
PENERAPAN *K-MEANS CLUSTERING* TINGKAT KETEPATAN WAKTU KELULUSAN MAHASISWA PROGRAM STUDI TEKNIK INFORMATIKA STMIK NURDIN HAMZAH JAMBI

Arisa Rahmini Azhara

Fakultas Ilmu Komputer, Program Studi Teknik Informatika

Universitas Nurdin Hamzah Jambi

Email : arisarahminiazhara9@gmail.com

Rangga Anantadira Nugraha

Fakultas Ekonomi dan Bisnis, Jurusan Sistem Informasi

Universitas Islam Negeri Syarif Hidayatullah Jakarta

Email : rangga.antdr@gmail.com

Abdul Piqri

Fakultas Ekonomi dan Bisnis, Jurusan Sistem Informasi

Universitas Islam Negeri Syarif Hidayatullah Jakarta

Email: abdul.piqri13@gmail.com

ABSTRAK

Pengukuran *K-Means Clustering* pada tingkat ketepatan waktu kelulusan mahasiswa Teknik Informatika merupakan penerapan teknik data mining untuk mengetahui ketepatan waktu kelulusan mahasiswa. Penelitian ini dilakukan dengan menggunakan alat bantu *RapidMiner* untuk mengelompokkan data siswa berdasarkan atribut indeks prestasi (IP) pada semester pertama sampai semester 6 dan lama studi. Tujuan penelitian ini penting bagi suatu perguruan tinggi karena merupakan salah satu syarat akreditasi. Metode pengelompokan yang digunakan adalah *K-Means Clustering* dengan 3 cluster yaitu *on time*, *not right* dan *stop out*. Untuk menampung data baru tersebut, dibangun aplikasi tambahan berbasis web dengan bahasa pemrograman PHP dan database PhpMyAdmin. Aplikasi tersebut digunakan untuk menampung data mahasiswa baru yang lulus. Keluaran dari aplikasi ini adalah laporan keseluruhan *cluster*, laporan siswa dan laporan perangkat. Diharapkan dengan penerapan *K-Means Clustering* pengukuran tingkat ketepatan waktu kelulusan mahasiswa Teknik Informatika dapat membantu memudahkan Program Studi Teknik Informatika dalam menentukan mahasiswa lulusan mana yang tepat dan mengetahui faktor penyebab mahasiswa tersebut mengalami keterlambatan kelulusan.

Kata Kunci : *Clustering, K-Means, PhpMyAdmin, RapidMiner, Web.*

ABSTRACT

K-Means Clustering measurement of the level of timeliness of student graduation from Informatics Engineering is an application for data mining techniques to determine the accuracy of student graduation time. This research was conducted by using RapidMiner assistive tools to cluster student data based on attributes of the achievement index (IP) in the first semester through the 6th semester and the duration of study. The purpose of this study is important for a university because it is one of the requirements for accreditation. The grouping method used is K-Means Clustering with 3 clusters, namely on time, not right and stop out. To accommodate the new data, an additional web-based application is built with the PHP programming language and the PhpMyAdmin database. The application is used to hold data on new students who graduate. The output of this application is an entire cluster report, student reports and device reports. It is expected that with the application of K-Means Clustering the measurement of the level of

timeliness of graduation of Informatics Engineering students can help facilitate the Study Program of Informatics Engineering in determining which students are the right graduates and know the factors causing the student to experience a delay in graduation.

Keywords : *Clustering, K-Means, PhpMyAdmin, RapidMiner, Web.*

1. PENDAHULUAN

Data *Clustering* merupakan salah satu metode *Data Mining* yang bersifat tanpa arahan. *Data mining* adalah sebagai proses ekstraksi informasi dari kumpulan data melalui penggunaan algoritma dan teknik yang melibatkan bidang ilmu statistik, mesin pembelajaran, dan sistem manajemen *database*. Dalam bagian *data mining* digunakan untuk ekstraksi informasi penting yang tersembunyi dari dataset yang besar. Dengan adanya *data mining* maka akan didapatkan suatu permata berupa pengetahuan didalam kumpulan data-data yang banyak jumlahnya.

Ketepatan waktu kelulusan mahasiswa merupakan hal yang penting bagi sebuah Perguruan Tinggi karena merupakan salah satu syarat akreditasi, tidak terkecuali Sekolah Tinggi Manajemen Informatika dan Komputer (STMIK) Nurdin Hamzah. STMIK NH adalah Perguruan tinggi yang menyelenggarakan pendidikan dalam ilmu bidang komputer yang terdapat di kota Jambi dan menawarkan berbagai program studi salah satunya yaitu Teknik Informatika. Program studi tersebut menerima kurang lebih 250 mahasiswa setiap tahunnya, namun jumlah mahasiswa tersebut tidak seimbang dengan jumlah mahasiswa yang lulus tepat pada waktunya. Dalam menjalankan proses perkuliahan, mahasiswa yang satu dengan mahasiswa lainnya memiliki perbedaan keadaan dalam proses perkuliahan seperti ada yang sambil kerja atau hanya berkuliah, dan adapun mahasiswa yang nilainya bermasalah dan ada yang tidak, yang menyebabkan tiap mahasiswa memiliki waktu kelulusan yang berbeda dan masih banyak lagi faktor yang mempengaruhi ketepatan waktu kelulusan. Ketepatan waktu kelulusan juga merupakan kepentingan setiap mahasiswa karena setiap mahasiswa mempunyai target atau tujuan yang ingin dicapai, yaitu berupa pekerjaan yang layak dan membanggakan orang tua.

Untuk dapat mengukur ketepatan waktu kelulusan mahasiswa maka pengukuran itu juga perlu di analisis lebih jauh karena merupakan faktor keberhasilan prodi atau pihak akademik dalam meningkatkan kualitas kelulusan mahasiswa tersebut. Salah satu alat bantu yang bisa membantu mengukur kelulusan tersebut adalah *RapidMiner*, aplikasi *RapidMiner* menyediakan berbagai metode dalam data mining yang bisa mempermudah dalam mengolah data yang cukup besar. Adapun metode yang dapat digunakan adalah *K-Means clustering*, dimana *K-Means* merupakan suatu metode penganalisaan data atau metode *Data Mining* yang melakukan proses dalam bentuk pemodelan. Kelebihan dari metode *K-Means* ini yaitu mudah untuk diimplementasikan dan dijalankan, waktu yang dibutuhkan untuk menjalankan pembelajaran ini relatif cepat, mudah untuk diadaptasi, dan umum digunakan. *Cross-Industry Standard Process for Data Mining (CRISP-DM)* yang dikembangkan tahun 1996 oleh analis dari beberapa industri seperti DaimlerChrysler, SPSS dan NCR, *CRISP DM* menyediakan standar proses data mining sebagai strategi pemecahan masalah secara umum dari bisnis atau unit penelitian. [1]

2. METODOLOGI PENELITIAN

2.1. Data mining

Data mining [2] adalah suatu proses ekstraksi atau penggalian data yang belum diketahui sebelumnya, namun dapat dipahami dan berguna dari database yang besar serta digunakan untuk membuat suatu keputusan bisnis yang sangat penting. Definisi lainnya tentang Data mining adalah proses menemukan pola yang menarik, dan pengetahuan dari data yang berjumlah besar. [3]

Pada prosesnya data mining akan mengekstrak informasi yang berharga dengan cara menganalisis adanya pola-pola ataupun hubungan keterkaitan tertentu dari data-data yang berukuran besar. Data mining berkaitan dengan bidang ilmu-ilmu lain, seperti Database System,

Data Warehousing, Statistic, Machine Learning, Information Retrieval, dan Komputasi Tingkat Tinggi. Selain itu data mining didukung oleh ilmu lain seperti Neural Network, Pengenalan Pola, Spatial Data Analysis, Image Database, Signal Processing.

2.2. Clustering

Clustering merupakan sebuah teknik pengelompokan sejumlah data obyek ke dalam *cluster* (group) sehingga dalam *cluster* akan berisi data yang semirip mungkin. [4]

Clustering adalah pengelompokan satu set objek data ke dalam beberapa kelompok atau *cluster* sehingga obyek dalam sebuah *cluster* memiliki jumlah kemiripan yang tinggi, tetapi sangat berbeda dengan objek di *cluster* lain. Ketidak miripan dan kesamaan dinilai berdasarkan nilai atribut yang menggambarkan objek dan sering melihat perilaku jarak *clustering* sebagai alat data mining dapat diterapkan pada berbagai bidang, seperti biologi, keamanan, intelijen bisnis dan pencarian web [3].

2.3. K-Means

K-Means merupakan salah satu metode pengelompokan data nonhirarki (sekatan) yang berusaha mempartisi data yang ada kedalam bentuk dua atau lebih kelompok . [5]

K-Means merupakan metode analisis kelompok yang mengarah pada pemartisan N objek pengamatan ke dalam K kelompok (*cluster*) di mana setiap objek pengamatan dimiliki oleh sebuah kelompok dengan mean (rata-rata) terdekat, mirip dengan algoritma *Expectation-Maximization* untuk *Gaussian Mixture* dimana keduanya mencoba untuk menemukan pusat dari kelompok dalam data sebanyak iterasi perbaikan yang dilakukan oleh kedua algoritma. [5]

3. HASIL DAN PEMBAHASAN

3.1 Proses Penyelesaian K-Means Clustering

Mengelompokkan data dengan algoritma K-Means dilakukan dengan cara menentukan jumlah *cluster*, hitung jarak terdekat dengan pusat *cluster*. Dalam penelitian ini peneliti mendapatkan 952 data dari tahun kelulusan 2010 sampai 2014 yang telah di normalisasi berdasarkan perangkatan mahasiswa ketepatan waktu kelulusannya menggunakan metode K-Means :

1. Dari himpunan data yang akan diklasterisasi, dipilih sejumlah k objek secara acak sebagai centroid awal.
2. Setiap objek yang bukan centroid dimasukkan ke kluster terdekat berdasarkan ukuran jarak tertentu.
3. Setiap centroid diperbaharui berdasarkan rata – rata dari objek yang ada di dalam setiap kluster.
4. Langkah kedua dan ketiga tersebut diulang – ulang (diiterasi) sampai semua centroid stabil atau konvergen, dalam arti semua centroid yang dihasilkan pada itersi sebelumnya.

Tabel 1. Tabel Mahasiswa dan nilai semester

No	NAMA MAHASISWA	S1	S2	S3	S4	S5	S6	lama studi
1	SUPRIYANTO	3,74	3,91	3,36	3,67	3,71	3,73	4
2	SEPTRI HARIANTO	1,48	0	0	0	0	0	0
3	GHUFRAN AAN	1,83	0	0	0	0	0	0
4	ANDRI YANTO	3,04	2,7	2,5	2,95	2,81	2,72	4
5	NOVENDRA PANJAITAN	2,26	2,93	2,05	2,57	2,51	2,39	0
6	SETIAWAN JODI	1,35	1,43	2,36	2,89	2,36	2,33	5
7	PRIBADI SETIAWAN	3,91	3,91	3,55	3,57	3,73	3,75	4
8	KURNIAWAN	3,17	3,17	2,91	3,38	3,14	3,16	4
9	AGUS DEDY SAPUTRA	3,22	3,3	3,5	3,62	3,43	3,43	4
10	ALHUSNI	2,65	2,57	2,23	2,48	2,6	2,61	0
11	ABDUL ROKHIM	3,26	3,17	2,68	2,86	3,02	2,99	4
12	BAYU SUHENDRA	3,52	3,13	3,27	2,52	3,03	3,07	4
13	BAYU PUTRA ANGGARA	0,52	0	0	0	0	0	0
14	RAHMAT HIDAYAT	2,3	0	0	0	0	0	0
15	NURASIA	3,78	3,61	3,64	3,52	3,65	3,64	4
16	ARYA PUTRA	3,52	3,39	3,82	3,62	3,58	3,56	4
17	HAMIDUN	3,78	3	3,45	3,62	3,47	3,46	4
18	CHYNDYA LOVINA	3,26	3,52	3,09	3,14	1,92	0	5
19	SULTAN HASANUDDIN	2,57	0,87	0,82	0	1,13	2,24	6
20	HERI SAPUTRA	3,57	3,09	3,14	3,29	3,3	3,28	4
21	M ABIMAYU MUHAIMIN	3,39	3	3,14	3,05	3,11	3,09	4
22	NUZUL SEPTIYADI	2,43	1,65	2,64	2,67	2,77	2,35	4
.....
950	KURNIAWAN RATANTO	0	0	0	0	0	0	0

Setelah data telah didapatkan dan akan di bagi menjadi 3 kelompok yaitu tepat waktu tidak tepat dan stop out dengan dipersiapkan *initial cluster center* terlebih dahulu dengan random seperti pada tabel 2 :

Tabel 2. Tabel *initial cluster center*

	Centroid Awal							tahun lulus
	Cluster	s1	s2	s3	s4	s5	s6	
tepat waktu	C1	3,06	2,94	2,91	3,04	3	3,01	3,85
tidak tepat	C2	2,44	2,26	2,09	1,89	2,07	2,01	5,9
Stop	C3	1,64	0,76	0,4	0,29	0,44	0,3	0

Setelah *initial cluster center* telah didapat, selanjutnya dilakukan perhitungan jarak antara nilai dengan rumus : $\sqrt{((NU-CU)^2 + (NJL-CJL)^2 + (NJP-CJP)^2 + (NLM-CLM)^2)}$ yang memiliki arti :

- NU = nilai normalisasi umur

- CU = nilai *initial cluster center* umur
- NJL = nilai normalisasi jumlah laki - laki
- CJL = nilai *initial cluster center* jumlah laki - laki
- NJP = nilai normalisasi jumlah perempuan
- CJP = nilai *initial cluster center* jumlah perempuan
- NLM = nilai normalisasi lama memakai
- CLM = nilai *initial cluster center* lama memakai

Setelah mendapatkan *cluster center* yang baru akan dilakukannya perulangan iterasi hingga mendapatkan *cluster center* yang baru bernilai sama dengan *cluster center*, hal tersebut dapat dilihat pada tabel 4 dan tabel 5.

Tabel 4. Tabel *cluster center* initial

		Centroid Awal							
	Cluster	s1	s2	s3	s4	s5	s6	tahun lulus	
tepat waktu	C1	3,06	2,94	2,91	3,04	3	3,01	3,85	
tidak tepat	C2	2,44	2,26	2,09	1,89	2,07	2,01	5,9	
Stop	C3	1,64	0,76	0,4	0,29	0,44	0,3	0	

Tabel 5. *cluster center* iterasi2

		Centroid Baru							
	Cluster	s1	s2	s3	s4	s5	s6	tahun lulus	
tepat waktu	C1	3,06	2,94	2,91	3,04	3	3,01	3,85	
tidak tepat	C2	2,44	2,26	2,09	1,89	2,07	2,01	5,9	
Stop	C3	1,64	0,76	0,4	0,29	0,44	0,3	0	

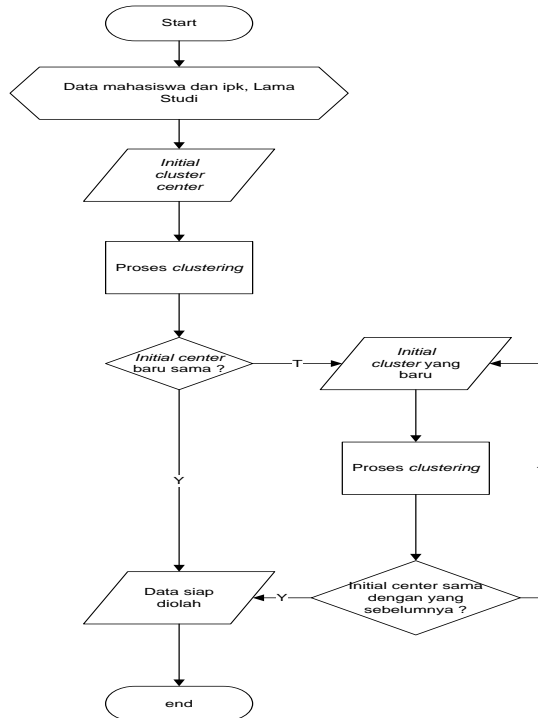
Setelah dilihat *cluster center* iterasi 2 memiliki nilai yang sama dengan *cluster center* 1. Dengan begitu proses pencarian dihentikan.

3. 2. Perancangan Sistem

3.2.1. Flowchart

Bagan Alir (*flowchart*) adalah bagan (*chart*) yang menunjukkan alir (*flow*) di dalam program atau prosedur sistem secara logika. Bagan alir digunakan terutama untuk alat Bantu komunikasi dan untuk dokumentasi.

Pada flowchart proses ini akan dijelaskan bagaimana proses dari data nilai yang telah dinormalisasikan akan diproses hingga mendapatkan *cluster* yang terbagi 3, yaitu *cluster* 1 yang termasuk dalam golongan tepat waktu, *cluster* 2 yang termasuk dalam golongan tidak tepat waktu dan *cluster* 3 yaitu termasuk kedalam stop out. Sebelum mendapatkan *cluster – cluster* tersebut diperlukannya sebuah *initial cluster center* yaitu nilai *initial cluster* centre pada iterasi yang pertama (perhitungan pertama kali) diberikan secara acak. Pada iterasi selanjutnya, nilai *initial cluster* centre (pengulangan ke-1 sampai dengan posisi normal/maksimal iterasi) diberikan dengan menghitung nilai rata-rata data pada setiap *clusternya*. Jika nilai *initial cluster* centre yang baru sama dengan nilai *initial cluster* centre yang baru maka proses iterasi dilanjutkan hingga nilai sama atau sampai dengan nilai maksimal iterasi yang telah ditetapkan sebelumnya (misalnya 950). Namun jika nilai *initial cluster* centre yang baru sama dengan *initial cluster* centre yang lama, proses pengelompokkan berhenti. Untuk lebih jelasnya flowchart proses ini dapat dilihat gambar 1.



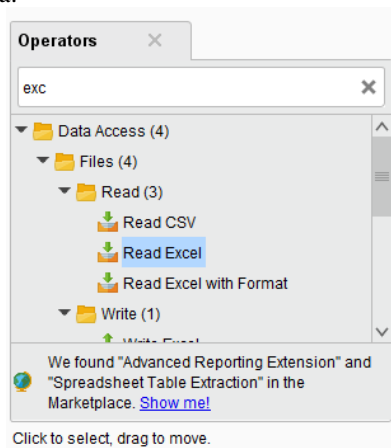
Gambar 1. Flowchart proses clustering

3.2.2. Antarmuka

Tahapan implementasi merupakan tahapan dimana aplikasi yang telah selesai dianalisa, dirancang, dan dibangun, akan diuji dan dioperasikan sesuai dengan fungsi dan tujuannya dari sesuai dengan yang diinginkan. Berikut ini adalah penjelasan mengenai implementasi klusterisasi k-means jenis pelanggan Batanghari Vision menggunakan *RapidMiner*.

a. Proses menghubungkan data

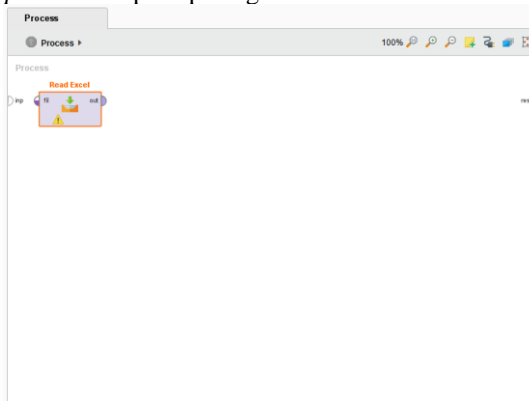
Pada proses ini dilakukan proses penghubung data dari Microsoft Excel dengan aplikasi *RapidMiner*. Berikut prosesnya.



Gambar 2. Operator perintah membaca excel

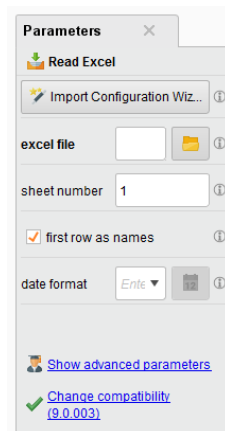
Pada gambar 2. terdapat folder *data access* yang memiliki folder *files* dan folder *write* yang memiliki perintah menambah data didalam masing – masing folder tersebut, dalam proses

kali ini peneliti menarik operator *read excel* yang berfungsi untuk membaca data *excel* untuk dimasukkan kedalam *RapidMiner* seperti pada gambar 3.



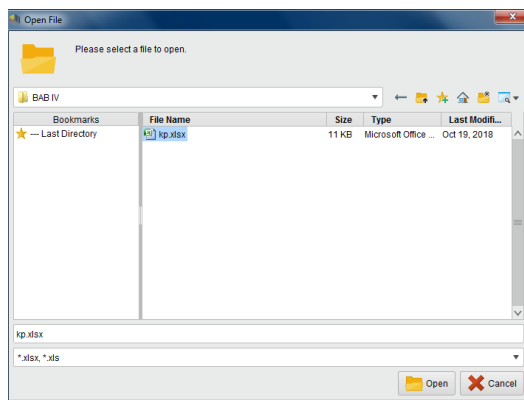
Gambar 3. Process operator yang digunakan

Gambar 3. diatas menunjukkan operator *read excel* berhasil dimasukkan kedalam proses.



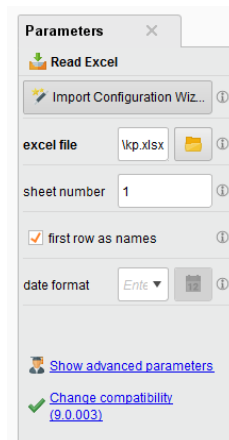
Gambar 4. Parameter untuk mengambil data excel

Setelah operator *read excel* terdapat dikolom proses untuk memasukkan data *excel* terlebih dahulu kita klik ikon folder yang terdapat diparameter. Hal ini dapat dilihat pada Gambar 4.



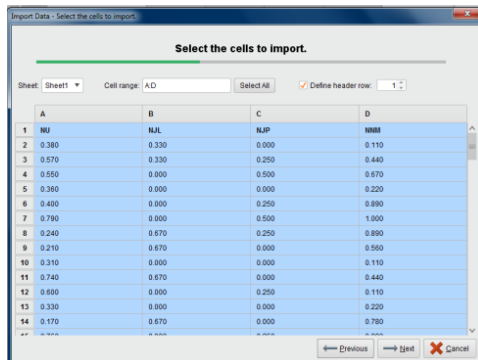
Gambar 5. Open file excel

Setelah mengklik ikon folder di parameter akan muncul jendela *open file* (Gambar 5) dimana peneliti dapat mencari datanya didalam foldernya dan memilih data yang akan dipakai.



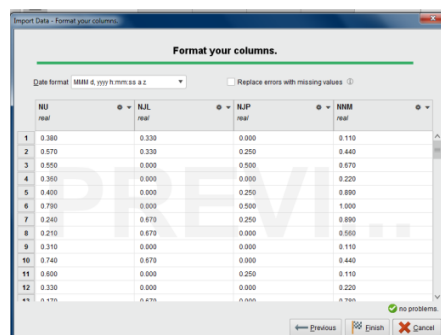
Gambar 6. Import configuration

Setelah data diambil peneliti mengklik *import configuration* yang terdapat di parameter untuk melakukan konfigurasi data yang dipakai untuk mengatur format dan pemangaturan dari excel diubah kedalam format dan pengaturan *RapidMiner*.(Gambar 6)



Gambar 7. Select cell yang akan di import

Pada gambar 7 menunjukkan kolom yang terdapat pada data excel peneliti beserta isinya untuk digunakan dalam proses data mining yang dilakukan didalam *RapidMiner*.



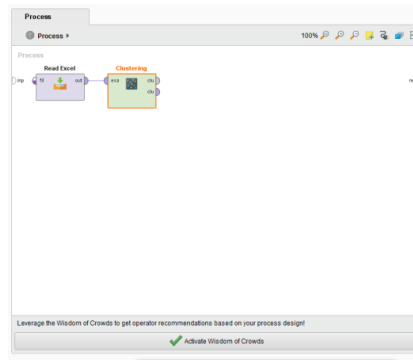
Gambar 8. Format kolom data

Setelah memilih data mana saja yang akan digunakan seperti gambar 8, kali ini dilakukan konfigurasi untuk menentukan format data dalam kolom yang akan digunakan.

Setelah mengikuti langkah dari gambar diatas data excel yang dimiliki sudah terhubung dengan *RapidMiner*.

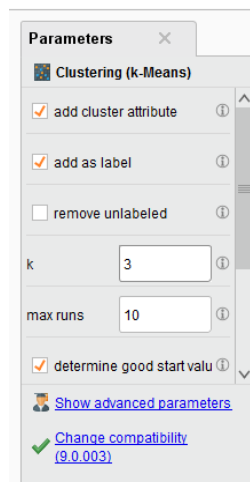
b. Proses pemilihan metode

Pada pemilihan metode ini peneliti memilih metode k-means dengan mengambil metodenya di *RapidMiner* pada pemilihan operator. Dapat dilihat pada Gambar 9 berikut ini :



Gambar 9. Metode k-means yang dipilih

Setelah data excel dapat digunakan kali ini peneliti mengambil operator *k-means* dengan cara yang sama dengan pengambilan operator *read excel* sebelumnya dan dimasukkan kedalam kolom proses dan dihubungkan dengan operator *read excel*.(Gambar 9)

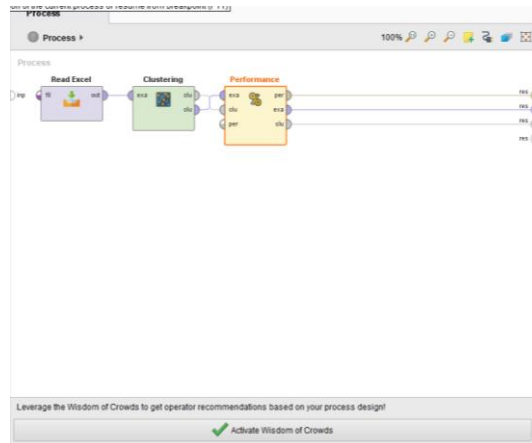


Gambar 10. Parameter metode k-means

Pada gambar 10 peneliti mengkonfigurasi dalam parameter proses k-means untuk menambah *cluster attribute* dan *add as label* yang bertujuan untuk menunjukkan hasil *cluster* yang akan dihasilkan nanti dengan dipanggil sebagai *label*

c. Proses akhir

Pada proses terakhir ini akan ditambah operator *Cluster Distance Performance* sehingga didapatkannya jarak antar cluster.

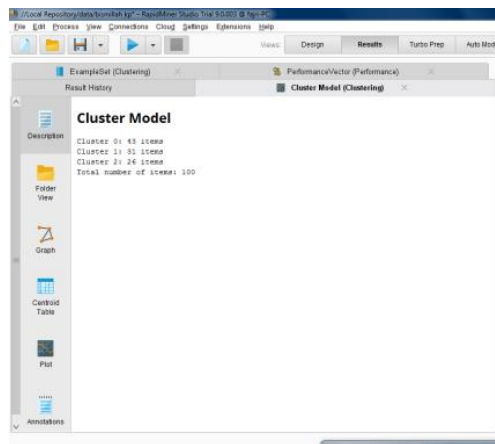


Gambar 11. Proses akhir

Setelah operator read excel, *clustering* dan performance dihubungkan seperti gambar 11, proses siap dijalankan.

d. Tampilan hasil proses

Pada tampilan hasil proses ini menunjukkan hasil olahan dengan menggunakan metode k-means. Dapat dilihat pada Gambar - gambar berikut ini :



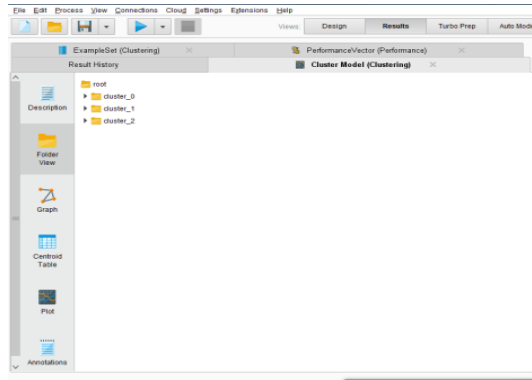
Gambar 12. Hasil pembagian cluster melalui description

Pada gambar 12 dapat dilihat bahwa 100 data yang dimiliki terbagi 3 cluster yang dalam pembagiannya :

- 43 data masuk kedalam *cluster_0*
- 31 data masuk kedalam *cluster_1*
- 26 data masuk kedalam *cluster_2*

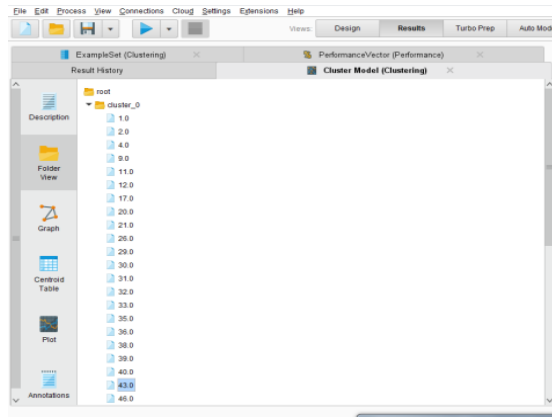
e. Tampilan folder cluster

Pada tampilan hasil proses ini menunjukkan hasil olahan dengan menggunakan metode k-means dalam bentuk folder. Gambar dapat dilihat berikut ini :



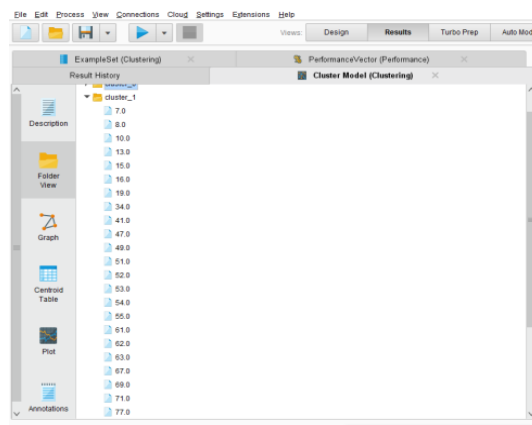
Gambar 13. Folder hasil pembagian *cluster* melalui *folder view*

Pada gambar 13 cluster – cluster yang terbagi 3 tadi dimasukkan kedalam 3 folder yang berbeda.

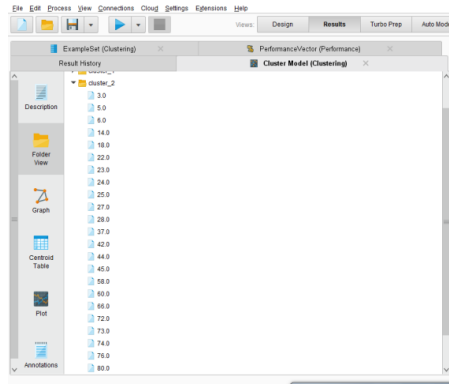


Gambar 14. Hasil pembagian *cluster_0* melalui *folder view*

Pada gambar 14 data – data yang termasuk kedalam *cluster_0* ada didalam folder tersebut, begitu juga data yang termasuk kedalam *cluster_1* dan *cluster_2*, hal ini dapat dilihat pada gambar 15 dan gambar 16



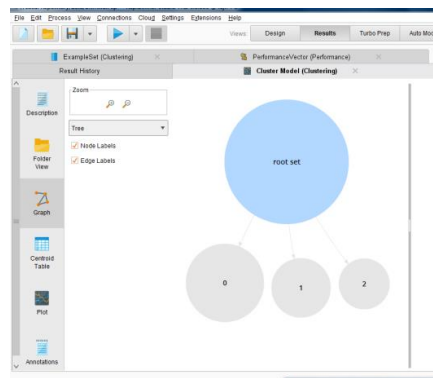
Gambar 15. Hasil pembagian *cluster_1* melalui *folder view*



Gambar 16. Hasil pembagian *cluster_2* melalui *folder view*

f. Tampilan grafik cluster

Pada tampilan hasil proses ini menunjukkan hasil olahan dengan menggunakan metode k-means dalam bentuk grafik dengan ukuran yang berbeda – beda dengan menunjukkan besar nilai yang terdapat pada cluster tersebut. Gambar dapat dilihat berikut ini :



Gambar 17. Grafik pembagian *cluster* dimenu *graph*

Pada gambar 17 grafik di atas menunjukkan dari 100 data yang digunakan *cluster_0* memiliki jumlah data yang lebih besar dibanding *cluster_1* dan *cluster_2*.

g. Tampilan tabel centroid

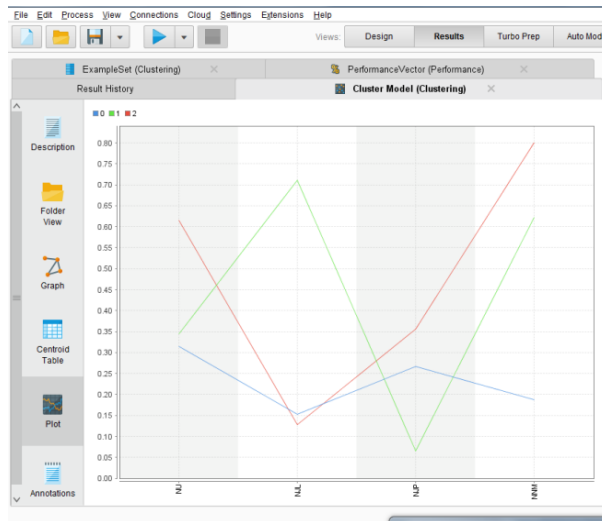
Pada tampilan hasil proses ini (Gambar 18) menunjukkan centroid yang dimiliki tiap *cluster*. Gambar dapat dilihat berikut ini :

Attribute	cluster_0	cluster_1	cluster_2
NU	0.314	0.345	0.615
NLI	0.153	0.712	0.127
NLP	0.267	0.065	0.356
NPAB	0.187	0.621	0.600

Gambar 18 Tabel *centroid*

h. Tampilan grafik batang

Pada tampilan hasil proses ini(Gambar 19) menunjukkan hasil olahan dengan menggunakan metode k-means dalam bentuk grafik batang. Gambar dapat dilihat berikut ini :



Gambar 19 Grafik batang cluster

Pada grafik di atas menunjukkan jarak cluster yang dimiliki tiap kolom data yang digunakan.

i. Tampilan cluster dan jarak cluster

Pada tampilan hasil proses ini menunjukkan hasil olahan dengan menggunakan metode k-means dalam bentuk data tiap individu dalam bentuk tabel yang memiliki kelompok *cluster* dan jarak tiap *cluster*(Gambar 3.20).

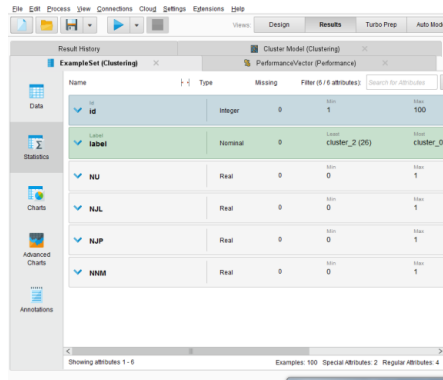
Row No.	id	label	M1	M2	M3	M4
1	1	cluster_0	0.380	0.330	0	0.110
2	2	cluster_0	0.570	0.330	0.250	0.440
3	3	cluster_2	0.550	0	0.500	0.670
4	4	cluster_0	0.360	0	0	0.220
5	5	cluster_2	0.400	0	0.250	0.890
6	6	cluster_2	0.790	0	0.500	1
7	7	cluster_1	0.240	0.670	0.250	0.890
8	8	cluster_1	0.210	0.670	0	0.560
9	9	cluster_0	0.310	0	0	0.110
10	10	cluster_1	0.740	0.670	0	0.440
11	11	cluster_0	0.600	0	0.250	0.110
12	12	cluster_0	0.330	0	0	0.220
13	13	cluster_1	0.170	0.670	0	0.780
14	14	cluster_2	0.760	0	0.250	0.890
15	15	cluster_1	0.400	0.670	0	0.110
16	16	cluster_1	0.330	0.670	0.250	0.780
17	17	cluster_0	0.140	0	0.500	0.560

Gambar 20 Hasil pembagian cluster dan jarak antar cluster

Hasil dari pembagian *cluster* yang dilakukan ditambah kedalam kolom bernama label yang menunjukkan pembagian *cluster* tiap data dan ditunjukkan jarak antara tiap data yang digunakan dalam kolomnya masing – masing.

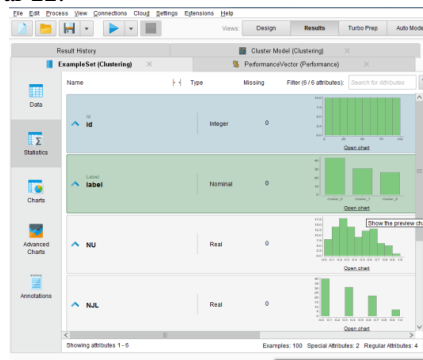
j. Tampilan statistik

Pada tampilan hasil proses ini menunjukkan hasil olahan dengan menggunakan metode k-means dalam bentuk statistik. Gambar dapat dilihat berikut ini :



Gambar 21. Statistik pada setiap kolom

Pada gambar 21 menunjukkan tiap kolom yang jika diklik akan menghasilkan statistik tiap kolom yang digunakan seperti gambar 22.



Gambar 22. Statistik pada setiap kolom

3. KESIMPULAN DAN SARAN

4.1 Kesimpulan

Setelah peneliti penerapan *K-Means Clustering* pengukuran tingkat ketepatan waktu kelulusan mahasiswa program studi teknik informatika STMIK Nurdin Hamzah, maka peneliti dapat menarik kesimpulan sebagai berikut :

1. Pengelompokan pengukuran tingkat ketepatan waktu kelulusan mahasiswa menggunakan metode *K-Means Clustering* dan untuk menampung data baru, dibuat aplikasi berbasis web dengan bahasa pemrograman PHP dan HTML.
2. Hasil output data baru aplikasi dapat disimpan dalam bentuk file excel sehingga dapat di impor langsung ke *RapidMiner* untuk melakukan *Clustering*.
3. Hasil *clustering* digunakan untuk penentuan ketepatan waktu kelulusan mahasiswa tersebut untuk mengetahui mahasiswa mana saja yang telah tepat waktu kelulusannya.
4. Hasil *Clustering* dengan metode K-Means pengelompokan ketepatan waktu kelulusan dari 950 data mahasiswa di dapatkan 575 mahasiswa kedalam *cluster* tepat waktu , 106 mahasiswa tidak tepat, 268 mahasiswa yang stop kuliah.
5. Hasil dari penerapan *K-Means Clustering* Pengukuran Ketepatan Waktu kelulusan ini yang dilakukan menggunakan *RapidMiner* data manual, klasifikasi terbesar terdapat pada *Cluster* 1 yaitu 58,5% yang ketepatan kelulusan mahasiswa dari tahun 2010 sampai dengan 2014.
6. Hasil pengelompokan pengukuran tingkat ketepatan waktu kelulusan mahasiswa dapat dimanfaatkan oleh prodi Teknik Informatika STMIK Nurdin Hamzah untuk mengamati

persentase tingkat kelulusan mahasiswa prodi TI yang tepat waktu, dapat menjadikan bahan evaluasi proses belajar mengajar kedepannya .

4.2 Saran

Dari kesimpulan diatas dan pembahasan bab sebelumnya peneliti menyarankan beberapa hal, yaitu :

1. Peneliti selanjutnya dapat melakukan penelitian ini dengan pengambilan data yang lebih luas untuk ketepatan waktu kelulusan mahasiswa di prodi TI STMIK Nurdin Hamzah.
2. Penelitian dapat dikembangkan dengan metode lain untuk melakukan *Clustering* minsalnya dengan metode pengklusteran tak terawasi *self organizing map* atau dengan metode *fuzzy subtractive clustering* untuk memperoleh *cluster* yang optimal.
3. Penelitian ini bisa lebih dikembangkan lagi dimasa mendatang dengan membuat aplikasi untuk pengukuran waktu kelulusan mahasiswa menggunakan bahasa pemrograman berbasis web, tidak menggunakan *tools* bantu lagi.

DAFTAR PUSTAKA

- [1] Kusrini and E. T. Luthfi (2009). *Algoritma Data Mining*. Yogyakarta: Andi,.
- [2] T. Connolly and C. Begg (2010). *Database Systems: A Practical Approach to Design, Implementation, and Management*. United States: Pearson.
- [3] J. Han, M. Kamber and J. Pei (2012). *Data Mining Concepts and Techniques*. United States: Elsevier.
- [4] Nugraha and J. A. Mita. 2014. "Data Mining dengan Metode Clustering untuk Pengolahan Informasi Persediaan Obat pada Puskesmas Pandanaran Semarang". *Fakultas Ilmu Komputer Udinus*, pp. 23-42.
- [5] E. Prasetyo (2012). *Data Mining Konsep dan aplikasi Menggunakan MATLAB*. Yogyakarta: Andi.
- [6] R. R. Putra and C. Wadisman. 2018, "Implementasi Data Mining Pemilihan Pelanggan Potensial Menggunakan Algoritma K-Means". *Journal of Information Technology and Computer Science (INTECOMS)*, pp. 72-77.