

KLASTERISASI DATA PENJUALAN ALAT TRANSPORTASI DENGAN RAPIDMINER MENGGUNAKAN METODE K-MEDOID

Wiwin Aprilyani¹, Kaslani², Edi Wahyudin³, Ryan Hamonangan⁴, Ruli Herdiana⁵

¹²³ Program Studi Komputerisasi Akuntansi, STMIK IKMI Cirebon

⁴⁵ Program Studi Teknik Informatika, STMIK IKMI Cirebon

Jalan Perjuangan No. 10B Karyamulya Kec. Kesambi Kota Cirebon, Jawa barat 45131

wiwinaprilyani1204@gmail.com

ABSTRAK

Penelitian ini bertujuan untuk menerapkan metode klusterisasi *K-medoid* menggunakan platform Rapidminer pada data penjualan alat transportasi. Melalui Rapidminer sebuah platform analisis data yang kuat, hal ini bertujuan untuk meningkatkan pemahaman terhadap preferensi dan perilaku pembelian konsumen dalam kaitannya dengan transportasi. Metode *K-medoid* dipilih karena kemampuannya dalam menghasilkan kelompok-kelompok data yang berbeda secara optimal tanpa bergantung pada bentuk atau jumlah kluster yang telah ditentukan sebelumnya. Setelah dilakukannya beberapa uji coba kluster pada data penjualan alat transportasi dengan record data terakhir berjumlah 904 record data menghasilkan nilai kluster yang terbaik yaitu dengan 2 kluster yang nilai *Davies Bouldin Index* berada pada -0,838 yang dimana nilai tersebut nilai terkecil diantara hasil uji coba kluster lainnya, dengan jumlah kluster 2 tersebut penjualan tertinggi pada perusahaan ini berada pada tahun 2005 dengan nilai 11,739 dan berada pada di kluster 0.

Kata kunci: *Penjualan, K-medoid, Klusterisasi, Rapidminer, Alat transportasi*

1. PENDAHULUAN

Dalam era digital ini penggunaan teknologi telah menjadi landasan utama bagi perkembangan bisnis di berbagai sector, termasuk industry Alat Transportasi. Data merupakan salah satu asset paling berharga bagi perusahaan untuk mengidentifikasi peluang dan menghadapi tantangan dalam lingkungan bisnis yang terus berubah. Khususnya dalam industry transportasi pemahaman yang mendalam terhadap pola penjualan Alat Transportasi dapat menjadi kunci keberhasilan perusahaan, Penjualan Alat Transportasi yang mencakup berbagai kendaraan seperti mobil, sepeda motor, pesawat, kapal, dan sebagainya adalah bagian integral dari ekosistem transportasi. Pengolahan data transaksi penjualan dengan komputasi tepat dapat memberikan manfaat bagi perusahaan agar penjualan lebih efektif dan efisien [1]. Salah satu Teknik perhitungan yang dapat diterapkan untuk menangani permasalahan tersebut diatas adalah dengan memanfaatkan Teknik data mining dengan menggunakan metode *Clustering*[2].

Berdasarkan data yang diperoleh tersebut dalam penjualan ini terjadi naik turun nya penjualan dalam tiap tahunnya, dalam proses itu juga akan mempengaruhi stok persediaan penjualan yang tersedia karena tidak tahu barang apa yang sedang laku keras di pasaran tersebut, maka dari itu dengan menggunakan algoritma *k-medoid*, akan tahu penjualan mana yang terbanyak pada tahun tersebut dengan memanfaatkan hasil klusterisasi itu dapat diketahui manajemen persediaan serta sumber daya yang terkait dengan penjualan alat transportasi tersebut.

Penelitian yang serupa yang pernah dilakukan seperti yang dilakukan oleh Abizar dkk pada tahun

2022 [3] yaitu Penerapan Algoritma *K-medoids* dalam Klusterisasi Penjualan Laptop. Penelitian tersebut bertujuan untuk mengelompokkan data penjualan laptop dengan menggunakan metode data mining algoritma *K-medoids clustering*. Data penjualan Laptop dikelompokkan berdasarkan kemiripan data tersebut sehingga data dengan karakteristik yang sama akan berada dalam 1 *cluster*. Berdasarkan perhitungan yang telah dilakukan oleh peneliti pada 1000 sample data dapat dikategorikan dalam 3 *cluster*. *Cluster 1* merupakan kategori penjualan laptop rendah yaitu 137 dan, kemudian *cluster 2* merupakan kategori penjualan laptop tertinggi yaitu 669 data dan *cluster 3* merupakan kategori penjualan laptop sedang yaitu 194 data dari 10 kategori penjualan laptop.

Dapat disimpulkan bahwa pengelompokan data penjualan laptop dengan menggunakan algoritma *k-medoid* sendiri bisa dilakukan untuk pengelompokan penjualan laptop, dikarenakan ingin mencari tahu penjualan terbanyak dari penjualan laptop itu maka dilakukannya *clustering* dengan menggunakan *k-medoids* dengan hasil yang diperoleh yaitu terdapat pada *cluster 2* merupakan paling banyak terjual karena spesifikasi dan harga laptop yang lebih terjangkau dibanding *cluster 1* dan *cluster 3*.

Berdasarkan uraian sebelumnya maka perlu dilakukan pengelompokan data pada penjualan alat transportasi ini dengan menggunakan *clustering*. *Clustering* sendiri ialah pengelompokan yang setiap kelompoknya memiliki karakteristik sendiri. Dengan menggunakan algoritma *k-medoid* penjualan alat transportasi tersebut akan terkelompok dengan nilai *cluster* yang menyesuaikan dengan nilai *Davies Bouldin Index* paling terkecil atau mendekati nilai 0.

2. TINJAUAN PUSTAKA

2.1. Data Mining

Penggalian data (Bahasa Inggris: data *mining*) adalah ekstraksi pola yang menarik dari data dalam jumlah besar. Suatu pola dikatakan menarik apabila pola tersebut tidak sepele, implisit, tidak diketahui sebelumnya dan berguna[4]. Ini melibatkan penggunaan Teknik komputasional untuk menyaring melalui data dalam jumlah besar dengan tujuan untuk mengidentifikasi pola yang konsisten, tren atau hubungan yang dapat digunakan untuk pengambilan keputusan atau prediksi di masa depan. Pemilihan algoritma ini harus sesuai dengan keseluruhan proses KDD (*Knowledge Discovery Databases*)[5].

KDD sendiri dapat dikatakan penyelesaian masalah dengan menganalisa data yang ada pada database dengan data tersimpan secara elektronik dan pencariannya dilakukan otomatis seperti pada komputer[6]. Alasan utama mengapa data mining sangat menarik perhatian industri informasi dalam beberapa tahun terakhir adalah karena ketersediaan data dalam jumlah yang sangat besar dan kebutuhan untuk mengubah data tersebut menjadi informasi dan pengetahuan berguna yang selaras dengan focus bidang ilmu yaitu melakukan kegiatan ekstraksi atau menambang pengetahuan dari data yang berukuran besar, informasi inilah yang nantinya akan berguna untuk pengembangan.[1]

Data mining sendiri banyak digunakan di berbagai industri seperti keuangan, pemasaran, ilmu pengetahuan, Kesehatan dan lainnya untuk memperlancar proses mengidentifikasi tren pasar, melakukan analisis risiko, dan memprediksi perilaku di masa depan. Keberhasilan dalam data mining sendiri tidak hanya bergantung pada algoritma yang digunakan, tetapi juga pada pemahaman yang mendalam tentang konteks domain dan kemampuan untuk menginterpretasikan hasilnya dengan benar.

2.2. Clustering

Clustering ialah Teknik analisis data yang digunakan untuk mengelompokkan sekumpulan data kedalam subset-subset yang mempunyai kesamaan karakteristik. Metode clustering ini memanfaatkan kesamaan antara objek-objek, dimana objek-objek yang mirip ditempatkan di tempat yang sama. Berbeda dengan metode klasifikasi, *clustering* tidak memiliki target variable[8]. Beragam metode yang dapat digunakan untuk mengukur nilai kesamaan objek yang akan dibandingkan, salah satunya ialah *Euclidean Distance*. *Euclidean Distance* ialah untuk menghitung jarak dua buah titik dengan mengetahui nilai jarak masing-masing atribut pada titik tersebut[9].

2.3. K-Medoid

Salah satu metode pengelompokan partisi yang biasa dikenal dengan *K-Medoid* bertujuan untuk menemukan satu set *K-cluster* diantara data yang memiliki objek yang paling menonjol dalam kumpulan data tersebut[9]

2.4. Rapid Miner

Rapid Miner sendiri ialah salah satu platform perangkat lunak yang menyediakan pemeriksaan penambangan informasi, penambangan teks dan analitik prediktif. *Rapid Miner* menggunakan Teknik deksriptif dan prediktif untuk menyajikan informasi kepada pemakai sehingga mereka dapat mengambil kesimpulan yang terbaik[10]. *Rapid Miner* ini sendiri dibuat dan ditulis menggunakan Bahasa pemrograman Java, maka dari itu *Rapid Miner* ini bisa dijalankan di berbagai macam system perangkat lunak.

3. METODE PERANCANGAN

3.1. Sumber Data

Sumber data penelitian ini berasal dari website public, data tersebut termasuk ke dalam data sekunder, data sekunder sendiri ialah data yang diperoleh peneliti dari sumber yang ada. Data yang diperoleh tersebut dipublikasikan oleh (<https://www.kaggle.com/datasets/kyanyoga/sample-sales-data>) sebagai salah satu portal data terbuka yang menyediakan data untuk diuji coba oleh pengguna dan bisa untuk menambah wawasan dari hasil uji coba tersebut. Data set yang diambil yaitu data penjualan Alat transportasi periode tahun 2003-2005 yang diambil pada tanggal 11 Agustus 2023 sebagai dataset pendukung pada penelitian ini.

3.2. Teknik Pengumpulan Data

Teknik pengumpulan data yang dilakukan ialah termasuk kedalam dokumentasi, data tersebut diperoleh dengan cara mengunduh data tersebut dari website (www.kaggle.com) yang sudah disediakan oleh pihak Kaggle.

3.3. Tahapan Perancangan



Gambar 1. Proses KDD

Tahapan perancangan yang digunakan dalam penelitian ini menggunakan rancangan Knowledge Discovery in Database. Berikut adalah tahapan perancangan KDD yang akan dilakukan:

3.3.1. Proses data selection

Pada proses data *selection* ini mempersiapkan data penjualan Alat transportasi ini yang berisi daftar transaksi penjualan. Pada data *selection* ini akan memilih data mana saja yang akan digunakan untuk proses *clustering*. Data mentah sebelum dilakukannya cleaning data berjumlah sebanyak 1000 record dan 26 atribut.

3.3.2. Pre-Processing

Dalam tahapan *Pre-processing* ini yaitu mengumpulkan data transaksi penjualan kemudian melakukan pembersihan data dari data yang kosong, terduplikat, atau data yang hilang, dan melakukan transformasi data agar sesuai dengan algoritma *clustering*.

3.3.3. Data Mining

Memasukan algoritma yang akan digunakan kedalam *tools rapidminer*, lalu masukan data yang sudah dilakukan pembersihan diawal tersebut, kemudian tentukan nilai *k* yang sehingga mendapatkan hasil DBI yang paling kecil atau terbaik dari hasil mengcluster tersebut. Kemudian interpretasi dan mengevaluasi hasil dari nilai DBI yang terbaik dari yang dihasilkan.

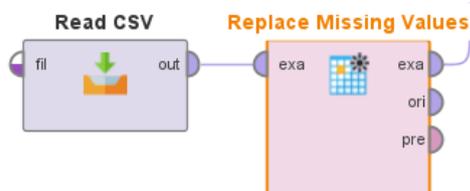
4. HASIL DAN PEMBAHASAN

4.1. Menerapkan data mining pada data penjualan Alat Transportasi dengan menggunakan metode Algoritma K-medoids.

ORDERNUM...	QUANTITYO...	PRICEEACH	ORDERLIN...	SALES	ORDERDATE	STATUS	QTR_ID
integer	integer	polynomial	integer	real	date	polynomial	integer
10223	37	100	1	3665.600	Feb 20, 2004	Shipped	1
10237	23	100	7	2333.120	Apr 5, 2004	Shipped	2
10361	20	72.55	13	1451.000	Dec 17, 2004	Shipped	4
10163	21	100	1	4860.240	Oct 20, 2003	Shipped	4
10270	21	100	9	4905.390	Jul 19, 2004	Shipped	3
10347	30	100	1	3844.700	Nov 29, 2004	Shipped	4
10120	29	96.34	3	2793.860	Apr 29, 2003	Shipped	2
10223	47	100	4	5422.390	Feb 20, 2004	Shipped	1
10361	26	51.15	8	1329.900	Dec 17, 2004	Shipped	4
10120	46	100	2	9284.860	Apr 29, 2003	Shipped	2
10223	49	100	3	9774.030	Feb 20, 2004	Shipped	1
10237	39	100	9	7023.900	Apr 5, 2004	Shipped	2
10237	25	48.05	8	1201.250	Nov 21, 2004	Shipped	4
10217	48	100	4	7020.480	Feb 4, 2004	Shipped	1
10258	26	100	12	4023.380	Jun 15, 2004	Shipped	2
10270	32	100	2	4302.080	Jul 19, 2004	Shipped	3
10347	27	100	2	4428.000	Nov 29, 2004	Shipped	4
10373	45	100	7	4439.900	Dec 5, 2004	Shipped	3

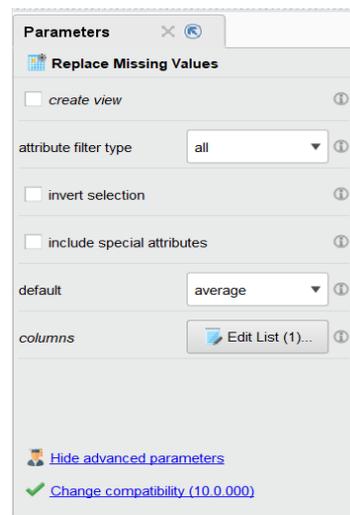
Gambar 2. Proses pemilihan data

Proses pemilihan data penjualan alat transportasi ini berlangsung di dalam tool RapidMiner 4.1.2. Pada gambar 2 diatas ialah proses awal dimasukkannya data penjualan Alat transportasi ke dalam *tools RapidMiner* dalam bentuk Read CSV. Data diatas masi dalam bentuk data mentah, maka dari itu harus dilakukan pembersihan pada data yang missing dengan menggunakan operator bernama *Replace Missing Values*.



Gambar 3. Replace Missing Values

Pada gambar 3 diatas dengan menggunakan *tools Replace Missing Values* bisa mempermudah untuk menghilangkan data yang *missing* atau terduplikat pada atribut tertentu yang terdapat pada data penjualan akan difilter missingnya terdapat di bagian *columns* di dalam *edit list*.

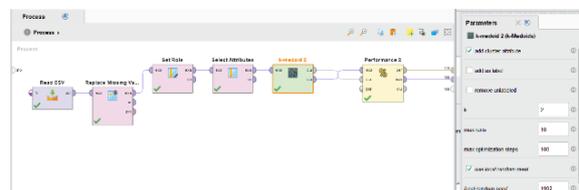


Gambar 4. Parameters replace missing values

Pada gambar 4 adalah tahap dimana untuk mengetahui nilai DBI yang dihasilkan oleh nilai *k* yang diuji pada parameter *k-medoids*. Parameter yang digunakan ini ialah untuk menentukan kriteria yang digunakan untuk evaluasi kerja, dan dalam "*main criterion*" ialah "*Average within centroid*".

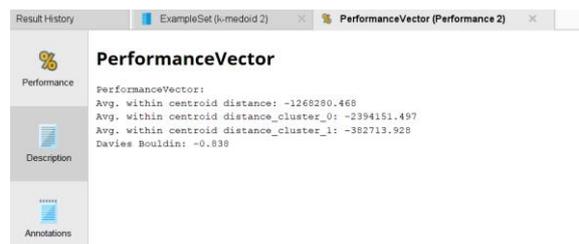
4.2 Mengetahui berapa hasil evaluasi terbaik dari cluster yang dihasilkan

Pada hasil di tahap ini akan berisikan beberapa uji *cluster* dan mendapatkan *cluster* dan nilai DBI terbaik. Dimulai dengan nilai *k* yang paling kecil yaitu 2 cluster.



Gambar 5. Uji cluster

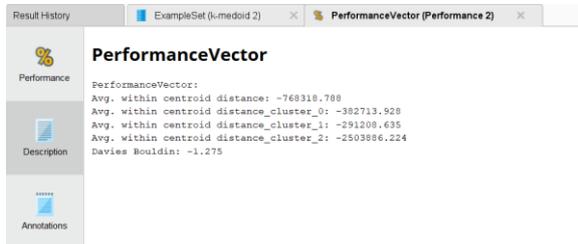
Pada gambar 5 adalah uji coba dengan menggunakan nilai *k* yang paling kecil yaitu 2 dan menggunakan parameter "*use local random seed*" untuk pengacakan nilai *cluster* yang akan diuji coba, pengacakan sendiri disini dapat digunakan untuk memilih nilai *k* berbeda pada awal algoritma.



Gambar 6. Hasil Runing performance dengan cluster 2

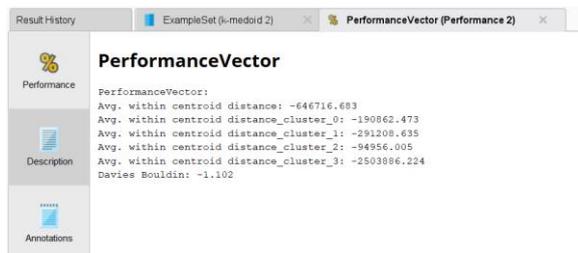
Pada gambar 6 diatas hasil running performance pada algoritma *k-medoid*. *Cluster* yang dihasilkan

ialah 2 yaitu cluster_0 dan cluster_1 dengan nilai jarak cluster_0 yaitu -2394151.497 dan nilai jarak cluster_1 yaitu -382713.928 dengan nilai *Davies Bouldin Index* -0,838.



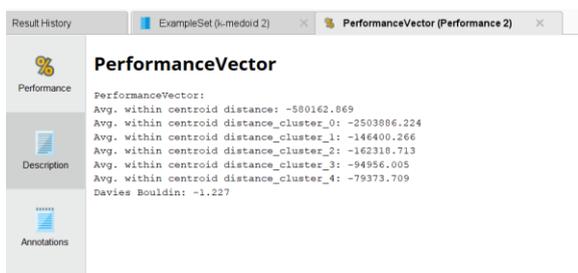
Gambar 7. Hasil Runing performance dengan cluster 3

Pada gambar 7 diatas menggunakan nilai k 3 dengan cluster yang dihasilkan ialah cluster_0, cluster_1 dan cluster_2 dengan nilai jarak masing masing cluster ialah cluster_0 -382713.928, cluster_1 -291208.635, cluster_2 -2503886.224, dengan nilai *Davies Bouldin Index* -1,275 yang diartikan nilai tersebut jauh dari nilai angka mendekati 0 atau yang paling terkecil. Maka dengan nilai k 3 tidak dapat dikatakan sebagai hasil cluster terbaik karena hasil *Davies Bouldin Index* jauh dari angka 0.



Gambar 8. Hasil Runing performance dengan cluster 4

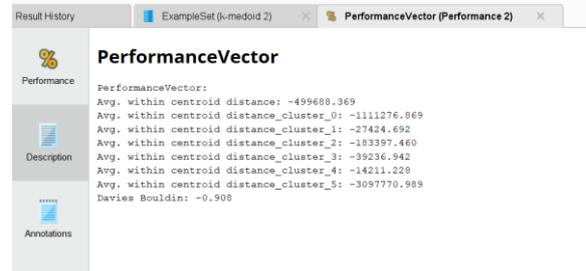
Pada gambar 8 diatas menggunakan nilai k 4 dengan cluster yang dihasilkan ialah cluster_0 -190862.473, cluster_1 -291208.635, cluster_2 -94956.005, cluster_3 -2503886.224. Nilai *Davies Bouldin Index* yang dihasilkan ialah -1.102 lebih kecil dari hasil cluster 3 tapi hasilnya cukup besar untuk mendapatkan hasil yang mendekati angka 0.



Gambar 9. Hasil Runing performance dengan cluster 5

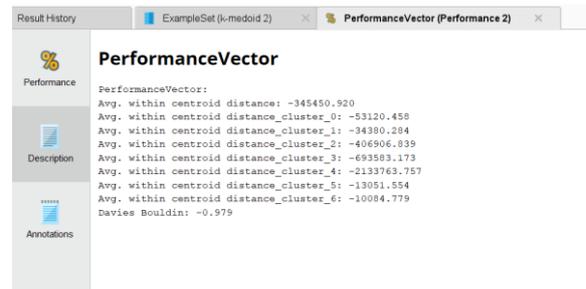
Pada gambar 9 diatas menggunakan nilai k 5 dan cluster yang dihasilkan dengan nilai jarak masing-masing yaitu cluster_0 -2503886.224, cluster_1 -1464400.266, cluster_2 -162318.713, cluster_3 -94956.005, cluster_4 -79373.709. dengan nilai *Davies*

Bouldin Index -1.227 yang diartikan bahwa cluster dengan nilai k 5 belum bisa dikatakan dengan cluster terbaik.



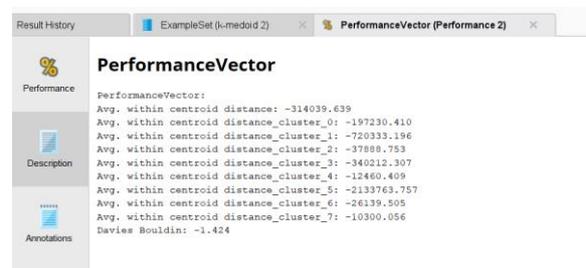
Gambar 10. Hasil Runing performance dengan cluster 6

Pada gambar 10 diatas nilai k yang digunakan ialah 6 yang nilai jarak masing-masing cluster adalah cluster_0 -1111276.869, cluster_1 -27424.692, cluster_2 -183397.460, cluster_3 -39236.942, cluster_4 -14211.228, cluster_5 -3097770.989, dengan nilai *Davies Bouldin Index* yaitu -0,908 lebih kecil dari cluster sebelumnya tetapi masih cukup besar dari cluster 2.



Gambar 11. Hasil Runing performance dengan cluster 7

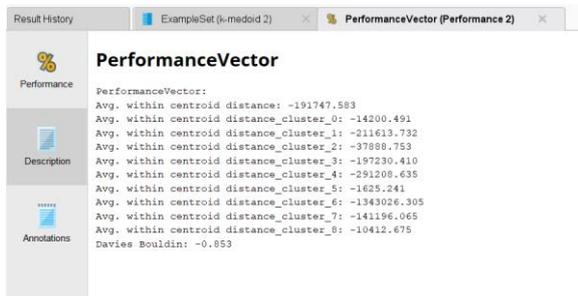
Pada gambar 11 diatas menggunakan nilai k 7 yang masing-masing mempunyai nilai jarak yaitu cluster_0 -53120.458, cluster_1 -34380.284, cluster_2 -406906.839, cluster_3 -693583.173, cluster_4 -2133763.757, cluster_5 -13051.554, cluster_6 -10084.779 dengan nilai *Davies Bouldin Index* yaitu -0.979 masi belum bisa dikatakan cluster terbaik karena masi cukup besar dari cluster 2.



Gambar 12. Hasil Runing performance dengan cluster 8

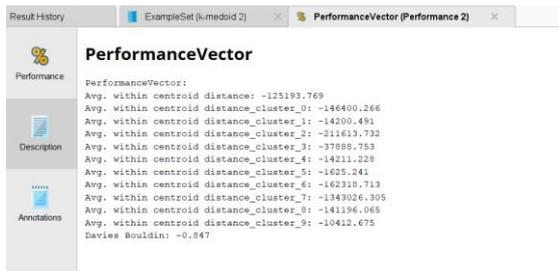
Pada gambar 12 diatas menggunakan nilai k 8 dan masing-masing memiliki nilai jaraknya yaitu dimulai dengan cluster_0 -197230.410, cluster_1 -720333.196, cluster_2 -37888.753, cluster_3 -340212.307, cluster_4 -12460.409, cluster_5 -

2133763.757, cluster_6 -26139.505, cluster_7 - 10300.056 dengan nilai *Davies Bouldin Index* yaitu -1.424 dengan nilai tersebut maka *cluster* dengan nilai k 8 belum bisa dikatakan sebagai *cluster* terbaik.



Gambar 13. Hasil Runing performance dengan cluster 9

Pada gambar 13 diatas menggunakan nilai k 9 *cluster* tersebut menghasilkan nilai jaraknya masing-masing diantaranya yaitu cluster_0 -14200.491, cluster_1 -211613.732, cluster_2 -37888.753, cluster_3 -197230.410, cluster_4 -291208.635, cluster_5 -1625.241, cluster_6 -1343026.305, cluster_7 -141196.065, cluster_8 -10412.675 dengan nilai *Davies Bouldin Index* yaitu -0,853 *cluster* dengan nilai 9 ini mendekati *cluster* terbaik, tetapi nilai yang dihasilkan masi cukup besar dari nilai DBI dengan nilai k 2.



Gambar 14. Hasil Runing performance dengan cluster 10

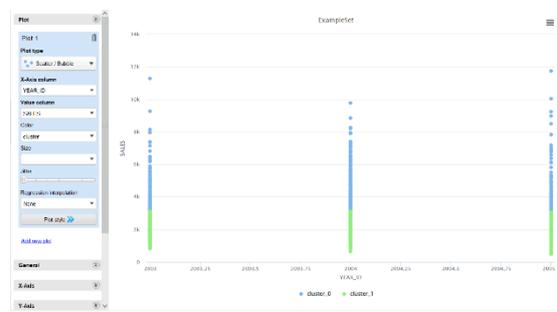
Pada gambar 14 diatas menggunakan nilai k 10 masing-masing *cluster* yang dihasilkan memiliki nilai jaraknya yaitu dimulai dari cluster_0 -146400.266, cluster_1 -14200.491, cluster_2 -211613.732, cluster_3 -37888.753, cluster_4 -14211.228, cluster_5 -1625.241, cluster_6 -162318.713, cluster_7 -1343026, cluster_8 -141196.065, cluster_9 -10412.675 dengan nilai *Davies Bouldin Index* -0,847 hasilnya masi belum bisa dikatakan *cluster* terbaik dikarenakan nilai *Davies Bouldin Index* yang dihasilkan pada nilai k 10 cukup besar dari nilai *Davies Bouldin Index* yang dihasilkan pada nilai k 2.

Berikut table hasil *Davies Bouldin Index* untuk menentukan *Cluster* yang terbaik dari uji coba *cluster* sebelumnya:

Tabel 1. hasil *Davies Bouldin Index*

No	Nama Cluster	Hasil <i>Davies Bouldin Index</i>
1	Cluster 2	-0,838
2	Cluster 3	-1,275
3	Cluster 4	-1,102
4	Cluster 5	-1,227
5	Cluster 6	-0,908
6	Cluster 7	-0,979
7	Cluster 8	-1,424
8	Cluster 9	-0,853
9	Cluster 10	-0,847

Berdasarkan table 1 diatas menunjukkan hasil keseluruhan setelah dilakukannya uji coba untuk mencari hasil *Davies Bouldin Index* terbaik pada data penjualan Alat transportasi. Dengan hasil akhir diatas bahwa evaluasi terbaik dan *cluster* yang dihasilkan merujuk kepada *cluster 2* sebagai nilai *Davies Bouldin Index* yang mendekati angka 0 dengan berada di angka -0,838.



Gambar 15. Visualisasi hasil cluster

Pada gambar 15 diatas ialah *visualitation* dari hasil *cluster* yang telah diuji yang menyatakan bahwa penjualan alat transportasi telah terjual pada tahun 2003, 2004 dan 2005, dari ketiga tahun tersebut penjualan terbanyak berada pada tahun 2005 sebanyak 11, 739 alat transportasi yang sudah terjual. Penjualan tertinggi tersebut berada di *cluster 0* dari 2 *cluster* yang telah diuji sebelumnya untuk mendapatkan nilai *Davies Bouldin Index* yang terkecil.

5. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian yang telah disampaikan pada bab sebelumnya maka dapat ditarik kesimpulan sebagai berikut: Hasil akhir yang diperoleh dari pengelompokan pada data penjualan alat transportasi dengan menggunakan algoritma *K-medoid* kemudian diolah melalui *tools Rapidminer* menghasilkan *cluster* terbaik dengan nilai k yaitu 2, dimana dari 2 *cluster* itu terdiri dari *cluster* dengan penjualan tertinggi dan *cluster* dengan penjualan terendah. *Cluster* dengan penjualan tertinggi berada di *cluster_1* pada tahun 2005 dengan jumlah 11.739 dan penjualan terendah berada di *cluster_0* pada tahun 2003 dengan jumlah 482. *Cluster* terbaik dan nilai *Davies Bouldin Index* yang di dapatkan dari hasil uji coba *cluster 2 -10* pada data penjualan alat transportasi

yaitu berada di cluster 2 dengan nilai *dbi/ Davies Bouldin Index* yaitu -0,838.

Diharapkan dengan adanya penelitian ini, perusahaan dapat memikirkan strategi untuk meningkatkan pemasaran dalam penjualan alat transportasi di tahun-tahun berikutnya agar mendapatkan keuntungan yang lebih dari tahun sebelumnya.

Berdasarkan hasil tugas akhir ini, ada beberapa saran yang dapat dilakukan agar pengembangan system data *mining* ini menjadi lebih baik yaitu penelitian ini ditunjukkan untuk *clustering* pada pengelompokan data penjualan alat transportasi menggunakan *K-Medoid*. Disarankan untuk pengembangan selanjutnya dapat dilakukannya proses implementasi data *mining* dengan metode lainnya seperti metode *Naive bayes* yang berguna untuk memprediksi penjualan pada tahun berikutnya.

DAFTAR PUSTAKA

- [1] m sadikin sulistywati, "Penerapan algoritma k-medoids untuk menentukan segmentasi pelanggan".
- [2] A. Supriyadi *et al.*, "PERBANDINGAN ALGORITMA K-MEANS DENGAN K-MEDOIDS PADA PENGELOMPOKAN ARMADA KENDARAAN TRUK BERDASARKAN PRODUKTIVITAS."
- [3] A. Zahrotul Kamalia, E. Aan Pradana, and N. Surojudin, "PENERAPAN ALGORITMA K-MEANS DALAM KLASTERISASI PENJUALAN LAPTOP," *Jurnal Teknologi Pelita Bangsa*, vol. 13, 2022.
- [4] B. Riyanto, "PENERAPAN ALGORITMA K-MEDOIDS CLUSTERING UNTUK PENGELOMPOKAN PENYEBARAN DIARE DI KOTA MEDAN (STUDI KASUS: KANTOR DINAS KESEHATAN KOTA MEDAN)," *KOMIK (Konferensi Nasional Teknologi Informasi dan Komputer)*, vol. 3, no. 1, Dec. 2019, doi: 10.30865/komik.v3i1.1659.
- [5] S. Nurajizah, "Analisa Transaksi Penjualan Obat menggunakan Algoritma Apriori," *JURNAL INOVTEK POLBENG - SERI INFORMATIKA*, vol. 4, 2019.
- [6] R. Tri Vulandari, "PENGELOMPOKAN TINGKAT KEAMANAN WILAYAH JAWA TENGAH BERDASARKAN INDEKS KEJAHATAN DAN JUMLAH POS KEAMANAN DENGAN METODE KLASTERING K-MEANS," *Jurnal Ilmiah SINUS*.
- [7] A. Nugraha, O. Nurdiawan, and G. Dwilestari, "PENERAPAN DATA MINING METODE K-MEANS CLUSTERING UNTUK ANALISA PENJUALAN PADA TOKO YANA SPORT," 2022.
- [8] F. Harahap, "TIN: Terapan Informatika Nusantara Perbandingan Algoritma K Means dan K Medoids Untuk Clustering Kelas Siswa Tunagrahita," 2021.
- [9] W. Kurniawan, A. Rifai, ; Windu Gata, and D. Gunawan, "Analisis Algoritma K-Medoids Clustering Dalam Menentukan Pemesanan Hotel," *JURNAL SWABUMI*, vol. 8, no. 2, 2020, [Online]. Available: <https://www.kaggle.com>
- [10] M. Arifandi, A. Hermawan, and D. Avianto, "IMPLEMENTASI ALGORITMA K-MEDOIDS UNTUK CLUSTERING WILAYAH TERINFEKSI KASUS COVID19 DI DKI JAKARTA," *Jurnal Teknologi Terapan* /, vol. 7, no. 2, 2021.