

K-MEANS CLUSTERING DATA NIK MAHASISWA BARU MENGUNAKAN ALTAIR AI-STUDIO

Yogi Bachtiar^{1*}, Rayung Wulan², Aswin Fitriansyah³

^{1,2,3}Program Studi Teknik Informatika, Universitas Indraprasta PGRI
yogi.bachtiar@gmail.com

Submitted January 3, 2025; Revised August 4, 2025; Accepted August 5, 2025

Abstrak

Di era digital yang serba terhubung ini, kebutuhan untuk memiliki identitas digital yang aman dan terpercaya menjadi semakin penting. Salah satu langkah besar yang diambil Indonesia dalam mencapai hal tersebut adalah dengan memanfaatkan NIK (Nomor Induk Kependudukan) sebagai single data identitas warga negara. NIK merupakan nomor unik yang diberikan kepada setiap warga negara Indonesia sejak kelahirannya. Tidak hanya itu, pemanfaatan NIK sebagai single data juga memungkinkan pemerintah untuk melakukan analisis data secara lebih efektif. Dengan adanya data terpusat, pemerintah dapat mengambil keputusan dan membuat kebijakan yang lebih tepat sasaran dan sesuai dengan kebutuhan masyarakat. Dalam penelitian ini, penulis menerapkan teknik data mining menggunakan algoritma K-Means. Pendekatan tersebut digunakan untuk mengelompokkan data berdasarkan jarak antar data, yang mengorganisir data ke dalam sejumlah kluster yang telah ditentukan. Tahapan penelitian dimulai dari Persiapan data, Transformasi data, Pemodelan data dan Evaluasi data. Hasil dari penelitian pengujian algoritma K-mean menggunakan Altair AI-Studio dapat disimpulkan bahwa sebaran mahasiswa baru Universitas Indraprasta PGRI tahun akademik 2023/2024 berdasarkan NIK tersebar pada 32 dari 37 Provinsi di Indonesia dengan populasi terbanyak dari provinsi DKI Jakarta yaitu 4.630 Mahasiswa dan Jawa Barat sebanyak 4.219 Mahasiswa yang masuk dalam kategori Cluster0.

Kata Kunci : Klasterisasi, K-Means, Altair Ai-Studio

Abstract

In this connected digital era, the need to have a secure and trusted digital identity is becoming increasingly important. One of the major steps taken by Indonesia in achieving this is by utilizing the NIK (Population Identification Number) as a single citizen identity data. NIK is a unique number given to every Indonesian citizen since their birth. Not only that, the utilization of NIK as single data also allows the government to conduct data analysis more effectively. With centralized data, the government can make decisions and make policies that are more targeted and in accordance with the needs of the community. In this research, the author applies data mining techniques using the K-Means algorithm. The approach is used to group data based on the distance between data, which organizes data into a predetermined number of clusters. The research stages start from data preparation, data transformation, data modeling and data evaluation. The results of the K-mean algorithm testing research using Altair AI-Studio can be concluded that the distribution of new Indraprasta PGRI University students for the 2023/2024 academic year based on NIK is spread across 32 of the 37 provinces in Indonesia with the largest population from DKI Jakarta province, namely 4,630 students and West Java as many as 4,219 students who are included in the Cluster0 category.

Keywords : Clustering, K-Means, Altair Ai-Studio

1. PENDAHULUAN

Di era digital yang serba terhubung ini, kebutuhan untuk memiliki identitas digital yang aman dan terpercaya menjadi

semakin penting. Salah satu langkah besar yang diambil Indonesia dalam mencapai hal tersebut adalah dengan memanfaatkan NIK (Nomor Induk Kependudukan)

sebagai single data identitas warga negara. NIK merupakan nomor unik yang diberikan kepada setiap warga negara Indonesia sejak kelahirannya. Nomor ini tidak hanya berfungsi sebagai identitas legal, tetapi juga menjadi kunci utama dalam proses verifikasi dan validasi data seseorang di berbagai sektor kehidupan. Dengan adanya NIK sebagai single data, proses administrasi dan layanan publik menjadi lebih efisien dan terintegrasi[1]. Sebagai contoh, saat seorang warga ingin mengurus dokumen kependudukan seperti Kartu Tanda Penduduk (KTP) atau akta kelahiran, mereka hanya perlu menyebutkan NIK, dan seluruh data terkait akan terhubung secara otomatis.

Selain itu, NIK juga berperan penting dalam layanan digital lainnya, seperti pembukaan rekening bank, pendaftaran asuransi, atau bahkan proses pemilu. Dengan menggunakan NIK, proses verifikasi identitas menjadi lebih akurat dan aman, mencegah kemungkinan terjadinya penipuan atau penyalahgunaan data. Tidak hanya itu, pemanfaatan NIK sebagai single data juga memungkinkan pemerintah untuk melakukan analisis data secara lebih efektif[2]. Dengan adanya data terpusat, pemerintah dapat mengambil keputusan dan membuat kebijakan yang lebih tepat sasaran dan sesuai dengan kebutuhan masyarakat.

Di Universitas Indraprasta PGRI, proses penerimaan mahasiswa baru selalu menjadi momen penting setiap tahunnya. Ribuan calon mahasiswa dari berbagai latar belakang dan daerah asal mengajukan pendaftaran, membawa harapan dan mimpi mereka untuk menempuh pendidikan tinggi. Namun, di balik ribuan data tersebut, tersimpan informasi berharga yang dapat membantu Universitas dalam mengelola dan melayani mahasiswa barunya dengan lebih baik. Untuk mengatasi tantangan ini, Universitas Indraprasta PGRI memutuskan untuk

menggunakan teknik klasterisasi data, sebuah metode analisis data yang mengategorikan objek (dalam hal ini, mahasiswa baru) ke dalam kelompok atau klaster berdasarkan kesamaan karakteristik tertentu. Dengan melakukan klasterisasi, Universitas dapat mengidentifikasi pola, tren, dan hubungan tersembunyi dalam data mahasiswa baru yang dimilikinya[3].

Pada penelitian sebelumnya telah dibahas mengenai “Klasterisasi Profil Mahasiswa Baru Menggunakan Algoritma K-Means Pada Fakultas Teknologi Informasi Universitas Budi Luhur” oleh Vanessa Melati Al Aqsa yang menghasilkan kesimpulan 3 buah klaster yang terdiri dari cluster satu adalah klaster Teknik Informatika, asal sekolah SMK, dengan kota asal Kecamatan Larangan, Karang Tengah, Kebayoran Lama, dan Kembangan dengan jumlah 240 items, klaster dua adalah klaster Teknik Informatika, asal sekolah SMK, dengan kota asal Kecamatan Pesanggrahan dan Pondok Aren adalah dengan jumlah 242 items, dan cluster tiga adalah cluster Teknik Informatika, asal sekolah SMK, dengan kota asal Ciledug dengan jumlah 205 items, dengan nilai Davies-Bouldin (DB) sebesar 0,470[4]. Berikutnya Nanda Arista Rizki dengan Judul “Implementasi Algoritma K-Means Untuk Mengelompokkan Mahasiswa Program Studi Pendidikan Matematika Berdasarkan Sumber Belajarnya” Penelitian tersebut menghasilkan 3 kelompok utama menurut penggunaan sumber belajarnya. Sumber belajar yang menjadi penentu dalam pembagian kelompok adalah buku elektronik dan jurnal. Kelompok pertama menggunakan buku elektronik dan jurnal, sedangkan kelompok ketiga tidak menggunakan keduanya. Sementara kelompok kedua hanya menggunakan buku elektronik[5]. Serta adapula penelitian dari Ronaldi Mbanimara dengan topik “Klasifikasi Pemetaan Penduduk Penerima Bantuan Renovasi Rumah Menggunakan

Algoritma K-Means” dengan menghasilkan tiga kelompok penduduk sesuai dengan pendapatan rata-rata dan status kepemilikan bangunan. Hasil tiga kelompok tersebut yaitu penduduk yang layak, kurang layak dan tidak layak menerima bantuan[6].

Penambangan Data

Data telah menjadi aset berharga bagi setiap organisasi. Namun, seringkali data yang dimiliki begitu besar dan kompleks, sehingga menyembunyikan informasi berharga yang tersembunyi di balik tumpukan data tersebut. Inilah di mana data mining berperan penting. Data mining, atau pertambangan data, adalah sebuah proses untuk mengekstrak pola, tren, dan pengetahuan yang berguna dari kumpulan data yang besar dan beragam[7]. Layaknya seorang penambang yang menggali bijih berharga dari dalam tanah, data mining memungkinkan kita untuk menggali informasi berharga yang tersembunyi di balik gunung data yang kita miliki. Proses data mining melibatkan serangkaian teknik dan algoritma canggih yang mampu menganalisis data dari berbagai sumber, seperti database, file teks, gambar, video, dan bahkan data streaming. Algoritma ini dirancang untuk menemukan pola, korelasi, aturan asosiasi, dan hubungan lain yang mungkin tidak terlihat oleh mata manusia[8].

Klasterisasi Data

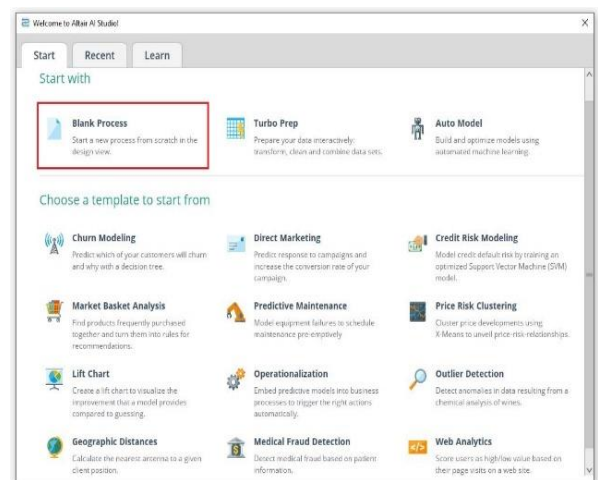
Klasterisasi data (clustering data) adalah proses pengelompokan objek data ke dalam beberapa kelompok atau cluster berdasarkan kesamaan karakteristik atau fitur yang dimiliki oleh data-data tersebut. Tujuan utama dari klasterisasi data adalah untuk mengidentifikasi pola, struktur, atau hubungan tersembunyi dalam dataset yang besar dan kompleks[9]. Dalam klasterisasi data, setiap cluster terdiri dari objek data yang mirip satu sama lain dalam cluster yang sama dan berbeda dengan objek data dalam cluster lainnya. Kesamaan ini diukur

berdasarkan metrik jarak atau kesamaan yang ditentukan, seperti jarak Euclidean, jarak Manhattan, atau korelasi[10].

Algoritma K-Means

Di dalam dunia analisis data, terdapat banyak teknik dan metode yang dapat digunakan untuk mengelompokkan atau mengklasifikasikan data. Salah satu algoritma yang populer dan banyak digunakan adalah K-Means Clustering[11]. Algoritma ini memungkinkan kita untuk mengelompokkan data berdasarkan kesamaan karakteristik atau fitur yang dimiliki oleh setiap data. Algoritma K-Means adalah salah satu algoritma clustering yang paling populer dan banyak digunakan dalam penambangan data (data mining) dan analisis data. Algoritma ini termasuk dalam kategori clustering partisi, di mana data dikelompokkan ke dalam K jumlah cluster yang telah ditentukan sebelumnya.[12]

Altair AI Studio



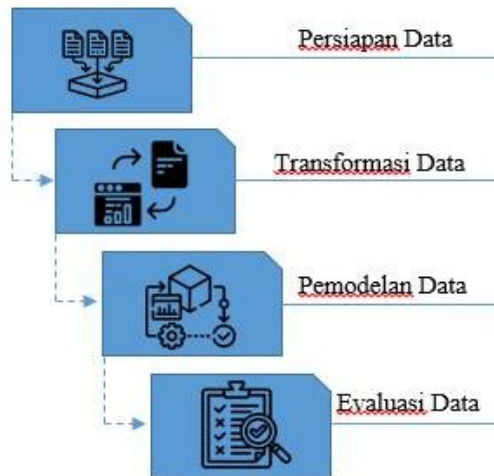
Sumber : Dokumen Pribadi

Gambar 1. Alatair Ai Studio

Klasterisasi data PMB ini dilakukan dengan bantuan tool Altair AI Studio yang merupakan hasil akuisisi dari aplikasi sejenis dengan nama RapidMiner. Proses ini berguna dalam analitik data canggih (AI) dan perangkat lunak pembelajaran mesin (*Machine Learning*).

2. METODE PENELITIAN

Alur Penelitian



Sumber : Dokumen Pribadi

Gambar 2. Alur Penelitian

Alur penelitian yang dilakukan yaitu dimulai dari Persiapan Data. Data awal yang bersumber dari data PMB Universitas Indraprasta PGRI tersebut dilakukan Transformasi dalam bentuk pengkodean dari data yang bersifat *VarChar* ke tipe data *Integer*. Langkah selanjutnya dilakukan Pemodelan data menggunakan Altair AI-Studio untuk menghasilkan *insight* data yang diharapkan. Berikutnya untuk memastikan penentuan klaster maka diperlukan Evaluasi data dengan Davies Bouldien Index (DBI)[13].

Teknik yang diterapkan dalam analisis statistik ini adalah analisis cluster, yaitu metode untuk menganalisis data dengan tujuan mengelompokkan data ke dalam beberapa kategori individu atau objek yang mempunyai karakteristik berbeda. Dengan demikian, individu atau objek tersebut tergabung dalam satu kelompok yang memiliki sifat-sifat yang relatif serupa. Kemiripan tersebut diukur berdasarkan jarak yang ada. Berikut merupakan tahapan dalam melakukan analisis cluster K-Means yaitu[14] :

- 1) Tentukan banyaknya cluster Penentuan banyaknya klaster pada K-Means

dengan menggunakan metode *rule of thumb*

$$K = \sqrt{\frac{n}{2}} \quad (1)$$

K = banyaknya cluster
n = banyaknya data

- 2) Membagi objek penelitian pada tahap K-cluster awal.
- 3) Lalu tiap objek dimasukkan ke suatu cluster yang memiliki rata-rata terdekat (jarak). Untuk menentukan jaraknya menggunakan *Euclidean*. Hitung ulang jarak untuk kelompok yang menerima objek baru dan yang mendapat pengurangan objek. Persamaan jarak *euclidean* adalah sebagai berikut :

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

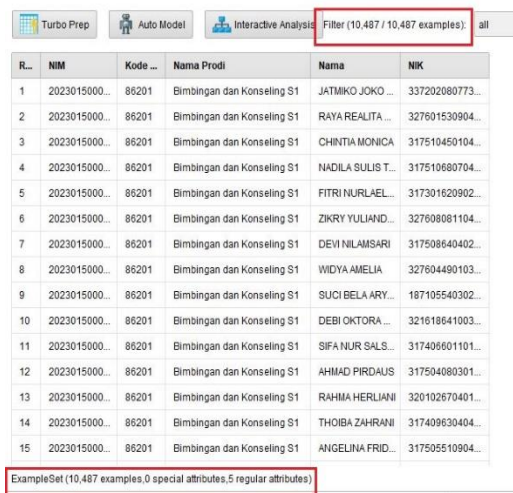
- 4) Ulangi tahapan ke-2 sampai tidak terjadi perpindahan objek antar klaster.

Hasil akhir dari penentuan objek berpindah ke klaster tertentu tidak berdasarkan dari penentuan K pertama. Hal ini menunjukkan bahwa perubahan terbesar terjadi pada perpindahan yang pertama saja[15].

3. HASIL DAN PEMBAHASAN

Persiapan Data

Data yang digunakan adalah data Mahasiswa baru Universitas Indraprasta PGRI tahun ajar Ganjil 2023/2024. Dengan total 10.487 *record* Data yang berisikan atribut NIM, KodeProdi, NamaProdi, Nama dan NIK.

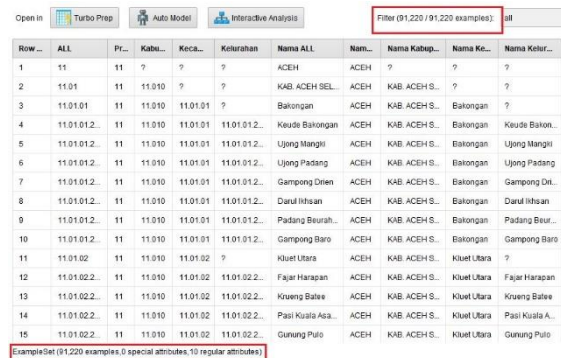


R.	NIM	Kode ...	Nama Prodi	Nama	NIK
1	2023015000...	86201	Bimbingan dan Konseling S1	JATMOKO JOKO ...	337202080773...
2	2023015000...	86201	Bimbingan dan Konseling S1	RAYA REALITA ...	327601530904...
3	2023015000...	86201	Bimbingan dan Konseling S1	CHINTIA MONICA	317510450104...
4	2023015000...	86201	Bimbingan dan Konseling S1	NADILA SULIS T...	317510680704...
5	2023015000...	86201	Bimbingan dan Konseling S1	FITRI NURLAEL...	317301620902...
6	2023015000...	86201	Bimbingan dan Konseling S1	ZKRY YULIAND...	327608081104...
7	2023015000...	86201	Bimbingan dan Konseling S1	DEVI NILAMSARI	317508640402...
8	2023015000...	86201	Bimbingan dan Konseling S1	WIDYA AMELIA	327604490103...
9	2023015000...	86201	Bimbingan dan Konseling S1	SUCI BELA ARY...	187105540302...
10	2023015000...	86201	Bimbingan dan Konseling S1	DEBI OKTORA ...	321618641003...
11	2023015000...	86201	Bimbingan dan Konseling S1	SIFA NUR SALS...	317406601101...
12	2023015000...	86201	Bimbingan dan Konseling S1	AHMAD PIRDAUS	317504080301...
13	2023015000...	86201	Bimbingan dan Konseling S1	RAHMA HERLIANI	320102670401...
14	2023015000...	86201	Bimbingan dan Konseling S1	THOIBA ZAHIRANI	317409630404...
15	2023015000...	86201	Bimbingan dan Konseling S1	ANGELINA FRID...	317505510904...

Sumber : Dokumen Pribadi

Gambar 1. Data Mahasiswa Baru

Data berikutnya adalah data kode NIK semua wilayah di Indonesia yang atributnya terdiri Kode Provinsi, Kabupaten, Kecamatan dan Kelurahan yang sesuai dengan peraturan Permendagri No. 72/2019 data hingga tingkat Kelurahan Desa. Data tersebut berisikan 91.220 record.



Row ...	ALL	Pr...	Kabu...	Keca...	Kelurahan	Nama ALL	Nam...	Nama Kabep...	Nama Ke...	Nama Kelur...
1	11	11	?	?	?	ACEH	ACEH	?	?	?
2	11.01	11	11.010	?	?	KAB. ACEH SEL.	ACEH	KAB. ACEH S...	?	?
3	11.01.01	11	11.010	11.01.01	?	Bakongan	ACEH	KAB. ACEH S...	Bakongan	
4	11.01.01.2	11	11.010	11.01.01	11.01.01.2	Kecude Bakongan	ACEH	KAB. ACEH S...	Bakongan	Kecude Bakon...
5	11.01.01.2	11	11.010	11.01.01	11.01.01.2	Ujung Mangli	ACEH	KAB. ACEH S...	Bakongan	Ujung Mangli
6	11.01.01.2	11	11.010	11.01.01	11.01.01.2	Ujung Padang	ACEH	KAB. ACEH S...	Bakongan	Ujung Padang
7	11.01.01.2	11	11.010	11.01.01	11.01.01.2	Gampong Drien	ACEH	KAB. ACEH S...	Bakongan	Gampong Dri...
8	11.01.01.2	11	11.010	11.01.01	11.01.01.2	Darul Ihsan	ACEH	KAB. ACEH S...	Bakongan	Darul Ihsan
9	11.01.01.2	11	11.010	11.01.01	11.01.01.2	Padang Beurah...	ACEH	KAB. ACEH S...	Bakongan	Padang Beur...
10	11.01.01.2	11	11.010	11.01.01	11.01.01.2	Gampong Baro	ACEH	KAB. ACEH S...	Bakongan	Gampong Baro
11	11.01.02	11	11.010	11.01.02	?	Kluet Utara	ACEH	KAB. ACEH S...	Kluet Utara	?
12	11.01.02.2	11	11.010	11.01.02	11.01.02.2	Fajar Harapan	ACEH	KAB. ACEH S...	Kluet Utara	Fajar Harapan
13	11.01.02.2	11	11.010	11.01.02	11.01.02.2	Krueng Balee	ACEH	KAB. ACEH S...	Kluet Utara	Krueng Balee
14	11.01.02.2	11	11.010	11.01.02	11.01.02.2	Pasi Kuala Aca...	ACEH	KAB. ACEH S...	Kluet Utara	Pasi Kuala A...
15	11.01.02.2	11	11.010	11.01.02	11.01.02.2	Gunung Pulo	ACEH	KAB. ACEH S...	Kluet Utara	Gunung Pulo

Sumber : Dokumen Pribadi

Gambar 2. Data Kode Wilayah

Transformasi Data

Dari kedua data seperti Gambar 3. dan Gambar 4. dilakukan *cleansing noise* dan *inner join* antar tabel. Sehingga didapatkan data gabungan yang siap diolah. Serta berikutnya *labeling* data yang menghasilkan data Uji NIK yang siap dilakukan pemodelan klasterisasi. Data

yang sudah layak untuk pemodelan kemudian akan dilakukan import.



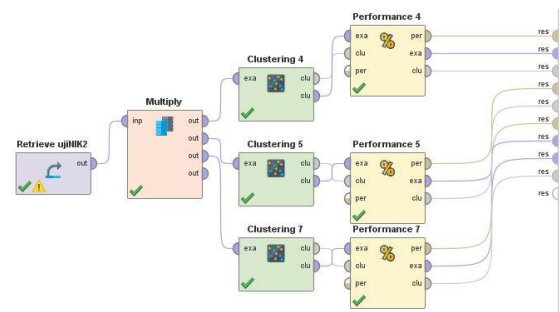
Row ...	Prov1	NIK	Kode Prodi	Prov	Kab	Kec	NIM	Kecamatan
1	BANTEN	3604151010...	23201	36	3604	360415	202345500...	20518
2	BANTEN	3674046011...	23201	36	3674	367404	202345500...	286301
3	BANTEN	3674030503...	23201	36	3674	367403	202345500...	286304
4	BANTEN	3674054305...	23201	36	3674	367405	202345500...	286302
5	BANTEN	3671091807...	23201	36	3671	367109	202345500...	280317
6	BANTEN	3603140705...	23201	36	3603	360314	202345500...	280321
7	BANTEN	3601121101...	23201	36	3601	360112	202345500...	280112
8	BANTEN	3674065109...	23201	36	3674	367406	202345500...	18401
9	BANTEN	3603122303...	23201	36	3603	360312	202345500...	20517
10	BANTEN	3603123110...	23201	36	3603	360312	202345500...	280312
11	DAERAH ISTL.	3471100604...	23201	34	3471	347110	202345500...	46009
12	DKI JAKARTA	3175052911...	23201	31	3175	317505	202345500...	18401
13	DKI JAKARTA	3175061303...	23201	31	3175	317506	202345500...	18408
14	DKI JAKARTA	3174011307...	23201	31	3174	317401	202345500...	18309
15	DKI JAKARTA	3173043112...	23201	31	3173	317304	202345500...	16205

Sumber : Dokumen Pribadi

Gambar 3. Data Uji NIK

Pemodelan dengan Altair AI-Studio

Dalam tahapan ini, penulis menerapkan teknik data mining menggunakan algoritma K-Means. Pendekatan tersebut digunakan untuk mengelompokkan data berdasarkan jarak antar data, yang mengorganisir data ke dalam sejumlah cluster yang telah ditentukan.



Sumber : Dokumen Pribadi

Gambar 4. Pemodelan K-Means

Data “UjiNIK” di *import* kedalam *canvas* Altair AI-Studio dengan operator *Retrieve*, kemudian ditambahkan juga operator *Multiply*, *Clustering K-Means* dan *Performance*. Operator *Multiply* digunakan sebagai alat untuk melakukan beberapa pemodelan dalam satu waktu *Running*. Dalam penelitian ini dimodelkan 3 klaster dan 3 performa. Antar *node* disambungkan

untuk melakukan pemrosesan yang pada akhirnya menghasilkan *Performance*, *Example dan Cluster Result*.

Evaluasi dan Hasil

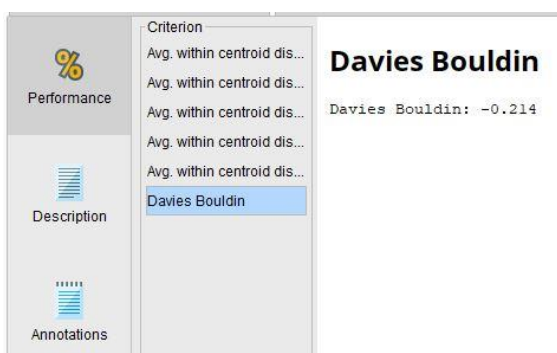
Dalam melakukan Evaluasi Klaster dan Performa diperlukan dasar pemilihan Klaster terbaik yang akan dilakukan analisis secara mendalam. Maka dari itu dilakukan beberapa kali percobaan.

Tabel 1. Davies Bouldin Index

Percobaan	Jumlah klaster	Nilai DBI
1	Klaster K=2	-0.421
2	Klaster K=3	-0.395
3	Klaster K=4	-0.214
4	Klaster K=5	-0.359
5	Klaster K=6	-0.356
6	Klaster K=7	-0.241
7	Klaster K=8	-0.290
8	Klaster K=9	-0.242

Sumber : Dokumen Pribadi

Penentuan Klaster terpilih tersebut mengikuti aturan Davies Bouldin Index (DBI) dengan memilih DBI terkecil pada setiap Klaster yang di uji coba[13]. Dari **Tabel 1.** didapati nilai DBI terkecil yaitu pada percobaan yang ke-3 / **Klaster K=4** dengan nilai Davies Bouldin Index : -**0.214**. Maka dari hasil tersebut K=4 akan dilakukan pemodelan *K-Means Clustering* dengan *Result* antara lain *Performance*, *ExampleSet* dan *Cluster Result*.

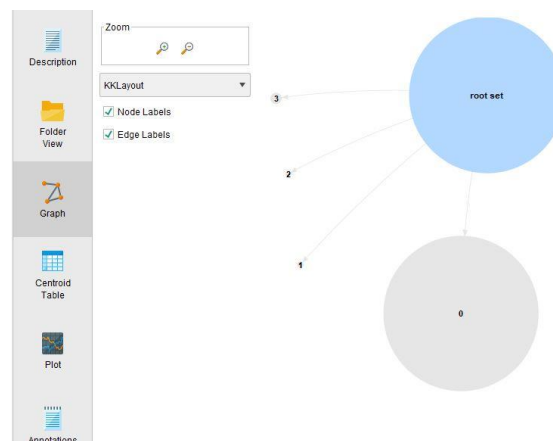


Sumber : Dokumen Pribadi

Gambar 5. Davies Bouldin Index Ke-3

Dari hasil Evaluasi Performa pemilihan Klaster K=4 diatas maka dilakukan pula evaluasi mengenai *Cluster Model*. Dengan

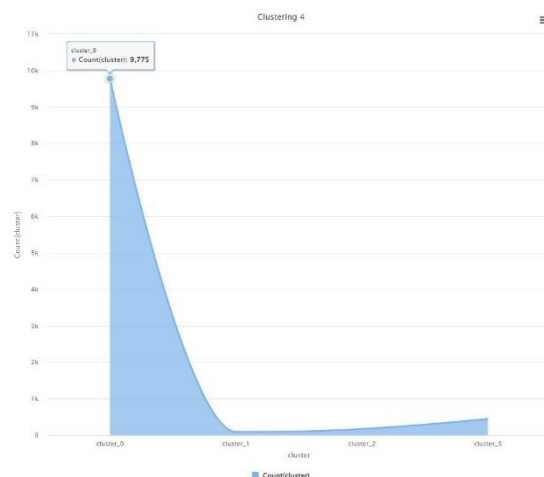
masih menggunakan K=4 sebagai Modelnya didapati hasil sebagai berikut :



Sumber : Dokumen Pribadi

Gambar 6. Graph Performa Cluster4

Pada Gambar 8. Menunjukkan terdapat 4 Klaster dengan *Cluster0* sebagai Klaster terbesar. Selanjutnya dilakukan pula pada pemodelan K=4 dengan anotasi node *Example* yang menghasilkan data *ExampleSet* sebagai berikut :



Sumber : Dokumen Pribadi

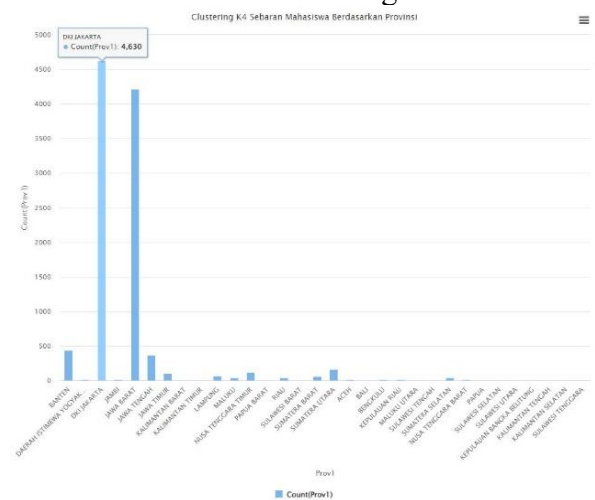
Gambar 7. Jumlah ExampleSet

Dari Gambar 9. Menunjukkan bahwa *Cluster0* berisikan data *ExampleSet* sebanyak 9.775 item data NIK. Dari 4 Klaster tersebut berikut data sebarannya.

Cluster	Count	Prov	Sum
Cluster0	6	Banten, Daerah Istimewa Yogyakarta, DKI Jakarta, Jawa Barat, Jawa Tengah, Jawa Timur.	9775
Cluster1	9	Maluku, Papua Barat, Sulawesi Barat, Maluku Utara, Sulawesi Tengah, Papua, Sulawesi Selatan, Sulawesi Utara, Sulawesi Tenggara.	90
Cluster2	7	Kalimantan Barat, Kalimantan Timur, Nusa Tenggara Timur, Bali, Nusa Tenggara Barat, Kalimantan Tengah, Kalimantan Selatan	169
Cluster3	10	Jambi, Lampung, Riau, Sumatera Barat, Sumatera Utara, Aceh, Bengkulu, Kepulauan Riau, Sumatera Selatan, Kepulauan Bangka Belitung.	445

Provinsi dengan sebaran Jambi, Lampung, Riau, Sumatera Barat, Sumatera Utara, Aceh, Bengkulu, Kepulauan Riau, Sumatera Selatan, Kepulauan Bangka Belitung. Jumlah mahasiswanya yaitu 445.

Dari hasil pemodelan *K-Means Clustering* dengan 4 klaster terpilih menghasilkan data jumlah sebaran Mahasiswa baru Universitas Indraprasta PGRI per provinsi di seluruh Indonesia sebagai berikut :



Gambar 8. Sebaran data Mahasiswa

4. SIMPULAN

Dari penelitian pengujian algoritma *K-Mean* menggunakan Altair AI-Studio diatas dapat disimpulkan bahwa sebaran mahasiswa baru Universitas Indraprasta PGRI tahun akademik 2023/2024 berdasarkan NIK tersebar pada 32 dari 37 Provinsi di Indonesia dengan populasi terbanyak dari provinsi DKI Jakarta yaitu 4.630 Mahasiswa dan Jawa Barat sebanyak 4.219 Mahasiswa yang masuk dalam kategori *Cluster0*. Dengan data sebaran tersebut diharapkan kampus dapat melihat lebih jelas pangsa pasar yang sudah dimilikinya dan dapat lebih agresif lagi dalam promosi diwilayah tersebut agar tidak diambil oleh pesaing sesama kampus diwilayah yang sama.

DAFTAR PUSTAKA

- [1] M. Subur, M. Martanto, and U. Hayati, "Clustering Penduduk Miskin Untuk Penerima Bantuan Sosial Menggunakan Algoritma K-Means," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 1, pp. 789–795, 2024, doi: 10.36040/jati.v8i1.8809.
- [2] E. Rahayu, N. Erawati, and R. Ananda, "Klasifikasi Kelayakan Warga Penerima BPNT dengan Algoritma k-Nearest Neighbor Classification of Eligibility for BPNT Recipients using the k - Nearest Neighbor Algorithm," *Sistemasi*, vol. 13, no. 1, pp. 107–113, 2024, [Online]. Available: <http://sistemasi.ftik.unisi.ac.id/index.php/stmsi/article/download/3250/615>
- [3] N. A. Rahmalinda and A. Jananto, "Penerapan Metode K-Means Clustering Dalam Menentukan Strategi Promosi Berdasarkan Data Penerimaan Mahasiswa Baru," *J. Tekno Kompak*, vol. 16, no. 2, pp. 163–175, 2022.
- [4] V. M. Al Aqsa, D. Mahdiana, Rusdah, and D. Anubhakti, "Klasterisasi Profil Mahasiswa Baru Menggunakan Algoritme K-Means Pada Fakultas Teknologi Informasi Universitas Budi Luhur," *Semin. Nas. Mhs. Fak. Teknol. Inf.*, vol. 2, no. 1, pp. 309–315, 2023.
- [5] N. A. Rizki, K. Kurniawan, I. K. Hasan, and N. Sampe, "Implementasi Algoritma K-Means Untuk Mengelompokkan Mahasiswa Berdasarkan Sumber Belajarnya," *Metik J.*, vol. 7, no. 2, pp. 62–67, 2023, doi: 10.47002/metik.v7i2.584.
- [6] R. Mbanimara and W. Saputro, "Klasifikasi Pemetaan Penduduk Penerima Bantuan Renovasi Rumah Menggunakan Algoritma K- Means Ronaldi," *J. Pendidik. dan Konseling*, vol. 4, no. 5, pp. 637–646, 2022.
- [7] N. Putri Apriliani, "Perlindungan Hukum Terhadap Korban Penyalahgunaan Data Pribadi (Studi Kasus Penyalahgunaan NIK dalam Proses Pendaftaran Bacaleg di Lampung)," *UNEW LAW Rev.*, vol. 6, no. 2, p. hlm 6130, 2023, [Online]. Available: <https://review-unes.com/https://creativecommons.org/licenses/by/4.0/>
- [8] N. Noviyanto and P. Ekasari, "Algoritma K-Means Untuk Klasterisasi Jabatan Fungsional Dosen Pada Perguruan Tinggi Swasta Di Lingkungan LLDikti Wilayah III," *Paradig. - J. Komput. dan Inform.*, vol. 24, no. 1, pp. 103–107, 2022, doi: 10.31294/paradigma.v24i1.1112.
- [9] L. Susanti, "Klasifikasi Tingkat Stress Pada Mahasiswa Teknik Informatika Dalam Melakukan Perkuliahan Metode Hybrid Menggunakan Algoritma Naive Bayes," *STRING (Satuan Tulisan Ris. dan Inov. Teknol.*, vol. 8, no. 3, pp. 243–248, 2024.
- [10] M. P. A. Ariawan, I. B. A. Peling, and G. B. Subiksa, "Prediksi Nilai Akhir Matakuliah Mahasiswa Menggunakan Metode K-Means Clustering (Studi Kasus : Matakuliah Pemrograman Dasar)," *J. Nas. Teknol. dan Sist. Inf.*, vol. 9, no. 2, pp. 122–131, 2023, doi: 10.25077/teknosi.v9i2.2023.122-131.
- [11] R. Wulan, T. T. Widaningsih, and F. Yanuar, "Clustering the K-means Algorithm with the Approach to Student Interpersonal Communication Patterns in Selecting Secondary Schools," *Fakt. Exacta*, vol. 16, no. 4, pp. 332–339, 2024, doi: 10.30998/faktorexacta.v16i4.20852.

- [12] R. F. N. Alifah and A. C. Fauzan, "Implementasi Algoritma K-Means Clustering Berbasis Jarak Manhattan untuk Klasterisasi Konsentrasi Bidang Mahasiswa," *ILKOMNIKA: Journal of Computer Science and Applied Informatics*, vol. 5, no. 1, pp. 31–41, 2023. doi: 10.28926/ilkomnika.v5i1.542.
- [13] D. L. Davies and D. W. Bouldin, "A Cluster Separation Measure," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-1, no. 2, pp. 224–227, 1979, doi: 10.1109/TPAMI.1979.4766909.
- [14] Z. F. Zuhana Realita Alfy; Ardhi Dinullah Baihaqie; A'ini, "Analisis Cluster K-Means Pada Indikator Indeks Pembangunan Teknologi, Informasi, Dan Komunikasi," *STRING (Satuan Tulisan Ris. dan Inov. Teknol.*, vol. 8, no. 2, pp. 183–188, 2023.
- [15] Y. Bachtiar, "Implementasi Algoritma Greedy Untuk Optimalisasi Jarak Tempuh Rute Ziarah Makam Walisongo," *Fakt. Exacta*, vol. 17, no. 3, p. 287, 2024, doi: 10.30998/faktorexacta.v17i3.23322.