kembali ke tahap persiapan data. Pada tahap *Evaluation*, dilakukan interpretasi terhadap hasil dari data mining yang dihasilkan dalam proses pemodelan pada tahap sebelumnya. Evaluasi dilakukan terhadap model yang diterapkan pada tahap sebelumnya dengan tujuan agar model yang ditentukan dapat sesuai dengan tujuan yang ingin dicapai dalam tahap pertama. Tahap akhir adalah *Deployment* yaitu penyusunan laporan terhadap hasil yang didapat dari evaluasi pada tahap sebelumnya atau dari proses data mining yang dilakukan secara keseluruhan.

Mengacu pada konsep *Knowledge Discovery in Database*, maka seluruh tahap penelitian ini dilakukan dengan tetap pada kerangka CRISP-DM. Pekerjaan awal dari studi ini adalah pra pmerosesan data. Data diproses menjadi beberapa kelompok atau klaster menggunakan algoritme K-Means, dan setelah aturan kluster telah terbentuk, maka informasi dihitung menggunakan rumus *euclidean distance* untuk menghitung jarak antar data sehingga dapat menentukan anggota kelompok dari setiap klaster. Namun demikian seluruh eksperimen

pemodelan diselesaikan dengan perangkat *Rapidminer* dan pengujian model kluster yang terbentuk dievaluasi menggunakan *Davies Bouldin Index (DBI)*.

## HASIL DAN PEMBAHASAN

Pada fase pemilihan data (*data selection*), dilakukan pemilihan dari beberapa tabel database yang tersedia, antara lain data master mahasiswa, data akademik dan data hasil studi pelacakan (*tracer study*) lulusan. Dari tabel-tabel data, maka data yang dianalisis terdiri dari sebelas atribut sebagaimana terlihat pada Tabel 1.

**Tabel 1**  
**Atribut yang dianalisis**

ATRIBUT	DESKRIPSI
NIM	Nomor Induk Mahasiswa;
Asal SMA	nama SLTA asal dari mahasiswa;
Kota SMA	kota tempat SMA mahasiswa bersekolah;
Usia Lulus	usia mahasiswa saat lulus;
IPS Semester 1	Indek Prestasi Semester ke-1;
IPS Semester 2	Indek Prestasi Semester ke-2;
IPS Semester 3	Indek Prestasi Semester ke-3;
IPS Semester 4	Indek Prestasi Semester ke-4;
IPS Semester 5	Indek Prestasi Semester ke-5;
IPK Akhir	Indeks Prestasi Kumulatif yang diperoleh saat lulus;
Waktu Tunggu	masa tunggu lulusan mendapatkan pekerjaan pertama kali

Sumber: Hasil penelitian, 2021

Pada fase pra-pemrosesan, dilakukan proses pembersihan data (*cleaning*) dari duplikasi dan data yang tidak konsisten. Adapun penghapusan dilakukan pada data yang tidak lengkap, mengandung inkonsistensi dan tidak valid berdasarkan wawancara kepada pihak yang melakukan studi pelacakan pada obyek penelitian.

Tahap selanjutnya merupakan fase transformasi data. Transformasi data dilakukan untuk menyesuaikan data agar dapat diproses lebih fleksibel. Transformasi yang dilakukan meliputi transformasi data kota SMA/SLTA menjadi satuan jarak (kilometer), transformasi data tanggal lahir dan tanggal kelulusan menjadi tahun lulus. Transformasi untuk data asal kota SMA/SLTA menggunakan nilai jarak berdasarkan geografis area wilayah kota/kabupaten di Jakarta dan sekitar Jakarta, dengan aturan sebagaimana terlihat pada Tabel 2.

**Tabel 2 Transformasi Data Jarak Lokasi SLTA**

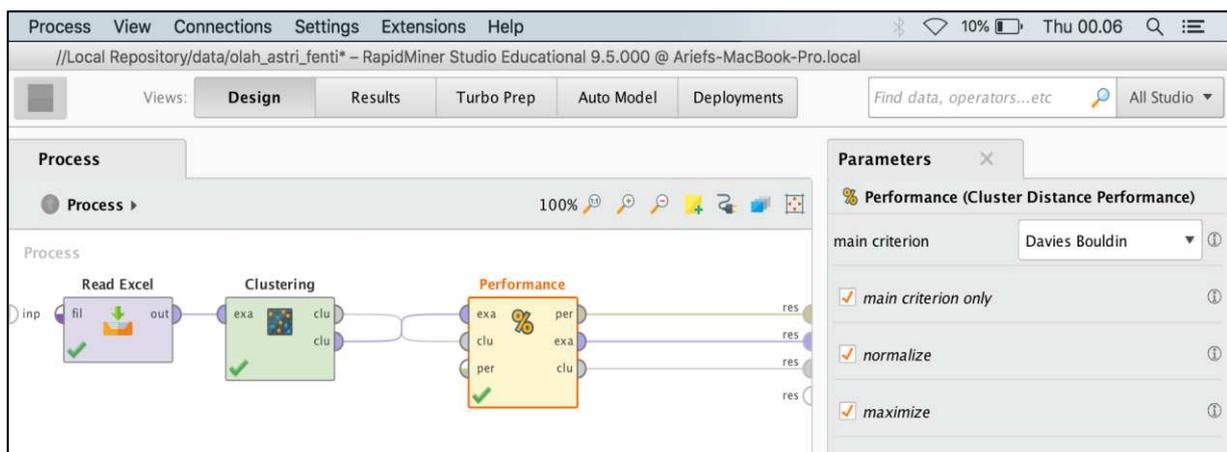
Kota	Transformasi
Jakarta	0
Tangerang Selatan	1

Kota	Transformasi
Bogor	2
Depok	2
Tangerang	2
Bekasi	2
Non-Jabodetabek	3

Sumber: Hasil penelitian, 2021

Pada Tabel 2 terlihat bahwa data wilayah kota/kabupaten dari lokasi SLTA sewaktu mahasiswa bersekolah ditransformasi dalam bentuk numerik, yaitu untuk jarak terdekat dinyatakan 0 untuk lokasi SLTA di wilayah Jakarta, sementara jarak kota lain yang terdekat dari Jakarta adalah Tangerang Selatan yang merupakan wilayah kota madya baru di sekitar Jakarta yang ditransformasi menjadi jarak sebesar 1. Untuk wilayah non Jakarta dan Tangsel ditransformasi menjadi jarak sebesar 2 untuk Jabodetabek dan non-Jabodetabek menjadi 3. Transformasi data juga dilakukan untuk mendapatkan usia lulusan dengan cara mengurangi atribut Tanggal Yudisium dengan atribut Tanggal Lahir sehingga didapatkan usia lulusan saat dinyatakan lulus/yudisium.

Proses pemodelan data mining dilakukan dengan tujuan untuk mencari informasi atau pola menggunakan algoritme K-Means. Dalam fase ini pemodelan diselesaikan dengan metode *Blackbox* pada aplikasi data mining yang telah dikenal yaitu *RapidMiner*. Pemodelan secara *Blackbox* berarti mekanisme analisis data akan bekerja berdasarkan fitur-fitur yang dimiliki oleh perangkat lunak *Rapidminer* dengan pemilihan algoritme yang ditentukan sebelum proses pemodelan dilakukan. Desain pemodelan pada palikasi *RapidMiner Studio* terlihat pada Gambar 3.



Gambar 3 Desain Pemodelan Klasterisasi pada Rapidminer Studio

Dari desain pemodelan sebagaimana terlihat pada Gambar 3 diketahui bahwa proses diawali dengan membaca file excel dengan fitur *Read Excel*, yang merupakan data pelatihan. Untuk

pemodelan, menggunakan metode *K-Means* dan sebagai tolok ukur validitas kluster yang terbentuk, digunakan *Davies-Bouldin Index* (DBI). Proses pemodelan dilakukan dengan percobaan yang membagi data mulai dari dua hingga sepuluh kluster. Pada fitur evaluasi (*evaluation*) pada *RapidMiner Studio*, dilakukan *setting* normalisasi (*normalize*) dan pemaksimalan (*maximize*) untuk memastikan proses evaluasi dilakukan pada model yang terbaik dengan nilai DBI yang positif. Hasil pemodelan ditunjukkan pada Tabel 3.

**Tabel 3 Hasil Pemodelan Klasterisasi**

No.	Jumlah kaster (k)	Nilai DBI
1.	2	0,336
2.	3	0,260
3.	4	0,280
4.	5	0,227
5.	6	0,221
6.	7	0,235
7.	8	0,235
8.	9	0,235
9.	10	0,229

Sumber: Hasil penelitian, 2021

Pada Tabel 3 dapat dilihat bahwa dengan berbagai percobaan yang dilakukan, mulai dari  $k=2$  hingga  $k=10$  ternyata model klasterisasi yang paling optimal, terbentuk pada pemodelan dengan jumlah kluster sebanyak enam ( $k=6$ ), yaitu dengan nilai DBI sebesar 0,221. Deskripsi masing-masing kluster yang terbentuk, ditampilkan pada Tabel 4.

**Tabel 4 Deskripsi Kluster Terbaik Yang Terbentuk**

Kluster	Jumlah anggota	Rata-rata			
		Jarak SLTA	Usia Lulus	IPK Akhir	Waktu Tunggu
I	67	2,24	21,13	3,53	1,69
II	39	1,10	21,31	3,41	-1,21
III	6	0,67	27,50	3,41	0,50
IV	33	0,33	21,73	3,53	3,00
V	71	0,25	21,23	3,59	1,21
VI	5	1,50	20,75	3,33	6,50

Sumber: Hasil penelitian, 2021

Berdasarkan Tabel 4 diketahui bahwa spesifikasi Kluster I adalah lulusan yang memiliki masa tunggu mendapat pekerjaannya adalah 1-2 bulan setelah diwisuda, dengan lokasi SLTA sebagian besar berada di luar Jabodetabek, dengan jumlah anggota sebanyak 67 orang. Dalam hal kualitas akademik, kluster ini memiliki IPK rata-rata di atas 3,50.

Klaster II adalah lulusan yang telah bekerja 1-2 bulan sebelum wisuda (memiliki masa tunggu bernilai negatif), dengan lokasi SLTA sebagian besar berada di wilayah penyangga DKI Jakarta, memiliki jumlah anggota sebanyak 39 orang. Dalam hal kualitas akademik, klaster ini memiliki IPK rata-rata di bawah 3,50.

Klaster III adalah lulusan yang memiliki masa tunggu paling singkat dalam memperoleh pekerjaannya yaitu sekitar 0-1 bulan sesudah wisuda, dengan lokasi sebagian besar berada di wilayah Jakarta, memiliki jumlah anggota sebanyak 6 orang. Klaster ini diisi oleh lulusan yang memiliki usia rata-rata 27,5 tahun, dan sebagian besar merupakan mahasiswa kelas sore. Dalam hal kualitas akademik, klaster ini memiliki IPK rata-rata di bawah 3,50.

Klaster IV adalah lulusan yang memiliki masa tunggu perolehan kerja sekitar 3 bulan sesudah wisuda, dengan lokasi SLTA sebagian besar berada di wilayah Jakarta, memiliki jumlah anggota sebanyak 33 orang. Dalam hal kualitas akademik, klaster ini memiliki IPK rata-rata di atas 3,50.

Klaster V adalah lulusan yang memiliki masa tunggu perolehan kerja lebih cepat dari klaster IV yaitu hanya sekitar 1-2 bulan sesudah wisuda, dengan lokasi SLTA sebagian besar berada di wilayah Jakarta. Klaster inilah yang paling banyak anggotanya, yaitu sekitar 32,27% dari lulusan atau 71 orang. Dari sisi kualitas akademik, kelompok ini merupakan klaster yang tertinggi nilai rata-rata IPK nya, yaitu sebesar 3,59.

Klaster VI adalah lulusan yang memiliki masa tunggu perolehan kerja paling lama yaitu sekitar 6 bulan sesudah wisuda, dengan lokasi SLTA sebagian berada di wilayah Jabodetabek, memiliki jumlah anggota sebanyak 4 orang. Klaster inilah yang paling sedikit anggotanya, sekitar 1,8% saja dari hasil studi pelacakan. Dalam hal kualitas akademik, klaster ini memiliki IPK rata-rata terendah yaitu 3,33.

Dalam hal evaluasi model, hampir sebagian besar klaster yang terbentuk memiliki nilai DBI terbaik pada rentang 0,22 hingga 0,23, yaitu pada pemodelan dengan 5, 6, 7, 8, 9 dan 10 klaster. Nilai DBI yang relatif kurang baik berada pada pemodelan dengan 2, 3 dan 4 klaster, yaitu antara 0,26 hingga 0,33. Dengan kondisi ini maka pembentukan klaster yang relatif ideal bagi lulusan Akademi Sekretari Budi Luhur, direkomendasikan pada posisi 5 atau 6 klaster, berdasarkan nilai-nilai evaluasi menggunakan *Davies Bouldin Index* (DBI).

Tahap akhir adalah interpretasi untuk menjelaskan makna dari setiap klaster yang terbentuk. Langkah ini dilakukan dengan cara membagi centroid dari iterasi terakhir dengan jumlah anggota setiap klasternya, sehingga dapat disimpulkan deskripsi dari setiap klaster. Dalam rangka memvisualisasi hasil evaluasi model dan mengetahui interpretasi pengetahuan, maka dibangun prototipe aplikasi klasterisasi yang berbasis web. Fitur awal dari prototipe aplikasi

adalah menampilkan informasi hasil seleksi data yang akan dimodelkan. Tampilan layar *Data Preview* atau pra-tinjau data dapat dilihat pada Gambar 4.

No	Student ID	High School	City	DOB	Year of Admission	Year of Graduation	Judisium Date	GPA	Job Waiting Time (Negative)	Job Waiting Time (Positive)
1	1221300021	Cyber Media	Jakarta	20/12/93	2012	2015	25 Agustus 2015	3.42	1	-
2	1221300039	SMK Muhammadiyah	Tangerang	04/04/93	2012	2015	25 Agustus 2015	2.99	1	-
3	1221300047	SMK Yadika 4	Tangerang	12/11/94	2012	2015	25 Agustus 2015	3.16	1	-
4	1221300096	SMAN 12	Tangerang	04/06/94	2012	2015	25 Agustus 2015	3.64	2	-
5	1221300104	SMKN 23 Jakarta	Jakarta	08/02/94	2012	2015	25 Agustus 2015	3.69	2	-
6	1221300138	SMK Yadika 4	Tangerang	05/04/94	2012	2015	25 Agustus 2015	3.09	-	1
7	1221300153	SMA Negeri 1 Pedes	Karawang	17/05/92	2012	2015	25 Agustus 2015	3.17	-	1
8	1221300179	SMK Jakarta 1	Jakarta	13/09/94	2012	2015	25 Agustus 2015	3.51	1	-

**Gambar 4 Tampilan Layar Pratinjau Data**

Dari gambar 4 terlihat bahwa prototipe aplikasi mampu membaca hasil pembacaan file excel sebagaimana perangkat lunak RapidMiner, dan dengan demikian maka proses klasterisasi dapat dilanjutkan. Proses klasterisasi merupakan fitur utama dari prototipe aplikasi ini. Hasil pembentukan klasterisasi dapat dilihat pada Gambar 5.

Student ID	Cluster	High School Distance	Age	GPA	Job Waiting Time
41	cluster_5	2	21	3.38	7
42	cluster_5	0	21	3.27	8
43	cluster_0	2	21	3.67	1
44	cluster_0	2	20	3.42	1
45	cluster_4	0	21	2.86	2
46	cluster_4	0	21	3.07	2
47	cluster_1	0	21	3.91	-1
48	cluster_1	0	21	3.27	-1
49	cluster_3	0	22	3.81	3

**Gambar 5 Tampilan Layar Hasil Klasterisasi**

Dari Gambar 5 terlihat bahwa prototipe aplikasi mampu menempatkan data lulusan ke dalam setiap klaster yang terbentuk untuk nilai k=6. Tahap akhir dari penelitian adalah menguji model prototipe aplikasi yang dibangun menggunakan 221 data populasi lulusan hasil pelacakan

studi. Proses validasi diselesaikan dengan cara memasukkan seluruh data secara random ke dalam aplikasi yang dikembangkan, untuk mendapatkan penempatan anggota klaster. Hasil pengujian aplikasi terlihat pada Tabel 5.

**Tabel 5 Hasil Pengujian Aplikasi**

Pemodelan dengan Rapidminer	Diklasterisasi dengan Tepat	Diklasterisasi dengan Tidak Tepat
Klaster 1	63	4
Klaster 2	37	2
Klaster 3	6	0
Klaster 4	29	4
Klaster 5	66	5
Klaster 6	4	1
Jumlah	205	16

Sumber: Hasil penelitian, 2021

Berdasarkan Tabel 5 diketahui bahwa dari 221 data yang telah diklasterisasi oleh aplikasi yang dikembangkan ternyata berhasil mengklaster dengan tepat pada 205 data, sementara peroses klasterisasi dengan hasil yang tidak tepat ada sebanyak 16 data. Dengan hasil tersebut maka akurasi pemodelan pada aplikasi yang dikembangkan adalah sebesar 92,76%, dengan nilai kesalahan penempatan klaster sebesar 7,24%.

## SIMPULAN

Dari seluruh tahapan penelitian yang dilakukan, maka dapat disimpulkan bahwa data akademik dan data hasil studi pelacakan dapat dielaborasi sebagai pemodelan klasterisasi menggunakan algoritme K-Means, pada nilai evaluasi *Davies Bouldin Index* sebesar 0,221 pada nilai  $k=6$ . Pemodelan klasterisasi yang dilakukan menggunakan algoritme K-Means berjalan dengan baik dan lancar pada perangkat lunak RapidMiner. Prototipe aplikasi yang dikembangkan berhasil melakukan pemodelan klasterisasi dengan nilai akurasi sebesar 92,76%.

Agar penelitian ini dapat dikembangkan pada masa mendatang maka analisis data dapat diperluas, misalnya pada bidang pekerjaan lulusan dan penghasilan awal pekerjaan. Hal ini perlu dilakukan bila klasterisasi akan difokuskan pada segmentasi lulusan bekerja berdasarkan seluruh hasil studi pelacakan. Selain itu, klasterisasi juga dapat menggambarkan mutu lulusan secara komprehensif, tidak sebatas pada data hasil studi pelacakan, namun juga pada data nilai mata kuliah-mata kuliah kompetensi, yang dapat digunakan untuk memetakan segmentasi lulusan berdasarkan Kerangka Kualifikasi Nasional Indonesia (KKNI) dan Standar Kompetensi Kerja Nasional Indonesia (SKKNI).

**Daftar Pustaka**

- Abdurrahman, G. (2016). Clustering Data Ujian Tengah Semester Data Mining Menggunakan Algoritma K Means. *Jurnal Sistem dan Teknologi Informasi Indonesia*, 1(2), 71–79.
- Hariyanto, M., & Shita, R. T. (2018). Penyakit DBD Menggunakan Metode Algoritma K-Means dan Metode, 1(1), 117–122.
- Novita Sari, V., Yupianti, Y., & Maharani, D. (2018). Penerapan Metode K-Means Clustering Dalam Menentukan Predikat Kelulusan Mahasiswa Untuk Menganalisa Kualitas Lulusan. *Jurteks*, 4(2), 133–140. <https://doi.org/10.33330/jurteks.v4i2.53>
- Pradnyana, G. A., & Permana, A. A. J. (2018). Sistem Pembagian Kelas Kuliah Mahasiswa Dengan Metode K-Means Dan K-Nearest Neighbors Untuk Meningkatkan Kualitas Pembelajaran. *JUTI: Jurnal Ilmiah Teknologi Informasi*, 16(1), 59. <https://doi.org/10.12962/j24068535.v16i1.a696>
- Priyatman, H., Sajid, F., & Haldivany, D. (2019). Klasterisasi Menggunakan Algoritma K-Means Clustering untuk Memprediksi Waktu Kelulusan Mahasiswa. *Jurnal Edukasi dan Penelitian Informatika (JEPIN)*, 5(1), 62. <https://doi.org/10.26418/jp.v5i1.29611>
- Rosmini, R., Fadlil, A., & Sunardi, S. (2018). Implementasi Metode K-Means Dalam Pemetaan Kelompok Mahasiswa Melalui Data Aktivitas Kuliah. *It Journal Research and Development*, 3(1), 22. [https://doi.org/10.25299/itjrd.2018.vol3\(1\).1773](https://doi.org/10.25299/itjrd.2018.vol3(1).1773)
- Setiawan, R. (2016). Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Untuk Menentukan Strategi Promosi Mahasiswa Baru (Studi Kasus: Politeknik LP3I Jakarta). *Jurnal Lentera ICT*, 3(1), 76–92.
- Sibuea, M. L., & Safta, A. (2017). Pemetaan Siswa Berprestasi Menggunakan Metode K-Means Clustering. *Jurteks*, 4(1), 85–92. <https://doi.org/10.33330/jurteks.v4i1.28>
- Wardhani, A. K. (2016). Implementasi Algoritma K-Means Untuk Pengelompokan Penyakit Pasien Pada Puskesmas Kajen Pekalongan. *Jurnal Transformatika*, 14, 30–37.
- Wibowo, A., Makruf, M., Virdyna, I., Venna, F.C., (2020). Penentuan Klaster Koridor TransJakarta dengan Metode Majority Voting pada Algoritma Data Mining. *Jurnal RESTI*, 5(3), 565-575.