

Penerapan Data Mining untuk memprediksi Nilai Akhir Matakuliah Menggunakan Metode K-Means Clustering

Application of Data Mining to predict Final Course Grades Using the K-Means Clustering Method

Alfrian C Talakua

Program Studi Teknik Informatika, Universitas Kristen Wira Wacana Sumba

E-mail: alfriantalakua@unkriswina.ac.id

KEYWORDS:

Data Mining, K-Means, Student grades

KATA KUNCI:

Data mining, K-means, Nilai Mahasiswa

ABSTRACT

Course grades are one indicator to measure a student's academic skills. Predicting students' final grades can help students and lecturers to know and measure students' learning progress before the grades are input into the academic section, so that if there are students who get less than optimal grades, preventative action can be taken early. K-means is used to solve this problem to make predictions in dividing student clusters based on course grades. The results of the research concluded that from the web development course with a total of 129 students, there were 2 clusters with a Davies Bouldin score = 1,046 in the division of 8 assessments and showed that there were only 17 students who had a good average score and there were 122 students with a good average not good.

ABSTRAK

Nilai matakuliah merupakan salah satu indikator untuk mengukur kecakapan akademik dari seorang mahasiswa. Prediksi nilai akhir mahasiswa dapat membantu mahasiswa dan dosen untuk mengetahui serta mengukur progres belajara dari mahasiswa sebelum nilai tersebut diinputkan pada bagian akademik, sehingga apabila terdapat mahasiswa yang mendapat nilai kurang maksimal dapat dilakukan tindakan pencegahan lebih awal. K-means digunakan untuk menyelesaikan permasalahan tersebut untuk melakukan prediksi dalam pembagian cluster mahasiswa berdasarkan nilai pada matakuliah. Hasil penelitian disimpulkan bahwa dari matakuliah pengembangan web dengan jumlah mahasiswa 129 terdapat 2 cluster dengan nilai Davies Bouldin = 1.046 dalam pembagian 8 penilaian dan menunjukkan bahwa hanya terdapat 17 mahasiswa yang memiliki nilai dengan rata-rata yang baik dan terdapat 122 mahasiswa dengan rata-rata yang kurang baik.

PENDAHULUAN

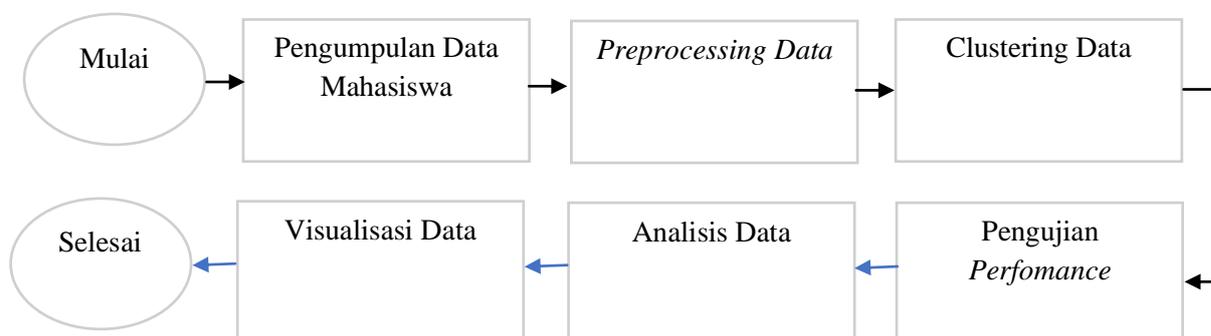
Kemajuan ilmu pengetahuan dan teknologi menuntut para lulusan dari Perguruan tinggi untuk dapat bekerja dengan maksimal pada bidang digital dan juga dapat dengan cepat beradaptasi dengan lingkungan kerja. Setiap lulusan dari perguruan tinggi harus siap bersaing dengan keterampilan bidang ilmu yang dipelajari selama berkuliah, untuk itu keefektifan metode yang digunakan oleh pengajar dalam melakukan mengevaluasi terhadap hasil belajar mahasiswa merupakan salah satu faktor yang menentukan tinggi atau rendahnya kualitas pendidikan. Nilai matakuliah merupakan salah satu indikator untuk mengukur kecakapan akademik dari seorang mahasiswa. Prediksi nilai akhir mahasiswa dapat membantu mahasiswa dan dosen untuk mengetahui serta mengukur progres belajara dari mahasiswa sebelum nilai tersebut diinputkan pada

bagian akademik, sehingga apabila terdapat mahasiswa yang mendapat nilai kurang maksimal dapat dilakukan tindakan pencegahan lebih awal. Hal ini perlu dilakukan juga untuk membantu memberi kualifikasi bagi pengajar selanjutnya jika matakuliah yang diambil merupakan matakuliah prasyarat sehingga pengajar selanjutnya dapat melakukan antisipasi lebih awal untuk sebelum masuk pada perkuliahan selanjutnya. Metodologi penilaian mahasiswa ditentukan oleh pengajar dengan menerapkan indikator dari peraturan akademik perguruan tinggi, aturan dari bagian akademik dapat menjadi acuan tentang bagaimana penilaian akan dilakukan oleh pengajar kepada mahasiswa disetiap matakuliah. Data hasil penilaian dari setiap matakuliah dapat digunakan pengajar untuk melakukan evaluasi keberhasilan pembelajaran yang telah. Hasil penilaian semester sebelumnya dan semester sekarang dapat digunakan untuk memprediksi nilai mahasiswa yang mengikuti perkuliahan semester depan. Dengan kata lain, dimungkinkan pengajar dapat melakukan memprediksi nilai akhir siswa pada semester ini untuk memasuki semester depan, matakuliah yang akan digunakan pada penelitian ini ialah matakuliah pengembangan web yang mana merupakan matakuliah prasyarat untuk matakuliah selanjutnya yaitu matakuliah pemrograman berbasis framework untuk semester depan. Metode data mining dapat digunakan untuk membantu membuat prediksi, menggunakan informasi dari masa lalu dan sekarang untuk membuat perkiraan secara sistematis tentang apa yang paling mungkin terjadi di masa depan, prediksi merupakan proses yang cukup rentan terhadap kesalahan. Namun dengan melakukan prediksi dapat mendapat nilai yang cukup dekat dengan hasil akhir dan proses ini juga akan membantu untuk diterapkan pada kasus yang sama dengan rentan waktu yang berbeda hingga mendapatkan pola yang tepat[1]. Pada Penelitian ini akan menggunakan data nilai mahasiswa dengan variabel perhitungan terhadap Nilai Kuis, Nilai UTS dan Nilai UAS untuk mengklasifikasi kemampuan mahasiswa dalam memahami matakuliah pengenalan web. Variabel prediksi tersebut akan digunakan untuk menghasilkan prediksi nilai akhir untuk data penilaian mahasiswa semester yang sedang berjalan. Pada penelitian ini menggunakan kajian pustaka dari beberapa penelitian sebelumnya. Penelitian yang dilakukan Maulana, dkk[2] menganalisis kinerja siswa dengan melakukan pemetaan pada faktor-faktor yang mempengaruhi kinerja siswa dengan melakukan analisis awal terlebih dahulu kemudian membuat model prediksi awal perlu dilakukan identifikasi variable-variabel yang mempengaruhi kinerja siswa. Sebaran pengetahuan dari guru ke siswa dapat diukur dengan melihat pengelompokan nilai UAS. Mencari tahu kluster mana yang benar-benar belajar mandiri atau tidak menjadi sangat penting mengingat situasi pandemi Covid yang sedang terjadi di Indonesia dan di seluruh dunia. Untuk memaksimalkan pembelajaran dan pemahaman siswa, K-Means dapat digunakan untuk memetakan kemampuan akademik siswa. Penelitian Sungkar dan Qurohman [3] dalam penelitiannya menggunakan algoritma C5.0 untuk memprediksi kelulusan pembelajaran mahasiswa. Kelulusan setiap mahasiswa dari suatu mata kuliah menunjukkan bahwa pengajarannya efektif dan mahasiswa mempertahankan materi dosen. Berdasarkan perilaku khas setiap siswa, dimungkinkan untuk memprediksi kapan mereka akan menyelesaikan kursus. Metode alternatif untuk menentukan pola kebiasaan berdasarkan data yang terkumpul adalah data mining. Penambangan data adalah proses penggalian informasi berharga dari kumpulan data sehingga bisnis, pemerintah, dan organisasi lain dapat menggunakannya untuk membuat keputusan. Mengklasifikasikan kumpulan data

memungkinkan untuk prediksi kelulusan menggunakan penambangan data. Algoritma C5.0 merupakan kemajuan dari algoritma C4.5 karena prosesnya. pada dasarnya sama, tetapi algoritma baru memiliki beberapa keunggulan dibandingkan algoritma lama. Keluaran dari algoritma C5.0 berupa pohon keputusan atau aturan yang dibuat berdasarkan nilai entropy atau gain. Prosedur prediksi dilakukan dengan memanfaatkan atribut Nilai Kehadiran, Nilai Tugas, Skor UTS, dan Skor UAS berdasarkan algoritma klasifikasi C5.0. Pohon keputusan dengan aturan adalah hasil dari proses klasifikasi algoritma C5.0. Algoritme C5.0 bekerja dengan baik, dengan tingkat akurasi tinggi sebesar 93,3%. Berdasarkan hasil dari pemaparan yang dilakukan oleh peneliti terdahulu, terlihat bahwa pentingnya nilai akademis bagi mahasiswa, pengajar dan perguruan tinggi sebagai sebuah tolak ukur performa akademis dalam meningkatkan kualitas akademis dari institusi, untuk mendapatkan nilai matakuliah yang baik cenderung mahasiswa dihadapkan dengan proses remedial atau pengulangan matakuliah yang cukup panjang dan dilakukan pada saat akhir semester, bahkan jika tidak lulus maka harus menunggu satu tahun berikutnya untuk mengulang matakuliah tersebut dan hambatan berikutnya tidak dapat mengambil matakuliah syarat pada semester depan, proses remedial ini juga memiliki beberapa kendala lain terbatasnya waktu, kondisi fisik dan mental mahasiswa. Tujuan utama dari penelitian ini adalah untuk mengembangkan model prediksi yang mampu memprediksi nilai akhir mahasiswa dengan tingkat akurasi yang tinggi. Prediksi yang akurat dapat memberikan informasi berharga kepada mahasiswa, dosen, dan perguruan tinggi dalam merencanakan langkah-langkah yang tepat untuk meningkatkan performa akademik mahasiswa dengan menggunakan metode k-means clustering untuk mengelompokkan mahasiswa yang perlu ditingkatkan performa akademis agar mendapatkan nilai matakuliah yang maksimal. K-Means ialah metode klasterisasi dengan basis jarak dengan melakukan pembagian data ke beberapa cluster serta algoritma ini bekerja di atribut numerik[4]. Algoritma k-means masuk partitioning clustering yang akan melakukan pemisahan data ke kluster pada daerah terpisah. Algoritma k-means terkenal dengan sistem yang mudah serta kemampuan mengkluster data besar serta data outlier secara cepat. Pada algoritma k-means, tiap data wajib masuk ke kluster tertentu serta dimungkinkan ke tiap data di kluster tertentu pada sebuah proses, difase setelahnya berpindah ke kluster lain[5]. Sehingga dengan menggunakan algoritma ini akan memudahkan dalam melakukan pengelompokan data terkait dengan masalah pada penelitian.

METODE PENELITIAN

Pada penelitian ini menggunakan studi pustaka diawal untuk mempelajari algoritma, pengumpulan data serta beberapa kajian dari penelitian sebelumnya, hasil dari kajian pustaka digunakan untuk menyusun tahapan pengolahan data yang dijelaskan sebagai berikut ;



Gambar 1. Tahapan Analisis Data

2.1 Pengumpulan Data Mahasiswa

Pada tahap ini dilakukan pengumpulan data mahasiswa yang mengambil matakuliah pengembangan web pada semester yang sedang berlangsung, data yang dikumpulkan berupa data nilai kuis, nilai Ujian Tengah semester dan ujian akhir semester, jumlah data dari empat kelas pengembangan web sebanyak 150 data yang akan ditampilkan pada tabel 1.

2.2 Preprocessing Data

Pada tahap ini dilakukan untuk mempersiapkan dataset yang diperlukan untuk melakukan clustering, proses ini dilakukan untuk mengambil data mahasiswa yang aktif hadir sehingga bagi mahasiswa tidak lagi masuk pada data set yang disiapkan pada data set akan dibagi dalam 8 atribut dengan pembagian 6 Nilai kuis 1 nilai UTS dan 1 nilai UAS.

Tabel 1. Dataset Mahasiswa

No.	Nama	Kuis 1	Kuis 2	Kuis 3	Kuis 4	Kuis 5	Kuis 6	UTS	UAS
1	ANAS***	80	70	60	40	40	30	35	35
2	YO***	80	50	50	60	40	50	40	35
3	IRNA I***	70	65	35	55	40	35	35	35
4	YUN***	65	45	60	35	40	35	45	30
5	AMBU ***	50	45	70	0	45	35	40	30
6	MAGDA****	55	35	0	35	45	55	30	35
7	FEBI****	0	65	30	35	40	50	35	30
8	AL**	35	35	50	40	45	30	30	40
9	ROS****	40	30	35	45	0	65	35	30
10	HILDE****	80	55	70	55	45	50	40	40
...
129	ZEF*****	70	65	35	40	40	35	40	55

2.3 Clustering data

Pada tahap ini akan dilakukan clustering data menggunakan algoritma K-Means dengan beberapa Langkah sebagai berikut ;

Langkah-langkah melakukan clustering dengan metode K-Means adalah sebagai berikut[6][7]:

- Pilih jumlah cluster k .
- Inisialisasi k pusat cluster ini bisa dilakukan dengan berbagai cara. Namun yang paling sering dilakukan adalah dengan cara random. Pusat-pusat cluster diberi nilai awal dengan angka-angka random.
- Alokasikan semua data/ objek ke cluster terdekat.

$$D(i, j) = \sqrt{(C_{1i} - C_{1j})^2 + (C_{2i} - C_{2j})^2 + \dots + (C_{ki} - C_{kj})^2}$$

Keterangan:

Dimana:

$D(i, j)$ = jarak data ke i ke pusat cluster j

C_{ki} = Data ke i pada atribut data ke k

C_{kj} = Data ke j pada atribut data ke k

- Hitung kembali pusat cluster dengan keanggotaan cluster yang sekarang. Pusat cluster adalah rata-rata dari semua data/ objek dalam cluster tertentu. Jika dikehendaki bisa juga menggunakan median dari cluster tersebut. Jadi rata-rata (mean) bukan satu-satunya ukuran yang bisa dipakai.
- Tugaskan lagi setiap objek memakai pusat cluster yang baru. Jika pusat cluster tidak berubah lagi maka proses clustering selesai. Atau, kembali ke langkah nomor 3 sampai pusat cluster tidak berubah lagi.

Berdasarkan tahapan diatas maka akan ditentukan nilai K yang digunakan untuk clustering ialah 2, 3, 4 dan 5 sehingga dalam beberapa percobaan pengujian akan membagi cluster dengan jumlah maksimal yaitu 5 cluster kemudian akan dilihat mana cluster yang baik berdasarkan nilai Davis Boilden yang mendekati 0 itu akan digunakan sebagai cluster yang memiliki jarak yang baik. Proses Pengolahan data akan menggunakan software rapidminer sehingga dapat mempermudah untuk proses pengolahan data dan visualisasi data.

2.4 Analisis Data

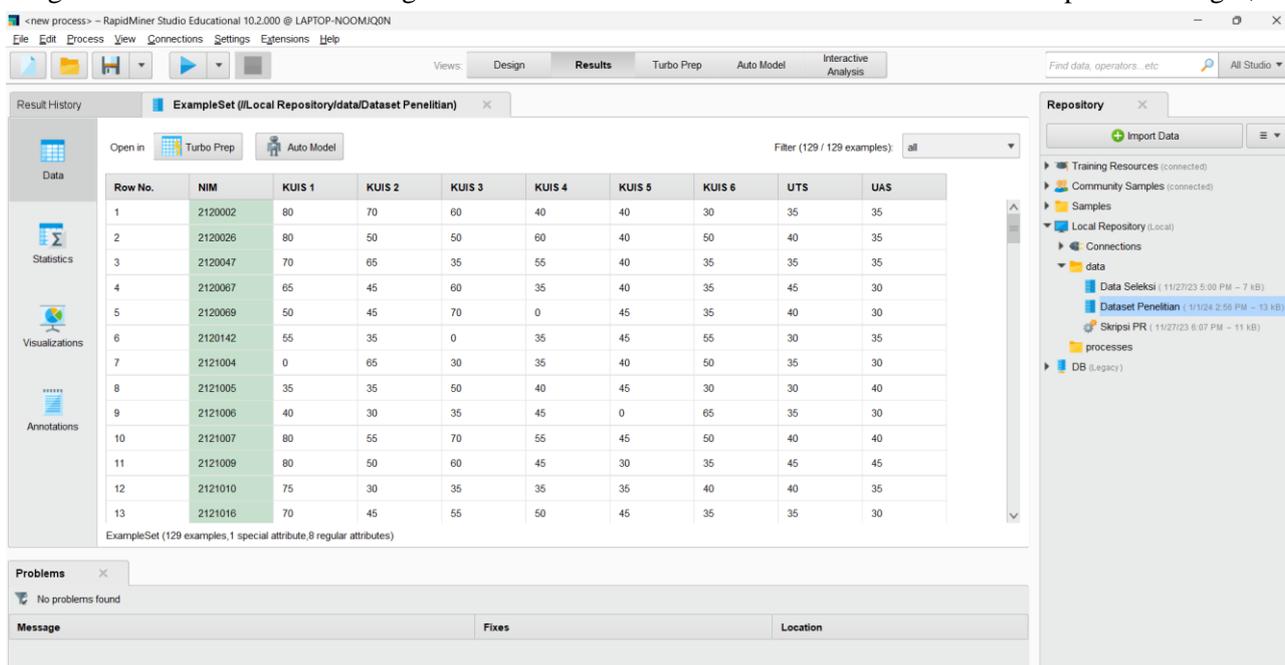
Pada tahap ini dilakukan proses analisis data untuk melihat jumlah data pada setiap cluster dan diurutkan berdasarkan ranking yang terbaik sehingga bisa terlihat urutan ranking dari setiap cluster yang dibuat. Proses ini juga akan membantu memberi rekomendasi bagi pengajar untuk melihat potensi akademik secara minimal dan maksimal dari setiap mahasiswa yang mengambil matakuliah pengembangan web.

2.5 Visualisasi Data

Pada tahap ini dilakukan visualisasi data untuk menampilkan data dalam bentuk tabel atau diagram sehingga dapat terlihat jumlah data yang tercluster dan jarak dari setiap data mahasiswa pada masing-masing cluster yang terbentuk.

HASIL DAN PEMBAHASAN

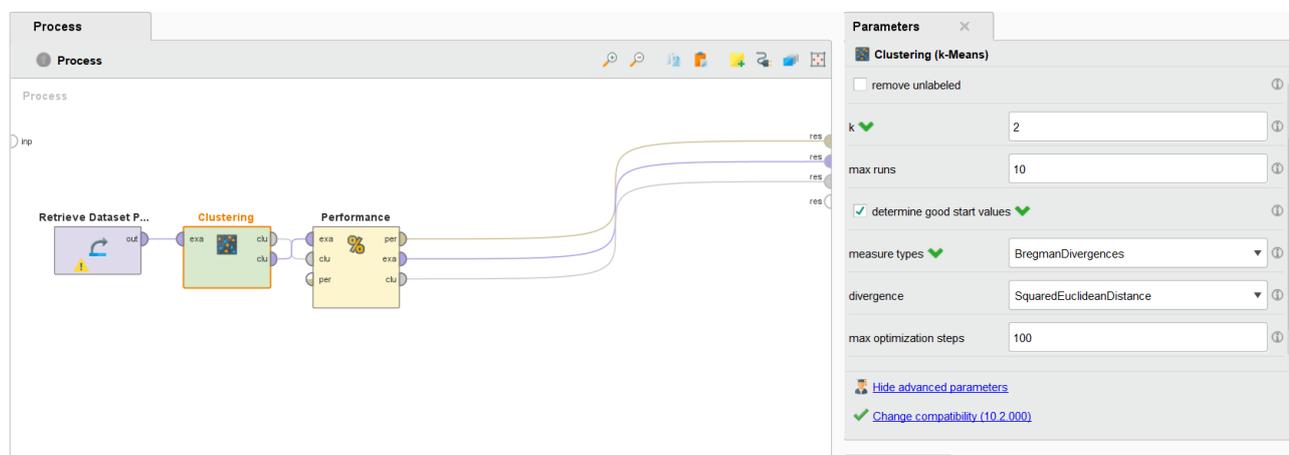
Proses pengolahan data dimulai dengan menginputkan dataset pada software Rapidminer, dataset diatur dengan mencantumkan NIM sebagai Label dan Nilai kuis serta UTS dan UAS memiliki tipe data interger;



Row No.	NIM	KUIS 1	KUIS 2	KUIS 3	KUIS 4	KUIS 5	KUIS 6	UTS	UAS
1	2120002	80	70	60	40	40	30	35	35
2	2120026	80	50	50	60	40	50	40	35
3	2120047	70	65	35	55	40	35	35	35
4	2120067	65	45	60	35	40	35	45	30
5	2120069	50	45	70	0	45	35	40	30
6	2120142	55	35	0	35	45	55	30	35
7	2121004	0	65	30	35	40	50	35	30
8	2121005	35	35	50	40	45	30	30	40
9	2121006	40	30	35	45	0	65	35	30
10	2121007	80	55	70	55	45	50	40	40
11	2121009	80	50	60	45	30	35	45	45
12	2121010	75	30	35	35	35	40	40	35
13	2121016	70	45	55	50	45	35	35	30

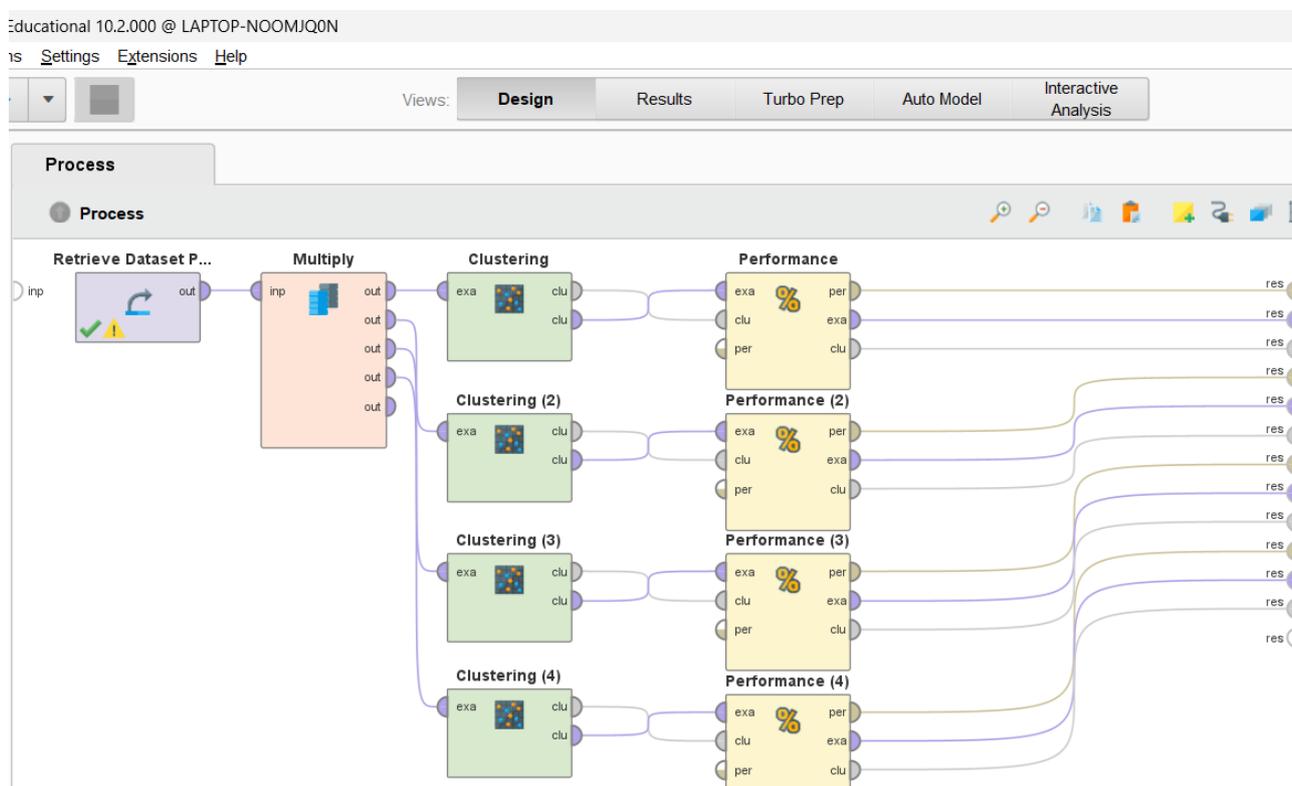
Gambar 2. Dataset Mahasiswa

Setelah proses input dataset proses selanjutnya akan memilih modul dengan menentukan nilai K awal ialah 2 dengan measure types menggunakan Bregman divergences dikarenakan data yang diolah semuanya merupakan data numeric, kemudian menggunakan maximum perpindahan titik centroid maximum 100 kali sebelum dilakukan cluster harus menambahkan pengukuran performance untuk melihat nilai pembagian cluster yang akan terbagi menjadi 2 Davies Bouldin = 1.046.



Gambar 3. Pengaturan Cluster

Pada proses pertama menggunakan Nilai K = 2 mendapat jumlah Cluster 0 = 112 Data dan Cluster 1 = 17 data dengan nilai Davies Bouldin = 1.046, nilai tersebut memiliki jarak yang cukup dekat dari setiap item namun untuk melihat apakah data analisis dengan nilai k = 2 itu cukup baik maka kita harus menguji lagi dengan beberapa nilai K dalam proses berikutnya kita akan mencoba dengan nilai k = 3, k = 4 dan k = 5 namun kita memerlukan 1 modul tambahan yaitu modul multiply data.



Gambar 4. Multiply data

Setelah melakukan cluster dengan nilai k yang maka ditemukan bahwa Cluster dengan Nilai K = 3 terbagi menjadi Cluster 0: 15 items , Cluster 1: 100 items ,Cluster 2: 14 items Total number of items: 129 dengan nilai **Davies Bouldin: 1.318**, sedangkan untuk Nilai K = 4 terbagi menjadi Cluster 0: 19 items, Cluster 1: 45 items, Cluster 2: 14 items,Cluster 3: 51 items Total number of items: 129 dengan nilai **Davies Bouldin: 1.452**, kemudian nilai K = 5 terbagi menjadi Cluster 0: 55 items Cluster 1: 14 items Cluster 2: 15 items Cluster 3: 44 items Cluster 4: 1 items Total number of items: 129 dengan nilai **Davies Bouldin: 1.450**, Sehingga dapat disimpulkan pada kasus ini data lebih baik dibagi menjadi 2 cluster karena memiliki jarak yang cukup dekat dengan nilai **Davies Bouldin = 1.046**. **Pembagian data Cluster dapat dilihat pada tabel berikut :**

Tabel 2. Data Cluster 0

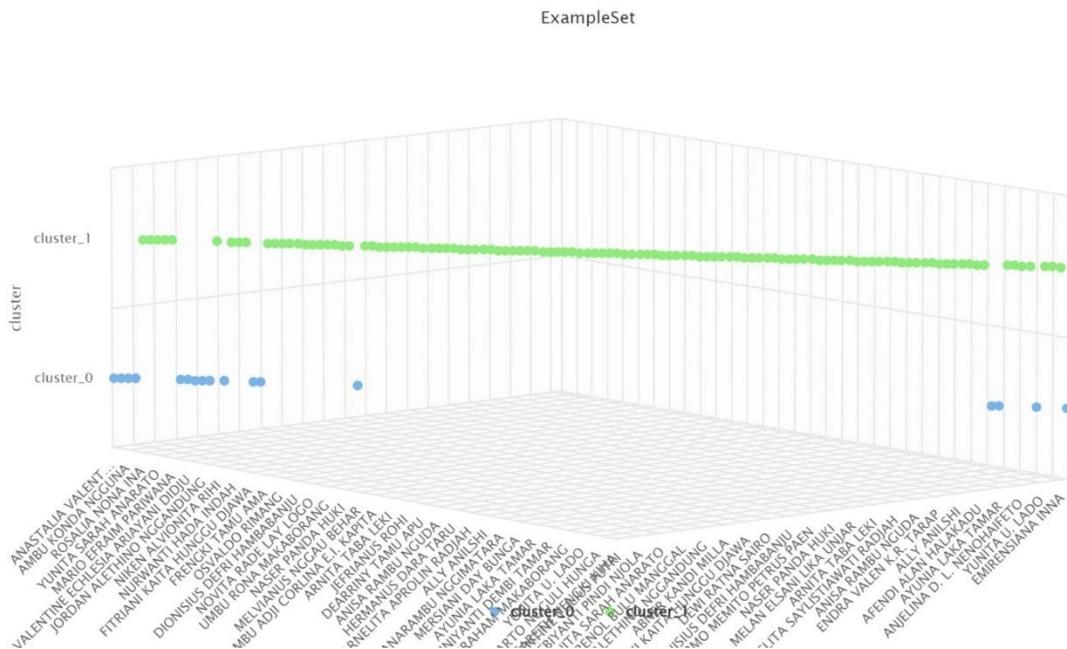
No.	Nama	Cluster	Kuis1	Kuis2	Kuis3	Kuis4	Kuis5	Kuis6	UTS	UAS
1	ANAS*****	cluster 0	80	70	60	40	40	30	35	35
2	YO***	cluster 0	80	50	50	60	40	50	40	35
3	IR**	cluster 0	70	65	35	55	40	35	35	35
4	YUN***	cluster 0	65	45	60	35	40	35	45	30
5	HILDE*****	cluster 0	80	55	70	55	45	50	40	40
6	JOISMA***	cluster 0	80	50	60	45	30	35	45	45
7	DEWI*****	cluster 0	75	30	35	35	35	40	40	35

8	YUN***	cluster_0	70	45	55	50	45	35	35	30
9	END**	cluster_0	65	30	40	50	45	55	30	30
10	YAN*****	cluster_0	75	55	50	45	35	0	40	40
11	PI**	cluster_0	80	50	50	50	60	60	70	50
12	VAL****	cluster_0	60	50	30	40	35	30	40	50
13	MIC***A	cluster_0	60	45	35	65	55	45	35	55
14	DU***N	cluster_0	70	45	60	45	35	40	55	60
15	FERD***	cluster_0	75	50	65	55	50	55	50	55
16	ARTO NG*	cluster_0	75	60	40	50	45	35	60	50
17	ZEFA****	cluster_0	70	65	35	40	40	35	40	55

Tabel 3. Data Cluster 1

No.	Nama	Cluster	Kuis1	Kuis2	Kuis3	Kuis4	Kuis5	Kuis6	UTS	UAS
1	AMB*	cluster_1	50	45	70	0	45	35	40	30
2	MAGD****	cluster_1	55	35	0	35	45	55	30	35
3	FEBI****	cluster_1	0	65	30	35	40	50	35	30
4	ALD*	cluster_1	35	35	50	40	45	30	30	40
5	ROSA***	cluster_1	40	30	35	45	0	65	35	30
6	ADI***	cluster_1	45	35	75	35	30	30	30	30
7	MAR**	cluster_1	60	35	40	40	40	35	35	35
8	IR**	cluster_1	60	40	35	40	40	35	25	35
9	REN**	cluster_1	50	35	35	35	25	40	30	35
10	SIN**	cluster_1	60	30	30	45	30	50	40	35
11	JUW***	cluster_1	30	50	50	40	20	40	30	30
12	MEYFR***	cluster_1	50	30	25	35	30	45	30	35
13	JOR***	cluster_1	45	35	30	40	35	40	45	50
14	ER**	cluster_1	40	35	30	30	35	40	45	35
15	MER***	cluster_1	45	30	35	45	30	35	40	50
...
112	ME***	cluster_1	65	35	25	25	35	35	35	35

Dari pembagian data cluster data dapat divisualisasikan agat dapat terlihat jelas jarak antar items dan pembagian cluster;



Gambar 5. Visualisasi data Cluster

KESIMPULAN

Berdasarkan Hasil penelitian perhitungan nilai akhir mahasiswa menggunakan K-means dapat disimpulkan bahwa dari matakuliah pengembangan web terdapat 2 cluster dengan nilai Davies Bouldin = 1.046 dalam pembagian 8 penilaian dan menunjukkan bahwa hanya terdapat 17 mahasiswa yang memiliki nilai dengan rata-rata yang baik dan terdapat 122 mahasiswa dengan rata-rata yang kurang baik hal ini dapat menjadi kendala kedepan untuk matakuliah selanjutnya sehingga diperlukan strategi yang harus dilakukan untuk menurunkan nilai pada cluster 1 agar tidak menjadi kendala pada proses perkuliahan di matakuliah selanjutnya.

DAFTAR PUSTAKA

- [1] 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.
- [2] I. Maulana and U. Rosalina, "Clustering Data Nilai Ujian Akhir Semester Menggunakan Algoritma Data Mining K-Means," *PERISKOP J. Sains dan Ilmu Pendidik.*, vol. 1, no. 2, pp. 76–85, 2020, doi: 10.58660/periskop.v1i2.10.
- [3] M. S. Sungkar and M. T. Qurohman, "Penerapan Algoritma C5.0 Untuk Prediksi Kelulusan Pembelajaran Mahasiswa Pada Matakuliah Arsitektur Sistem Komputer," *J. Media Inform. Budidarma*, vol. 5, no. 3, p. 1166, 2021, doi: 10.30865/mib.v5i3.3116.
- [4] N. Afiasari, N. Suarna, and N. Rahaningsi, "Implementasi Data Mining Transaksi Penjualan Menggunakan Algoritma Clustering dengan Metode K-Means," *J. SAINTEKOM*, vol. 13, no. 1, pp. 100–110, 2023, doi: 10.33020/saintekom.v13i1.402.
- [5] G. Gustientiedina, M. H. Adiya, and Y. Desnelita, "Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan," *J. Nas. Teknol. dan Sist. Inf.*, vol. 5, no. 1, pp. 17–24, 2019, doi: 10.25077/teknosi.v5i1.2019.17-24.
- [6] Y. Santoso, "Penerapan Data Mining Untuk Nilai Akhir Mata Kuliah Menggunakan Algoritma K-Means Clustering," *SNIA (Seminar Nas. Inform. dan ...)*, pp. 27–30, 2021, [Online]. Available: <https://snia.unjani.ac.id/web/index.php/snias/article/view/285%0Ahttps://snia.unjani.ac.id/web/index.php/snias/article/download/285/116>
- [7] F. Handayani, "Aplikasi Aplikasi Data Mining Menggunakan Algoritma K-Means Clustering untuk Mengelompokkan Mahasiswa Berdasarkan Gaya Belajar," *J. Teknol. dan Inf.*, vol. 12, no. 1, pp. 46–63, 2022, doi: 10.34010/jati.v12i1.6733.