



Penerapan Metode K-Means Pada Data Penjualan Untuk Mendapatkan Produk Terlaris Di PT. Titian Nusantara Boga

Tri Wahyudi^{1*}, Naini Sa'adah², Devi Puspitasari³

¹Teknik Informatika, ^{2,3}Sistem Informasi, Sekolah Tinggi Ilmu Komputer Cipta Karya Informatika, Jakarta, Indonesia

Email Korespondensi : triwahyudi100390@gmail.com,

Abstrak– PT. Titian Nusantara Boga adalah perusahaan berkembang yang bergerak di bidang ritel dengan menjual produk makanan. Namun seiring berkembangnya zaman, membuat persaingan semakin ketat karena banyak perusahaan yang bergerak di bidang yang sama. Penelitian ini membahas tentang penerapan data mining, untuk meningkatkan penjualan pada PT. Titian Nusantara Boga dengan mengelompokkan produk ke dalam kategori sangat laris, cukup laris dan kurang laris menggunakan algoritma K-Means Clustering dengan Metode CRISP-DM yang melalui proses *business understanding, data understanding, data preparation, modelling, evaluation, dan deployment*. Pengujian ini dilakukan menggunakan aplikasi *RapidMiner* 9.10 dengan jumlah 3 atribut dan 793 *record* data kemudian berhasil membentuk 3 cluster, yaitu cluster pertama mendapatkan hasil 128 produk, cluster kedua mendapatkan 622 produk dan cluster ketiga mendapatkan 43 produk. Kemudian dilakukan uji validasi dengan *Davies Bouldin Index* dan mendapatkan nilai DBI pada cluster pertama 0,679, cluster kedua 0,816 dan cluster ketiga 0,837. Semakin kecil nilai DBI yang dihasilkan maka kualitasnya semakin baik, maka dari nilai DBI yang dihasilkan dapat diketahui cluster pertama masuk ke dalam kategori sangat laris, yaitu dengan 128 produk dan nilai DBI 0,679.

Kata Kunci: Data Mining, Clustering, K-Means, RapidMiner, Penjualan

Abstract– PT. Titian Nusantara Boga is a growing company engaged in the retail sector by selling food products. However, as time goes by, competition is getting tougher because many companies are engaged in the same field. This study discusses the application of data mining, to increase sales at PT. Titian Nusantara Boga by grouping products into categories that are very in demand, moderate in demand and less in demand using the K-Means Clustering algorithm with the CRISP-DM method which goes through a process of business understanding, data understanding, data preparation, modeling, evaluation, and deployment. This test was carried out using the *RapidMiner* 9.10 application with a total of 3 attributes and 793 data records and then managed to form 3 clusters, namely the first cluster got 128 products, the second cluster got 622 products and the third cluster got 43 products. Then a validation test was carried out with the *Davies Bouldin Index* and the DBI value in the first cluster was 0.679, the second cluster was 0.816 and the third cluster was 0.837. The smaller the DBI value produced, the better the quality, so from the resulting DBI value it can be seen that the first cluster is in the very best-selling category, namely with 128 products and a DBI value of 0.679..

Keywords: Data Mining, Clustering, K-Means, RapidMiner, Sales

I. PENDAHULUAN

Makanan adalah salah satu kebutuhan manusia yang sangat penting, dengan adanya teknologi informasi yang berkembang banyak memunculkan ide-ide dari masyarakat sehingga menghasilkan atau terciptanya banyak jenis-jenis makanan dan juga kebutuhan customer yang terus meningkat, sehingga menjadi peluang bisnis yang menjanjikan bagi pengusaha - pengusaha kuliner atau masyarakat yang baru ingin memulai usaha. Namun untuk meningkatkan penjualan tidaklah mudah, banyak cara yang harus dilakukan seperti menyusun strategi dan pengambilan keputusan yang tepat, agar nantinya dapat menguntungkan dan dapat unggul dalam persaingan

Tujuan dari tentukannya strategi penjualan yaitu agar bisa mengurangi ketidakpastian atau resiko yang mungkin terjadi dari masing - masing strategi yang diterapkan yaitu dengan dilakukannya analisis yang mendalam. Namun dalam dunia bisnis tidak semuanya berjalan dengan lancar, seperti halnya pada PT. Titian Nusantara Boga dimana tidak semua produk yang dijual laris terjual, beberapa diantaranya yaitu daya jual yang belum sepenuhnya terbagi dengan optimal. Dalam pengembangan strategi pemasaran tim marketing pada PT. Titian Nusantara Boga kesulitan untuk mengetahui

produk mana yang lebih banyak terjual dan mana yang kurang diminati, sehingga data transaksi pada PT. Titian Nusantara Boga masih kurang tertata dengan baik.

Di era globalisasi, pesatnya perkembangan sistem informasi merupakan suatu fakta yang dapat dimanfaatkan. Kecanggihan sistem informasi ini semakin bertambah dengan penggunaan komputer yang sudah banyak digunakan dalam berbagai bidang kehidupan, misalnya bidang pendidikan, kesehatan, hiburan, terutama bidang bisnis yang kesemuanya itu memerlukan penggunaan sistem informasi. Penggunaan sistem informasi dalam dunia bisnis penjualan produk dapat memberikan data yang valid.[1]

Mengidentifikasi pola penjualan dengan melihat tren pembelian konsumen, jika dianalisis dan disikapi dengan baik, dapat membantu menentukan produk mana yang paling laris dan mana yang tidak, sehingga pengendalian persediaan dapat dilakukan dan dapat dijadikan masukan bagi pelaku usaha dalam menyusun strategi pemasaran.[2]

Menerapkan penggunaan Data Mining adalah salah satu cara menerapkannya, Data mining merupakan penggalian data yang tersembunyi dari database. Proses clustering adalah proses pengelompokan berdasarkan prinsip kesamaan antar kelas yang mengurangi kesamaan



antar kelas. Berbagai algoritma clustering telah dikembangkan untuk mencapai kinerja yang baik.[3]

Obyek/data yang dikelompokkan dalam satu kelompok memiliki sifat yang sama berdasarkan kriteria tertentu. Salah satu tugas analisis data adalah mengklasifikasikan atau mengelompokkan data ke dalam beberapa kategori, kelompok atau klaster. Data yang diperoleh melalui metode ini dikelompokkan menjadi beberapa klaster berdasarkan minat beli konsumen.[4]

Data yang diperoleh dengan metode ini dapat dikelompokkan menjadi beberapa klaster berdasarkan kesamaan materi, seperti data dengan karakteristik yang sama dalam satu klaster dan data dengan karakteristik yang berbeda pada klaster lain dikelompokkan dengan karakteristik yang sama.[5]

Algoritma K-Means telah dipelajari oleh beberapa peneliti untuk memenuhi kebutuhan berbagai disiplin ilmu. Salah satunya dibidang penjualan atau bisnis. Dalam penelitian ini, algoritma K-Means digunakan. Hal ini karena penerapannya sangat sederhana, waktu yang dibutuhkan untuk menyelesaikan prosesnya tidak lama, serta mudah untuk diubah dan dipelajari.[6]

Penelitian terkait pada data mining dalam pengelompokan dengan metode *k-means* yang pernah dilakukan oleh Oki Oktaviarna Tensao¹, I Nyoman Yudi Anggara Wijaya², Ketut Queena Fredlina³ (2022) hasil dari *cluster* diatas bahwa *cluster 2* dan *cluster 3* hanya dilakukan pemasaran dengan periklanan yang tertera pada tabel promotion mix.[7]

Penelitian yang dilakukan oleh Gita Aprilianur¹, Elvin Leander Hadisaputro² (2022) hasil dari cluster diatas memilih 3 grup acak sebagai centroid awal. Setelah data tiap kelompok tidak berubah, terlihat bahwa hasil akhirnya ada 24 produk yang laku keras, 59 produk laris, dan 17 produk kurang laku. Kemudian, penerapan metode *K-Means* pada Rapidminer dilakukan dengan memasukkan data yang akan diubah menjadi database di Ms.Excel, data tersebut akan terhubung ke *Rapidminer Tools* dan diproses. dan membentuk *K-means*. Setelah itu, *Rapidminer* akan memproduksi produk mana yang permintaannya tinggi, permintaannya tinggi, dan permintaannya rendah.[8]

Dalam penelitian Deni Triyansyah¹, Devi Fitriah² (2018) hasil yang telah didapatkan adalah nilai Davies Bouldin Index yang diperoleh pada clustering pencarian reseller yang potensial adalah 0.271 dan 0.196 pada klustering pencarian tipe sepatu yang laris.[9]

Dengan dilakukannya penelitian ini diharapkan dapat membantu pihak PT. Titian Nusantara Boga dalam menentukan penjualan produk terlaris yang tepat untuk meningkatkan laba penjualan dan dapat bersaing dengan kompetitor. Dalam hal ini yang membuat penulis tertarik untuk membuat analisa data mining pada PT. Titian Nusantara Boga

II. METODE PENELITIAN

CRISP-DM

CRISP-DM adalah metode pemodelan proses pengembangan data yang banyak digunakan oleh para

profesional untuk memecahkan masalah. Proses penelitian membahas enam tahapan CRISP-DM, yaitu pemahaman bisnis, pemahaman data, persiapan data, pemodelan, evaluasi, dan implementasi[10]

Clustering

Pada dasarnya clustering adalah metode untuk mengkategorikan atau pengelompokan sekelompok objek sesuai dengan atribut yang sama atau karakteristik dengan data-data lainnya.[11]

Algoritma pengelompokan K-means adalah alat pengelompokan paling populer yang digunakan dalam aplikasi ilmiah dan industri. K mewakili masing-masing cluster k sama dengan rata-rata (atau rata-rata tertimbang) dari titik-titiknya, yang disebut massa. Pusat massa cluster adalah rata-rata dari koordinat semua titik di cluster.[12]

K-Means

K-Means merupakan metode data mining dengan proses awal yang mengambil beberapa komponen populasi untuk digunakan sebagai titik sentral dalam cluster awal. Identifikasi kelompok dalam suatu cluster dapat dilakukan dengan menghitung jarak setiap objek dari titik pusat. Algoritma clustering K-Means dapat membagi data berdasarkan jarak antar data ke dalam kelompok-kelompok yang telah ditentukan. Algoritma ini mengandalkan fungsi untuk mengukur data dengan karakteristik yang sama. Jarak dihitung menggunakan fungsi Euclidean. Kemudian data tersebut dimasukkan ke dalam kelompok dengan jarak terdekat.[13]

Adapun prosedur perhitungan Algoritma K-Means Clustering analysis diuraikan sebagai berikut:

- Menentukan banyaknya cluster

Dalam menentukan banyaknya cluster diharuskan tidak lebih dari jumlah kriteria yang ada.

- Menentukan titik pusat cluster (centroid)[14]

Berikut adalah langkah – langkah melakukan clustering :

$$D(x, y) = \sqrt{(X_1 - Y_1)^2 + (X_2 - Y_2)^2}$$

Dimana:

- D adalah jarak
- X adalah data
- Y adalah centroid.[15]

Penelitian ini menggunakan metode CRISP-DM dan menggunakan algoritma K-Means untuk mengelompokkan produk. Alasan menggunakan metode ini adalah karena relatif sederhana dan mudah diterapkan, dapat digunakan untuk dataset dalam jumlah besar dan telah digunakan secara luas untuk menyelesaikan berbagai persoalan komputasi.

III. HASIL DAN PEMBAHASAN

Dalam pengujian ini dapat dilakukan dengan menggunakan metode Cross-Industry Standard Process for Data Mining (CRISP-DM) yang memiliki 6 tahapan yaitu:

- Pemahaman Bisnis (*Business Understanding*)**



PT. Titian Nusantara Boga adalah sebuah perusahaan berkembang yang menjual makanan snack. Usaha tersebut dirintis pada tahun 2012 yang berlokasi di Jl. Tebet Raya No.86, Jakarta Selatan. Pada penelitian ini bertujuan untuk meningkatkan penjualan berdasarkan pengelompokkan produk.

1) Tentukan Tujuan Bisnis

Dilakukannya penelitian ini yaitu bertujuan untuk menentukan produk yang dijual dengan mengelompokkan atau mengkategorikan berdasarkan produk yang paling laris, cukup laris dan kurang laris, sehingga dapat dijadikan sebagai dasar untuk meningkatkan penjualan. Untuk menentukan produk mana yang lebih diminati oleh pelanggan berdasarkan kategori sangat laris, cukup laris dan kurang laris, maka diperlukan pengambilan keputusan yang tepat.

2) Menilai Situasi

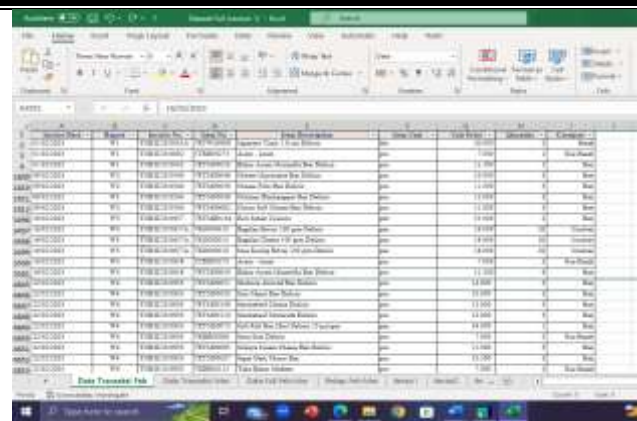
PT. Titian Nusantara Boga merupakan suatu perusahaan yang menjual berbagai macam kebutuhan snack seperti roti, kue, pastry & jajanan tradisional yang ditargetkan untuk kepentingan perusahaan atau institusi dan masyarakat umum seperti untuk acara sosial, meeting, training dan lain – lain. Namun PT. Titian Nusantara Boga saat ini belum melakukan pemanfaatan data transaksi dengan baik Maka dari itu untuk bisnis ini perlu memahami strategi penjualannya seperti dalam tujuan bisnis dan kemudian menerjemahkannya ke dalam tujuan data mining.

b. Pemahaman Data (Data Understanding)

Pada tahap ini berdasarkan data yang telah didapat, pemahaman terhadap kebutuhan data dengan mncapai tujuan dalam menentukan strategi penjualan yang efektif dan efisien. Dataset yang didapat dari transaksi penjualan yang merupakan report harian yang berbentuk dokumen berformat excel dan memiliki 12.204 record data, data yang digunakan 2 bulan terakhir yaitu bulan Februari sampai dengan Maret 2023.

1) Pengumpulan Data Awal

Pada tahap pengumpulan data awal yaitu dilakukan pengumpulan data yang dibutuhkan untuk mendukung proses tahap pemahaman data. Sumber data awal yang akan digunakan dan diolah dalam penelitian ini adalah dataset transaksi penjualan pada PT. Titian Nusantara Boga bulan Februari sampai dengan Maret 2023 dan data tersebut merupakan berupa dalam bentuk excel.



Gambar 1 Pengumpulan Data Awal

2) Pendeskripsian Data

Dalam tahap pendeskripsian data yaitu dilakukan pemahaman dataset transaksi penjualan dengan mendeskripsikan dataset yang diperoleh, seperti atribut atau label yang terdapat pada dataset tersebut. Data yang diperoleh peneliti yaitu 12.204 record data, kemudian memiliki 9 atribut yaitu Invoice Date, Report, Invoice No, Item No, Item Description, Item Unit, Unit Price, Quantity dan Category.

3) Evaluasi Pemilihan Data

Pada tahap ini untuk mendukung proses evaluasi pemilihan data, sebelum masuk ke proses pengolahan data, data dievaluasi terlebih dahulu agar proses pengolahan data lebih mudah dilakukan, yaitu dilakukannya perekapan data dengan cara data yang ter record di blok untuk membuat tabel baru di excel dengan menggunakan pivot table. Dilakukannya perekapan data agar data yang akan diolah lebih efektif dan efisien yaitu tidak terlalu banyak dan lebih mudah diolah, tetapi hasil rekapan data merubah atribut awal yang semula ada 9 atribut dengan 12.204 record data setelah direkap yaitu menjadi 4 atribut saja dengan 793 record data . Berikut hasil rekapan data transaksi penjualan bulan Februari sampai dengan Maret’

Tabel 1 Rekap Data Penjualan Per Tiga Bulan

Repor t	Item Description	Februar i	Mare t
W1	Arem - Arem	833	255
W1	Bagelan Butter 100 grm Delicio	10	25
W1	Bagelan Butter Isi 4 Delicio	86	40
W2	Cheese Stick Puff Isi 5 pcs	134	297
W2	Chicken Blackpepper Bun Delicio	49	64
W2	Chicken Blackpepper Medium Delicio	20	32
W2	Chicken Floss Medium	13	14



Delicio			
W2	Choco Lava Cake Delicio	20	15
W3	Rainbow Choco Chips Delicio	23	20
W3	Raisin Bun Delicio	75	55
W3	Raisin Medium Delicio	56	35
W3	Redbean Roll Bun Delicio	38	32
W3	Rilakuma Delicio	6	6
W3	Risol Ayam Delicio	676	744
W3	Risol Sapi Delicio	235	251
W4	Soes Buah Delicio	304	156
W4	Soes Cake Mini Delicio	60	66
W4	Soes Choco Eclair Delicio	140	111
W4	Sosis Solo Delicio	528	22
W4	Srikaya Cream Cheese Bun Delicio	72	26
W4	Srikaya Cream Cheese Medium Delicio	45	85
W4	Super Dark Choco Bun	100	67
W4	Super Dark Choco Medium Swiss Roll Greenbean	45	45
W4	Pandan Slice Cake 2.5 cm Delicio	104	40
....

4) Pemilihan Atribut

Dalam proses pemahaman data yaitu menentukan atribut yang digunakan untuk proses pengolahan data karena telah disesuaikan dengan fokus penelitian. Karena data yang akan diolah adalah hasil dari rekapan data bulan Februari sampai dengan Maret maka atribut yang akan digunakan berubah dari 9 atribut menjadi 4 atribut saja, yaitu Report, Item Description, Februari, Maret namun untuk masuk ke proses data mining hanya 3 atribut saja berdasarkan data yang sudah direkap yang digunakan untuk diolah datanya.

c. Persiapan Data (Data Preparation)

Persiapan data pada tahap ini yaitu data mentah awal berupa dataset transaksi penjualan yang nanti akan diterapkan ke dalam alat pemodelan, dan kemudian akan dilakukan proses data mining.

1) Seleksi data

Pada data awal terdapat 9 atribut dengan 12.204 record data, setelah dilakukan perekapan data menjadi 4 atribut dengan 793 record data. Kemudian dalam penelitian ini akan dipilih 3 atribut saja yaitu Item Description, Februari, Maret. Berikut data berdasarkan atribut yang digunakan dapat dilihat pada tabel 4.

Tabel 2 Atribut Yang Digunakan

Atribut	Tipe	Keterangan
Item Description	ID	Nama produk
Februari	Integer	Kuantitas produk di bulan Februari
Maret	Integer	Kuantitas produk di bulan Maret

2) Data Praproses

Pada tahap ini yaitu dilakukannya pengecekan untuk memastikan bahwa data transaksi penjualan yang dipilih layak untuk diolah. Setelah dilakukan pengecekan data satu per satu dan tidak ada data yang bermasalah setelah dibersihkan, maka data pemakaian digabungkan dalam satu cluster data tersebut yang akan diuji pada proses algoritma k-means sehingga terbentuk 3 cluster yang digunakan.

d. Modelling

Modelling merupakan salah satu tahap pada crisp-dm yaitu dengan memilih teknik data mining dan menentukan algoritma yang akan digunakan. Pemodelan data mining dalam penelitian ini menggunakan excel untuk perhitungan rumus jarak Euclidean Distance dan menggunakan aplikasi RapidMiner versi 9.10 untuk software yang digunakan untuk mengolah datanya. Berikut langkah – langkah pengujian algoritma K-Means, yaitu sebagai berikut.

- 1) Tentukan jumlah cluster.
- 2) Pilih titik K atau centroid secara acak.

Tabel 3 Pusat Centroid Awal

Centroid Ke	Februari	Maret
C1	182	200
C2	700	94
C3	676	744

- 3) Hitung varian dan tempatkan centroid baru dari setiap cluster, pada setiap data akan dihitung jarak terdekat dengan setiap centroid yang sudah ditentukan pada tahap sebelumnya dengan menggunakan rumus Euclidean. Kemudian dipilih data centroid secara acak yang dapat dilihat pada Tabel 4. Data yang dipilih pada pusat titik awal yaitu data ke225, 43, 533. Berikut adalah perhitungan dengan menggunakan persamaan Euclidean Distance:

Cluster1

$$D(x, y) = \sqrt{(8331_1 - 182_1)^2 + (2552_2 - 200_2)^2} = 653,3192175$$

Cluster 2

$$D(x, y) = \sqrt{(8331_1 - 700_1)^2 + (2552_2 - 94_2)^2} = 208,8300745$$



Cluster 3

$$D(x, y) = \sqrt{(8331_1 - 676_1)^2 + (2552_2 - 744)^2} = 513,5854359$$

- 4) Mengklasifikasi data berdasarkan kedekatannya dengan centroid. Berikut adalah hasil perhitungan jarak atau hasil pengklusteran iterasi pertama pada setiap data yang dibentuk berdasarkan pusat centroid yang dipilih secara acak.

Tabel 4 Hasil Pengklusteran Iterasi ke-1

Da ta Ke	C1	C2	C3	Clus ter	Jarak Tepend ek
1	653,319 2175	208,830 0745	513,5854 359	2	208,830 0745
2	245,375 2229	693,441 4179	980,0596 921	1	245,375 2229
3	186,590 4606	616,370 0187	918,5401 461	1	186,590 4606
4	234,522 92	676,630 623	969,3301 811	2	234,522 92
5	149,036 9082	614,052 1151	883,7782 527	1	149,036 9082
6	211,101 8711	660,051 5131	945,9175 44	1	211,101 8711
7	159,539 9636	581,447 3321	886,9143 138	1	159,539 9636
8	241,760 2118	674,999 2593	976,2888 917	1	241,760 2118
9	259,113 8746	697,323 4544	993,9416 482	3	259,113 8746
10	181,474 5161	632,026 1071	915,9939 956	3	181,474 5161
11	1298,89 222	1551,03 4816	991,3329 41	1	991,332 941
12	196,725 6974	688,035 6095	924,3078 491	1	196,725 6974
13	266,790 1797	704,770 1753	1001,611 202	1	266,790 1797
14	138,708 3271	515,193 1677	604,0033 112	1	138,708 3271
15	230,288 9489	692,529 4218	964,0046 68	1	230,288 9489
16	126 8101	406,078 8101	656,7800 241	1	126
17	106,531 6854	575,222 5656	839,9744 044	1	106,531 6854
18	203,823 453	657,473 9539	938,6479 638	1	203,823 453
79	99,8248	537,059	821,7864	1	99,8248
3	4661	587	686	3	4661

- 5) Hitunglah kembali pusat cluster dengan anggota cluster yang sekarang. Pusat cluster ialah nilai rata-rata dari semua data objek dalam cluster tertentu.

Tabel 5 Hasil Pembaruan *Centroid* Iterasi ke-1

Cluster	Februari	Maret	Produk
---------	----------	-------	--------

C1	85,87042254	81,48309859	710
C2	645,8064516	181	31
C3	559,3846154	962,0384615	52

- d. Hitung lagi setiap objek memakai pusat cluster yang baru. Jika pusat cluster tidak berubah lagi maka proses kluster selesai. Atau, Kembali ke langkah nomor 3 sampai pusat cluster tidak berubah lagi.

Tabel 6 Hasil Pembaruan *Centroid* Iterasi ke-2

Cluster	Februari	Maret	Produk
C1	76,135174	73,9171512	688
C2	588,26786	253,553571	56
C3	505,14286	988,489796	49

Tabel 7 Hasil Pembaruan *Centroid* Iterasi ke-3

Cluster	Februari	Maret	Produk
C1	72,516987	70,1565731	677
C2	552,79452	289,342466	73
C3	479,72093	1043,53488	43

Tabel 8 Hasil Pembaruan *Centroid* Iterasi ke-4

Cluster	Februari	Maret	Produk
C1	71,279346	68,7265973	673
C2	529	307,148148	81
C3	493,74359	1086,10256	39

Tabel 9 Hasil Pembaruan *Centroid* Iterasi ke-5

Cluster	Februari	Maret	Produk
C1	69,794294	65,5315315	666
C2	498,8022	324,285714	91
C3	505,61111	1120,44444	36

Tabel 10 Hasil Pembaruan *Centroid* Iterasi ke-6

Cluster	Februari	Maret	Produk
C1	68,462121	63,1818182	660
C2	481,3299	324,268041	97
C3	505,61111	1120,44444	36

Tabel 11 Hasil Pembaruan *Centroid* Iterasi ke-7

Cluster	Februari	Maret	Produk
C1	66,752294	61,5917431	654
C2	468,13592	319,15534	103
C3	505,61111	1120,44444	36

Tabel 12 Hasil Pembaruan *Centroid* Iterasi ke-8

Cluster	Februari	Maret	Produk
C1	64,8936826	61,0308166	649
C2	451,700935	307,121495	107
C3	530,486486	1108,62162	37

Tabel 13 Hasil Pembaruan *Centroid* Iterasi ke-9

Cluster	Februari	Maret	Produk
C1	62,2043682	59,7581903	641



C2	429,692982	294,350877	114
C3	558,368421	1095,5	38

Tabel 14 Hasil Pembaruan *Centroid* Iterasi ke-10

Cluster	Februari	Maret	Produk
C1	59,8862559	58,0853081	633
C2	417,622951	287,647541	122
C3	558,368421	1095,5	38

Tabel 15 Hasil Pembaruan *Centroid* Iterasi ke-11

Cluster	Februari	Maret	Produk
C1	59,5221519	58,0110759	632
C2	416,585366	286,162602	123
C3	558,368421	1095,5	38

Tabel 16 Hasil Pembaruan *Centroid* Iterasi ke-12

Cluster	Februari	Maret	Produk
C1	58,8444444	57,8190476	630
C2	414,288	283,48	125
C3	558,368421	1095,5	38

Tabel 17 Hasil Pembaruan *Centroid* Iterasi ke-13

Cluster	Februari	Maret	Produk
C1	58,418124	57,863275	629
C2	408,944	278,92	125
C3	569,564103	1082,79487	39

Tabel 18 Hasil Pembaruan *Centroid* Iterasi ke-14

Cluster	Februari	Maret	Produk
C1	57,7799043	57,6172249	627
C2	402,65873	274,087302	126
C3	577,825	1070,725	40

Tabel 19 Hasil Pembaruan *Centroid* Iterasi ke-15

Cluster	Februari	Maret	Produk
C1	57,7799043	57,6172249	627
C2	402,65873	274,087302	126
C3	577,825	1070,725	40

Dari hasil perhitungan diatas sampai pada tahap pembaruan centroid terakhir tidak berubah, didapat hasil pengelompokkan produk ke dalam cluster 1, 2 dan 3. Cluster 1 berhasil mendapatkan 627 produk, cluster 2 126 produk dan cluster 3 mendapatkan 40 produk.

e. Evaluation

Evaluasi pada tahap ini akan dilakukan dengan menerapkan metode *Davies Bouldin Index* (DBI). Evaluasi ini dilakukan analisa atau pengukuran ketepatan terhadap pemodelan yang telah dilakukan. Kemudian menetapkan apakah terdapat permasalahan penting dari bisnis atau penelitian yang tidak tertangani dengan baik. Dimana metode ini adalah penguji kualitas data berdasarkan kluster yang dihasilkan. Evaluasi yang ditujukan adalah untuk mengetahui pemodelan yang dilakukan apakah sudah

tepat dan sesuai diterapkan pada kasus penelitian ini serta sudah sesuai rencana awal penelitian. Kemudian hasil dari evaluasi tersebut yaitu menentukan langkah berikutnya apakah bisa dilanjut atau diulang dari awal karena tidak sesuai dengan rencana awal penelitian. Untuk itu penelitian ini dilakukan perhitungan Separasi dengan *Sum of Square Within Cluster* (SSW), *Sum-of-square between-cluster* (SSB) dan *Ratio* lebih dahulu secara manual diexcel.

1) SSW (Sum of Square Within Cluster)

Berikut Perhitungan SSW:

1. Pertama hitung data terhadap pembaruan *centroid* terakhir (lihat tabel 4.24) dengan menggunakan rumus *Euclidean Distance* Cluster1

$$D(x, y) = \sqrt{(833_1 - 402,658731)2_1)^2 + (2552_2 - 274,0873022_2)^2} = 430,7643597$$

2. Kemudian lanjutkan untuk perhitungan jarak *cluster* 2 sampai *cluster* 3.

Tabel 20 Hasil Perhitungan Jarak Cluster 1 Terhadap *Centroid* Akhir

Data Ke-	Feb	Mar	Cluster	Jarak
2	10	25	1	57,85155672
3	86	40	1	33,26770827
4	27	24	1	45,57982358
13	1	4	1	78,09458584
15	9	48	1	49,71891067
17	125	110	1	85,22028159
18	44	50	1	15,74509059
36	30	40	1	32,89513179
37	1	3	1	78,78450854
38	44	74	1	21,40750063
39	30	30	1	39,17185461
40	2	5	1	76,68096295
41	3	50	1	55,30696186
44	20	22	1	51,92213283
....

3. Kemudian lanjutkan perhitungan sampai dengan cluster ke -3.
4. Selanjutnya jumlahkan hasil perhitungan jarak pada masing - masing setiap cluster kemudian hasilnya dibagi jumlah data centroid akhir.

Tabel 21 Hasil Perhitungan SSW

SSW	Jumlah Jarak	Produk	Hasil
SSW1	43996,05459	627	70,16914608
SSW2	27912,85022	126	221,5305573
SSW3	19421,97522	40	485,5493805

2) SSB (Sum-of-square between-cluster)



Pada tahap ini dilakukan lagi perhitungan jarak menggunakan data centroid terakhir antara cluster satu dengan cluster yang lainnya dengan menggunakan rumus *Euclidean Distance*.

SSB 1.2

$$D(x,y) = \sqrt{(57,77990431 - 402,658731)^2 + (57,61722492 - 274,0873022)^2} = 407,1863193$$

SSB 1.3

$$D(x,y) = \sqrt{(57,77990431 - 577,8251)^2 + (57,61722492 - 1070,7252)^2} = 1138,786313$$

SSB 2.3

$$D(x,y) = \sqrt{(402,658731 - 577,8251)^2 + (274,0873022 - 1070,725)^2} = 815,6683423$$

Tabel 22 Hasil Perhitungan SSB

SSB	Cluster1	Cluster2	Cluster3
SSB1	0	407,1863193	1138,786313
SSB2	407,1863193	0	815,6683423
SSB3	1138,786313	815,6683423	0

3) Rasio (Ratio)

Selanjutnya dilakukan perhitungan rasio, yaitu sebagai berikut:

$$R1 = \frac{70,16914608 + 221,5305573}{407,1863193} = 0,716378939$$

$$R2 = \frac{70,16914608 + 485,5493805}{407,1863193} = 1138,786313$$

$$R3 = \frac{221,5305573 + 485,5493805}{815,6683423} = 0,866871866$$

Tabel 23 Hasil Perhitungan Rasio

Rasio	Cluster1	Cluster2	Cluster3	R-Max
R1	0	0,716378	0,487991	0,716378
R2	0,716378	0	0,866871	0,866871
R3	0,487991	0,866871	0	0,866871

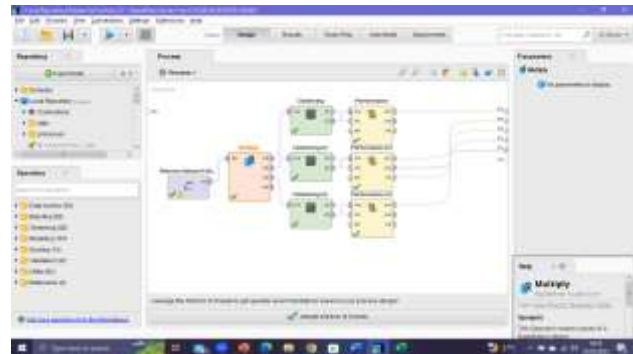
4) DBI (Davies Bouldin Index)

Kemudian hitunglah jumlah R-Max untuk

mendapatkan hasil DBI, Berikut hasil perhitungan DBI :

$$DBI = \frac{0,716378939 + 0,866871866 + 0,866871866}{3} = 0,816707557 / 0,817$$

Setelah itu dilakukan uji data dengan aplikasi *RapidMiner* untuk membandingkan nilai DBI antara cluster 1 dengan cluster yang lainnya. Agar dapat mengelompokkan produk dalam kategori sangat laris, cukup laris dan kurang laris. Berikut hasil pengujian data menggunakan *RapidMiner*



Gambar 2 Tampilan Design



Gambar 3 Tampilan Hasil Cluster Model

f. Deployment

Dalam tahap ini deployment adalah tahap akhir dalam pembuatan laporan hasil kegiatan data mining. Laporan akhir ini berhasil membentuk 3 cluster yaitu cluster 1 menghasilkan 627 produk, Cluster 2 menghasilkan 126 produk dan cluster 3 menghasilkan 40 produk. Kemudian dilakukan uji validasi menggunakan rumus DBI dan menghasilkan nilai 0,817. Seperti terlihat pada gambar 4



Gambar 4 Tampilan Hasil Cluster Model

IV. KESIMPULAN

Dari hasil penelitian yang telah dilakukan, maka dapat disimpulkan bahwa metode *k-means* dapat diterapkan. Pada PT. Titian Nusantara Boga untuk menentukan penjualan produk terlaris berdasarkan 793 record data yang didapat dan mendapatkan 3 atribut. Metode ini dapat digunakan untuk membantu pihak perusahaan dalam melihat tingkat penjualan mereka.

Pada penelitian ini menggunakan algoritma *k-means* dengan metode CRISP-DM untuk melakukan pengujian datanya, melalui 6 tahapan, yaitu *Business Understanding, Data Understanding, Data Preparation, Modelling dan Deployment*.

Berdasarkan hasil pengujian yang telah dilakukan, penelitian ini berhasil mengelompokkan produk yaitu cluster 1 mendapatkan hasil 128 produk, cluster 2 622 produk dan cluster 3 43 produk. Kemudian dilakukan uji validasi dengan menggunakan DBI (*Davies Bouldin Index*) dimana cluster 1 mendapatkan hasil nilai DBI 0,679, cluster 2 menghasilkan nilai DBI 0,816 dan cluster 3 menghasilkan nilai DBI 0,837. Maka dari hasil nilai DBI diatas yang termasuk kedalam kategori sangat laris adalah cluster 1, kategori cukup laris cluster 2, kategori kurang laris cluster 3. Hasil tersebut bisa dikategorikan karena jika nilai DBI semakin mendekati angka 0 maka kualitasnya semakin baik.

V. REFERENSI

[1] M. H. Fakhriza and K. Umam, "Analisis Produk Terlaris Menggunakan Metode K-Means Clustering Pada "Pt.Sukanda Djaya," *JIKA (Jurnal Inform.,* vol. 5, no. 1, p. 8, 2021, doi: 10.31000/jika.v5i1.3236.

[2] Normah, B. Rifai, S. Vambudi, and R. Maulana, "Penerapan Data Mining Metode K-Means Clustering Untuk Analisa Penjualan Pada Toko Fashion Hijab Banten," *J. Tek. Komput. AMIK BSI,* vol. 8, no. 2, pp. 174–180, 2022, doi: 10.31294/jtk.v4i2.

[3] F. Indriyani and E. Irfiani, "Clustering Data Penjualan pada Toko Perlengkapan Outdoor Menggunakan Metode K-Means," *JUITA J. Inform.,* vol. 7, no. 2, p. 109, 2019, doi: 10.30595/juita.v7i2.5529.

[4] 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.

[5] M. H. Siregar, "Data Mining Klasterisasi Penjualan Alat-Alat Bangunan Menggunakan Metode K-Means (Studi Kasus Di Toko Adi Bangunan)," *J. Teknol. Dan Open Source,* vol. 1, no. 2, pp. 83–91, 2018, doi: 10.36378/jtos.v1i2.24.

[6] H. Al Rasyid, B. F. K. Soebari, and D. S. Y. Kartika, "Implementasi Algoritma K-Means Clustering Untuk Pengelompokan Penjualan Produk Pada Online Shop Toko Gizi," *Pros. Semin. Nas. Teknol. dan Sist. Inf. 2022 Surabaya,* vol. 1, no. September, pp. 10–11, 2022.

[7] Oki Oktaviarna Tensao, I Nyoman Yudi Anggara Wijaya, and Ketut Queena Fredlina, "Analisa Data Mining dengan Algoritma K-Means Clustering Untuk Menentukan Strategi Promosi Mahasiswa Baru Pada STMIK Primakara," *Inf. (Jurnal Inform. dan Sist. Informasi),* vol. 14, no. 1, pp. 1–17, doi: 10.37424/informasi.v14i1.135.

[8] G. Aprilianur and E. L. Hadisaputro, "Penerapan Data Mining Menggunakan Metode K-Means Clustering Untuk Analisa Penjualan Toko Myam Hijab Penajam," *J. JUPITER,* vol. 14, no. 1, pp. 161–170, 2022.

[9] D. Triyansyah and D. Fitrihanah, "Analisis Data Mining Menggunakan Algoritma K-Means Clustering Untuk Menentukan Strategi Marketing," *J. Telekomun. dan Komput.,* vol. 8, no. 3, p. 163, 2018, doi: 10.22441/incomtech.v8i3.4174.

[10] D. Astuti, "Penentuan Strategi Promosi Usaha Mikro Kecil Dan Menengah (UMKM) Menggunakan Metode CRISP-DM dengan Algoritma K-Means Clustering," *J. Informatics, Inf. Syst. Softw. Eng. Appl.,* vol. 1, no. 2, pp. 60–72, 2019, doi: 10.20895/inista.v1i2.71.

[11] B. Harahap, "Penerapan Algoritma K-Means Untuk Menentukan Bahan Bangunan Laris (Studi Kasus Pada UD. Toko Bangunan YD Indarung)," *Reg. Dev. Ind. Heal. Sci. Technol. Art Life,* pp. 394–403, 2019, [Online]. Available: <https://ptki.ac.id/jurnal/index.php/readystar/article/view/82>

[12] M. Silalahi, "Analisis Clustering Menggunakan



- Algoritma K-Means Terhadap Penjualan Produk Padapt Batamas Niaga Jaya,” *Comput. Based Inf. Syst. J.*, vol. 02, pp. 20–35, 2018.
- [13] T. R. Ayu Pangestu, “PENERAPAN DATA MINING MENGGUNAKAN ALGORITMA K-MEANS PENGELOMPOKAN PELANGGAN BERDASARKAN KUBIKASI AIR TERJUAL MENGGUNAKAN WEKA,” vol. 11, no. 3, pp. 67–71, 2021.
- [14] A. Rohmah, F. Sembiring, and A. Erfina, “IMPLEMENTASI ALGORITMA K-MEANS CLUSTERING ANALYSIS UNTUK MENENTUKAN HAMBATAN PEMBELAJARAN DARING (STUDI KASUS : SMK YASPIM GEGERBITUNG),” pp. 290–298, 2021.
- [15] H. Nicodemus Turnip and H. Fahmi, “Hendra Turnip, 2 Hasanul Fahmi (Penerapan Data Mining Pada Penjualan Kartu Paket Internet Yang Banyak Diminati Konsumen Dengan Metode K-Means) Penerapan Data Mining Pada Penjualan Kartu Paket Internet Yang Banyak Diminati Konsumen Dengan Metode K-Means,” *JIKOMSI J. Ilmu Komput. dan Sist. Inf.*, vol. 4, no. 2, pp. 36–41, 2021.