

Penerapan Data Mining untuk Menentukan Jumlah Pencari Kerja Terdaftar Berdasarkan Umur dan Pendidikan Menggunakan *K-Means Clustering* (Studi Kasus di Dinas Tenaga Kerja Dan Transmigrasi Provinsi Bengkulu)

Achmad Fikri Sallaby¹ Eko Suryana²

^{1,2} Universitas Dehasen

Jl. Meranti Sawah Lebar Bengkulu, 38228 Indonesia

fikrisallaby@unived.ac.id¹

ekosuryana@unived.ac.id²

Abstract— *The number of registered job seekers consists of applicants based on age and education. Each has a different number of registrants, with registrants by age having multiple categories and by education having variations in education owned by registered job seekers. This study applies Data Mining by using Clustering method to determine the number of job seekers based on the level of the number of applicants in the Office of Manpower and Transmigration of Bengkulu Province. The algorithm used is K-Means Clustering, where data are grouped based on the same characteristics will be entered into the same group and the data set entered into the group does not overlap. The test is done with Rapid Miner application 5.3. Rapid Miner is a Data Mining software that can be used to access several methods in Data Mining, so it can generate clusters in the grouping of data on the number of registered job seekers.*

Abstrak— *Jumlah pencari kerja terdaftar terdiri dari pendaftar berdasarkan umur dan pendidikan. Masing masing memiliki jumlah pendaftar yang berbeda, dengan pendaftar berdasarkan umur memiliki beberapa kategori dan berdasarkan pendidikan memiliki variasi pendidikan yang dimiliki pencari kerja terdaftar. Penelitian ini menerapkan Data Mining dengan menggunakan metode Clustering untuk menentukan jumlah pencari kerja berdasarkan tingkat jumlah pendaftar di Dinas Tenaga Kerja dan Transmigrasi Provinsi Bengkulu. Algoritma yang digunakan yaitu K-Means Clustering, di mana data dikelompokkan berdasarkan karakteristik yang sama akan dimasukkan ke dalam kelompok yang sama dan set data yang dimasukkan ke dalam kelompok tidak tumpang tindih. Pengujian dilakukan dengan aplikasi Rapid Miner 5.3. Rapid Miner merupakan software Data Mining yang dapat digunakan untuk mengakses beberapa metode yang ada dalam Data Mining, sehingga dapat menghasilkan cluster-cluster dalam pengelompokan data jumlah pencari kerja yang terdaftar.*

Keywords— *Data Mining, K-Means Clustering, Pencari Kerja, Rapid Miner.*

I. Pendahuluan

Undang undang Nomor 7 Tahun 1981 tentang wajib lapor ketenagakerjaan (Lembaran Negara Republik Indonesia Tahun 1981 Nomor 39, Tambahan Lembaran Negara Republik Indonesia Nomor 3201) setiap pemberi kerja di daerah wajib menyampaikan informasi lowongan pekerjaan kepada Dinas Tenaga Kerja dan Transmigrasi dan Instansi yang bertanggung jawab di bidang ketenagakerjaan Kabupaten / Kota

Data Mining merupakan bagian dari proses *Knowledge Discovery in Database* (KDD) Dimana proses tersebut mengacu pada proses yang luas untuk menemukan pengetahuan dalam data dan menekankan aplikasi khusus pada metode *Data Mining* [5].

K-Means clustering merupakan algoritma yang mudah dan sederhana. Algoritma *Clustering*

memiliki daya tarik yang luas dan digunakan untuk analisis data eksplorasi [1]. Peningkatan algoritma *Clustering K-Means* menggunakan waktu yang lebih baik dan akurat. Analisis *cluster* merupakan pengelompokan satu set objek dengan benda yang sama yang lebih mirip satu sama lain [3].

Penggunaan *Data Mining* dalam pengelompokan para pencari kerja terdaftar dikarenakan *Data Mining* sudah banyak diterapkan dalam berbagai bidang. Salah satunya metode *clustering* khususnya metode *K-Means*. *Data Mining*. Hasil pengolahan dengan *Data Mining* berupa sebuah pengetahuan (*knowledge*) baru yaitu pengelompokan data jumlah pencari kerja terdaftar, sehingga dapat.

II. Metode Penelitian

Pada penelitian ini menggunakan metode *Data Mining K-Means Clustering*. Tahapan tahapan *Data Mining* yaitu, Pembersihan data, Integrasi data, seleksi data, transformasi data, proses *mining*, Evaluasi Pola dan presentasi pegetahuan [2]. *K-means clustering* merupakan salah satu metode data *clustering non-hirarki* yang mengelompokkan data dalam bentuk satu atau lebih *cluster*/kelompok. Data-data yang memiliki karakteristik yang sama dikelompokkan dalam satu *cluster*/kelompok dan data yang memiliki karakteristik yang berbeda dikelompokkan dengan *cluster*/kelompok yang lain sehingga data yang berada dalam satu *cluster*/kelompok memiliki tingkat variasi yang kecil [4]. Pengelompokan (*clustering*) merupakan bagian dari ilmu *Data Mining* yang bersifat tanpa arahan (*unsupervised*). *Clustering* adalah proses pembagian data ke dalam kelas atau *cluster* berdasarkan tingkat kesamaannya [6]. Adapun langkah-langkah pada proses *clustering* dengan menggunakan algoritma *K-Means*, adalah sebagai berikut:

1. Menentukan Jumlah *Cluster*

Jumlah *cluster* yang ditentukan untuk mengelompokkan data pada penelitian ini sebanyak 3 *cluster*.

2. Menentukan *Centroid*

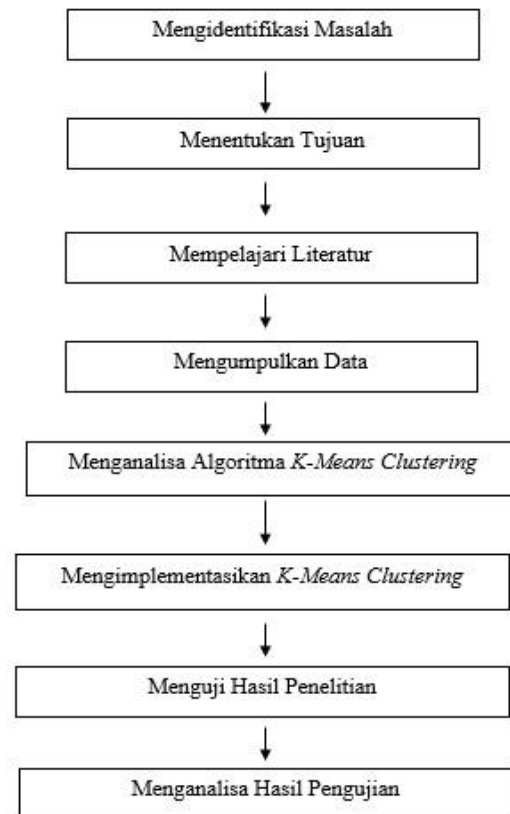
Pusat awal *cluster* (*centroid*) ditentukan secara *random* atau acak.

3. Menghitung Jarak dari *Centroid*

Menghitung jarak antara titik *centroid* dengan titik tiap objek dengan menggunakan *Euclidian Distance*.

4. Alokasikan masing-masing objek ke *centroid* terdekat.

Untuk mengalokasikan objek ke dalam masing-masing *cluster* dengan cara mengelompokkan berdasarkan jarak minimum objek ke pusat *cluster*. Dalam penelitian ada urutan kerangka kerja yang harus diikuti, urutan kerangka kerja ini merupakan langkah-langkah yang dilakukan dalam penulisan. Adapun kerangka kerja yang digunakan dalam penulisan ini adalah seperti terlihat pada gambar 1.



Gbr. 1 Kerangka Kerja

III. Hasil dan Pembahasan

Pengumpulan data dan data yang digunakan pada penelitian ini adalah data pencari kerja yang terdaftar di Dinas Tenaga Kerja dan Transmigrasi Provinsi Bengkulu. *Cluster* atau pengelompokan yang ingin dibentuk yaitu tingkat jumlah pencari kerja terdaftar yang berdasarkan jumlah pencari kerja yang terdaftar bulan Desember dan jumlah pencari kerja pada bulan sebelumnya pada tahun 2015 dari tingkat jumlah pencari kerja terdaftar berdasarkan Umur dan Pendidikan. Hasil adalah data rekapitulasi yang digunakan dapat dilihat pada tabel 3.1 Tabel Berdasarkan Umur dan Tabel 3.2 Berdasarkan Pendidikan

Tabel 1
Sampel Data Umur

Bulan	Kelompok Umur	Pencaker Terdaftar bulan Desember	Jumlah Pencaker bulan lalu
Desember	15 - 29	13	1299
	20 - 29	112	11757
	30 - 44	9	5798
	45 - 54	1	144
	55+	0	5

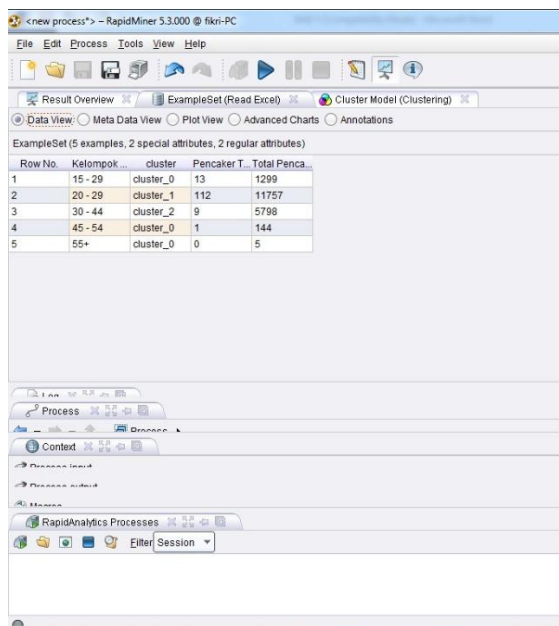
Tabel 2
 Sampel Data Pendidikan

Jenis Pendidikan	Sisa Pendaftar Bulan lalu	Yang terdaftar Bulan Desember
(1)	(2)	(3)
SEKOLAH DASAR		
BH DAN TIDAK TAMAT SD	10	0
SD	197	5
SUB TOTAL	207	5
PENDIDIKAN MENENGAH PERTAMA		
SLTP UMUM	123	0
SLTP KEJURUAN	1	0
SETINGKAT SLTP	916	0
SUB TOTAL	1040	0
PENDIDIKAN MENENGAH ATAS		
S.M.U	1608	37
SUB TOTAL	1608	37

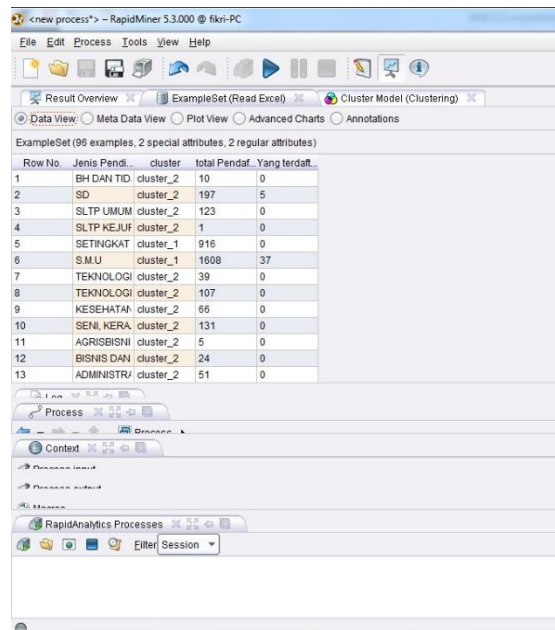
Hasil pengelompokan data jumlah pencari kerja berdasarkan umur dan pendidikan dengan *Rapid Miner* :

A. *Data View*

merupakan *sheet* untuk menampilkan data yang telah diolah secara keseluruhan lengkap dengan *cluster*. *Data view* dapat lihat pada gambar 2 dan 3.



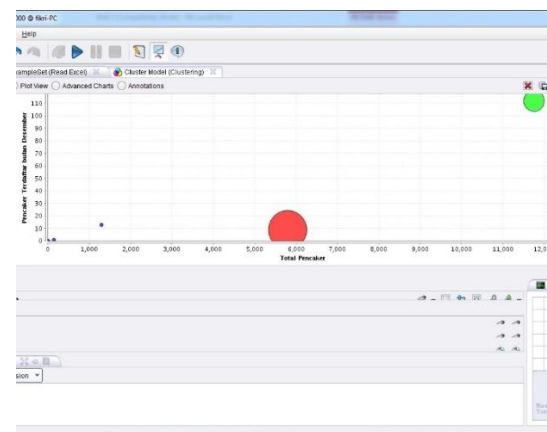
Gbr. 2 Tampilan data *View* umur



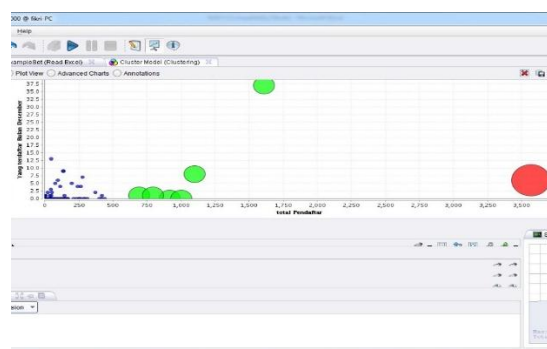
Gbr. 3 Tampilan data *view* pendidikan

B. *Plot View*

merupakan *sheet* di *Result Perspective* untuk menampilkan hasil pengelompokan data yang telah diolah secara keseluruhan lengkap dengan klasternya dari *example set (read excel)* dalam bentuk diagram *Scatter*. Tampilannya dapat dilihat di gambar 4 dan 5.



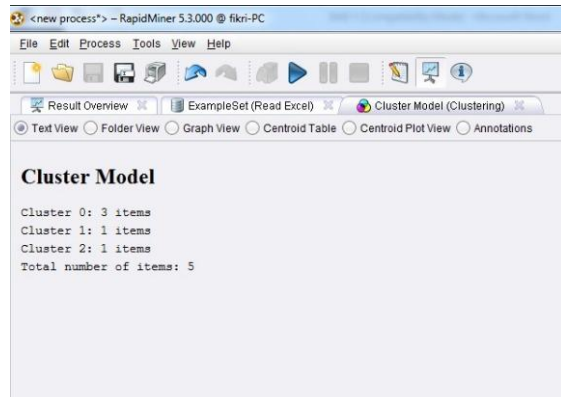
Gbr. 4 Tampilan *Plot View* umur



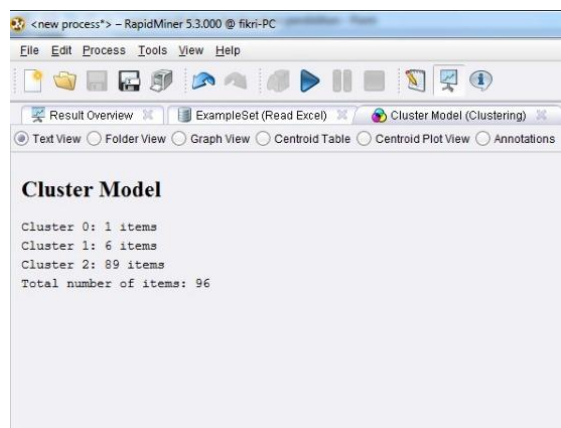
Gbr. 5 Tampilan *Plot View* Pendidikan

C. Clustering Model

merupakan *sheet* untuk menampilkan *database* yang telah berhasil diolah secara keseluruhan. Dilihat dari *cluster model (clustering)*. Dapat dilihat pada gambar 6 dan 7.



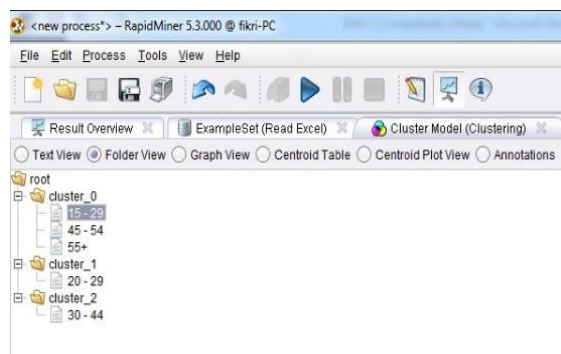
Gbr. 6 Tampilan Cluster Model Umur



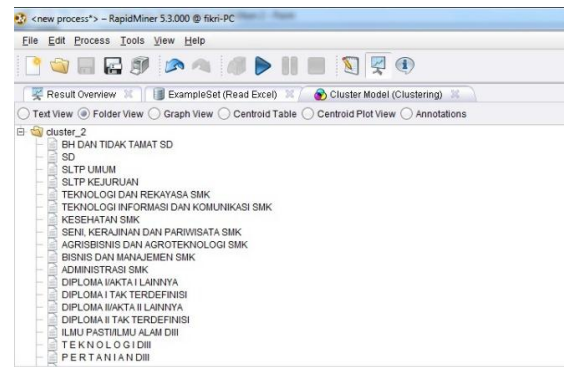
Gbr. 7 Tampilan Clustering Model Pendidikan

D. Folder View

merupakan *sheet* untuk menampilkan *database* yang telah berhasil diolah secara keseluruhan. dilihat dari *cluster model (clustering)* dapat dilihat pada gambar 8 dan 9.



Gbr. 8 Tampilan folder view umur



Gbr. 9 Tampilan folder view Pendidikan

IV. Penutup

Kesimpulan yang didapatkan dari penelitian ini adalah pengelompokan data pelanggan adalah sebanyak 3 *cluster*. Dimana Pada Data Jumlah Pencari Kerja Berdasarkan Umur *cluster 0* adalah kelompok data jumlah pencari kerja dengan jumlah terendah, *cluster 1* yaitu kelompok data jumlah pencari kerja dengan Jumlah tertinggi dan *Cluster 2* dengan jumlah Pendaftar yang sedang. Sedangkan pada data berdasarkan pendidikan *Cluster 0* adalah kelompok dengan jumlah pendaftar terbanyak dan pada *Cluster 1* dengan jumlah pendaftar sedang dan pada *Cluster 2* dengan jumlah pendaftar rendah.

Referensi

- [1] Bondu Venkateswarlu and Prof G.S.V.Prasad Raju. (2013). "Mine Blood Donors Information through Improved K-Means Clustering." *International Journal of Computational Science and Information Technology (IJCSITY) Vol.1, No.3.*
- [2] Budanis Dwi Meilani dan Nofi Susanti, (2014), "Aplikasi Data Mining Untuk Menghasilkan Pola Kelulusan Siswa Dengan Metode Naïve Bayes." Ed. *Jurnal LINK Vol 21/No.2*
- [3] Er. Nikhil Chaturvedi and Er. Anand Rajavat. (2013). "An Improvement in K-mean Clustering Algorithm Using Better Time and Accuracy." *International Journal of Programming Language and Applications (IJPLA) Vol. 3 No.4.*
- [4] Johan, O, O. (2013). "Implementasi Algoritma K-Means Clustering Untuk Menentukan Strategi Marketing President University" Ed. *Jurnal Ilmiah Teknik Industri, Vol. 12.*
- [5] Priyadharsini and Thanamani. (2014). "An Overview of Knowledge Discovery Database and Data mining Techniques." *An ISO 3297: 2007 Certified Organization Vol. 2.*
- [6] Tutik Khotimah. (2014). "Pengelompokan Surat Dalam Al Qur'an Menggunakan Algoritma K-Means." Ed. *Jurnal Simetris, Vol 5 No 1*