



Penerapan *Information Gain* dan Algoritma *K-Means* Untuk Klusterisasi Kedisiplinan Pegawai Menggunakan *Rapidminer*

Zulkarnaen Noor Syarif*

Teknologi Sistem Informasi, Ilmu Komputer
Universitas Budi Luhur

Email: zulkarnaen.ns@gmail.com

*) *Corresponding Author*

ABSTRACT

One aspect of the discipline of an employee in an agency can be seen from the side of attendance. The level of employee attendance is closely related to an employee's disciplinary assessment. The level of employee discipline can be seen by looking at the hours of attendance or check in attendance, so that with these parameters you will get early, on time and late entry. This study explores data on attendance by using the k-means clustering algorithm. Before calculating the k-means clustering algorithm, attribute selection using information gain is expected to reduce attributes with small weights. The calculations are performed using Rapidminer software. The results showed that the attribute that had the greatest influence was the percentage of late entry with a weight of 0.783. Clustering using the k-means algorithm produces three clusters with the performance value of the Davies Bouldin Index (DBI) -0.645. Cluster zero has fifteen members, cluster one has thirty-six members, and cluster two has fifty-two members. Cluster zero is a cluster that has a low level of discipline, cluster one is a cluster that has a high level of discipline, while cluster two is a cluster that has a moderate level of discipline.

Keywords: *information gain, discipline, rapidminer, k-means, davies bouldin index (DBI).*

ABSTRAK

Salah satu aspek disiplin seorang pegawai disebuah instansi dapat dilihat dari sisi presensi kehadiran. Tingkat kehadiran pegawai sangat erat kaitannya penilaian disiplin seorang pegawai. Tingkat disiplin pegawai dapat dilihat dengan melihat jam datang atau checkin absensi, sehingga dengan parameter tersebut akan didapatkan masuk lebih awal, tepat waktu dan keterlambatan. Penelitian ini menggali data mengenai absensi kehadiran dengan menggunakan algoritma klusterisasi k-means. Sebelum proses perhitungan algoritma klusterisasi k-means terlebih dahulu seleksi atribut dengan menggunakan information gain yang diharapkan dapat mereduksi atribut yang memiliki bobot kecil. Perhitungan dilakukan dengan menggunakan perangkat lunak Rapidminer. Hasil penelitian menunjukkan atribut yang memiliki pengaruh terbesar adalah persentase jumlah masuk terlambat dengan bobot 0.783. Klusterisasi dengan algoritma k-means menghasilkan tiga kluster dengan nilai performa Davies Bouldin Index (DBI) -0.645. Klaster nol memiliki lima belas anggota, klaster satu memiliki tiga puluh enam anggota, dan klaster dua memiliki lima puluh dua

anggota. Klaster nol merupakan klaster yang memiliki tingkat disiplin rendah, klaster satu adalah klaster yang memiliki tingkat disiplin tinggi, sedangkan klaster dua merupakan klaster yang memiliki tingkat disiplin sedang.

Keywords: *information gain, disiplin, rapidminer, k-means, davies bouldin index (DBI).*

A. PENDAHULUAN

1. Latar Belakang

Setiap lembaga menginginkan pegawainya dapat mengoptimalkan dalam bekerja karena hal ini dapat mempengaruhi kinerja seorang pegawai. Salah satu kriteria yang digunakan untuk mengukur tingkat kedisiplinan seorang pegawai dapat dilihat dari sisi presensi kehadiran. Presensi merupakan data yang berisikan tingkat kehadiran pegawai dalam sebuah instansi. Kehadiran pegawai sangat mempengaruhi kedisiplinan pegawai saat bekerja. Pegawai yang memiliki tingkat disiplin yang tinggi adalah pegawai yang selalu datang tepat waktu, tidak banyak alpa, dan tidak sering izin saat masuk kerja. Data *mining* digunakan untuk mencari pola-pola yang dapat dipahami manusia yang menjelaskan, dan data *mining* telah memberikan banyak produk yang berguna bagi masyarakat luas. Data *mining* bukanlah suatu bidang yang sama sekali baru, akan tetapi salah satu kesulitan dalam mendefinisikan data *mining* adalah kenyataan bahwa data *mining* mewarisi banyak aspek dan teknik dari bidang-bidang ilmu yang sudah mapan terlebih dahulu. Dalam mengelompokkan presensi pegawai dapat menggunakan algoritma klasterisasi K-Means. Dengan menggunakan klasterisasi dapat mengelompokkan pegawai berdasarkan tingkat kedisiplinannya berdasarkan presensi kehadiran pegawai.

2. Permasalahan

Banyaknya jumlah transaksi presensi pegawai yang dihasilkan menyebabkan kesulitan personil kepegawaian dalam mengelola dan mengklaster tingkat kedisiplinan pegawai.

3. Tujuan

Penelitian ini bertujuan untuk mengetahui penerapan seleksi fitur dan normalisasi data untuk mencari model klasterisasi kedisiplinan pegawai berdasarkan presensi kehadiran dengan menggunakan *software* Rapidminer.

4. Tinjauan Pustaka

a. Information Gain

Information gain adalah salah satu metode dalam penyeleksian fitur yang sering digunakan oleh peneliti dalam menentukan batas dari kepentingan suatu atribut. Untuk memperoleh nilai *information gain* yaitu dari nilai *entropy* sebelum pemisahan dikurangi dengan nilai *entropy* setelah pemisahan. Didalam pengukuran nilai ini hanya digunakan sebagai tahap awal dalam menentukan atribut yang akhirnya akan digunakan dalam proses pengklasifikasian suatu algoritma. Terdapat tiga tahapan dalam pemilihan fitur dengan *information gain* yaitu: a) melakukan perhitungan nilai *information gain* untuk setiap atribut dalam dataset yang original; b) menentukan batas (*threshold*) yang diinginkan. Dengan menentukan batasan ini akan memungkinkan atribut yang berbobot sama dengan batas atau lebih besar yang nantinya akan dipertahankan serta membuang atribut yang berada dibawah batas; 3) melakukan perbaikan Data set dengan menggunakan pengurangan atribut.

b. Seleksi Fitur

Seleksi fitur yaitu suatu proses dalam memilih sejumlah fitur yang termasuk *subset* dari fitur yang terdahulu atau fitur aslinya sehingga fitur-fitur yang diperoleh adalah fitur-fitur yang berpengaruh (signifikan) terhadap akurasi klasifikasi. Dalam proses seleksi fitur dapat mengaruhi jumlah fitur dan *noise* yang dinilai kurang relevan dan berlebihan [2], sehingga diharapkan dapat meningkatkan keakuratan. Salah satu strategi yang dipakai dalam melakukan pengurangan dimensi terhadap fitur-fitur yang digunakan dalam proses data *mining* merupakan pengertian dari proses seleksi fitur.

c. Klasterisasi K-Means

Data *mining* adalah suatu proses dalam menemukan keterkaitan yang berarti, pola dan kecenderungan dengan memeriksa dalam sekumpulan besar data yang tersimpan didalam penyimpanan dengan menggunakan teknik pengenalan berupa pola seperti teknik matematika dan statistik.

Klasterisasi merupakan suatu metode dalam menganalisa data, yang biasa diinputkan sebagai salah satu metode data *mining*, yang bertujuan untuk mengelompokkan data dengan karakteristik dan kesuatu wilayah yang sama dan data dengan karakteristik kewilayan lain yang berbeda. Klasterisasi dengan pendekatan partisi dan klastering dengan pendekatan hirarki merupakan metode yang digunakan dalam mengembangkan metode klusterling. Klusterling dengan pendekatan partisi merupakan suatu pengelompokan data dengan cara memilah data yang dianalisa kedalam kluster-kluster yang ada. Algoritma k-means digunakan untuk menghasilkan karakteristik kelompok dan memiliki akurasi cukup tinggi. Metode K-Means adalah salah satu metode yang digunakan dalam melakukan klusterling dengan partisi. Metode K-Means ini secara umum melakukan proses pengelompokkan dengan beberapa tahapan yaitu:

- a. Menentukan jumlah kluster
- b. Mengalokasikan data secara random ke kluster yang ada
- c. Menghitung rata-rata setiap kluster dari data yang tergabung didalamnya
- d. Semua data dialokasikan Kembali ke kluster terdekat
- e. Proses tahap c diulang kembali sehingga tidak ada perubahan atau perubahan yang terjadi masalah sudah berada dibawah *threshold* atau sama.

d. Davies-bouldin index

Davies-bouldin index adalah salah satu metode dalam mengevaluasi internal yang mengukur pengevaluasian cluster dalam suatu metode pengelompokkan yang didasari pada nilai kohesi dan sparasi. Jumlah dari kedekatan data terhadap centroid dari kluster yang diikuti merupakan definisi dari kohesi.

e. Rapidminer

Rapidminer yaitu suatu *tool* didalam mengoperasikan data mining yang digunakan dalam menganalisis informasi. Rapidminer digunakan didalam penelitian, Pendidikan, *rapid prototyping*, pengembangan aplikasi dan aplikasi dalam bidang industri. Pada tahun 2001 proyek *rapidminer* dimulai dan proyek tersebut dilakukan oleh Ralf Klinkenberg, Simon Fischer dan Ingo Mierswa pada *Artificial Intelligence Group* of Katharina Morik yang berada di Dortmund University Of Technology. Data *Cleaning*, data *transformation*, *Oftimization*, *validation and visualization* merupakan aplikasi yang open source. Adapun untuk melihat visualisasinya yaitu bisa dalam bentuk *bar*, *scatter plot*, *Pie Chart* dan lain-lain [6]. Rapidminer memiliki operator data *mining* kurang lebih 500 operator, didalamnya sudah termasuk operator input, output, data *pre processing* dan visualisasi.

B. METODE

1. Pengumpulan Data

Langkah awal pada penelitian ini adalah pengumpulan data, data yang digunakan adalah data absensi tahun 2020 serta data masa kerja dan pensiun diambil dari instansi SEAMEO BIOTROP yang diperoleh secara terpisah. *Dataset* absensi diperoleh berupa format yang berekstensi sql hasil *generate* dari *database server* MySQL yang memiliki 6 (enam) tabel. Sedangkan dataset masa kerja dan pensiun diperoleh dengan format extension xlsx yang terdiri dari 27 (dua puluh tujuh) atribut. Pada tahap berikutnya semua atribut akan dihitung dengan pendekatan *information gain* untuk memperoleh kepentingan dari atribut terhadap klusterisasi. Selanjutnya setelah mengetahui nilai bobot dari semua atribut terhadap hasil klusterisasi maka atribut yang tidak perlu dapat direduksi dari *dataset* kemudian *dataset* diperbaharui.

Data absen merupakan data absen selama tahun 2020 yang terdiri dari 6 tabel, terdiri dari activities, migrations, password_resets, users, users2, dan votes. Dari 6 tabel tersebut yang digunakan hanya 2 (dua) yaitu activities yang memiliki dan users2.

Dari kedua sumber dataset tersebut atribut diseleksi terlebih dahulu secara manual, kemudian menggunakan query join tabel untuk mendapatkan atribut yang diinginkan. Untuk memperkaya atribut menggunakan formula spreadsheet dalam mendapatkan atribut label masuk. Sehingga diperoleh 16 atribut yang terdiri dari nik, jenis kelamin, status, departemen, jenjang pendidikan, tahun masuk, usia, masa kerja, tahun pensiun, jumlah masuk lebih awal, jumlah masuk tepat waktu, % masuk lebih awal, % masuk tepat waktu, % jumlah masuk terlambat, label masuk. Secara lebih terperinci spesifikasi dari dataset absensi pegawai dapat dilihat pada tabel 1 berikut:

Tabel 1: Meta data absensi pegawai

No.	Atribut	Type	Keterangan
1.	nik	Polynomial (id)	-
2.	jenis_kelamin	Binominal	L, P
3.	status	Polynomial	PNS, NON PNS
4.	departemen	Polynomial	
5.	jenjang_pendidikan	Polynomial	S3, S2, S1, D3, D2, D1, SLTA, SLTP, SD
6.	tahun_masuk	Numerik	-
7.	usia	Numerik	-
8.	masa_kerja	Numerik	-
9.	tahun_pensiun	Numerik	-
10.	jumlah_masuk_lebih_awal	Numerik	Checkin lebih awal 45 menit
11.	jumlah_masuk_tepat_waktu	Numerik	Checkin diantara jam 07:31 sd 08:15
12.	jumlah_masuk_terlambat	Numerik	Checkin lebih dari jam 08.15
13.	%_masuk_lebih_awal	Numerik	-
14.	%_masuk_tepat_waktu	Numerik	-
15.	%_jumlah_masuk_terlambat	Numerik	-
16.	Label Masuk	Polynomial (label)	Sangat Disiplin, Disiplin, Cukup Disiplin, Kurang Disiplin, Tidak Disiplin

Atribut label masuk diperoleh dari ketentuan pada tabel 2 berikut:

Tabel 2: Persentase keterlambatan

Range % keterlambatan	Label Masuk
0% - 20%	Sangat Disiplin
21% - 40%	Disiplin
41% - 60%	Cukup Disiplin
61% - 80%	Kurang Disiplin
81% - 100%	Tidak Disiplin

2. Eksperimen

Pada tahap eksperimen dilakukan dengan menggunakan perangkat lunak Rapidminer. Dalam proses seleksi fitur dilakukan dengan menggunakan fasilitas yang sudah tersedia dalam perangkat lunak Rapidminer. Sehingga pada tahapan eksperimen ini tidak dilakukan pengkodean ataupun kegiatan membuat pengkodean dalam bahasa pemrograman tertentu untuk menghitung algoritma. Dikarenakan dalam perangkat lunak Rapidminer sudah tersedia banyak algoritma data mining dan cukup *drag and drop* dalam penggunaannya. Pada tahapan ini dilakukan perhitungan *dataset* dengan menggunakan pendekatan *information gain*. Dalam proses ini diharapkan mendapatkan hasil berupa bobot untuk setiap atribut dalam dataset tersebut. Selanjutnya atribut yang memiliki nilai kecil atau dapat diterjemahkan tidak memiliki pengaruh besar dapat direduksi dari *dataset* untuk memperoleh hasil klasterisasi.

Setelah atribut diperoleh dengan mereduksi atribut-atribut yang memiliki bobot lebih kecil dari 0.01 kemudian proses *binning* atau menormalisasikan data-data berikut untuk kebutuhan algoritma klusterisasi K-means:

- a. Normalisasi data jenjang pendidikan seperti pada tabel 3 berikut

Tabel 3. Normalisasi data jenjang pendidikan

Jenjang Pendidikan	Nilai
SD	1
SLTP	2
SLTA	3
D1	4
D2	5
D3	6
S1	7
S2	8
S3	9

- b. Normalisasi data departemen seperti pada tabel 4 berikut

Tabel 4. Normalisasi data departemen

Departemen	Nilai
CBD	1
Cleaning Service	2
FAD	3
FMD	4
GAD	5
HRMD	6
KMD	7
PPC	8
ReD	9
SLD	10

- c. Normalisasi data label masuk seperti pada tabel 5 berikut

Tabel 5. Normalisasi data label masuk

Label Masuk	Nilai
Sangat Disiplin	5
Disiplin	4
Cukup Disiplin	3
Kurang Disiplin	2
Tidak Disiplin	1

Setelah mendapatkan dataset yang telah dinormalisasi langkah selanjutnya dilakukan perhitungan dengan menggunakan algoritma klusterisasi K-means dan performa *Davies Bouldin Index* (DBI) untuk mendapatkan hasil kluster yang terbaik.

3. Evaluasi

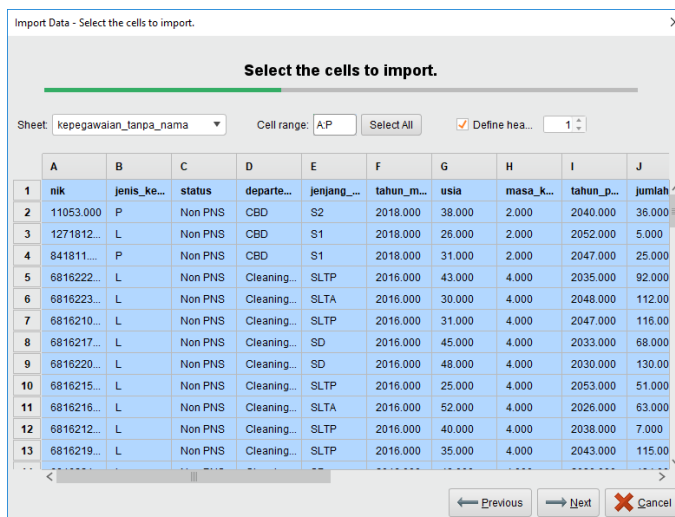
Setelah semua bobot atribut diperoleh dan diketahui dengan pendekatan seleksi fitur *information gain* maka tahapan berikutnya adalah mengevaluasi dengan cara membandingkan hasil dengan dataset sebelumnya (16 atribut) dan dengan menggunakan *dataset* yang telah terpilih. Evaluasi ini dilakukan hanya untuk memastikan bahwa usulan atribut yang digunakan atau pengurangan atribut memiliki pengaruh terhadap hasil klusterisasi. Selanjutnya dilakukan perhitungan algoritma kluster K-means dan performa *Davies Bouldin Index* untuk memperoleh kluster yang terbaik. Pola informasi yang dihasilkan dari proses tersebut ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan.

C. HASIL DAN PEMBAHASAN

1. HASIL

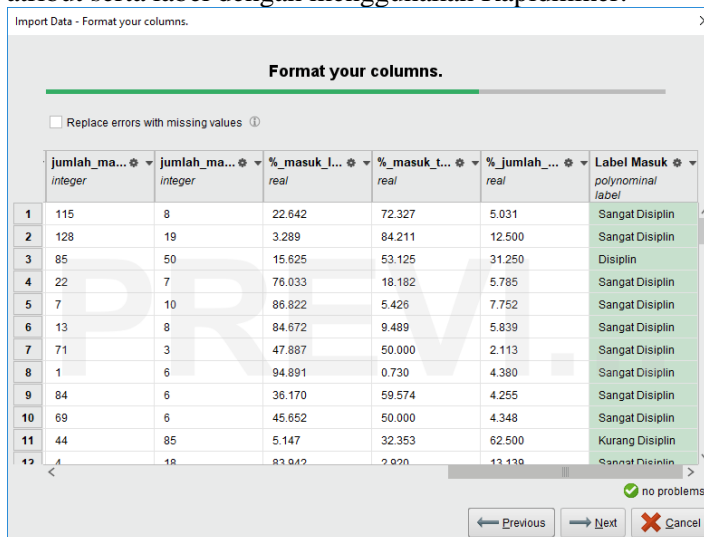
a. Hasil Seleksi Fitur dengan Rapidminer

Untuk proses seleksi fitur dengan menggunakan perangkat lunak rapidminer dapat dilakukan dengan beberapa langkah. Langkah berikut secara rinci akan dijelaskan dalam sub bab berikut. Pertama buka terlebih dahulu Rapidminer kemudian import data file spreadsheet yang ada. File *spreadsheet* dipilih karena dataset yang ada bertipe file *xlsx*.



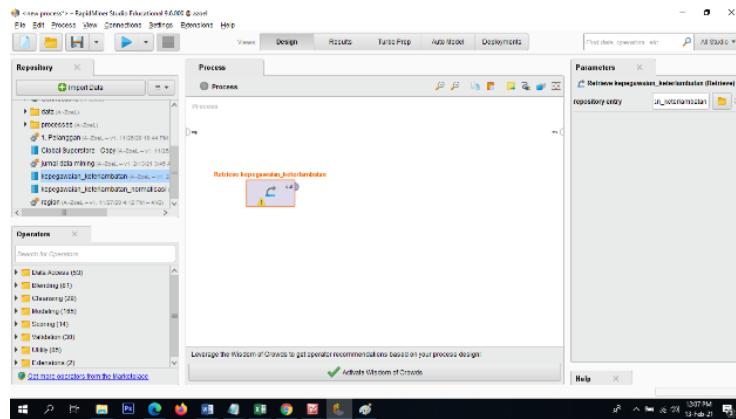
Gambar 1. Proses *import* data dari *spreadsheet*

Langkah selanjutnya konfigurasi import dari *dataset* yang akan dimasukkan atribut beserta tipenya dan pemilihan label atau atribut tujuan yaitu label disiplin. Gambar 2 merupakan pemilihan tipe atribut serta label dengan menggunakan Rapidminer.



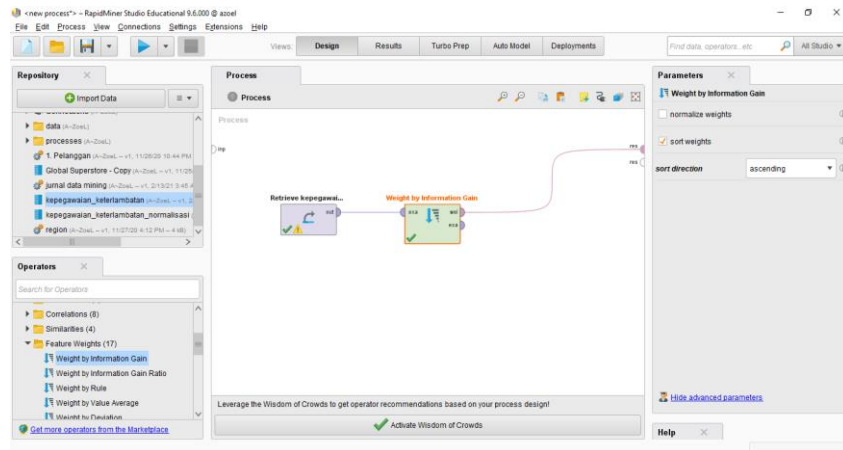
Gambar 2. Pemilihan jenis dan tipe atribut

Setelah import berhasil dilakukan ke dalam Rapidminer masukkan *dataset* tersebut pada bagian *repository* ke bagian proses dengan cara *klik and drag*, seperti gambar 3.



Gambar 3. Dataset dimasukkan ke bagian proses

Setelah *dataset* dimasukkan ke bagian proses gunakan *information gain* untuk menghitung bobot dari keseluruhan atribut yang ada. *Information gain* dapat digunakan dengan cara *drag and drop* algoritma dari operator *modelling* dan *atribut feature weights* kemudian pilih *weight by information gain*. Alternatif lain dapat dilakukan dengan cara mencari menggunakan tombol *search* pada bagian atas operator.



Gambar 4. Konfigurasi algoritma *information gain* pada Rapidminer

Setelah tersusun sempurna maka saat dijalankan akan muncul hasil seperti gambar 5.

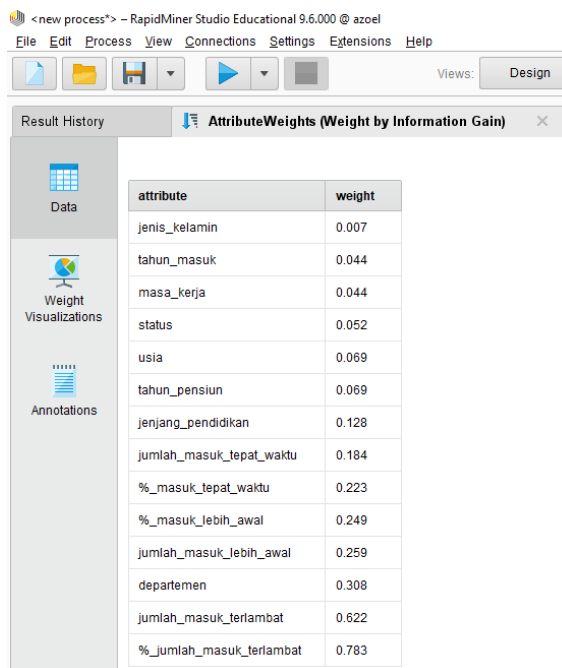
Berikut adalah tabel 6 merupakan hasil perhitungan nilai *information gain* seluruh dataset disiplin pegawai. Tampak bahwa *jenis_kelamin*, *tahun_masuk*, *masa_kerja*, *status*, *usia*, dan *tahun_pensiun* memiliki nilai dibawah 0.1. Dan nilai gain tertinggi yaitu atribut *%_jumlah_masuk_terlambat* yang memiliki nilai gain sebesar 0.783.

Tabel 6. Hasil perhitungan *information gain*

attribute	weight
jenis_kelamin	0.007
tahun_masuk	0.044
masa_kerja	0.044
status	0.054
usia	0.069
tahun_pensiun	0.069
jenjang_pendidikan	0.128
jumlah_masuk_tepat_waktu	0.184
%_masuk_tepat_waktu	0.223
%_masuk_lebih_awal	0.249

jumlah_masuk_lebih_awal	0.259
departemen	0.308
jumlah_masuk_terlambat	0.622
%_jumlah_masuk_terlambat	0.783

Dari tabel tersebut diatas atribut yang memiliki bobot dibawah 0.1 akan direduksi atau dihilangkan untuk dilakukan proses perhitungan algoritma klasterisasi kmeans.



Gambar 5. Seleksi hasil fitur

b. Proses *Binning*

Proses Binning disebut juga proses normalisasi, yaitu proses dalam mentransformasi dari sebuah nilai dari data non-numerik menjadi data yang dapat dikalkulasi [9]. Proses transformasi nilai data ini sangat diperlukan dalam pengimplementasian algoritma k-mean untuk proses clustering sebuah data. Dataset absen pegawai mengalami proses *binning*, dan nilainya berubah menjadi nilai desimal sesuai dengan jumlah nilai yang dimiliki setiap atribut. Aturan untuk proses binning dalam penelitian ini adalah sebagai berikut:

1. Untuk atribut yang memiliki dua nilai, setiap nilai di-set dengan desimal 1 atau 0 sesuai dengan nilai atribut.
2. Untuk atribut dengan lebih dari dua nilai, nilai dari atribut itu akan bernilai increment atau menaik dengan nilai sebelumnya sesuai dengan jumlah nilai dalam atribut tersebut, nilai dimulai dari angka 1.

Pada *dataset* absen pegawai terdapat proses *binning* dilakukan dengan menggunakan fungsi formula yang tersedia pada aplikasi *spreadsheet*, adapun atribut yang dilakukan proses *binning* terdiri dari:

1. Atribut data jenjang pendidikan
 Atribut dari nilai jenjang pendidikan memiliki lebih dari 1 sehingga dalam menormalisasi data untuk atribut ini menggunakan nilai *increment*. Sehingga atribut jenjang pendidikan yang memiliki sembilan nilai yang terdiri dari SD, SLTP, SLTA, D1, D2, D3, S1, S2, S3 maka menjadi 1, 2, 3, 4, 5, 6, 7, 8, 9.
2. Atribut data departemen
 Atribut dari nilai departemen memiliki lebih dari satu sehingga dalam menormalisasi data untuk atribut ini menggunakan nilai *increment* atau menaik. Sehingga atribut jenjang pendidikan yang memiliki sembilan nilai yang terdiri dari CBD, *Cleaning*

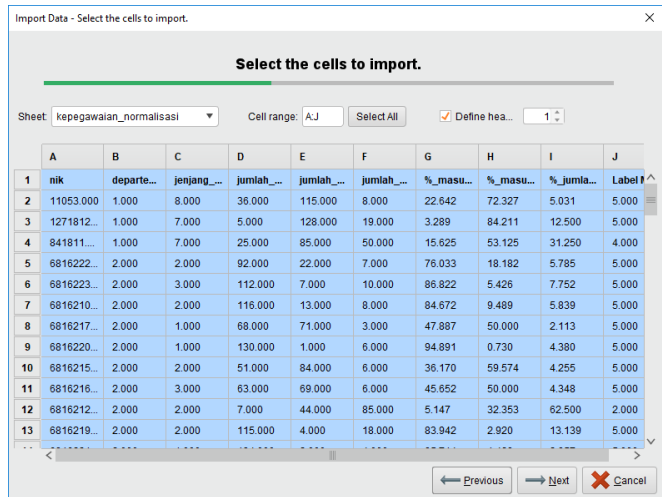
Service, FAD, FMD, GAD, HRMD, KMD, PPC, ReD, SLD maka nilai dari atribut departemen menjadi 1, 2, 3, 4, 5, 6, 7, 8, 9, 10.

3. Atribut data label masuk

Menormalisasi data untuk atribut ini menggunakan nilai *decreased* atau menurun. Sehingga atribut label masuk yang memiliki lima nilai yang terdiri dari Sangat Disiplin, Disiplin, Cukup Disiplin, Kurang Disiplin, Tidak Disiplin maka menjadi 5, 4, 3, 2, 1.

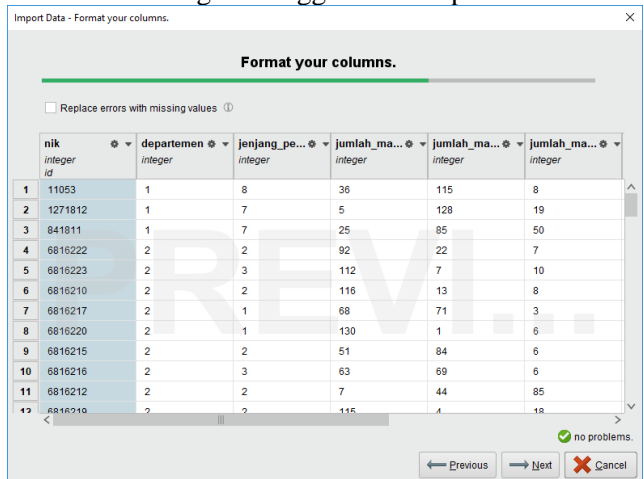
c. **Proses Klasterisasi Algoritma K-Means dengan Rapidminer**

Untuk proses klasterisasi algoritma k-means dengan menggunakan perangkat lunak Rapidminer dapat dilakukan dengan beberapa langkah. Buka terlebih dahulu Rapidminer kemudian *import* data file *spreadsheet* yang sudah direduksi atributnya dan dinormalisasi.



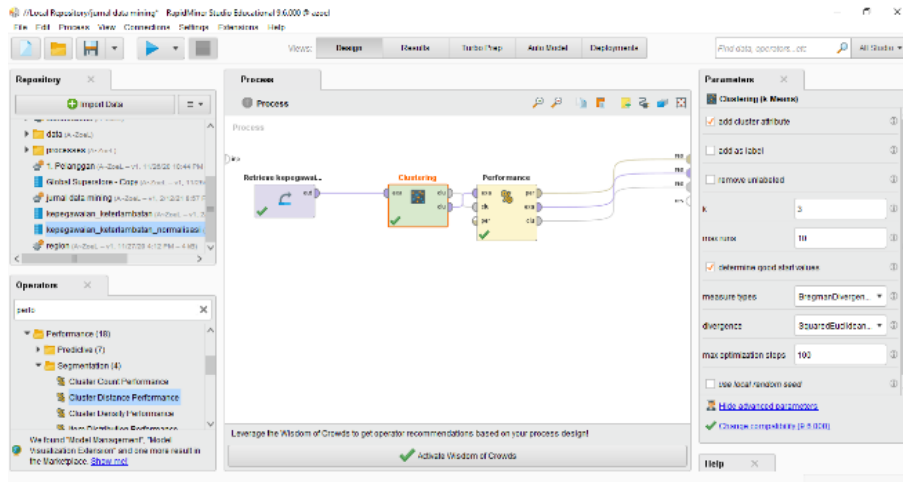
Gambar 6. Import file yang sudah dinormalisasi dan reduksi

Langkah selanjutnya konfigurasi import dari dataset yang akan dimasukkan atribut beserta tipenya dan pemilihan label atau atribut tujuan yaitu label disiplin. Gambar 7 merupakan pemilihan tipe atribut serta label dengan menggunakan Rapidminer.



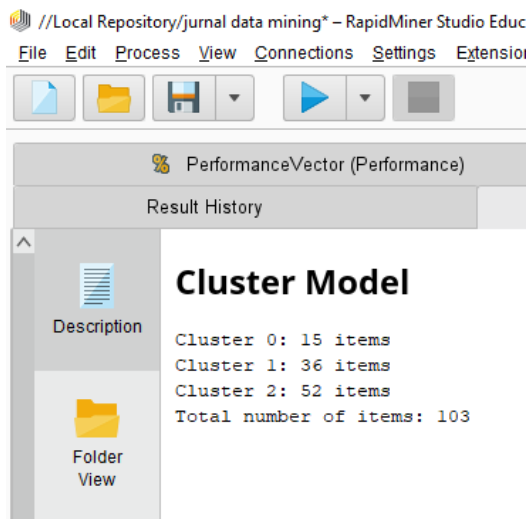
Gambar 7. Pemilihan jenis dan tipe atribut ternormalisasi

Setelah import berhasil dilakukan ke dalam rapidminer masukkan dataset tersebut pada bagian repository ke bagian proses dengan cara *click and drag* serta atur susunan algoritma seperti gambar 8.



Gambar 8. Susunan Algoritma K-means dan DBI

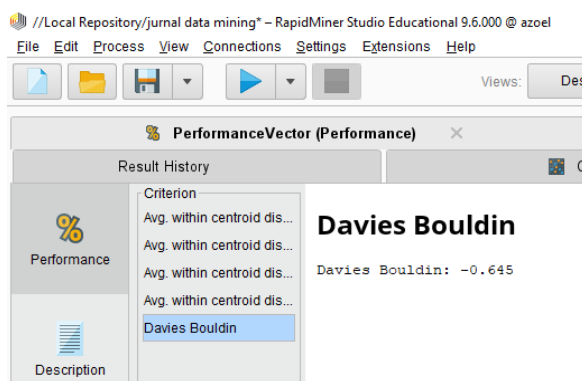
Setelah tersusun sempurna maka saat dijalankan akan muncul hasil seperti gambar 9 berikut.



Gambar 9. Hasil *cluster* algoritma k-means

Pada gambar 9 menghasilkan 3 kluster, yang terdiri dari kluster nol terdiri dari lima belas anggota anggota, kluster satu terdiri dari 36 anggota, dan kluster dua terdiri 52 anggota.

Untuk melihat hasil performa DBI pilih tab Performance Vector (Performance), maka akan tampil seperti gambar 10.



Gambar 10. Performa Algoritma K-Mean

2. PEMBAHASAN

Setelah pemodelan dilakukan di Rapidminer dengan baik dan benar tahap selanjutnya adalah mengevaluasi hasil keluaran. Setelah dilakukan percobaan dengan beberapa kluster maka hasil kluster terbaik yang diperoleh adalah tiga kluster dengan nilai DBI sebesar -0.645. Berikut adalah beberapa percobaan yang telah dilakukan.

Table 7: Nilai DBI pengujian kluster

K	Nilai DBI
2	-0.682
3	-0.645
4	-0.760
5	-0.740

Centroid untuk masing-masing atribut untuk masing-masing kluster dapat dilihat pada tabel 8 berikut:

Tabel 8. Centroid Setiap Kluster

Attribute	cluster_0	cluster_1	cluster_2
departemen	7.133	5.111	6.346
jenjang_pendidikan	5.800	4.083	5.423
jumlah_masuk_lebih_awal	12	109.694	30.481
jumlah_masuk_tepat_waktu	41.867	28.667	106.731
jumlah_masuk_terlambat	84.467	8.028	18.269
%_masuk_lebih_awal	8.494	75.532	19.938
%_masuk_tepat_waktu	29.356	18.972	68.537
%_jumlah_masuk_terlambat	62.150	5.496	11.525
Label Masuk	2.333	5	4.827

Visualisasi pada gambar 11 menunjukkan bahwa kluster nol merupakan kluster yang memiliki tingkat disiplin rendah, kluster satu adalah kluster yang memiliki tingkat disiplin tinggi, sedangkan kluster dua merupakan kluster yang memiliki tingkat disiplin sedang. Sedangkan pada visualisasi pada gambar 12 menunjukkan bahwa kluster nol merupakan kluster yang memiliki tingkat absen masuk lebih awal paling tinggi adalah 45 kali, sedangkan kluster satu absen masuk lebih awal diantara 71 sampai dengan 145 kali. Dan kluster dua absen masuk awal paling tinggi adalah 68 kali.

D. KESIMPULAN

Berdasarkan hasil penelitian, dapat disimpulkan bahwa *information gain* dapat digunakan untuk menentukan pemilihan atribut dalam klusterisasi. Atribut yang memiliki pengaruh terbesar adalah %_jumlah_masuk_terlambat dengan bobot 0.783. Sementara atribut yang memiliki bobot dibawah 0.1 adalah jenis_kelamin (bobot 0.007), tahun_masuk (bobot 0.044), masa_kerja (bobot 0.044), status (bobot 0.054), usia (bobot 0.069), dan tahun_pensiun (bobot 0.069). Setelah fitur tersebut digunakan dan dihitung dengan menggunakan algoritma klusterisasi K-means menghasilkan tiga kluster dengan nilai performa Davies Bouldin Index (DBI) -0.645. Kluster nol memiliki lima belas anggota, kluster satu memiliki tiga puluh enam anggota, dan kluster dua memiliki lima puluh dua anggota. Kluster nol merupakan kluster yang memiliki tingkat disiplin rendah, kluster satu adalah kluster yang memiliki tingkat disiplin tinggi, sedangkan kluster dua merupakan kluster yang memiliki tingkat disiplin sedang.

E. DAFTAR PUSTAKA

- [1] A. F. Muhammad, "Klasterisasi Proses Seleksi Pemain Menggunakan Algoritma K-Means (Study Kasus : Tim Hockey Kabupaten Kendal)," *Jur. Tek. Inform. FIK UDINUS*, pp. 1–5, 2015.
- [2] D. H. Murti, N. Suciati, and D. J. Nanjaya, "Clustering Data Non-Numerik Dengan Pendekatan Algoritma K-Means Dan Hamming Distance Studi Kasus Biro Jodoh," *JUTI J. Ilm. Teknol. Inf.*, vol. 4, no. 1, p. 46, 2005, doi: 10.12962/j24068535.v4i1.a245.
- [3] Dio Prima Mulya, "Analisa Dan Implementasi Association Rule Dengan Algoritma Fp-Growth Dalam Seleksi Pembelian Tanah Liat," *Teknol. dan Sist. Inf. Bisnis*, vol. 1, no. 1, pp. 47–57, 2019.
- [4] Kusrini and E. T. Luthfi, *Algoritma Data Mining*, 1st ed. Yogyakarta: C.V ANDI OFFSET, 2009.
- [5] M. L. Sibuea and A. Safta, "Pemetaan Siswa Berprestasi Menggunakan Metode K-Means Clustering," *Jurteksis*, vol. 4, no. 1, pp. 85–92, 2017, doi: 10.33330/jurteksis.v4i1.28.
- [6] R. K. Dinata, H. Novriando, N. Hasdyna, and S. Retno, "Reduksi Atribut Menggunakan Information Gain untuk Optimasi Cluster Algoritma K-Means," *J. Edukasi dan Penelit. Inform.*, vol. 6, no. 1, p. 48, 2020, doi: 10.26418/jp.v6i1.37606
- [7] Suyanto, *Data Mining Untuk Klasifikasi dan Klasterisasi Data*, 1st ed. Bandung: INFORMATIKA BANDUNG, 2019.
- [8] U. Udayana *et al.*, "Seleksi Fitur Dalam Klasifikasi Genre Musik," *J. Ilmu Komput.*, vol. 10, no. 1, pp. 19–26, 2018.
- [9] Y. R. Sari, A. Sudewa, D. A. Lestari, and T. I. Jaya, "Penerapan Algoritma K-Means Untuk Clustering Data Kemiskinan Provinsi Banten Menggunakan Rapidminer," *CESS (Journal Comput. Eng. Syst. Sci.)*, vol. 5, no. 2, p. 192, 2020, doi: 10.24114/cess.v5i2.18519.