

Submit: 20 Februari 2023 | Diterima : 22 Februari 2023 | Terbit : 05 Maret 2023

# CLUSTERING DATA PENCARI KERJA MENURUT TINGKAT PENDIDIKAN MENGUNAKAN ALGORITMA K-MEANS

Siti Rohaniyah<sup>1</sup>, Ade Irma Purnamasari<sup>2</sup>

<sup>1)2)</sup> STMIK IKMI Cirebon, Indonesia

<sup>1)</sup>[sitirohaniyah18@gmail.com](mailto:sitirohaniyah18@gmail.com), <sup>2)</sup>[Irma2947@yahoo.com](mailto:Irma2947@yahoo.com),

## Abstrak

Pencari kerja dalam melakukan proses melamar untuk mendapatkan pekerjaan sesuai keinginan terhadap perusahaan yang diharapkan, tentunya harus sesuai dengan kebijakan yang ditentukan oleh perusahaan dalam merekrut para pelamar. Permasalahan yang sering terjadi hingga saat ini yakni masih banyak pencari kerja yang belum mendapatkan pekerjaan akibat terkendala faktor-faktor seperti tingkat pendidikan serta daerah tersebar yang belum tersedia lapangan pekerjaan, masing-masing pencari kerja yang melamar ke berbagai perusahaan tertentu tak luput memiliki kualifikasi berdasarkan tingkat pendidikan yang bervariasi yang dimiliki oleh pencari kerja di wilayah Jawa Barat. Data yang diambil untuk penelitian ini bersumber dari Dinas Tenaga Kerja dan Transmigrasi melalui *website* Open Data Jabar, data tersebut memiliki 579 *record* serta 8 atribut. Data yang akan digunakan yakni data pencari kerja pada tahun 2021 atribut yang dipilih yakni nama kabupaten/kota, tingkat pendidikan, jumlah pencari kerja dan tahun. Proses pengolahan data dilakukan dengan *Data Mining* atau yang sering dikenal dengan *Knowledge Discovery in Database*, metode yang digunakan pada Tugas Akhir ini adalah metode *clustering* menggunakan Algoritma *K-Means* untuk mengelompokkan suatu item ke dalam kelompok dimana *k* (kluster) merupakan jumlah kelompok yang diinginkan. Tujuan dilakukannya Tugas Akhir ini yakni untuk mengelompokkan data pencari kerja di wilayah Jawa Barat serta memisahkan data pencari kerja kedalam *cluster* berbeda untuk objek yang memiliki ciri yang berbeda. Pengujian ini dilakukan menggunakan *Software RapidMiner*, *RapidMiner* merupakan *software Data Mining* yang dapat digunakan dengan beberapa metode yang tersedia dalam *Data Mining*. Kontribusi dari pengelompokan data pencari kerja ini membantu pemerintah memberikan pekerjaan secara merata kepada para pencari kerja agar tidak terjadi penumpukan data pencari kerja pada setiap tahunnya. Hasil pengelompokan ini menghasilkan 6 cluster (*k6*), *Cluster\_5* merupakan jumlah pencari kerja yang terbanyak yang terdiri dari 56 anggota, masing-masing anggota terdiri dari DIPLOMA I/II/III/AKADEMI dengan jumlah pencari kerja terbanyak berjumlah 1062 orang, PROFESI dengan jumlah pencari kerja terbanyak berjumlah 20 orang, SD dengan jumlah pencari kerja terbanyak berjumlah 714 orang dan TIDAK/BELUM TAMAT SD dengan jumlah pencari kerja terbanyak berjumlah 2 orang. Sehingga dapat disimpulkan jumlah pencari kerja yang mendominasi yakni pencari kerja dengan tingkat pendidikan DIPLOMA I/II/III/AKADEMI yang berjumlah 1062 orang.

## Kata kunci :

Pencari kerja, Data Mining, Algoritma K-Means, cluster, RapidMiner

## PENDAHULUAN

Pekerjaan adalah kebutuhan setiap Warga Negara Indonesia yang sesudah memasuki usia Angkatan kerja untuk memperoleh pekerjaan dan mendapatkan penghasilan yang layak. Masing-masing pencari kerja mempunyai hak untuk mendapatkan peluang dalam memperoleh mata pencaharian atau pekerjaan, serta setara dengan keahlian dan keterampilan (Nurchahyo, 2021). Seiring dengan perkembangan dunia industri serta maraknya para pencari kerja khususnya di wilayah Jawa Barat ini, sedangkan lowongan pekerjaan yang tersedia tidak sebanding menyebabkan banyak lulusan dari berbagai pendidikan tingkat rendah sampai tinggi yang masih belum mendapatkan pekerjaan.

Pencari kerja merupakan Angkatan kerja yang masih menganggur serta belum mendapatkan pekerjaan tetap, solusi dalam menangani permasalahan pencari kerja yang semakin meningkat yakni dengan memperbesar lowongan pekerjaan secara merata (Ningsih & Abdullah, 2021). Salah satu permasalahan yang terjadi saat ini banyak pencari kerja yang hendak melamar pekerjaan terkendala akibat faktor-faktor seperti persyaratan mengenai

\*penulis korespondensi



tingkat pendidikan yang belum memenuhi kriteria serta daerah yang tersebar belum merata dalam menyediakan lapangan pekerjaan. Tugas Akhir ini menerapkan Data Mining yakni proses pengelompokan data Pencari kerja di wilayah Jawa Barat menggunakan Algoritma *K-Means Clustering*, untuk mengetahui terbagi kedalam beberapa kelompok para pencari kerja yang belum mendapatkan pekerjaan.

Data Mining merupakan salah satu dari proses *Knowledge Discovery in Database* (KDD), *Knowledge Discovery in Database* terdapat beberapa bagian yakni pembersihan data (*data cleaning*), integrasi data (*data integration*), pemilihan data (*data selection*), transformasi data (*data transformation*), evaluasi pola (*pattern evaluasi*), serta presentasi pengetahuan (*knowledge presentation*) (Wijaya & Dwiasnati, 2020). Dalam Data Mining terbagi kedalam 3 bagian yakni Klasifikasi, *Asossiation* dan *Clustering*. *Clustering* merupakan pengelompokan data yang dilakukan tidak berdasarkan kelas data tersebut kedalam kelas objek yang sama sesuai dengan topik yang diambil dalam suatu penelitian, untuk itu menindaklanjuti permasalahan mengenai data pencari kerja ini diperlukan metode *clustering* menggunakan Algoritma *K-Means*.

Algoritma *K-Means* adalah metode *clustering* berbasis jarak yang membagi data kedalam sejumlah *cluster*, metode ini hanya bekerja pada atribut numerik. *K-Means* melakukan pengukuran jarak dengan *Euclidian Distance* secara iteratif memastikan masing-masing *record* dari *cluster* asal (Aditya et al., 2020). Prosedur diawali dengan memilih *k* dengan inisial catatan sebagai pusat *cluster* serta memastikan setiap *record* yang sangat dekat dengan *cluster*. Kemudian *record* baru ditambahkan ke *cluster*, *cluster* dihitung ulang kembali untuk mendeskripsikan anggota baru dalam suatu *cluster*. Prinsip utama metode *K-Means* yakni Menyusun *K* buah (*centroid*) atau rata-rata (*mean*) yang terdapat dari sekumpulan data.

Tugas Akhir ini bertujuan untuk mengelompokan data pencari kerja berdasarkan Tingkat Pendidikan khususnya di wilayah Jawa Barat menggunakan Algoritma *K-Means* melalui metode *clustering* dengan menerapkan Data Mining menggunakan *software RapidMiner* agar menghasilkan data yang akurat dan relevan, serta menghasilkan *output* Tugas Akhir yang bersifat kuantitatif yakni fakta yang tersedia setelah Tugas Akhir ini selesai.

Rumusan Masalah:

Berdasarkan pemaparan latar belakang diatas maka tugas akhir merumuskan beberapa masalah sebagai berikut. Bagaimana cara melakukan pengelompokan data pencari kerja berdasarkan Tingkat Pendidikan menggunakan Algoritma *K-Means*?

Berapa *k optimum* yang didapatkan setelah melakukan uji coba menggunakan *software RapidMiner* pada data pencari kerja?

Batasan Masalah:

Batasan masalah bertujuan untuk membatasi ruang lingkup penelitian, sehingga Tugas Akhir ini terfokus pada. Data yang akan digunakan pada Tugas Akhir ini bersumber dari Dinas Tenaga Kerja dan Transmigrasi melalui *website* Open Data Jabar, yakni Data Pencari Kerja pada tahun 2021.

Metode yang digunakan dalam Tugas Akhir ini yakni metode *clustering* menggunakan Algoritma *K-Means*.

Atribut yang digunakan dalam *clustering* data pencari kerja yakni pendidikan tertinggi, jumlah pencari kerja dan tahun.

Tujuan Tugas Akhir:

Tujuan Tugas Akhir ini berdasarkan rumusan masalah yang tertera pada tugas akhir tersebut yakni sebagai berikut.

Mengelompokan data pencari kerja melalui penerapan Data Mining menggunakan Algoritma *K-Means* dengan *software RapidMiner*.

Mengetahui *k optimum* setelah dilakukan uji coba dengan *software RapidMiner* pada pengelompokan data pencari kerja di wilayah Jawa Barat.

## TINJAUAN PUSTAKA

Peluang Kerja merupakan banyaknya orang yang dapat tertampung untuk bekerja terhadap suatu instansi atau perusahaan tertentu. Peluang kerja ini menampung seluruh tenaga kerja yang ada apabila lapangan pekerjaan yang memadai atau setara dengan banyaknya tenaga kerja yang tersedia, salah satu indikator adanya pengangguran dalam perekonomian yakni pencari kerja. Pencari kerja merupakan Angkatan kerja yang sedang menganggur serta mencari pekerjaan dan belum mendapatkan pekerjaan tetap sehingga masih ingin berpindah tempat baik diluar maupun didalam negeri (Harahap & Nawawi, 2022). Pencari kerja memiliki kemampuan, keterampilan, tingkat pendidikan dan sikap pribadi yang berbeda, adanya perbedaan antara pencari kerja dan lowongan pekerjaan menyebabkan bahwa tidak setiap pelamar selalu cocok serta dapat diterima mengisi lowongan yang tersedia. Program pemerintah yang dapat membantu pencari kerja yakni dengan tersedianya badan tenaga kerja pemerintah yang memberikan informasi mengenai lowongan pekerjaan, cara lainnya melalui program pelatihan publik untuk mempermudah penyaluran tenaga kerja. Selain itu banyaknya pencari kerja hingga saat ini, mengakibatkan data

\*penulis korespondensi



pencairi kerja semakin menumpuk pada setiap tahunnya maka diperlukan pengolahan ilmu pengetahuan untuk pengelolaan data dalam jumlah besar yakni salah satunya dengan Data Mining atau sering kali dikenal dengan Knowledge Discovery in Database (KDD).

Berdasarkan Penelitian terdahulu yang dilakukan oleh (Pa et al., 2022) yang berjudul “Pengelompokan Data Pencari Kerja Terdaftar Berdasarkan Umur dan Pendidikan Menggunakan Metode *K-Means Clustering* Di Dinas Tenaga Kerja Dan Perindustrian Perdagangan Kota Binjai”. Penelitian ini dilakukan dengan Mengelompokan Data Pencari Kerja berdasarkan variabel Umur, Pendidikan, dan Wilayah, Metode penelitian ini berisi Identifikasi Masalah, Kajian Teori, Pengumpulan Data, Pengujian dan Implementasi serta Evaluasi. Hasil penelitian terdahulu menghasilkan 3 cluster, yaitu cluster 1 merupakan kelompok data pencari kerja berdasarkan umur, pendidikan, wilayah yang paling banyak dengan total 10 data dan terletak pada group umur 21 tahun, untuk kelompok pendidikan SMA dan wilayah Binjai Barat. Cluster 2 merupakan kelompok data pencari kerja yang terdaftar berdasarkan umur, pendidikan dan wilayah yang banyak dengan total 8 data dan terletak pada group umur 22 tahun, untuk kelompok pendidikan SMA dan wilayah Binjai Timur, serta Cluster 3 merupakan kelompok data pencari kerja berdasarkan umur, pendidikan dan wilayah dengan total 10 data dan terletak pada group umur 21 tahun, untuk kelompok pendidikan SMK dan wilayah Binjai Utara. Tugas Akhir yang saat ini dilakukan memiliki perbedaan variabel yang digunakan dengan penelitian terdahulu, dimana pengelompokan Data Pencari Kerja mengacu pada Umur dan Pendidikan. Sedangkan yang diambil untuk Tugas Akhir ini Pengelompokan Data Pencari Kerja berdasarkan Tingkat Pendidikan, fakta yang memperkuat dilakukannya Tugas Akhir ini yakni perbedaan attribute yang digunakan dapat memperoleh hasil yang berbeda dengan penelitian sebelumnya.

### METODE PENELITIAN

Teknik Pengumpulan Data yang dilakukan pada Tugas Akhir ini yakni terdiri dari studi Pustaka dan studi dokumentasi antara lain sebagai berikut.

#### Studi Pustaka

Proses pengumpulan data dilakukan melalui kegiatan studi Pustaka yakni dengan cara mengumpulkan berbagai ilmu pengetahuan serta referensi dari berbagai macam sumber Pustaka seperti referensi dari karya ilmiah maupun jurnal ilmiah yang bersangkutan dengan topik yang akan diteliti (Utami, 2020). Pada studi Pustaka ini referensi yang digunakan untuk Tugas Akhir yakni dengan mengambil langsung sumber jurnal dari internet.

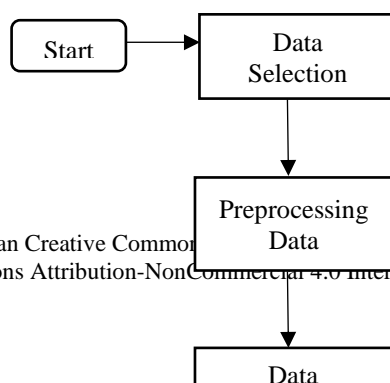
#### Studi Dokumentasi

Teknik studi dokumentasi yang dilakukan dengan cara mengambil data dari sumber data yang bersangkutan pada data pencari kerja tersebut, sumber data ini berasal dari Dinas Tenaga Kerja dan Transmigrasi pada *website* Open Data Jabar dengan nama dataset yang didapat yakni *Clustering* pencari kerja berdasarkan jenis kelamin dan tingkat pendidikan.

A	B	C	D	E	F
kode_provinsi	nama_provinsi	kode_kabupaten_kota	nama_kabupaten_kota	jenis_kelamin	pendidikan_tertinggi
32	JAWA BARAT	3201	KABUPATEN BOGOR	LARI - LARI	DIPLOMA I (D1)/IBU/IAKADEMI/DIPLOMA I (D1)/IBU/ACADEMY
32	JAWA BARAT	3201	KABUPATEN BOGOR	PEREMPUNAN	DIPLOMA I (D1)/IBU/IAKADEMI/DIPLOMA I (D1)/IBU/ACADEMY
32	JAWA BARAT	3201	KABUPATEN BOGOR	LARI - LARI	LAINNYA
32	JAWA BARAT	3201	KABUPATEN BOGOR	PEREMPUNAN	LAINNYA
32	JAWA BARAT	3201	KABUPATEN BOGOR	LARI - LARI	PROFESI
32	JAWA BARAT	3201	KABUPATEN BOGOR	PEREMPUNAN	PROFESI
32	JAWA BARAT	3201	KABUPATEN BOGOR	LARI - LARI	53
32	JAWA BARAT	3201	KABUPATEN BOGOR	PEREMPUNAN	53
32	JAWA BARAT	3201	KABUPATEN BOGOR	LARI - LARI	SEKOLAH DASAR/PRIMARY SCHOOL
32	JAWA BARAT	3201	KABUPATEN BOGOR	PEREMPUNAN	SEKOLAH DASAR/PRIMARY SCHOOL
32	JAWA BARAT	3201	KABUPATEN BOGOR	LARI - LARI	SEKOLAH MENENGAH ATAS (KEJURUJAN)/SENIOR HIGH SCHOOL (VOCATIONAL)
32	JAWA BARAT	3201	KABUPATEN BOGOR	PEREMPUNAN	SEKOLAH MENENGAH ATAS (KEJURUJAN)/SENIOR HIGH SCHOOL (VOCATIONAL)
32	JAWA BARAT	3201	KABUPATEN BOGOR	LARI - LARI	SEKOLAH MENENGAH ATAS (UMUM)/SENIOR HIGH SCHOOL (GENERAL)
32	JAWA BARAT	3201	KABUPATEN BOGOR	PEREMPUNAN	SEKOLAH MENENGAH ATAS (UMUM)/SENIOR HIGH SCHOOL (GENERAL)
32	JAWA BARAT	3201	KABUPATEN BOGOR	LARI - LARI	SEKOLAH MENENGAH PERTAMA/JUNIOR HIGH SCHOOL
32	JAWA BARAT	3201	KABUPATEN BOGOR	PEREMPUNAN	SEKOLAH MENENGAH PERTAMA/JUNIOR HIGH SCHOOL
32	JAWA BARAT	3201	KABUPATEN BOGOR	LARI - LARI	TIDAK/BELUM PERNAH SEKOLAH/NO SCHOOLING
32	JAWA BARAT	3201	KABUPATEN BOGOR	PEREMPUNAN	TIDAK/BELUM PERNAH SEKOLAH/NO SCHOOLING
32	JAWA BARAT	3201	KABUPATEN BOGOR	LARI - LARI	TIDAK/BELUM TAMAT SO/NOT YET COMPLETED PRIMARY SCHOOL
32	JAWA BARAT	3201	KABUPATEN BOGOR	PEREMPUNAN	TIDAK/BELUM TAMAT SO/NOT YET COMPLETED PRIMARY SCHOOL
32	JAWA BARAT	3201	KABUPATEN BOGOR	LARI - LARI	UNIVERSITAS/UNIVERSITY
32	JAWA BARAT	3201	KABUPATEN BOGOR	LARI - LARI	UNIVERSITAS/UNIVERSITY

Gambar 1. Dataset pencari kerja yang didapat dari Open Data Jabar

Tahapan perancangan dalam melakukan Tugas Akhir ini yakni dengan Data Mining atau Knowledge Discovery in Database (KDD), proses Data Mining yang dilakukan pada tahap perancangan kali ini dibuat dalam bentuk flowchart/diagram alir sehingga secara garis besar akan lebih mudah untuk dipahami. Flowchart yang telah dibentuk dapat dilihat pada diagram alir dibawah ini.



\*penulis korespondensi



This is an Creative Commons Attribution-NonCommercial 4.0 International License.

## Gambar 2. Diagram alir Tahap Perancangan

Berdasarkan diagram alir diatas dapat dijelaskan lebih spesifik lagi mengenai proses Data Mining yang terbagi kedalam 5 (lima) tahapan, diantaranya sebagai berikut:

### Data selection

Tahap pertama yakni pengambilan sample data yang dibutuhkan dalam Tugas Akhir ini, data yang akan digunakan yakni Data Pencari Kerja di Provinsi Jawa Barat pada tahun 2021. Dataset ini mempunyai 8 atribut yakni kode provinsi, nama provinsi, kode kabupaten kota, nama kabupaten kota, jenis kelamin, pendidikan tertinggi, jumlah pencari kerja, satuan dan tahun. Atribut yang dipakai yakni pendidikan tertinggi, jumlah pencari kerja serta tahun.

### Preprocessing data

Tahap kedua yakni preprocessing data dengan melakukan pengecekan data pencari kerja yang tidak konsisten seperti data yang masih *missing value* serta menghilangkan atribut yang tidak digunakan, pada data pencari kerja ini atribut yang digunakan hanya 3 (tiga) yakni pendidikan tertinggi, jumlah pencari kerja dan tahun.

### Data transformation

Tahap ketiga dataset yang akan diolah dilakukan transformasi terlebih dahulu agar *RapidMiner* dapat menganalisa data yang akan diproses, untuk itu perlu melakukan perubahan data pada attribute pendidikan tertinggi yang semula berbentuk *polynomial* diubah menjadi numerik agar dapat dilakukan proses Data Mining pada tools *RapidMiner*.

### Data mining

Tahap selanjutnya yakni melakukan pengelompokan Data Pencari Kerja menggunakan metode dari data mining yakni Algoritma *K-Means*. Pada pengolahan data tersebut dilakukan pemodelan untuk proses *clustering*, pengolahan data dilakukan menggunakan tools *RapidMiner*.

### Evaluasi

Setelah proses *k-means* dilakukan, tahap terakhir yang dilakukan yakni menginterpretasikan data pencari kerja untuk menentukan hasil *cluster* yang paling baik ke dalam bentuk *cluster* yang mudah dipahami.

## HASIL PENELITIAN DAN DISKUSI

Berikut ini adalah hasil yang diperoleh dari pengelompokan dari pengelompokan data pencari kerja berdasarkan tingkat pendidikan menggunakan Algoritma *K-Means* pada tools *RapidMiner* serta mengacu pada tahapan KDD sebagai berikut.

\*penulis korespondensi



*Data selection*

Tahap pertama dalam proses penyelesaian pengelompokan data pencari kerja yakni pemilihan data. Data yang akan digunakan adalah data pencari kerja pada tahun 2021 yang bersumber dari *website* Open Data Jabar, attribute yang digunakan yakni nama kabupaten/kota, pendidikan tertinggi, jumlah pencari kerja dan tahun. Dataset pencari kerja tersebut dapat dilihat pada tabel dibawah ini.

Tabel 1 Datsset awal

Kode provinsi	Nama provinsi	Kode kabupaten kota	Nama kabupaten kota	Jenis kelamin	Pendidikan tertinggi	Jumlah pencari kerja	Satuan	Tahun
32	JAWA BARAT	3201	KABUPATEN BOGOR	LAKI-LAKI	DIPLOMA I/II/III/AKADEMI	217	ORANG	2021
32	JAWA BARAT	3201	KABUPATEN BOGOR	PEREMPUAN	DIPLOMA I/II/III/AKADEMI	229	ORANG	2021
32	JAWA BARAT	3201	KABUPATEN BOGOR	LAKI-LAKI	S3	1	ORANG	2021
32	JAWA BARAT	3201	KABUPATEN BOGOR	PEREMPUAN	S3	2	ORANG	2021
.....	.....	.....	.....	.....	.....	.....	.....	.....
32	JAWA BARAT	3279	KOTA BANJAR	LAKI-LAKI	SMP	177	ORANG	2021
32	JAWA BARAT	3279	KOTA BANJAR	PEREMPUAN	SMP	216	ORANG	2021
32	JAWA BARAT	3279	KOTA BANJAR	LAKI-LAKI	UNIVERSITAS	84	ORANG	2021
32	JAWA BARAT	3279	KOTA BANJAR	PEREMPUAN	UNIVERSITAS	49	ORANG	2021

Setelah memilih data pencari kerja yang akan digunakan, data pencari kerja akan diproses kembali dengan menghilangkan attribute yang tidak dipakai, attribute yang akan digunakan yakni nama kabupaten/kota, pendidikan tertinggi, jumlah pencari kerja dan tahun yang akan diproses menggunakan *tools RapidMiner*.

Tabel 2 Pemilihan data

Nama kabupaten/kota	Pendidikan tertinggi	Jumlah pencari kerja	Tahun
KABUPATEN BOGOR	SMA	4454	2021
KABUPATEN BOGOR	DIPLOMA I/II/III/AKADEMI	229	2021
KABUPATEN BOGOR	SD	16	2021
KABUPATEN BOGOR	SMP	139	2021
.....	.....	.....	.....
KOTA BANJAR	SMK	0	2021
KOTA BANJAR	TIDAK/BELUM TAMAT SD	0	2021
KOTA BANJAR	PROFESI	0	2021
KOTA BANJAR	UNIVERSITAS	84	2021

*Preprocessing data*

Tahap kedua adalah tahap *preprocessing* data pada data pencari kerja yakni menghapus attribut yang tidak digunakan agar tidak menghambat proses *Data Mining*, pada data pencari kerja tersebut tidak ada *missing value*. Attribut yang tidak ada *missing value* dapat dilihat pada gambar dibawah ini.

\*penulis korespondensi



Name	Type	Missing	Statis...	Filter (6 / 6 attributes):	Average
id	Integer	0	Min 1	Max 216	108.5
cluster	Nominal	0	Least cluster_0 (5)	Most cluster_5 (56)	Values cluste
NAMA KABUPATEN/KOTA	Real	0	Min -1.776	Max 1.657	Average 0
Tingkat PENDIDIKAN	Real	0	Min -1.524	Max 1.524	Average 0
Jumlah Pencari Kerja	Real	0	Min -0.466	Max 5.663	Average 0
Tahun	Real	0	Min 0	Max 0	Average 0

Gambar 3. Pengecekan missing value

*Transformation data*

Tahap ketiga yakni transformasi data, data yang akan ditransformasi yakni atribut pendidikan tertinggi yang semula masih berbentuk polynominal dilakukan inisialisasi menjadi numerik agar dapat dilakukan proses Data Mining. Berikut ini tabel *attribute* yang sudah di inisialisasi terlebih dahulu.

Tabel 3 Inisialisasi pada attribute pendidikan tertinggi

Pendidikan tertinggi	Inisialisasi
DIPLOMA I/II/III/AKADEMI	1
PROFESI	2
TIDAK/BELUM TAMAT SD	3
SD	4
SMA	5
SMK	6
SMP	7
UNIVERSITAS	8

Tabel 4 Transformasi data berdasarkan tingkat pendidikan

Nama Kabupaten/Kota	Pendidikan Tertinggi	Jumlah Pencari Kerja	Tahun
KABUPATEN BOGOR	1	446	2021
KABUPATEN BOGOR	2	0	2021
KABUPATEN BOGOR	3	0	2021
KABUPATEN BOGOR	4	30	2021

\*penulis korespondensi

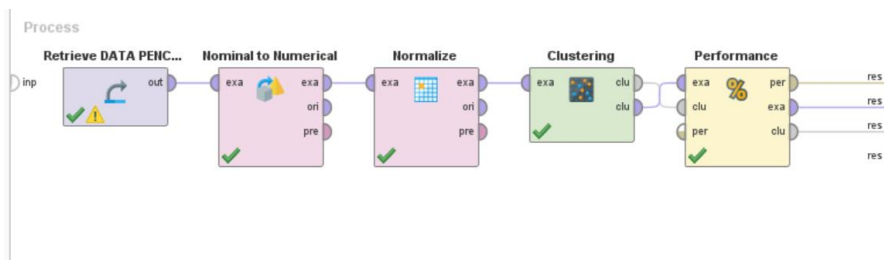


This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Nama Kabupaten/Kota	Pendidikan Tertinggi	Jumlah Pencari Kerja	Tahun
KABUPATEN BOGOR	5	8298	2021
KABUPATEN BOGOR	6	0	2021
KABUPATEN BOGOR	7	266	2021
KABUPATEN BOGOR	8	890	2021
....	....	....	....
KOTA BANJAR	1	120	2021
KOTA BANJAR	2	0	2021
KOTA BANJAR	3	0	2021
KOTA BANJAR	4	154	2021
KOTA BANJAR	5	1108	2021
KOTA BANJAR	6	0	2021
KOTA BANJAR	7	393	2021
KOTA BANJAR	8	133	2021

Data mining

Tahap selanjutnya yakni proses *Data Mining* yang dilakukan pada pengelompokan data pencari kerja berdasarkan tingkat pendidikan menggunakan *tools RapidMiner*, berikut ini merupakan susunan proses clustering menggunakan Algoritma *K-Means* pada *tools RapidMiner*.



Gambar 4. Proses clustering pada tools RapidMiner

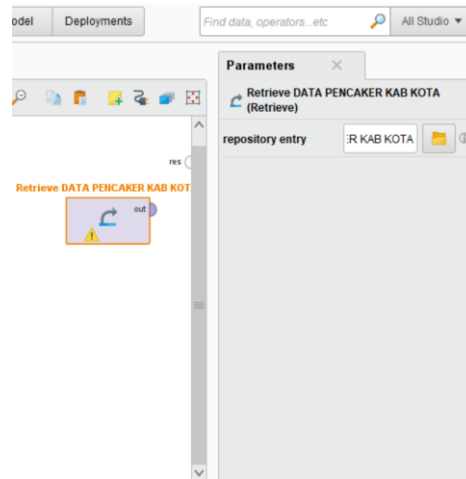
Sebelum proses *Data Mining* dilakukan berikut ini merupakan penjelasan susunan operator yang digunakan dalam *tools RapidMiner* untuk proses pengelompokan data pencari kerja menggunakan Algoritma *K-Means* yakni sebagai berikut:

*Retreive*

Operator *retrieve* pada gambar dibawah ini digunakan untuk mengakses data pencari kerja yang disimpan di *local repository* serta memuatnya ke dalam proses, *retrieve* digunakan untuk memulai proses *Data Mining* serta dapat menilai data tersebut sudah layak untuk tahap selanjutnya pada *Data Mining*. Operator *retrieve* memuat objek *RapidMiner* ke dalam proses, objek ini berupa sekumpulan data dan dapat berupa koleksi atau model.

\*penulis korespondensi

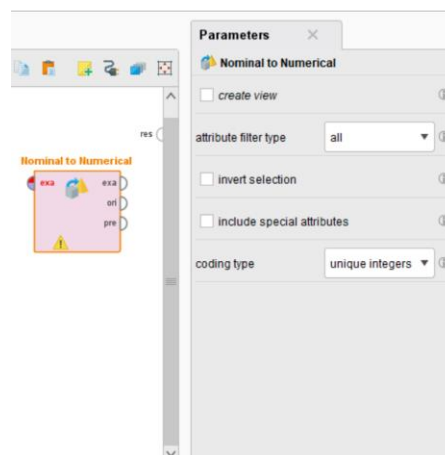




Gambar 5. Operator retrieve

### *Nominal to Numerical*

Operator *Nominal to Numerical* digunakan untuk mengubah jenis *attribute* yang non-numerik yakni pada nama kabupaten/kota serta pendidikan tertinggi yang akan diubah kedalam bentuk numerik agar proses data mining tidak terhambat, selain itu operator ini tidak hanya mengubah jenis atribut tetapi juga memetakan semua nilai *attribute* menjadi nilai numerik.



Gambar 6. Operator nominal to numerical

### *Normalize*

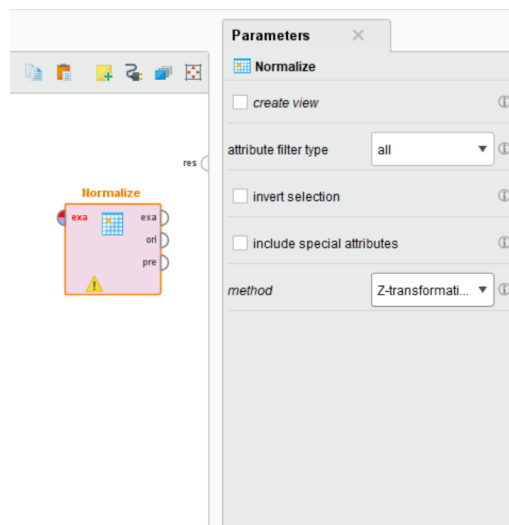
Operator ini menormalkan nilai attribute yang dipilih, normalisasi digunakan untuk menskalakan nilai sehingga menghasilkan rentang tertentu serta berguna untuk membandingkan attribute dengan jarak *Euclidian* semua attribut yang memiliki skala yang sama untuk perbandingan yang adil. Pada data pencari kerja ini dilakukan normalisasi menggunakan metode range *Z-transformation* karena merupakan standarisasi penggunaan *range* antar data.

\*penulis korespondensi



This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

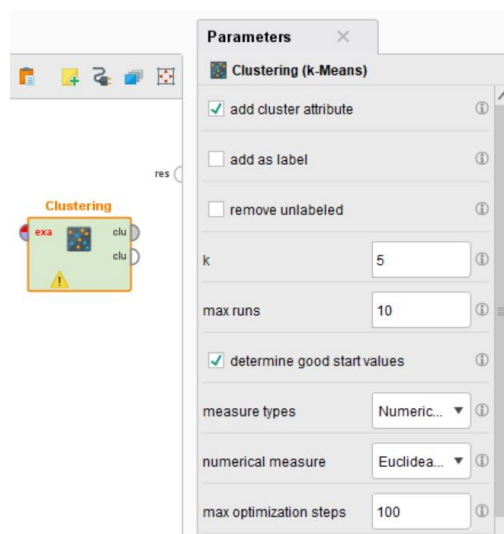




Gambar 7. Operator Normalize

### Clustering

Operator *clustering* merupakan operator utama dalam proses pengelompokan data pencari kerja, yakni melakukan *clustering* menggunakan Algoritma *K-Means*, pada data pencari kerja dilakukan *clustering* dengan menentukan cluster yakni k2 sampai dengan k10. Setelah melakukan *running* ssebanyak 10 kali kemudian tentukan cluster mana yang paling baik mendekati 0 agar mengetahui cluster mana yang paling akurat hasilnya.



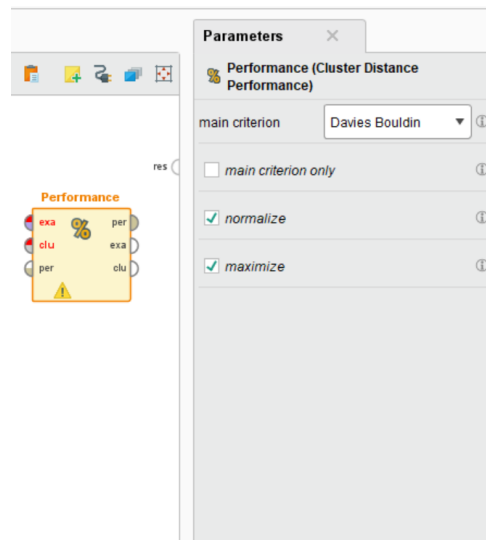
Gambar 8. Operator clustering(k-means)

### Performance

Operator ini digunakan untuk evaluasi kinerja metode clustering berbasis centroid yakni menggunakan Davies Bouldin Index (DBI), Davies Bouldin Index digunakan untuk mengukur kedekatan data antar data dalam satu kelompok data. Nilai maksimal Davies Bouldin Index yakni 1, semakin kecil nilai Davies Bouldin Index maka kemiripan data dalam satu kelompok semakin berdekatan. Pada data pencari kerja ini dilakukan performance menggunakan Davies Bouldin Index untuk mengetahui performance data pencari kerja yang paling dekat jaraknya.

\*penulis korespondensi





Gambar 9. Operator performance

Evaluasi

Tahap terakhir dari keseluruhan proses data mining menggunakan *tools RapidMiner* yakni evaluasi hasil *clustering* menggunakan *Davies Bouldin Index (DBI)*, sehingga muncul hasil pengelompokan yakni (*cluster*) terbaik yang didapat setelah melakukan *running* sampai dengan 10 kali pada operator *clustering* didapat DBI yang paling baik yakni dapat dilihat pada tabel dibawah ini.

Tabel 5 Hasil cluster DBI

Perulangan	Cluster	DBI
1	K2	0.326
2	K3	0.257
3	K4	0.212
4	K5	0.195
5	K6	0.185
6	K7	0.208
7	K8	0.221
8	K9	0.219
9	K10	0.213

Berdasarkan hasil dari implementasi Algoritma *K-Means* yang telah dilakukan pada data pencari kerja berdasarkan tingkat pendidikan untuk mengetahui *k optimum* dilakukan *running* sebanyak 10 kali sehingga didapat *k optimum* yakni berada pada (*k6*) dengan nilai DBI 0.185. Nilai DBI tersebut dapat dilihat pada gambar dibawah ini.

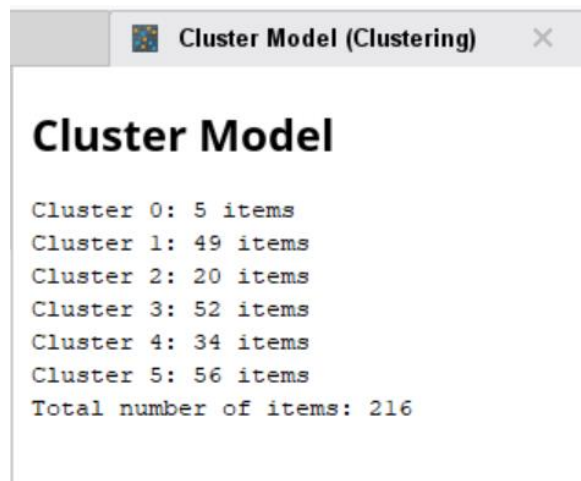
\*penulis korespondensi





Gambar 10. Nilai Davies Bouldin

Setelah mengetahui nilai *Davies Bouldin* dibawah ini merupakan gambar *cluster* model pengelompokan data pencari kerja berdasarkan tingkat pendidikan yakni sebagai berikut.



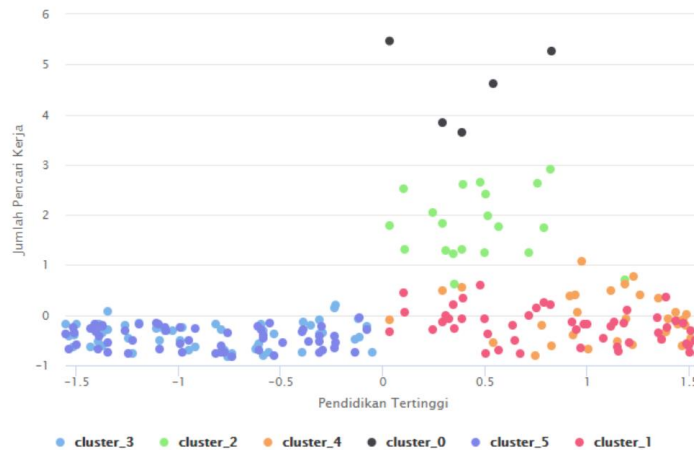
Gambar 11. Cluster model pengelompokan data pencari kerja berdasarkan tingkat pendidikan

Berdasarkan Cluster model diatas setelah melakukan running sebanyak 10 kali menggunakan perhitungan Davies Bouldin Index (DBI) didapat cluster ( $k_6$ ) dengan nilai DBI 0.185, *cluster* yang terbentuk berjumlah 6 yang terdiri dari *cluster* 0 (C0), *cluster* 1 (C1), *cluster* 2 (C2), *cluster* 3 (C3), *cluster* 4 (C4) dan *cluster* 5 (C5).

Setelah mengetahui hasil cluster menggunakan perhitungan DBI selanjutnya hasil klasterisasi data pencari kerja berdasarkan Tingkat Pendidikan disajikan menggunakan visualizations dalam RapidMiner dengan tipe plot Scatter/Bubble sebagai berikut.

\*penulis korespondensi





Gambar 12. Visualizations dengan tipe plot scatter/buuble pada data pencari kerja

Cluster\_0 yang berwarna hitam terdiri dari 5 anggota, Cluster\_1 yang berwarna pink terdiri dari 49 anggota, Cluster\_2 yang berwarna hijau terdiri dari 20 anggota, Cluster\_3 yang berwarna biru terdiri dari 52 anggota, Cluster\_4 yang berwarna jingga terdiri dari 34 anggota dan Cluster\_5 yang berwarna ungu terdiri dari 56 anggota. Cluster\_5 merupakan jumlah pencari kerja terbanyak yang terdiri dari 56 anggota, masing-masing anggota terdiri dari: DIPLOMA I/II/III/AKADEMI dengan jumlah pencari kerja terbanyak berjumlah 1062 orang, PROFESI dengan jumlah pencari kerja terbanyak berjumlah 20 orang, SD dengan jumlah pencari kerja terbanyak berjumlah 714 orang dan TIDAK/BELUM TAMAT SD dengan jumlah pencari kerja terbanyak berjumlah 2 orang. Sehingga dapat disimpulkan jumlah pencari kerja yang mendominasi yakni pencari kerja dengan tingkat pendidikan DIPLOMA I/II/III/AKADEMI yang berjumlah 1062 orang.

Berikut ini merupakan tabel hasil klasterisasi pencari kerja berdasarkan Tingkat Pendidikan.

Tabel 6 Hasil klasterisasi pencari kerja yang mendominasi pendidikan SMA

No.	Cluster	Nama Kabupaten/Kota	Pendidikan Tertinggi	Jumlah Pencari Kerja	Tahun
1.	Cluster_0	Kabupaten Cianjur	SMA	22819	2021
2.	Cluster_0	Kabupaten Cirebon	SMK	19133	2021
3.	Cluster_0	Kabupaten Purwakarta	SMA	16990	2021
4.	Cluster_0	Kabupaten Karawang	SMK	20476	2021
5.	Cluster_0	Kabupaten Bekasi	SMA	16333	2021

Tabel 7 Hasil klasterisasi pencari kerja yang mendominasi pendidikan UNIVERSITAS

No.	Cluster	Nama Kabupaten/Kota	Pendidikan Tertinggi	Jumlah Pencari Kerja	Tahun
1.	Cluster_1	Kabupaten Purwakarta	SMK	0	2021
2.	Cluster_1	Kabupaten Karawang	SMP	1409	2021
3.	Cluster_1	Kabupaten Karawang	UNIVERSITAS	2400	2021
4.	Cluster_1	Kabupaten Bekasi	SMP	446	2021
5.	Cluster_1	Kabupaten Bekasi	UNIVERSITAS	883	2021
....	....	....	....	....	....
48.	Cluster_1	Kota Banjar	SMP	393	2021
49.	Cluster_1	Kota Banjar	UNIVERSITAS	133	2021

Tabel 8 Hasil klasterisasi pencari kerja yang mendominasi pendidikan SMK

\*penulis korespondensi



No.	Cluster	Nama Kabupaten/Kota	Pendidikan Tertinggi	Jumlah Pencari Kerja	Tahun
1.	Cluster_2	Kabupaten Bogor	SMA	8298	2021
2.	Cluster_2	Kabupaten Sukabumi	SMA	8728	2021
3.	Cluster_2	Kabupaten Sukabumi	SMK	12057	2021
4.	Cluster_2	Kabupaten Garut	SMA	6622	2021
5.	Cluster_2	Kabupaten Garut	SMK	11036	2021
6.	Cluster_2	Kabupaten Tasikmalaya	SMK	7681	2021
....	....	....	....	....	....
19.	Cluster_2	Kabupaten Karawang	SMA	11576	2021
20.	Cluster_2	Kota Bekasi	SMK	9613	2021

Tabel 9 Hasil klasterisasi pencari kerja yang mendominasi pendidikan SD

No.	Cluster	Nama Kabupaten/Kota	Pendidikan Tertinggi	Jumlah Pencari Kerja	Tahun
1.	Cluster_3	Kabupaten Bogor	DIPLOMA I/II/III/AKADEMI	446	2021
2.	Cluster_3	Kabupaten Sukabumi	SD	511	2021
3.	Cluster_3	Kabupaten Sukabumi	PROFESI	3	2021
4.	Cluster_3	Kabupaten Cianjur	DIPLOMA I/II/III/AKADEMI	950	2021
5.	Cluster_3	Kabupaten Majalengka	SD	1828	2021
....	....	....	....	....	....
51.	Cluster_3	Kabupaten Subang	TIDAK/BELUM TAMAT SD	0	2021
52.	Cluster_3	Kabupaten Subang	SD	1504	2021

Tabel 10 Hasil klasterisasi pencari kerja yang mendominasi pendidikan SMP

No.	Cluster	Nama Kabupaten/Kota	Pendidikan Tertinggi	Jumlah Pencari Kerja	Tahun
1.	Cluster_4	Kabupaten Bogor	SMP	266	2021
2.	Cluster_4	Kabupaten Bogor	UNIVERSITAS	890	2021
3.	Cluster_4	Kabupaten Cianjur	SMP	5035	2021
4.	Cluster_4	Kabupaten Cianjur	UNIVERSITAS	1308	2021
5.	Cluster_4	Kabupaten Garut	SMP	1987	2021
....	....	....	....	....	....
33.	Cluster_4	Kabupaten Purwakarta	SMP	3355	2021
34.	Cluster_4	Kabupaten Purwakarta	UNIVERSITAS	792	2021

Tabel 11 Hasil klasterisasi pencari kerja yang mendominasi pendidikan DIPLOMA I/II/III/AKADEMI

No.	Cluster	Nama Kabupaten/Kota	Pendidikan Tertinggi	Jumlah Pencari Kerja	Tahun
1.	Cluster_5	Kabupaten Purwakarta	DIPLOMA I/II/III/AKADEMI	302	2021
2.	Cluster_5	Kabupaten Purwakarta	SD	714	2021
3.	Cluster_5	Kabupaten Bekasi	PROFESI	20	2021

\*penulis korespondensi



No.	Cluster	Nama Kabupaten/Kota	Pendidikan Tertinggi	Jumlah Pencari Kerja	Tahun
4.	Cluster_5	Kabupaten Bekasi	DIPLOMA I/II/III/AKADEMI	341	2021
5.	Cluster_5	Kota Bandung	DIPLOMA I/II/III/AKADEMI	1062	2021
....	....	....	....	....	....
55.	Cluster_5	Kota Cirebon	TIDAK/BELUM TAMAT SD	2	2021
56.	Cluster_5	Kota Banjar	SD	154	2021

### KESIMPULAN

Berdasarkan Hasil yang diperoleh dari Implementasi Algoritma *K-Means* pada data pencari kerja di wilayah Jawa Barat berdasarkan tingkat pendidikan menghasilkan 6 cluster (k6) dengan nilai DBI 0.185 diantaranya sebagai berikut: *Cluster\_0* terdiri dari 5 anggota, *Cluster\_1* terdiri dari 49 anggota, *Cluster\_2* terdiri dari 20 anggota, *Cluster\_3* terdiri dari 52 anggota, *Cluster\_4* terdiri dari 34 anggota dan *Cluster\_5* terdiri dari 56 anggota. *Cluster\_5* merupakan anggota terbanyak yakni berjumlah 56 anggota, masing-masing anggota terdiri dari : DIPLOMA I/II/III/AKADEMI dengan jumlah pencari kerja terbanyak berjumlah 1062 orang, PROFESI dengan jumlah pencari kerja terbanyak berjumlah 20 orang, SD dengan jumlah pencari kerja terbanyak berjumlah 714 orang dan TIDAK/BELUM TAMAT SD dengan jumlah pencari kerja terbanyak berjumlah 2 orang. Sehingga dapat disimpulkan pencari kerja yang mendominasi yakni pencari kerja dengan tingkat pendidikan DIPLOMA I/II/III/AKADEMI yang berjumlah 1062 orang. Berdasarkan pengelompokan data pencari kerja yang telah dilakukan dalam menentukan *k optimum* yang terbaik menghasilkan *cluster* (k6) dengan nilai DBI 0.185, sehingga semakin kecil nilai *Davies Bouldin Index* (DBI) maka kemiripan data dalam satu kelompok semakin berdekatan. Beberapa saran yang dapat dilakukan untuk Tugas Akhir selanjutnya yakni , untuk Tugas Akhir selanjutnya dapat menggunakan Algoritma lain seperti *K-Medoids* atau *Fuzzy C-Means* serta pada Tugas Akhir berikutnya untuk menambahkan datanya lebih banyak lagi.

### REFERENSI

- Aditya, A., Jovian, I., & Sari, B. N. (2020). Implementasi K-Means Clustering Ujian Nasional Sekolah Menengah Pertama di Indonesia Tahun 2018/2019. *Jurnal Media Informatika Budidarma*, 4(1), 51. <https://doi.org/10.30865/mib.v4i1.1784>
- Harahap, T. R., & Nawawi, Z. M. (2022). Pelayanan Permasalahan Dan Penempatan Tenaga Kerja Pada Dinas Tenaga Kerja (Disnaker) Kota Medan. *Balance Jurnal Akuntansi Dan ...*, 1(1), 96–109. <https://jurnal.risetilmiah.ac.id/index.php/jam/article/view/2%0Ahttps://jurnal.risetilmiah.ac.id/index.php/jam/article/download/2/2>
- Ningsih, W., & Abdullah, F. (2021). Analisis Perbedaan Pencari Kerja dan Lowongan Kerja Sebelum dan Pada Saat Pandemi Covid-19 di Kota Malang. *Journal of Regional Economics Indonesia*, 2(1), 42–56. <https://doi.org/10.26905/jrei.v2i1.6181>
- Nurchayho, N. (2021). Perlindungan hukum tenaga kerja berdasarkan peraturan perundang-undangan di Indonesia. *Jurnal Cakrawala Hukum*, 12(1), 69–78. <https://doi.org/10.26905/idjch.v12i1.5781>
- Pa, P., Pardede, A. M. H., & Rahmadani, S. (2022). Pengelompokan Data Pencari Kerja Terdaftar Berdasarkan Umur Dan Pendidikan Menggunakan Metode K-means Clustering Di Dinas Tenaga Kerja dan Perindustrian Perdagangan Kota Binjai. *Jurnal Informatika Kaputama (JIK)*, 6(3).
- Utami, farathika putri. (2020). Pengaruh Indeks Pembangunan Manusia (IPM), Kemiskinan dan Pengangguran Terhadap Pertumbuhan Ekonomi di Provinsi Aceh. *Jurnal Samudra Ekonomika*, 4(2), 101–113. <https://ejournalunsam.id/index.php/jse/article/view/2303>
- Wijaya, H. D., & Dwiasnati, S. (2020). Implementasi Data Mining dengan Algoritma Naïve Bayes pada Penjualan Obat. *Jurnal Informatika*, 7(1), 1–7. <https://doi.org/10.31311/ji.v7i1.6203>

\*penulis korespondensi



This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.