

KOMBINASI METODE *K-MEANS* DAN *DECISION TREE* DENGAN PERBANDINGAN KRITERIA DAN *SPLIT DATA*

Elly Muningsih

Sistem Informasi, Universitas Bina Sarana Informatika

Jl. Kramat Raya No. 98 Kwitang, Senen, Jakarta Pusat

Email : elly.emh@bsi.ac.id

Abstract

Data mining is a process of looking for patterns or pulling large and selected data information using certain techniques or methods. The K-Means and Decision Tree methods are part of the Data Mining technique. This study will combine the K-Means method to cluster data into three clusters then the results of the clustering will be classified using the Decision Tree Method with a comparison of the Gain Ratio, Information Gain and Gini Index criteria. The data is processed into two, namely training data and testing data with a percentage of 70:30, 80:20 and 90:10. The results of the research are to find out which criteria produce the best decision tree and performance based on the highest accuracy value from each data group. The data is taken from the UCI Repository with a total of 811 records and 52 attributes. From the data processing carried out, it is known that for the 70:30 data split, the accuracy value with the Gain Ratio, Information Gain and Gini Index criteria gets the same value, namely 97.53. The Gain Ratio and Gini Index criteria produce the highest accuracy value, which is 98.15% for 80:20 split data. While Information Gain got the highest accuracy value of 98.77% for 90:10 data split.

Keyword : *data mining, clustering, k-means, classification, decision tree*

Abstrak

Data mining merupakan suatu proses mencari pola atau menarik informasi data besar dan terpilih menggunakan teknik atau metode tertentu. Metode K-Means dan Decision Tree merupakan bagian dari teknik Data Mining. Penelitian ini akan mengkombinasikan metode K-Means untuk melakukan clustering data menjadi tiga cluster kemudian hasil dari clustering akan di klasifikasi menggunakan Metode Decision Tree dengan perbandingan kriteria Gain Ratio, Information Gain dan Gini Index. Data yang diolah di bagi (split) menjadi dua yaitu data training dan data testing dengan prosentase 70:30, 80:20 dan 90:10. Hasil dari penelitian untuk menemukan kriteria mana yang menghasilkan pohon keputusan dan performa terbaik berdasarkan nilai akurasi tertinggi dari masing-masing kelompok data. Data diambil dari Repository UCI dengan jumlah 811 record dan 52 atribut. Dari pengolahan data yang dilakukan diketahui untuk split data 70:30, nilai akurasi dengan kriteria Gain Ratio, Information Gain dan Gini Index mendapatkan nilai sama yaitu 97,53. Kriteria Gain Ratio dan Gini Index menghasilkan nilai akurasi tertinggi yaitu 98,15% untuk split data 80:20. Sedangkan Information Gain mendapat nilai akurasi tertinggi 98,77% untuk split data 90:10.

Kata kunci : *data mining, clustering, k-means, klasifikasi, decision tree*

1. Pendahuluan

Data mining adalah suatu proses untuk mengekstraksi informasi yang spesifik dari data kemudian disajikan dalam bentuk informasi yang relevan dan berguna agar dapat dimanfaatkan untuk menyelesaikan berbagai masalah [1]. *Data mining* juga dapat diartikan sebagai suatu proses mencari pola atau menarik informasi data besar dan terpilih menggunakan teknik atau metode tertentu agar ditemukan suatu pola yang sebelumnya tidak diketahui [2]. Menurut [3] dan [4] ada 6 (enam) teknik data mining yaitu Deskripsi, Estimasi, Prediksi, Klastering, Klasifikasi dan Asosiasi. Dalam penelitian ini teknik *data mining* yang digunakan adalah klastering (*clustering*) dan klasifikasi (*Clasiffication*).

Clustering merupakan suatu metode yang digunakan untuk membagi *dataset* menjadi beberapa kelompok yang memiliki kemiripan atau karakteristik yang sama [5]. Salah satu metode dalam *clustering*

adalah metode *K-Means*. Metode *K-Means* merupakan salah satu dari metode pengelompokan data non-hierarki yang mempartisi data kedalam bentuk dua kelompok atau lebih dimana data yang berkarakteristik sama akan dimasukkan kedalam satu kelompok yang sama, sedangkan data yang memiliki karakteristik berbeda akan dikelompokkan kedalam kelompok lainnya [6]. Metode *K-Means* disebutkan juga merupakan salah satu metode dalam fungsi *clustering* atau pengelompokan yang mengacu pada pengelompokkan data, *observasi* atau kasus berdasar kemiripan objek yang diteliti [7].

Klasifikasi sendiri merupakan suatu proses untuk menemukan model yang menjelaskan konsep atau kelas data yang bertujuan untuk memperkirakan kelas dari suatu objek yang label-nya tidak diketahui dimana model itu sendiri dapat berupa aturan “jika-maka”, berupa pohon keputusan (*decision tree*), formula matematika atau *neural network* [8]. Proses klasifikasi didasarkan pada empat komponen [9] yaitu kelas (suatu variabel

dependen kategorikal yang merepresentasikan “label” yang terdapat pada obyek), *predictor* (variabel independen yang direpresentasikan oleh karakteristik data), *training dataset* (merupakan satu *set* data yang berisi nilai dari kedua komponen di atas yang digunakan untuk menentukan kelas yang sesuai berdasarkan *predictor*) dan *testing dataset* (data baru yang akan diklasifikasikan oleh model yang telah dibuat dan akurasi klasifikasi dievaluasi)

Metode yang bisa digunakan dalam klasifikasi salah satunya adalah *Decision Tree*. *Decision Tree* baik digunakan untuk klasifikasi atau prediksi [10]. *Decision Tree* atau dikenal juga dengan pohon keputusan merupakan suatu teknik dalam *data mining* yang digunakan untuk mengeksplorasi data dengan cara membagi kumpulan data besar menjadi himpunan *record* yang lebih kecil dengan memperhatikan variabel tujuannya [11]. *Decision Tree* membangun *tree* (pohon) dimulai dari data pada simpul akar (*root node*) kemudian pilih sebuah atribut dan formulasikan sebuah *logical test* pada atribut tersebut, lakukan percabangan pada setiap hasil dari *test*, terus bergerak ke *subset* yang memenuhi simpul anak cabang (*internal node*) yang sesuai, lakukan proses *rekursif* dan ulangi hingga dahan-dahan dari *tree* memiliki contoh dari satu kelas tertentu [2].

Penelitian ini akan mengkombinasikan metode *K-Means* untuk melakukan pengelompokan data menjadi 3 *cluster* kemudian hasil dari *clustering* tersebut akan di klasifikasi menggunakan Metode *Decision Tree* dengan perbandingan kriteria *Gain Ratio*, *Information Gain* dan *Gini Index*. Data yang diolah sebelumnya akan di bagi (*split*) menjadi 2 yaitu data *training* dan data *testing* dengan perbandingan prosentase 70:30, 80:20 dan 90:10. Data diambil dari *Repository UCI* dengan jumlah 811 record dan 52 atribut. Pengolahan data dilakukan menggunakan aplikasi pengolah data *RapidMiner*. Hasil dari penelitian untuk menemukan kriteria mana yang menghasilkan pohon keputusan dan performa terbaik berdasarkan nilai akurasi tertinggi dari masing-masing kelompok data.

2. Metode Penelitian

Metode yang digunakan dalam penelitian ini adalah metode *CRISP-DM* (*Cross Industry Standard Process for Data Mining*). Metode *CRISP-DM* memiliki beberapa tahapan, yaitu mengumpulkan *dataset*, memilih atribut yang relevan, membangun model *clustering* dan model klasifikasi, menggunakan model *clustering* untuk pengelompokan data dan model klasifikasi perbandingan kriteria *Gain Ratio*, *Information Gain* dan *Gini Index* dan terakhir adalah evaluasi model [12].

2.1 Dataset

Dataset yang digunakan dalam penelitian ini adalah data transaksi yang diambil dari *UCI Machine Learning (Repository UCI)* dengan jumlah data awal 811 dan atribut 104. Data berisi informasi jumlah penjualan produk dalam kurun waktu 52 minggu, data minimal dan maksimal serta data nilai normal. Informasi data

transaksi yang digunakan adalah *Product_Code* (P1, P2, P3,.....P819) dimana ada data *Product_Code* yang hilang, data 52 minggu (W0, W1, ..., W51), data minimal penjualan (*MIN*), data maksimal penjualan (*MAX*) dan data nilai normalisasi mingguan (*Normalised 0, Normalised 1, ..., Normalised51*).

2.2 Preprocessing Data

Dari *dataset* yang ada dilakukan pemilihan atribut yang akan digunakan yaitu *Product_Code* (P1, P2, P3 dan seterusnya) untuk spesial atribut sebanyak 811 dan atribut data mingguan (W0... W51) untuk regular atribut sebanyak 52 yang nantinya akan di proses dimana tipe data pada atribut yang digunakan adalah integer.

2.3 Modelling dan Evaluasi

Pada tahapan ini, dilakukan beberapa langkah untuk *modelling* menggunakan *tools RapidMiner* :

1. Modelling Metode Clustering K-Means

Pada tahapan ini dilakukan *clustering* menggunakan metode *K-Means* dengan jumlah *cluster* sebanyak 3.

2. Modelling Metode Klasifikasi Decision Tree

Pada *modelling* metode klasifikasi ini, langkah-langkah yang dilakukan adalah :

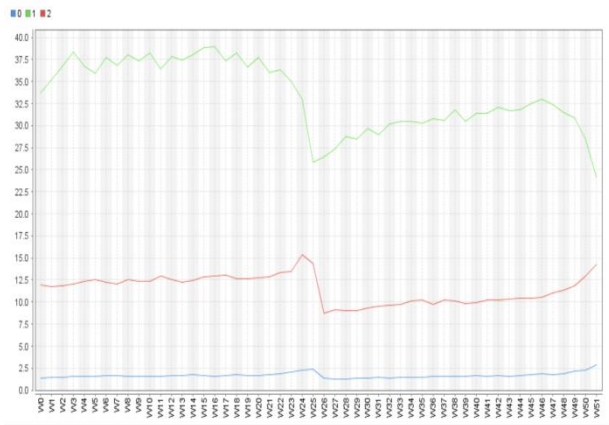
- Data dibagi menjadi dua yaitu data *training* dan data *testing* dengan perbandingan 70 : 30, 80 : 20 dan 90 : 10.
- Untuk tiap-tiap pembagian data di buat modelnya dengan metode klasifikasi *Decision Tree* kemudian pada setting parameter *criterion* dipilih *Gain Ratio*, *Information Gain* atau *Gini Index*.
- Tambahkan operator *Cross Validation* dan *Performance* untuk evaluasi dan mengetahui perbandingan tingkat akurasi masing-masing kriteria dengan pembagian data yang berbeda-beda.
- Dari pengolahan yang dilakukan akan diketahui pembagian data *training* dan data *testing* terbaik juga pemilihan *criterion* terbaik yang menghasilkan akurasi tertinggi.

3. Hasil dan Diskusi

Pengolahan data dilakukan menggunakan komputer personal *Intel Core i5*, 8 GB RAM, sistem operasi *Windows 10* dan *RapidMiner 9.5.001*.

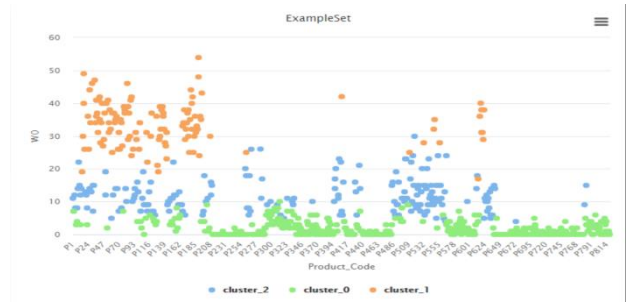
3.1 Penerapan Metode K-Means

Pengolahan data dengan metode *clustering K-Means* menggunakan jumlah *cluster* 3 menghasilkan anggota kelompok untuk masing-masing *cluster* yaitu *cluster_0* sebanyak 490 anggota, *cluster_1* sebanyak 124 anggota dan *cluster_2* sebanyak 197 anggota ditampilkan pada Gambar 1. dibawah ini :



Gambar 1. Anggota tiap cluster

Dari gambar yang ditampilkan sebelumnya dapat dijelaskan bahwa anggota kelompok *cluster_0* adalah data mingguan dengan jumlah penjualan terkecil, *cluster_1* adalah data mingguan dengan jumlah penjualan terbesar dan *cluster_2* adalah data mingguan dengan jumlah penjualan sedang. Sedangkan representasi penyebaran anggota kelompok untuk data mingguan W0 ditampilkan pada Gambar 2. berikut ini :

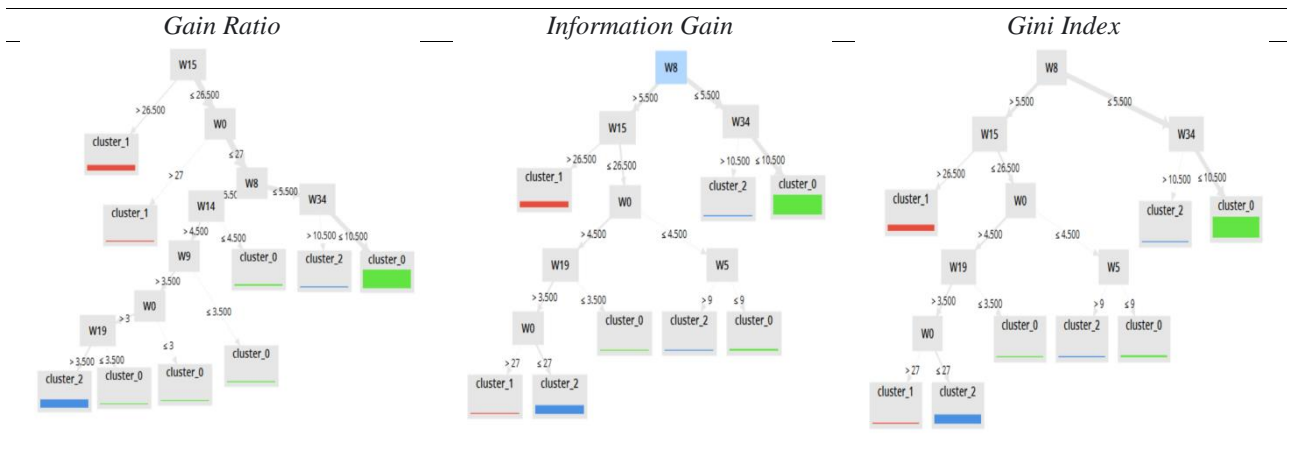


Gambar 2. Penyebaran anggota cluster

3.2. Pohon Keputusan Metode *Decision Tree*

1. Perbandingan Kriteria Data *Testing* dengan *Split Data 70:30*

Pada Gambar 3. dibawah ini ditampilkan pohon keputusan yang dihasilkan dari perbandingan kriteria *Gain Ratio*, *Information Gain* dan *Gini Index* dengan split data sebesar 70:30. Dari gambar dapat dijelaskan bahwa pada kriteria *Gain Ratio*, terlihat atribut W15 menjadi simpul akar dan hanya memiliki satu simpul cabang ke kiri yaitu atribut W0. Sedangkan kriteria *Information Gain* dan *Gini Index* memiliki hasil pohon keputusan yang sama dimana simpul akarnya adalah atribut W8 dan simpul akar ini memiliki dua cabang simpul ke kanan dan kiri. Cabang simpul kanan adalah atribut W15 dan simpul kiri adalah atribut W34.

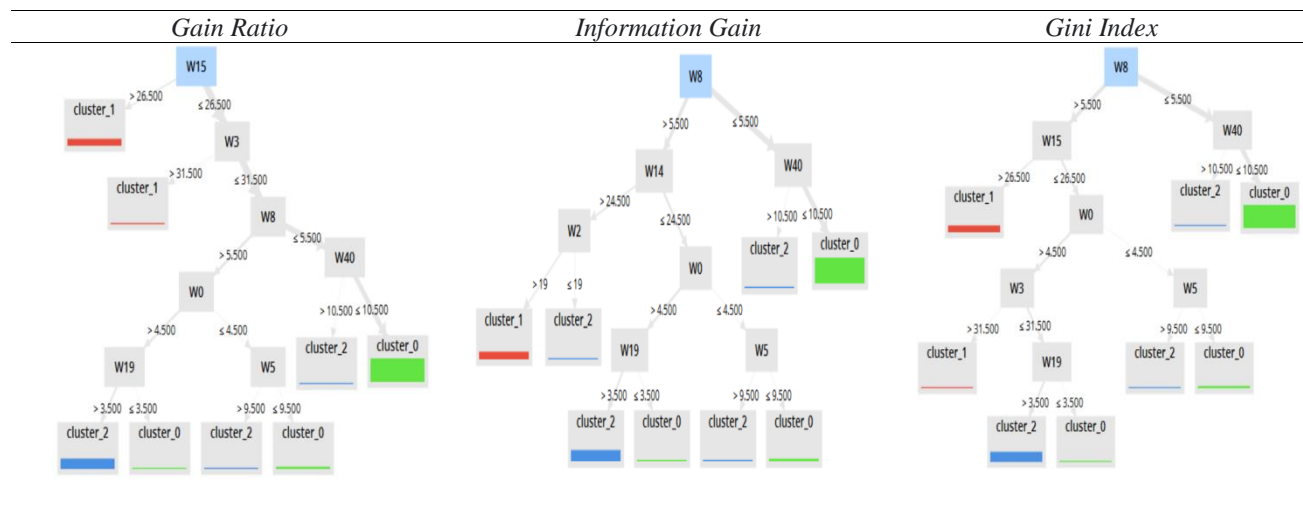


Gambar 3. Perbandingan Kriteria Data *Testing* dengan *split data 70:30*

2. Perbandingan Kriteria Data *Testing* dengan *Split Data 80:20*

Pada Gambar 4. dibawah ini ditampilkan pohon keputusan yang dihasilkan dari perbandingan kriteria *Gain Ratio*, *Information Gain* dan *Gini Index* dengan split data sebesar 80:20. Dari gambar dapat dijelaskan bahwa pada kriteria *Gain Ratio*, terlihat atribut W15 menjadi simpul akar dan hanya memiliki satu simpul cabang ke kiri yaitu atribut W3. Sedangkan kriteria

Information Gain dan *Gini Index* memiliki hasil pohon keputusan yang sedikit berbeda, dimana simpul akarnya sama yaitu atribut W8. Simpul akar kriteria *Information Gain* memiliki cabang simpul kanan adalah atribut W14 dan simpul kirinya adalah atribut W40. Sedangkan simpul akar kriteria *Gini Index* memiliki cabang simpul kanan adalah atribut W15 dan simpul kirinya adalah atribut W40. Cabang simpul kiri untuk kriteria *Information Gain* dan *Gini Index* adalah sama yaitu atribut W40.

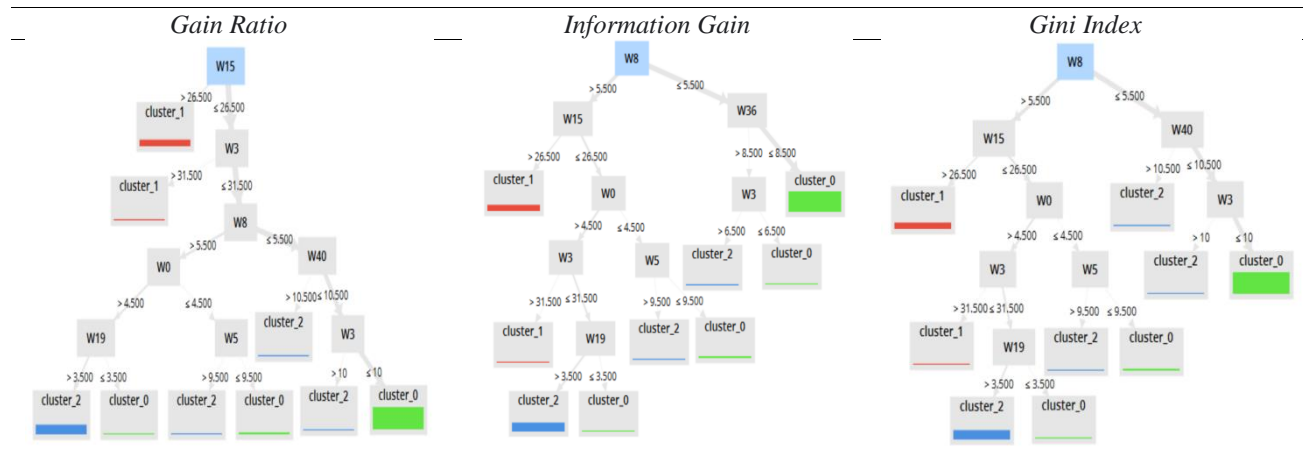


Gambar 4. Perbandingan Kriteria Data Testing dengan split data 80:20

3. Perbandingan Kriteria Data Testing dengan Split Data 90:10

Pada Gambar 5. dibawah ini ditampilkan pohon keputusan yang dihasilkan dari perbandingan kriteria *Gain Ratio*, *Information Gain* dan *Gini Index* dengan split data sebesar 90:10. Dari gambar dapat dijelaskan bahwa pada kriteria *Gain Ratio*, terlihat atribut W15 menjadi simpul akar dan hanya memiliki satu simpul cabang ke kiri yaitu atribut W3. Sedangkan kriteria

Information Gain dan *Gini Index* memiliki hasil pohon keputusan yang sedikit berbeda, dimana simpul akarnya sama yaitu atribut W8. Simpul akar kriteria *Information Gain* memiliki cabang simpul kanan adalah atribut W15 dan simpul kirinya adalah atribut W36. Sedangkan simpul akar kriteria *Gini Index* memiliki cabang simpul kanan adalah atribut W15 dan simpul kirinya adalah atribut W40. Cabang simpul kanan untuk kriteria *Information Gain* dan *Gini Index* adalah sama yaitu atribut W15.



Gambar 4. Perbandingan Kriteria Data Testing dengan split data 90:10

3.3 Nilai Akurasi

Untuk mengukur akurasi digunakan *Confusion Matrix*. Berikut ini ditampilkan perhitungan berdasarkan data testing untuk kriteria *Gain Ratio*, *Information Gain* dan *Gini Index*.

Perhitungan Berdasarkan Data Testing dengan Split Data 70:30.

1. Pada kriteria *Gain Ratio*, untuk *cluster_2* yang diprediksi dan benar ada 54, diprediksi *cluster_0* ternyata *cluster_2* ada 4, dan diprediksi *cluster_1* ternyata *cluster_2* ada 1. Untuk *cluster_0* yang diprediksi dan benar ada 146, yang diprediksi

cluster_2 ternyata *cluster_0* ada 1. Untuk *cluster_1* yang di prediksi dan benar ada 37. Nilai akurasi pada kriteria *Gain Ratio* sebesar 97,53%

2. Pada kriteria *Information Gain* dan *Gini Index* didapatkan hasil sama yaitu untuk *cluster_2* yang diprediksi dan benar ada 55, diprediksi *cluster_0* ternyata *cluster_2* ada 3, dan diprediksi *cluster_1* ternyata *cluster_2* ada 1. Untuk *cluster_0* yang diprediksi dan benar ada 145, yang diprediksi *cluster_2* ternyata *cluster_0* ada 2. Untuk *cluster_1* yang di prediksi dan benar ada 37. Nilai akurasi pada kriteria *Information Gain* dan *Gini Index* sebesar 97,53%

accuracy: 97.53%

	true cluster_2	true cluster_0	true cluster_1	class precision
pred. cluster_2	54	1	0	98.18%
pred. cluster_0	4	146	0	97.33%
pred. cluster_1	1	0	37	97.37%
class recall	91.53%	99.32%	100.00%	

Gambar 5. Confusion Matrix Gain Ratio (70:30)

accuracy: 97.53%

	true cluster_2	true cluster_0	true cluster_1	class precision
pred. cluster_2	55	2	0	96.49%
pred. cluster_0	3	145	0	97.97%
pred. cluster_1	1	0	37	97.37%
class recall	93.22%	98.64%	100.00%	

Gambar 6. Confusion Matrix Information Gain (70:30)

accuracy: 97.53%

	true cluster_2	true cluster_0	true cluster_1	class precision
pred. cluster_2	55	2	0	96.49%
pred. cluster_0	3	145	0	97.97%
pred. cluster_1	1	0	37	97.37%
class recall	93.22%	98.64%	100.00%	

Gambar 7. Confusion Matrix Gini Index (70:30)

Perhitungan Berdasarkan Data Testing dengan Split Data 80:20.

1. Pada kriteria *Gain Ratio* dan *Gini Index* memiliki hasil yang sama yaitu *cluster_2* yang diprediksi dan benar ada 37, diprediksi *cluster_0* ternyata *cluster_2* ada 2. Untuk *cluster_0* yang diprediksi dan benar ada 97, yang diprediksi *cluster_2* ternyata *cluster_0* ada 1. Untuk *cluster_1* yang di prediksi dan benar ada 25. Nilai akurasi pada kriteria *Gain Ratio* sebesar 98,15%
2. Pada kriteria *Information Gain*, untuk *cluster_2* yang diprediksi dan benar ada 37, diprediksi *cluster_0* ternyata *cluster_2* ada 2, dan diprediksi *cluster_1* ternyata *cluster_2* ada 1. Untuk *cluster_0* yang diprediksi dan benar ada 97, yang diprediksi *cluster_2* ternyata *cluster_0* ada 1. Untuk *cluster_1* yang di prediksi dan benar ada 23 dan diprediksi *cluster_2* ternyata *cluster_1* ada 2. Nilai akurasi pada kriteria *Information Gain* ini sebesar 96,91%

accuracy: 98.15%

	true cluster_2	true cluster_0	true cluster_1	class precision
pred. cluster_2	37	1	0	97.37%
pred. cluster_0	2	97	0	97.98%
pred. cluster_1	0	0	25	100.00%
class recall	94.87%	98.98%	100.00%	

Gambar 8. Confusion Matrix Gain Ratio (80:20)

accuracy: 96.91%

	true cluster_2	true cluster_0	true cluster_1	class precision
pred. cluster_2	37	1	2	92.50%
pred. cluster_0	2	97	0	97.98%
pred. cluster_1	0	0	23	100.00%
class recall	94.87%	98.98%	92.00%	

Gambar 9. Confusion Matrix Information Gain (80:20)

accuracy: 98.15%

	true cluster_2	true cluster_0	true cluster_1	class precision
pred. cluster_2	37	1	0	97.37%
pred. cluster_0	2	97	0	97.98%
pred. cluster_1	0	0	25	100.00%
class recall	94.87%	98.98%	100.00%	

Gambar 10. Confusion Matrix Gini Index (80:20)

Perhitungan Berdasarkan Data Testing dengan Split Data 90:10.

1. Pada kriteria *Gain Ratio* dan *Gini Index* memiliki hasil sama yaitu *cluster_2* yang diprediksi dan benar ada 19 dan diprediksi *cluster_0* ternyata *cluster_2* ada 1. Untuk *cluster_0* yang diprediksi dan benar ada 48 dan yang diprediksi *cluster_2* ternyata *cluster_0* ada 1. Untuk *cluster_1* yang di prediksi dan benar ada 12. Nilai akurasi pada kriteria *Gain Ratio* sebesar 97,53%
2. Pada kriteria *Information Gain*, untuk *cluster_2* yang diprediksi dan benar ada 20. Untuk *cluster_0* yang diprediksi dan benar ada 48 dan diprediksi *cluster_2* ternyata *cluster_0* ada 1. Untuk *cluster_1* yang di prediksi dan benar ada 23 dan diprediksi *cluster_2* ternyata *cluster_1* ada 2. Nilai akurasi pada kriteria *Information Gain* ini sebesar 98,77%.

accuracy: 97.53%

	true cluster_2	true cluster_0	true cluster_1	class precision
pred. cluster_2	19	1	0	95.00%
pred. cluster_0	1	48	0	97.96%
pred. cluster_1	0	0	12	100.00%
class recall	95.00%	97.96%	100.00%	

Gambar 11. Confusion Matrix Gain Ratio (90:10)

accuracy: 98.77%

	true cluster_2	true cluster_0	true cluster_1	class precision
pred. cluster_2	20	1	0	95.24%
pred. cluster_0	0	48	0	100.00%
pred. cluster_1	0	0	12	100.00%
class recall	100.00%	97.96%	100.00%	

Gambar 12. Confusion Matrix Information Gain (90:10)

accuracy: 97.53%

	true cluster_2	true cluster_0	true cluster_1	class precision
pred. cluster_2	19	1	0	95.00%
pred. cluster_0	1	48	0	97.96%
pred. cluster_1	0	0	12	100.00%
class recall	95.00%	97.96%	100.00%	

Gambar 13. Confusion Matrix Gini Index (90:10)

Dari penjelasan sebelumnya dapat di kumpulkan jadi Tabel 1. dibawah ini yang menampilkan kriteria, split data dan nilai akurasi. Jadi dapat dijelaskan bahwa jenis kriteria dan *split* data mempengaruhi nilai akurasi dari suatu pengolahan data yang memiliki tipe data integer.

Tabel 1. Perbandingan Nilai Akurasi

Kriteria	Nilai Akurasi (<i>Split</i> Data)		
	70:30	80:20	90:10
<i>Gain Ratio</i>	97,53	98,15	97,53
<i>Information Gain</i>	97,53	96,91	98,77
<i>Gini Index</i>	97,53	98,15	97,53

Dari kriteria *Information Gain* dengan *split* data 90:10 didapatkan *rule* untuk data pada penelitian ini yaitu :

```
w8 > 5.500
|   w14 > 24.500
|   |   w2 > 19: cluster_1 {cluster_2=0,
|   |   |   cluster_0=0, cluster_1=99}
|   |   |   w2 ≤ 19: cluster_2 {cluster_2=2,
|   |   |   |   cluster_0=0, cluster_1=0}
|   |   |   w14 ≤ 24.500
|   |   |   |   w0 > 4.500
|   |   |   |   |   w19 > 3.500: cluster_2
|   |   |   |   |   |   {cluster_2=151, cluster_0=0, cluster_1=0}
|   |   |   |   |   |   w19 ≤ 3.500: cluster_0
|   |   |   |   |   |   |   {cluster_2=0, cluster_0=2, cluster_1=0}
|   |   |   |   |   |   |   |   w0 ≤ 4.500
|   |   |   |   |   |   |   |   |   w5 > 9.500: cluster_2 {cluster_2=2,
|   |   |   |   |   |   |   |   |   |   cluster_0=0, cluster_1=0}
|   |   |   |   |   |   |   |   |   |   w5 ≤ 9.500: cluster_0 {cluster_2=0,
|   |   |   |   |   |   |   |   |   |   |   cluster_0=20, cluster_1=0}
w8 ≤ 5.500
|   w40 > 10.500: cluster_2 {cluster_2=3,
|   |   cluster_0=0, cluster_1=0}
|   |   w40 ≤ 10.500: cluster_0 {cluster_2=0,
|   |   |   cluster_0=370, cluster_1=0}
```

4. Kesimpulan

Dari pengolahan data yang sudah dilakukan yaitu mengkombinasikan metode *clustering K-Means* dan metode klasifikasi *Decision Tree* dengan membandingkan kriteria *Gain Ratio*, *Information Gain*, *Gini Index* dan pembagian (*split*) data 70:30, 80:20, 90:20 diketahui bahwa nilai akurasi tertinggi dihasilkan dari kriteria *Information Gain* dengan *split* data *training* dan *testing* sebesar 90:10 yaitu sebesar 98,77%. Pada penelitian berikutnya bisa dilakukan pengolahan dengan menambahkan dan menguji beberapa data yang berbeda baik dari jenis atribut maupun tipe datanya.

Daftar Pustaka

[1] Wong, N. P., Damanik, F. N. S., Jaya, E. S., & Rajaya, R. (2019). *Perbandingan Algoritma C4.5 dan Classification and Regression Tree (CART) Dalam Menyeleksi Calon Karyawan*. 20(1), 11–18.

[2] Elmande, Y., & Widodo, P. (2012). Pemilihan Criteria Splitting dalam Algoritma Iterative Dichotomiser 3 (ID3) untuk Penentuan Kualitas Beras: Studi Kasus Pada Perum Bulog Divre

Lampung. *Jurnal TELEMATIKA MKOM*, 4(1), 10. Retrieved from [http://demo.pohonkeputusan.com/files/PEMILIHAN_CRITERIA_SPLITTING_DALAM_ALGORITMA_ITERATIVE_DICHOTOMISER_3_\(ID3\)_UNTUK_PENENTUAN_KUALITAS_BERAS.pdf](http://demo.pohonkeputusan.com/files/PEMILIHAN_CRITERIA_SPLITTING_DALAM_ALGORITMA_ITERATIVE_DICHOTOMISER_3_(ID3)_UNTUK_PENENTUAN_KUALITAS_BERAS.pdf)

[3] Larose, D. T., & Larose, C. D. (2014). *Discovering Knowledge in Data: An Introduction to Data Mining: Second Edition*. In *Discovering Knowledge in Data: An Introduction to Data Mining: Second Edition* (Vol. 9780470908). <https://doi.org/10.1002/9781118874059>

[4] Noviyanto, N. (2020). Penerapan Data Mining dalam Mengelompokkan Jumlah Kematian Penderita COVID-19 Berdasarkan Negara di Benua Asia. *Paradigma - Jurnal Komputer Dan Informatika*, 22(2), 183–188. <https://doi.org/10.31294/p.v22i2.8808>

[5] Muningsih, E., Nur, H. M., Dwi Imaniawan, F. F., Saifudin, Handayani, V. R., & Endiarto, F. (2020). Comparative Analysis on Dimension Reduction Algorithm of Principal Component Analysis and Singular Value Decomposition for Clustering. *Journal of Physics: Conference Series*, 1641(1). <https://doi.org/10.1088/1742-6596/1641/1/012101>

[6] Gustientiedina, G., Adiya, M. H., & Desnelita, Y. (2019). Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan. *Jurnal Nasional Teknologi Dan Sistem Informasi*, 5(1), 17–24. <https://doi.org/10.25077/teknosi.v5i1.2019.17-24>

[7] Muningsih, E. (2015). Penerapan Metode Fuzzy C-Means dengan Model Fuzzy RFM (Studi Kasus): Clustering Pelanggan Potensial Online Shop). *Semantik*, 157–161.

[8] Wirdasari, D. (2013). ANALISA TEKNIK PENENTUAN ATRIBUT DALAM MEMBUAT POHON KEPUTUSAN PADA PENAMBANGAN DATA. *Jurnal Ilmiah Saintikom Sains Dan Komputer*, 12(2).

[9] Gorunescu, F. (2011). Data mining: Concepts, models and techniques. In *Intelligent Systems Reference Library* (Vol. 12). <https://doi.org/10.1007/978-3-642-19721-5>

[10] Rahayu, E. S., Satria, R., & Supriyanto, C. (2015). Penerapan Metode Average Gain, Threshold Pruning dan Cost Complexity Pruning Untuk Split Atribut Pada Algoritma C4.5. *Journal of Intelligent Systems*, 1(2), 91–97.

[11] Zega, S. A. (2014). Penggunaan Pohon Keputusan untuk Klasifikasi Tingkat Kualitas Mahasiswa Berdasarkan Jalur Masuk Kuliah. *Seminar Nasional Aplikasi Teknologi Informasi (SNATI) Yogyakarta*, 7–13.

[12] Sastry, S. H., & Babu, P. M. S. P. (2013). *Implementation of CRISP Methodology for ERP Systems*. 2(05), 203–217. Retrieved from <http://arxiv.org/abs/1312.2065>