IMPLEMENTASI *DATA MINING* PADA DATA PENJUALAN PAKAIAN MENGGUNAKAN ALGORITMA *K-MEANS* DENGAN *OPTIMIZE PARAMETER GRID*

Dila Aryani, Bambang Irawan, Agus Bahtiar
Teknik Informatika, STMIK IKMI Cirebon
Jl. Perjuangan No.10B, Karyamulya, Kec. Kesambi, Kota Cirebon, Jawa Barat, Indonesia
aryaaryani315@gmail.com

ABSTRAK

Perkembangan teknologi telah berdampak besar pada sektor perdagangan dan bisnis, khususnya *dalam e-commerce.* Menurut *website https://www.statista.com/topics/871/online-shopping* E-commerce Worldwide - Statistics & Facts Pada tahun 2019, sekitar 1,92 miliar orang terlibat dalam transaksi *e-commerce*, dengan Indonesia mencatatkan sebagai pengguna tertinggi pada tahun 2018. Pakaian berkontribusi besar pada perkembangan bisnis di Indonesia. Toko rizki_collection123 merupakan salah satu pelaku bisnis yang berjualan di *E-commerce*. Tantangan utama yang dihadapi oleh Toko rizki_collection123 adalah belum mendapatkannya informasi dari data penjualan untuk meningkatkan strategi bisnis. Tujuan utama dilakukan penelitian terhadap data penjualan untuk mengetahui nilai k optimal berdasarkan parameter, maka diperlukan operator *optimize parameter grid* untuk mempercepat pengelompokan data penjualan untuk mengetahui informasi karakteristik yang ada pada dataset, dengan *ParameterMixed Measure*. Penelitian ini menggunakan Metode *Data Mining* Algoritma *K-Means Clustering* dengan *Optimize Parameter Grid*. Penelitian ini menggunakan Teknik analisis data *Knowledge Discovery in Database (KDD).* Algoritma *K-Means* dipilih karena cukup efektif dalam pengelompokan data yang besar seperti data hasil penjualan. *Tools* yang digunakan pada penelitian ini yaitu *RapidMiner v.10.2.* Hasil *clustering* berdasarkan *Davies Bouldin Index* terendah terbentuknya 3 *cluster* dengan *Measure type BregmanDivergences* pada DBI= 0,035. *Cluste*r 0 memiliki sebanyak 2859 item atau sekitar 41,15% dari total transaksi 6946, sedangkan *Cluster* 1 memiliki 2392 item atau sekitar 34,44%. *Cluster* 2 memiliki 1695 item atau sekitar 24,41%

Kata kunci: *Data Mining, Clustering, K-Means, Optimize Parameter Grid, Knowledge Discovery in Database.*

1. PENDAHULUAN

Perkembangan dan kemajuan teknologi berjalan seiring dengan kebutuhan Masyarakat sangat berdampak pada sektor perdagangan dan bisnis [1]. *E-commerce* menjadi solusi transaksi yang sangat praktis hanya memerlukan perangkat elektronik dan koneksi internet [2]. Pada tahun tahun 2019 sekitar 1,92 miliar orang terlibat dalam transaksi jual-beli melalui *e-commerce* [3]. Dengan Indonesia mencatatkan sebagai pengguna tertinggi pada tahun 2018 dan pertumbuhan mencapai 78% [4]. Pentingnya pakaian sebagai kebutuhan manusia yang berfungsi sebagai penutup dan pelindung tubuh manusia. Menjadikan permintaan akan pakaian semakin tinggi, dan pakaian sangat berkontribusi pada perkembangan perdagangan dan bisnis di Indonesia [5].

Toko rizki_collection123 merupakan salah satu pelaku bisnis yang berjualan di *E-commerce* yang menjual pakaian. Toko rizki_collection123 bergabung pada platform *E-commerce Shopee* sekitar 4 tahun lalu yaitu tepatnya pada tahun 2020. Permasalahan yang dihadapi oleh Toko rizki_collection123 merupakan contoh nyata dari tantangan yang diahadapi oleh banyak penjual Pakaian di *E-commerce* saat ini. Dalam era digital yang serba cepat ini, dinamika pasar berubah dengan sangat cepat dan pelaku bisnis harus mampu beradaptasi dengan perubahan tersebut. Salah satu tantangan utama adalah penumpukan stok barang. Hal ini tidak hanya memenuhi ruang penyimpanan, tetapi juga dapat menimbulkan kerugian jika barang tersebut tidak laku terjual. Oleh karena itu, manajemen stok yang efisien sangan penting dalam bisnis *e-commerce*. Persaingan harga yang ketat juga menjadi tantangan lainnya. Dengan banyaknya pelaku bisnis di platform *E-commece*, konsumen memiliki banyak pilihan. Oleh karena itu, penetapan harga yang tepat sangat penting untuk menarik konsumen sekaligus memastikan keuntungan. Selain itu, perubahan preferensi konsumen yang cepat juga menjadi tantangan. Konsumen saat ini lebih berpengetahuan dan memiliki akses ke informasi yang luas, sehingga preferensi pembeli dapat berubah dengan cepat. Maka dari itu, untuk mengatasi tantangan-tantangan ini, diperlukan pemahaman yang mendalam tentang data penjualan. Sayangnya, banyak pebisnis termasuk Toko rizki_collection123, belum memanfaatkan potensi data penjualan mereka sepenuhnya.

Metode yang dipilih untuk melakukan analisis pada dataset ini yaitu, menggunakan Metode *Data Mining Algoritma K-Means* Clustering dengan Teknik analisis *Knowledge Discovery in Database (KDD). Knowledge Discovery in Database* adalah Suatu rangkaian aktivitas yang melibatkan proses pengumpulan dan pemanfaatan data historis dengan tujuan mengidentifikasi pola atau hubungan dalam kumpulan data yang memiliki skala besar[6].

Data Mining adalah proses menemukan pola yang menarik dan tersembunyi dari Kumpulan data yang besar [7]. *Clustering* adalah metode pengelompokan data yang dilakukan berdasarkan kemiripan karakteristik data [6]. *Algoritma K-Means* adalah *algoritma* yang banyak digunakan dalam *cluster*ing karena kemudahan dan efisiensinya yang diakui oleh *IEEE* sebagai salah satu dari *algoritma data mining* teratas [8]. *Algoritma K-Means Clustering* juga dikenal sebagai metode pengelompokan data yang relatif lebih sederhana untuk diimplementasikan [9]. Untuk menyelesaikan permasalahan penelitian, studi ini mengambil pendekatan *Knowledge Discovery In Database (KDD),* pendekatan ini mencakup serangkaian tahapan, dengan tujuan mengidentifikasi hubungan dalam dataset yang memiliki ukuran besar [6]. Evaluasi *clustering* yang digunakan, *Davies-Bouldin Index* yaitu metode untuk menentukan efisiensi atau jumlah *cluster* yang ideal dalam proses pengelompokan, Dimana semakin kecil nilai hasil perhitungan *Davies-Bouldin* maka semakin akurat nilai *cluster (K)* yang digunakan [10].

Tujuan utama dari penelitian ini adalah menerapkan teknologi data mining. khusunya algoritma K-Means, pada data penjualan Toko rizki_collection123. Dengan melakukan penelitian, diharapkan dapat diketahui dari dalam data penjualan yang dapat membantu toko dalam merumuskan strategi bisnis yang lebih efektif. Signifikasi penelitian ini terletak pada potensinya untuk mengisi kesenjangan pengetahuan tentang bagaimana teknologi data mining digunakan dalam konteks bisnis e-commerce, sehingga dapat memberikan kontribusi pada bidang informatika dengan menunjukan bagaimana *algoritma* dan teknologi yang dikembangkan dalam bidang ini dapat diterapkan untuk membantu pebisnis.

Hasil analisa clustering menggunakan data penjualan pakaian di platfom e-commerce menggunakan algoritma k-means dengan optimize parameter grid diharapkan dapat memberikan wawasan mendalam terkait dinamika penjualan dan perilaku pelanggan. Sehingga dapat memberikan pemahaman yang lebih baik tentang preferensi pelanggan dan pola penjualan pada platfrom e-commerce. Dengan pemahaman ini, pebisnis dapat mengoptimalkan strategi terkait pengelolaan stok, personalisasi pengalaman pelanggan, dan penyesuaian kampanye pemasaran.

2. TINJAUAN PUSTAKA

2.1. *Data Mining*

Data mining merupakan suatu proses yang menggunakan Teknik statistic, matematika, *Artificial Intelligence*, dan *machine learning* untuk mengekstraksi serta mengidengtifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai basis data besar [11]. Secara umum, data mining merupakan suatu proses analisis dan eksplorasi data yang melibatkan volume besar untuk menemukan pola-pola yang memiliki signifikasi [12]. Data Mining memiliki fungsi untuk melakukan *clustering*, selain melakukan tugas lain seperti memberikan deskripsi, estimasi, prediksi, klasifikasi dan mengidentifikasi asosiasi[13]. Pada penelitian ini proses data mining dilakukan pada *software RapidMiner v.10.2.*

2.2. *Clustering*

Clustering adalah suatu proses penting dalam rangkaian Data Mining yang mengidentifikasi kelompok atau *cluster* alami dalam basis data dengan mengoptimalkan suatu kriteria pengelompokan[14]. Tugas utama dari pengelompokan adalah mencapai tingkat kemiripan tinggi di dalam suatu kelas dan tingkat kemiripan rendah antar kelas, sehingga jarak antar kelas sebesar mungkin, dan jarak antara sampel di dalam suatu kelas dan pusat kelas sekecil Mungkin [15]. *Clustering* yang paling banyak digunakan adalah K-Means, merujuk pada fakta bahwa algoritma K-Means adalah metode pengelompokan yang paling umum dan sering digunakan[16]

2.3. Algoritma *K-Means*

Algoritma pengelompokan (*cluster*ing) K-Means awalnya dikembangkan oleh Hartigan pada tahun 1975, dan disajikan algoritma lengkap dengan rute implementasinya pada tahun 1979. Algoritma ini menjadi bagian krusial dari bidang unsupervised machine learning[17]. K-means terbukti efisien dan umunya digunakan dalam situasi Dimana data tidak memeiliki label, terutama untuk tugas-tugas *cluster*ing [18]. K-means merupakan salah satu algoritma partisi yang paling popular dan digunakan secara luas karena efisiensinya serta kesederhanaan implementasinya[19]. Algoritma ini melakukan perhitungan jarak antara setiap objek data dengan pusat *cluster,* kemudian mengelompokkan objek data yang memiliki jarak terdekat dengan pusat cluster ke dalam cluster yang sama [20]

2.4. *DAVIES BOULDIN* INDEX

Davies bouldin index merupakan metrik yang banyak digunakan untuk mengevaluasi kinerja pengelompokan dengan cara membagi *cluster* [21]. Evaluasi *cluster* dengan davies bouldin index menggunakan skema evaluasi internal, Dimana hasil pengelompokan dapat dinilai berdasarkan jumlah kedekatan data *cluster*[22]. Teknik evaluasi *clustering* dengan *Davies Bouldin* Index memerlukan perbandingan nilai (DBI) dalam setiap pengelompokan. Nilai DBI yang paling kecil atau mendekati 0, namun tidak bersifat negative, dianggap sebagai hasil *cluster*ing yang paling optimal[23].

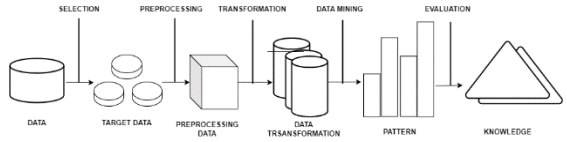
PERSAMAAN MATEMATIKA DBI:

$$DBI = \frac{1}{K}\sum_{i=1}^{K} max_{i \neq j}\left(\frac{d(ci,cj)}{\sigma i + \sigma j}\right)$$

Keterangan:
K adalah jumlah *cluster*
σ adalah simpangan baku *cluster* ke i
$d(ci,cj)$ adalah jarak antara pusat *cluster* I dan j
ci adalah pusat *cluster* ke-i

3. METODE PENELITIAN

Metode yang digunakan dalam penelitian ini adalah metode penelitian Kuantitatif. Penelitian kuantitatif merupakan tipe penelitian yang terorganisir secara sistematis dan terstruktur dari awal hingga akhir [24]. Setiap proses yang dilakukan menggunakan data berbentuk angka mengggunakan teknik statistik [25]. Penelitian ini menggunakan data penjualan dari toko rizki_collection123 selama empat bulan mulai dari bulan juni sampai september yang diambil pada tahun 2023. Dalam *Data mining* tahapan *Knowledge Discovery in Database (KDD)* adalah aktivitas yang mencakup pengumpulan dan pemanfaatan data historis guna mengidentifikasi keteraturan, pola, dan hubungan dalam kumpulan data yang berskala besar [26]. Tahapan KDD, mencakup Seleksi Data, *Preprocessing* dan Pembersihan Data, Transformasi Data, Data Mining, serta Evaluasi/Interpretasi atau dapat dilihat pada gambar sebagai berikut:



Gambar 1. Tahapan *Kowledge Discovery in Database (KDD)*

3.1. Seleksi Data

Tahap ini melibatkan pemilihan data yang relevan untuk analisis dari database yang besar. Pada tahap ini akan ditentukan ID serta pemilihan atribut pada data toko rizki_collection123.

3.2. Preprocessing

Tahap ini digunakan untuk mengubah skala data menjadi rentang tertentu atau membawa data ke dalam skala yang seragam. Setelah memilih atribut serta data yang digunakan memilili ID maka selanjutnya adalah menyiapkan data sebelum lanjut di proses dengan cara normalisasi untuk menyesuaikan skala pada data tersebut.

3.3. Transformasi

Tahap ini melibatkan transformasi data, seperti reduksi dimensi atau diskritisasi, untuk mengubah data menjadi bentuk yang sesuai untuk proses data mining. Selanjutnya adalah mengubah tipe data agar data dapat diproses, pada tahap ini semua data yang bersifat non-numerik diubah ke tipe numerik.

3.4. Data Mining

Tahap ini melibatkan penerapan teknik data mining, untuk mengelompokkan data menjadi kelompok-kelompok yang memiliki karakteristik serupa. Tahap ini merupakan tahap pemrosesan data dimana data dikelompokkan dengan algoritma k-means menggunakan *software RapidMiner v.10.2.*

3.5. Evaluasi

Tahap ini melibatkan evaluasi pola yang telah diidentifikasi berdasarkan metrik tertentu. Setelah data diproses maka tahap terakhir adalah evaluasi dengan menampilkan hasil dari proses data mining.

3.6. Optimize Parameter Grid

Optimize parameter grid adalah proses mencari kombinasi parameter yang optimal dalam algoritma machine learning dan teknik pemodelan lainya. Parameter ini digunakan untuk mencari kombinasi yang memberikan hasil terbaik berdasarkan kriteria yang ditetapkan dari parameter yang ingin dioptimalkan.

4. HASIL DAN PEMBAHASAN

Hasil penelitian Implementasi Data Mining Pada Data Penjualan Pakaian Menggunakan Algoritma K-Means Dengan Optimize Parameter Grid adalah sebagai berikut.

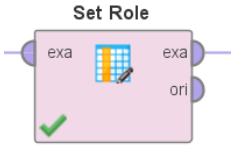
4.1. Data

Dataset yang digunakan dalam penelitian ini terdiri dari 6946 entri data dan 33 atribut data penjualan pakaian. Dataset diperoleh dari toko pakaian rizki_collection123 dengan alamat toko *https://shopee.co.id/rizki_collection123.* Data yang diperoleh berbentuk dokumen soft file dengan format excel yang dapat di akses melalui *https://docs.google.com/spreadsheets/d/12Z8u1Uzf_9tGahwqIsbDeQAWHRhBJ-Fr/edit?usp=sharing&ouid=117495451121510487552&rtpof=true&sd=true*. Setelah mengimport data, selanjutnya gunakan Operator Retrieve, proses ini memuat data penjualan yang tersimpan dari repositori ke dalam proses, seperti gambar dibawah ini.

Gambar 2. *Operator Retrieve* pada *RapidMiner*

4.2. Seleksi Data

Pada tahap ini yang harus dilakukan untuk pertama, yaitu menentukai ID dalam dataset menggunakan operator *Set Role* seperti yang terlihat pada terlihat seperti gambar dibawah.



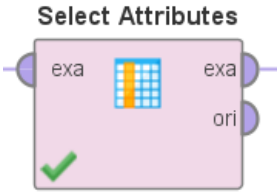
Gambar 3. *Operator Set Role pada RapidMiner*

Operator *Set Role* digunakan untuk mengubah peran dari salah satu atribut menjadi id seperti tabel berikut:

Tabel 1. Parameter Set Role

No.	Parameter	Information
1	Atribute name	No. Pesanan
2	Target role	Id

Tahap *selection data* selanjutnya yaitu menggunkan operator *Select Attributes*, operator select attributes dapat dilihat pada gambar berikut.



Gambar 4. Operator *Select Attributes* pada *RapidMiner*

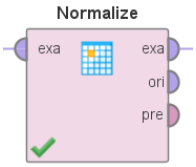
Pada *attribute filter type* atur menjadi *subset* dan dipilih atribut yang akan digunakan saja, seperti yang terlihat pada table dibawah.

Tabel 2. Parameter *Select Attributes*

No	Select Attributes	Information
1.	No. Pesanan	Special Atribut
2.	Biaya Administrasi	Regular Atribut
3.	Biaya Layanan (termasuk PPN 11%)	Regular Atribut
4.	Diskon Voucher Ditanggung Penjual	Regular Atribut
5.	Gratis Ongkir dari Shopee	Regular Atribut
6.	Harga Asli Produk	Regular Atribut
7.	Jasa Kirim	Regular Atribut
8.	Metode pembayaran pembeli	Regular Atribut
9.	Nama Kurir	Regular Atribut
10.	Ongkir yang Diteruskan oleh Shopee ke Jasa Kirim	Regular Atribut
11.	Ongkir Dibayar Pembeli	Regular Atribut
12.	Total Diskon Produk	Regular Atribut
13.	Total Penghasilan	Regular Atribut
14.	Username (Pembeli)	Regular Atribut
15.	Waktu Pesanan Dibuat	Regular Atribut

4.3. *Preprocessing*

Preprocessing Data, adalah proses perubahan data mentah kedalam format yang lebih mudah dipahami dan efisien. Preprocessing memiliki komponen kritikal yang tidak dapat diabaikan. Dalam tahap preprocessing pada penelitian ini digunakan *Normalize*. *Normalize* merupakan proses penting untuk menyesuaikan skala nilai atribut sehingga setiap atribut memiliki bobot yang sama dalam penelitian ini. Pada tahap ini digunakan operator *Normalize* seperti gambar dibawah.



Gambar 5. Operator Normalize

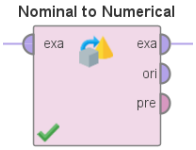
Untuk *Parameter Normalize* pada *attribute filter type* diatur menjadi *subset* dan dipilih atribut yang akan digunakan saja, seperti yang ada pada table dibawah.

Tabel 3. Parameter *Normalize*

No	Deskripsi	Informasi
1.	Attribute Filter type	Subset
2.	Selected attributes	Biaya Administrasi
3.		Biaya Layanan (termasuk PPN 11%)
4.		Diskon Voucher Ditanggung Penjual
5.		Gratis Ongkir dari Shopee
6.		Harga Asli Produk
7.		Ongkir Dibayar Pembeli
8.		Ongkir yang Diteruskan oleh Shopee ke Jasa Kirim
9.		Total Diskon Produk
10.		Total Penghasilan
11.	Method	Range transformation
12.	min	0.0
13.	max	1.0

4.4. Transformation Data

Dalam Langkah *Transformation Data*, untuk mengubah data yang bersifat nominal menjadi tipe numerik, digunakan operator *Nominal to Numerical* seperti yang terlihat pada gambar dibawah.



Gambar 5. Operator *Nominal to Numerical pada RapidMiner*

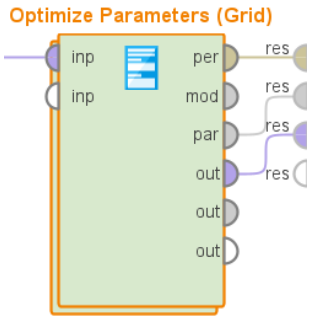
Untuk melihat parameter yang digunakan dalam operator *Nominal to Numerical* dapat dilihat tabel berikut.

Tabel 3. Parameter pada *Nominal to Numerical*

No.	Deskripsi	Informasi
1.	Attribute Filter type	subset
2.	Selected attributes	Jasa Kirim
3.		Metode Pembayaran Pembeli
4.		Nama Kurir
5.		No. Pesanan
6.		Username (Pembeli)
7.		Waktu Pesanan Dibuat

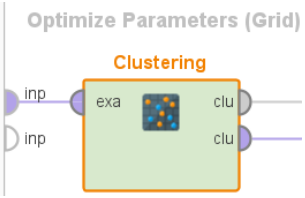
4.5. Data Mining

Operator yang digunakan yaitu operator Optimize Parameter (Grid), operator ini digunakan terlebih dulu sebelum menggunakan Operator K-Means. *Operator Optimize Parameter (Grid)* adalah Teknik yang digunakan untuk mencari kombinasi parameter terbaik untuk suatu model. Cara kerjanya yaitu dengan mencoba semua kombinasi parameter yang Mungkin dari parameter yang telah ditentukan. Operator Optimize parameter (Grid) dapat dilihat seperti berikut.



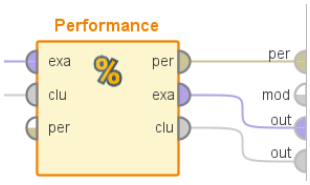
Gambar 6. Operator *Optimize Parameter (Grid)* pada RapidMiner

Dikarenakan operator Optimize Parameter (Grid) termasuk dalam kategori Operator SubProses, maka operator K-Means diintegrasikan didalam operator Optimize Parameter (Grid), seperti gambar berikut.



Gambar 7. Operator K-Means pada RapidMiner

Abaikan parameter operator K-Means, Setelah menggunakan operator K-Means. Langkah selanjutnya melibatkan penggunaan operator *Cluster Distance Performance*. Operator ini digunakan untuk mengevaluasi kinerja *cluster*ing, tujuan utamanya adalah mengukur sejauh mana objek dalam satu *cluster* serupa dan sejauh mana *cluster* berberda satu sama lain. Operator *Cluster Distance Performance* dapat dilihat pada gambar berikut.



Gambar 8. Operator Cluster Distance Performance pada RapidMiner

Abaikan parameter operator *Cluster* Distance Performance, Untuk melihat parameter yang digunakan pada operator Parameter Optimize Parameter (Grid), dapat dilihat pada table dibawah

Tabel 4. Parameter Optimize Parameter (Grid)

Parameter	Operators	Select Parameters	Value
Edit Parameter Setting	*Clustering* (K-Means)	*Clustering* k	Min: 2
			Max 20
			Steps: 100
			Scale: Linear
		Measure type	*MixedMeasures*
			Numerical Measures
			BregmanDivergences
	Performance (*Cluster* Distance Performance)	Max_optimization_steps	10
		Main_criterion	Davies Bouldin
		Normalize	True

4.6. Evaluasi

Setelah proses data mining telah dijalnkan sesuai tahapan dari *Knowledge Discovery in Database (KDD)* pada *software RapidMiner v.10.2.* maka berikut adalah evaluasi yang dilakukan terhadapa hasil eksperimen dari dataset penjualan pakian diperoleh dari *BregmanDivergences* menghasilkan *Davies Bouldin* sebagai berikut.

Tabel 5. Evaluasi hasil *Optimize Parameter Grid*

Iteration	*Clustering.* K	main_criterion	Measure_type	Davies Bouldin
1	3	Davies Bouldin	BregmanDivergences	0,035
2	4	Davies Bouldin	NumericalMeasures	0,036
1	5	Davies Bouldin	MixedMeasures	0,036
9	9	Davies Bouldin	MixedMeasures	0,036
4	6	Davies Bouldin	NumericalMeasures	0,036
7	10	Davies Bouldin	NumericalMeasures	0,036
7	7	Davies Bouldin	BregmanDivergences	0,036
5	8	Davies Bouldin	BregmanDivergences	0,036
4	2	Davies Bouldin	NumericalMeasures	0,036
1	2	Davies Bouldin	BregmanDivergences	0,037
12	10	Davies Bouldin	BregmanDivergences	0,038

ParameterSet

```
Parameter set:

Performance:
PerformanceVector [
-----Avg. within centroid distance: -26006.424
-----Avg. within centroid distance_cluster_0: -27036.460
-----Avg. within centroid distance_cluster_1: -29490.614
-----Avg. within centroid distance_cluster_2: -20301.790
*****Davies Bouldin: -0.035
]
Clustering.k    = 3
Performance.main_criterion      = Davies Bouldin
Clustering.measure_types        = BregmanDivergences
Performance.normalize   = true
Clustering.max_optimization_steps       = 1
```

Gambar 9. parameterSet *cluster*

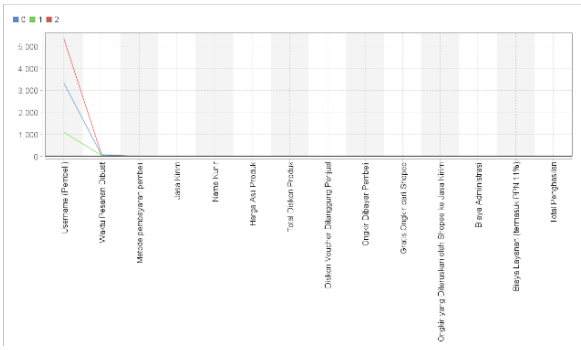
Setelah diketahui pembagian *cluster* dari *cluster* model, hasil evaluasi menggunakan *Davies Bouldin Index (DBI),* diperoleh nilai *cluster optimal* sebesar 0,035 yang terdapat pada K=3. Seperti yang ada Di gambar 9. Berdasarkan hasil *cluster* model yang pada cluster dengan nilai DBI terendah yaitu, terbentuk menjadi k=3, dan pembagian hasil *clustering* terlihat pada gambar berikut:

Cluster Model

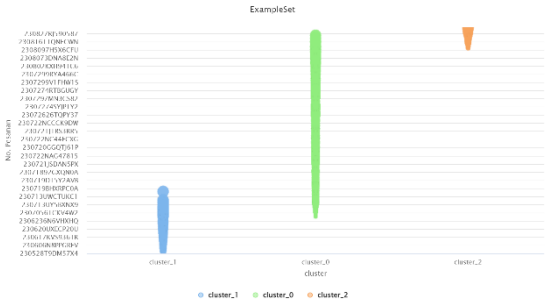
```
Cluster 0: 2859 items
Cluster 1: 2392 items
Cluster 2: 1695 items
Total number of items: 6946
```

Gambar 10. Hasil *Cluster Model*

Setelah diketahui pembagian *cluster* dari *cluster* model, dapat dilihat grafik plot berdasarkan *cluster 1*, *cluster 2* dan *cluster 3* dengan setiap atribut seperti pada gambar berikut.



Gambar 11. Grafik *Plot Cluster Model*



Gambar 12. Visualisasi pembagian *cluster*

Dari grafik plot diketahui presentasi *cluster 0*, *cluster 1* dan *cluster 2* yang paling mempengaruhi dalam bentuk *cluster*, terdapat pada atribut Username (Pembeli) dan Waktu Pemesanan Dibuat yang memiliki perbandingan yang sangat signifikan dibandingkan dengan atribut lainnya.

5. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian Implementasi *Data Mining* Pada Data Penjualan Pakaian Menggunakan Algoritma *K-Means* Dengan *Optimize Parameter Grid*, maka dapat ditarik kesimpulan sebagai berikut:

Berdasarkan hasil penelitian Implementasi Data Mining Pada Data Penjualan Pakaian Menggunakan Algoritma K-Means Dengan Optimize Parameter Grid maka dapat ditarik kesimpulan sebagai berikut: Jumlah *Cluster* optimal berdasarkan *Davies Bouldin Index* terendah, yaitu pada K=3, dengan Dbi= 0,035. *Parameter measure type* yang berpengaruh untuk menghasilkan nilai *cluster* terbaik adalah *Mixed Measure BregmanDivergences*, Dbi= 0,035. jumlah iterasi yang diperlukan oleh *algoritma K-Means* untuk mencapai konvergensi adalah pada iterasi ke-1.

Berdasarkan hasil *cluster*ing, terdapat 3 *cluster* yang terbentuk dari keseluruhan data yang berjumlah 6946. *Cluster* 0 memiliki sebanyak 2859 item atau skitar 41,15% dengan transaksi pada bulan Juli sebanyak 2757, bulan Agustus 99 transaksi, dan bulan september 3 transaksi dari total 2859 transaksi. Sedangkan *cluster* 1 memiliki sebanyak 2392 item atau sekitar 34,44% dengan transkasi pada bulan Mei sebanyak 12, bulan juni sebanyak 1091, bulan juli sebanyak 1235, lalu bulan Agustus 51 dan bulan September 3 transaksi. *Cluster* 2 memiliki 1695 item atau sekitar 24,41% dengan transaksi pada bulan juli sebanyak 543, bulan Agustus sebanyak 1104, dan bulan September 48 transaksi.

Dalam pengembangan hasil penelitian ini, perlu dilakukan analisis lebih lanjut terkait variabel yang mungkin memiliki dampak signifikan yang belum sepenuhnya dijelaskan. Validasi data menjadi langkah krusial untuk memastikan integritas data dan mengidentifikasi variabilitas yang perlu diperhatikan. Disarankan juga untuk mempertimbangkan pengembangan atau perbaikan model yang digunakan, dengan memilih metode analisis yang lebih canggih atau model yang lebih kompleks. Kesimpulan perlu lebih diperinci, khususnya dalam merinci implikasi praktis dari temuan penelitian dan memberikan rekomendasi aksi yang konkret. Ajakan untuk kolaborasi dalam penelitian lanjutan dapat meningkatkan pemahaman holistik, dan diseminasi informasi hasil penelitian kepada komunitas ilmiah dan praktisi terkait akan mendukung pengembangan pengetahuan di bidang ini secara lebih luas.

DAFTAR PUSTAKA

[1] A. M. Almaududi Ausat, E. S. Astuti, and W. Wilopo, "Analisis Faktor yang Berpengaruh pada Adopsi E-Commerce dan Dampaknya Bagi Kinerja UKM di Kabupaten Subang," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 9, no. 2, p. 333, 2022, doi: 10.25126/jtiik.2022925422.

[2] M. Solihat and D. Sandika, "e-COMMERCE DI INDUSTRI 4.0," *J. Ilm. Bisnis dan Ekon. Asia*, vol. 16, no. 2, pp. 273–281, 2022, doi: 10.32815/jibeka.v16i2.967.

[3] R. Siagian, P. S. Pahala Sirait, and A. Halima, "E-Commerce Customer Segmentation Using K-Means Algorithm and Length, Recency, Frequency, Monetary Model," *J. Informatics Telecommun. Eng.*, vol. 5, no. 1, pp. 21–30, 2021, doi: 10.31289/jite.v5i1.5182.

[4] S. Rahman, F. Fadrul, Y. Yusrizal, R. Marlyna, and M. M. Momin, "Improving the Satisfaction and Loyalty of Online Shopping Customers Based on E-Commerce Innovation and E-Service Quality," *Gadjah Mada Int. J. Bus.*, vol. 24, no. 1, pp. 56–81, 2022, doi: 10.22146/gamaijb.58783.

[5] M. Mardalius and T. Christy, "Mapping of Potential Customers As a Clothing Promotion Strategy Using K-Means Clustering Algorithm," *… J. Ilmu Pengetah. Dan …*, vol. 6, no. 1, pp. 67–72, 2020, doi: 10.33480/jitk.v6i1.1414.

[6] A. Wibowo, Moh Makruf, Inge Virdyna, and Farah Chikita Venna, "Penentuan Klaster Koridor TransJakarta dengan Metode Majority Voting pada Algoritma Data Mining," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 3, pp. 565–575, 2021, doi: 10.29207/resti.v5i3.3041.

[7] T. Hardiani, "Analisis Clustering Kasus Covid 19 di Indonesia Menggunakan Algoritma K-Means," *J. Nas. Pendidik. Tek. Inform.*, vol. 11, no. 2, pp. 156–165, 2022, doi: 10.23887/janapati.v11i2.45376.

[8] Gustiendina, Mh. Adiya, and Y. Desnelita, "Jurnal Nasional Teknologi dan Sistem Informasi Attribution-NonCommercial 4.0 International. Some rights reserved Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan Pada RSUD Pekanbaru," *J. Nas. Teknol. dan Sist. Inf.*, vol. 5, no. 1, pp. 17–24, 2019, [Online]. Available: https://teknosi.fti.unand.ac.id/index.php/teknosi/article/view/773

[9] S. Ramadhani, D. Azzahra, and T. Z, "Comparison of K-Means and K-Medoids Algorithms in Text Mining based on Davies Bouldin Index Testing for Classification of Student's Thesis," *Digit. Zo. J. Teknol. Inf. dan Komun.*, vol. 13, no. 1, pp. 24–33, 2022, doi: 10.31849/digitalzone.v13i1.9292.

[10] I. F. Ashari, E. Dwi Nugroho, R. Baraku, I. Novri Yanda, and R. Liwardana, "Analysis of Elbow, Silhouette, Davies-Bouldin, Calinski-Harabasz, and Rand-Index Evaluation on K-Means Algorithm for Classifying Flood-Affected Areas in Jakarta," *J. Appl. Informatics Comput.*, vol. 7, no. 1, pp. 89–97, 2023, doi: 10.30871/jaic.v7i1.4947.

[11] E. Luthfi and A. W. Wijayanto, "Analisis perbandingan metode hirearchical, k-means, dan k-medoids clustering dalam pengelompokkan indeks pembangunan manusia Indonesia," *Inovasi*, vol. 17, no. 4, pp. 761–773, 2021, doi: 10.30872/jinv.v17i4.10106.

[12] M. P. A. Ariawan, I. B. A. Peling, and G. B. Subiksa, "Prediksi Nilai Akhir Matakuliah Mahasiswa Menggunakan Metode K-Means Clustering (Studi Kasus : Matakuliah Pemrograman Dasar)," *J. Nas. Teknol. dan Sist. Inf.*, vol. 9, no. 2, pp. 122–131, 2023, doi: 10.25077/teknosi.v9i2.2023.122-131.

[13] T. R. Rivanthio and M. Ramdhani, "Penerapan Teknik Clustering Data Mining untuk Memprediksi Kesesuaian Jurusan Siswa (Studi Kasus SMA PGRI 1 Subang)," *Fakt. Exacta*, vol. 13, no. 2, p. 125, 2020, doi: 10.30998/faktorexacta.v13i2.6588.

[14] F. Ros, R. Riad, and S. Guillaume, "PDBI: A partitioning Davies-Bouldin index for clustering evaluation," *Neurocomputing*, vol. 528, pp. 178–199, 2023, doi: 10.1016/j.neucom.2023.01.043.

[15] X. Wang, Z. Shao, Y. Shen, and Y. He, "Research on fast marking method for indicator diagram of pumping well based on K-means clustering," *Heliyon*, vol. 9, no. 10, p. e20468, 2023, doi: 10.1016/j.heliyon.2023.e20468.

[16] A. Dogan and D. Birant, "Machine learning and data mining in manufacturing," *Expert Syst. Appl.*, vol. 166, p. 114060, 2021, doi: 10.1016/j.eswa.2020.114060.

[17] W. Liu, P. Zou, D. Jiang, X. Quan, and H. Dai, "Zoning of reservoir water temperature field based on K-means clustering algorithm," *J. Hydrol. Reg. Stud.*, vol. 44, no. June, p. 101239, 2022, doi: 10.1016/j.ejrh.2022.101239.

[18] M. A. Mahdi, K. M. Hosny, and I. Elhenawy, "Scalable Clustering Algorithms for Big Data: A Review," *IEEE Access*, vol. 9, pp. 80015–80027, 2021, doi: 10.1109/ACCESS.2021.3084057.

[19] N. H. M. M. Shrifan, M. F. Akbar, and N. A. M. Isa, "An adaptive outlier removal aided k-means clustering algorithm," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 34, no. 8, pp. 6365–6376, 2022, doi: 10.1016/j.jksuci.2021.07.003.

[20] D. Ramdhan, G. Dwilestari, R. D. Dana, A. Ajiz, and K. Kaslani, "Clustering Data Persediaan Barang Dengan Menggunakan Metode K-Means," *MEANS (Media Inf. Anal. dan Sist.*, vol. 7, no. 1, pp. 1–9, 2022, doi: 10.54367/means.v7i1.1826.

[21] Y. A. Wijaya, D. A. Kurniady, E. Setyanto, W. S. Tarihoran, D. Rusmana, and R. Rahim, "Davies

Bouldin Index Algorithm for Optimizing Clustering Case Studies Mapping School Facilities," *TEM J.*, vol. 10, no. 3, pp. 1099–1103, 2021, doi: 10.18421/TEM103-13.

[22] B. Surarso and R. Gernowo, "Implementation Of K-Medoids Clustering For High Education Accreditation Data," *J. Ilm. KURSOR*, vol. 10, no. 3, pp. 119–128, 2020.

[23] I. F. Ashari, R. Banjarnahor, D. R. Farida, S. P. Aisyah, A. P. Dewi, and N. Humaya, "Application of Data Mining with the K-Means Clustering Method and Davies Bouldin Index for Grouping IMDB Movies," *J. Appl. Informatics Comput.*, vol. 6, no. 1, pp. 07–15, 2022, doi: 10.30871/jaic.v6i1.3485.

[24] J. Eska, M. Fitri Larasati, P. Studi Sistem Informasi, and S. Tinggi Manajemen Informatika dan Komputer Royal Kisaran, "Application of K-Means Clustering Method To Cluster Students' English Skill Jason English Course," *J. Tek. Inform.*, vol. 3, no. 3, pp. 479–485, 2022, [Online]. Available: https://doi.org/10.20884/1.jutif.2022.3.3.167

[25] D. Sinta Saputri, G. Maha Putra, M. Fitri Larasati, P. Studi Sitem Informasi, and S. Tinggi Manajemen Informatika dan Komputer Royal Kisaran, "Implementation of the K-Means Clustering Algorithm for the Covid-19 Vaccinated Village in the Ujung Padang Sub-District," *J. Tek. Inform.*, vol. 3, no. 2, pp. 261–267, 2022, [Online]. Available: https://doi.org/10.20884/1.jutif.2022.3.2.165

[26] S. Handoko, F. Fauziah, and E. T. E. Handayani, "Implementasi Data Mining Untuk Menentukan Tingkat Penjualan Paket Data Telkomsel Menggunakan Metode K-Means Clustering," *J. Ilm. Teknol. dan Rekayasa*, vol. 25, no. 1, pp. 76–88, 2020, doi: 10.35760/tr.2020.v25i1.2677.