

Prediksi Penjualan Produk Pada Toko CTR Outdoor Menggunakan Algoritma K-Means dan Algoritma C4.5

Marwondo¹⁾, Eko Bambang²⁾

^{1,2)} Informatika, Fakultas Teknologi dan Informatika, Universitas Informatika dan Bisnis Indonesia
Email: marwondo@unibi.ac.id; ekobambang.n@student.unibi.ac.id

Abstrak

Mengetahui produk yang paling laku dijual serta yang kurang laku merupakan suatu hal yang sangat membantu dalam keberhasilan suatu usaha. Dengan diketahuinya kedua hal tersebut, maka pelaku usaha dapat melakukan pengadaan persediaan yang tepat. Data penjualan yang telah dimiliki dapat ditambang sehingga menghasilkan informasi tersebut dengan lebih akurat. Data yang ada diklasterisasi menggunakan K-Mean dan dilakukan prediksi menggunakan algoritma C4.5. Penerapan K-Means dalam pengelompokan data penjualan dapat menghasilkan rekomendasi barang yang sangat laku, laku, kurang laku, dan tidak laku. Algoritma C4.5 dapat menghasilkan kesimpulan bahwa faktor yang paling banyak mempengaruhi penjualan adalah warna, sedangkan faktor yang tidak terlalu mempengaruhi adalah ukuran (*size*). Hasil prediksi ini dapat digunakan oleh pemilik usaha untuk mengelola persediaan barang agar tidak menumpuk.

Kata Kunci: C4.5, Klasterisasi, K-Means, Penjualan, Prediksi

Abstract

Knowing which slowest moving products and which fast moving is something that is very helpful in the success of a business. By knowing these two things, business actors can procure the right supplies. Existing sales data can be mined to produce more accurate information. The existing data is clustered using K-Mean and predictions are made using the C4.5 algorithm. The application of K-Means in grouping sales data can produce recommendations for items that are selling well, selling well, not selling well, and not selling well. The C4.5 algorithm can produce the conclusion that the factor that influences sales the most is color, while the factor that does not influence it too much is size. The results of this prediction can be used by business owners to manage inventory so that it does not pile up.

Keywords: C4.5, Clustering, K-Means, Prediction, Selling

1 PENDAHULUAN

Data merupakan sekumpulan fakta mentah berupa simbol, angka atau dan kata-kata. Data terbagi menjadi dua tipe, yaitu data primer dan data sekunder. Data primer adalah data yang belum pernah dikumpulkan sebelumnya dan dikumpulkan untuk kepentingan penelitian saja. Data primer mengacu pada data yang berasal dari peneliti

untuk pertama kali. Data sekunder adalah data yang sudah ada sebelumnya, dikumpulkan oleh lembaga dan organisasi penelitian sebelumnya. Pentingnya data sebagai komponen utama dalam sistem menjadi hal yang perlu diperhatikan karena data merupakan dasar dalam menyediakan informasi dan menentukan kualitas informasi dengan cepat, akurat, dan relevan. Data dalam

dunia bisnis juga berperan dalam pertumbuhan dan perkembangan bisnis.

Toko CTR Outdoor merupakan toko yang menjual berbagai produk Outdoor seperti Sepatu, Jas Hujan, Jaket, Waist Bag, Tas. Lokasi Toko CTR Outdoor terletak di Kabupaten Bandung Barat. Toko CTR Outdoor masih terkendala dalam penjualan produk yang tidak laku yang mengakibatkan penumpukan produk di gudang. Hal ini tentunya berpengaruh pada perputaran modal usaha dikarenakan barang yang masih tertahan di Gudang dalam jangka waktu yang lama. Penumpukan persediaan ini dapat diatasi jika toko dapat mengetahui dengan pasti apa saja barang yang kurang laku sampai dengan yang paling laku. Persediaan barang sedianya dilakukan berdasarkan kondisi tersebut.

Untuk mengetahui barang yang paling laku sampai dengan yang kurang laku, dapat dilakukan klusterisasi data penjualan yang ada. Berbagai metode klusterisasi dapat digunakan, salah satunya adalah K-Means. Penerapan metode K-Means dalam pengelompokan data penjualan dapat menghasilkan rekomendasi barang yang laris, Kurang laris dan cukup laris (Indriyani & Irfiani, 2019). Siregar, M. H. (2018) menyatakan bahwa Metode K-Means dapat memberi gambaran bahwa cluster 1 adalah cluster yang penjualan kurang laris pada toko adi bangunan. Sedangkan klaster 2 adalah klaster yang dikategorikan barang yang laris di pasaran pada kelompok ini hanya dua barang yang menunjukkan tingkat penjualan yang tinggi. Menurut Hadi, F., & Diana, Y (2020) menyatakan bahwa pada iterasi yang dilakukan mulai iterasi 1 sampai iterasi 5 didapatkan bahwa pada iterasi 1 mempunyai pola cluster centroid yang berbeda dengan pola klaster iterasi 2. Sedangkan pada iterasi selanjutnya seperti iterasi 3 juga mempunyai pola *cluster centroid* yang berbeda dengan iterasi 1 dan 2.

Metode prediksi yang dapat digunakan untuk sistem prediksi penjualan salah satunya adalah Algoritma C4.5. Konsep dari Algoritma C4.5 adalah mengubah data menjadi pohon keputusan dan aturan-aturan

keputusan (*rule*). C4.5 adalah algoritma yang cocok untuk masalah klasifikasi dan data mining. C4.5 memetakan nilai atribut menjadi class yang dapat diterapkan untuk klasifikasi baru. Berdasarkan penjelasan diatas maka dibuatlah sistem prediksi penjualan menggunakan Algoritma C4.5 untuk CTR OUTDOOR dengan harapan agar dapat memudahkan prediksi penjualan.

Penelitian ini diharapkan akan bermanfaat menambah pengetahuan dan variasi penelitian dalam penerapan teknik-teknik data mining. Hasil dari penelitian ini diharapkan dapat dijadikan sebagai perbandingan bagi peneliti lain yang tertarik dalam penerapan teknik – teknik data mining pada area tersebut. Hasil penelitian ini juga dapat bermanfaat bagi penjual lainnya sebagai referensi. Model prediksi yang diperoleh dari penelitian ini juga dapat dipergunakan oleh penjual yang memiliki kesamaan dalam produknya, sebagai pendukung untuk meningkatkan penjualan serta meminimalisir adanya kerugian yang besar.

2 KAJIAN PUSTAKA

2.1 K-Means Clustering

Data Mining adalah proses yang memanfaatkan teknik-teknik statistik, matematika, dan kecerdasan buatan untuk mengekstrak dan mengidentifikasi informasi dan *knowledge* selanjutnya (atau pola-pola) yang berasal dari sekumpulan data yang sangat besar. Berbagai macam pola tersebut bisa dalam bentuk aturan bisnis, kesamaan-kesamaan, korelasi, tren, atau model-model prediksi. Kebanyakan literatur mendefinisikan data mining sebagai “proses yang rumit untuk mengidentifikasi pola-pola yang valid, baru, memiliki potensi bermanfaat, dan bisa dipahami, terhadap data yang disimpan di dalam database yang terstruktur”. Data diorganisir dalam baris-baris yang terstruktur menurut kategori, ordinal/berurutan, dan variable-variabel yang berkesinambungan. Data mining juga dikenal dengan istilah lain yaitu *Knowledge Discovery in Database dan Data Analysis*. Pengertian data mining Menurut Suntoro (2019)

mendefinisikan Data Mining adalah proses untuk mendapatkan informasi yang berguna dari basis data yang besar dan perlu diekstraksi agar menjadi informasi baru dan dapat membantu dalam pengambilan keputusan. Menurut Suyanto (2017:262) “k-means merupakan algoritma klasterisasi yang memiliki ide dasar sederhana dengan cara meminimalkan Sum of Squared Error (SSE) antara objek-objek data dengan sejumlah k centroid”.

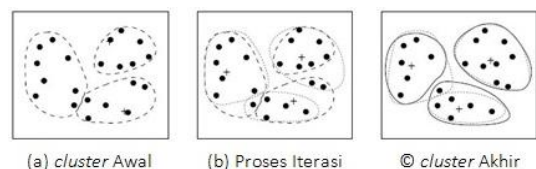
Salah satu metode data mining yang sering digunakan yaitu algoritma K-Means Clustering, untuk mengidentifikasi kelompok alami dari sebuah kasus yang didasarkan pada pengelompokan data yang memiliki kemiripan sehingga dengan pengelompokan tersebut dapat dilakukan analisis (Vulandari, 2017). Data-data dipilih menjadi beberapa kelompok dengan kriteria yang telah ditentukan lalu dikumpulkan menjadi satu dalam sebuah klaster, dimana setiap klaster ini memiliki *centroid* atau titik pusat (Suntoro, 2019).

Clustering adalah proses pengelompokan sejumlah data atau objek ke dalam klaster (*group*) sehingga setiap dalam klaster tersebut akan berisi data yang semirip mungkin dan berbeda dengan objek dalam klaster yang lainnya. Sampai saat ini, para ilmuwan masih terus melakukan berbagai usaha untuk melakukan perbaikan model klaster dan menghitung jumlah klaster yang optimal sehingga dapat dihasilkan klaster yang paling baik. Ada dua metode pengklasteran yang kita kenal, yaitu *hierarchical clustering* dan *partitioning*. Dalam *clustering*, data yang memiliki kesamaan dimasukkan ke dalam klaster yang sama, sedangkan data yang tidak memiliki kesamaan dimasukkan dalam klaster yang berbeda. Adapun menurut Rui Xu dan Donald (2009) *Clustering* (pengelompokan data) mempertimbangkan sebuah pendekatan penting untuk mencari kesamaan dalam data dan menempatkan data yang sama ke dalam kelompok-kelompok. *Clustering* membagi kumpulan data ke dalam beberapa 8 kelompok dimana kesamaan dalam sebuah kelompok

adalah lebih besar daripada diantara kelompok-kelompok.

K-Means Clustering adalah, K dimaksudkan sebagai konstanta jumlah klaster yang diinginkan, Means dalam hal ini berarti nilai suatu rata-rata dari suatu grup data yang dalam hal ini didefinisikan sebagai klaster, sehingga K-Means Clustering adalah suatu metode penganalisaan data atau metode data mining yang melakukan proses pemodelan tanpa supervisi (*unsupervised*) dan merupakan salah satu metode yang melakukan pengelompokan data dengan sistem partisi. Metode K-Means berusaha mengelompokkan data yang ada kedalam beberapa kelompok, dimana data dalam satu kelompok mempunyai karakteristik yang sama satu sama lainnya dan mempunyai karakteristik yang berbeda dengan data yang ada didalam kelompok yang lain. Pengertian K-means menurut beberapa ahli yaitu, menurut Suyanto (2017) K-means merupakan algoritma klasterisasi yang memiliki ide dasar sederhana dengan cara meminimalkan Sum of Squared Error (SSE) antara objek-objek data dengan sejumlah *centroid*.

Pada penelitian ini menggunakan metode Partitional clustering yaitu data dikelompokkan ke dalam sejumlah klaster tanpa adanya struktur hirarki antara satu dengan yang lainnya. Pada metode partitional clustering setiap klaster memiliki titik pusat klaster (*centroid*) dan secara umum metode ini memiliki fungsi tujuan yaitu meminimumkan jarak (*dissimilarity*) dari seluruh data ke pusat klaster masing-masing. Contoh metode partitional clustering: K-Means.



Sumber: Han dkk. (2012).
Gambar 1. Proses Pengklasteran Objek Menggunakan algoritma K-Means

2.2 Algoritma C4.5

C4.5 adalah algoritma yang sudah banyak dikenal dan digunakan untuk klasifikasi data yang memiliki atribut-atribut numerik dan kategorikal. Algoritma C4.5 diperkenalkan oleh J. Ross Quinlan yang merupakan pengembangan dari Algoritma ID3, dimana pengembangan dilakukan dalam hal bisa mengatasi missing data, bisa mengatasi data kontinyu, dan pruning. Algoritma tersebut digunakan untuk membentuk pohon keputusan. Pohon keputusan dianggap sebagai salah satu pendekatan yang paling populer. Dalam klasifikasi pohon keputusan terdiri dari sebuah node yang membentuk akar. Node akar tidak memiliki input. Node lain yang bukan sebagai akar tetapi memiliki tepat satu input disebut node internal atau *test node*, sedangkan node lainnya dinamakan daun, Daun mewakili nilai target yang paling tepat dari salah satu class. Hasil dari proses klasifikasi yang berupa aturan-aturan dapat digunakan untuk memprediksi nilai atribut bertipe diskret dari *record* yang baru.

Secara umum algoritma C4.5 untuk membangun pohon keputusan adalah sebagai berikut:

1. Pilih atribut sebagai akar.
2. Buat cabang untuk tiap-tiap nilai.
3. Bagi kasus dalam cabang.
4. Ulangi proses untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama.

Model yang didapatkan dari metode *decision tree* C4.5 kemudian dilakukan pengujian menggunakan *k-fold cross validation*. *Cross validation* adalah bentuk sederhana dari teknik statistik. Jumlah *fold* standar untuk memprediksi tingkat *error* dari data adalah dengan menggunakan *10-fold cross validation* (Witten, et al, 2011: 153).

Sebuah sistem yang melakukan klasifikasi tentunya diharapkan bisa memprediksi dengan benar. Namun tidak dapat dipungkiri jika pada penerapannya sistem tidak akan bisa 100% benar. Hal ini yang mengharuskan pengukuran kinerja dari sistem tersebut untuk mengetahui berbagai aspek pengujiannya. Menentukan suatu model

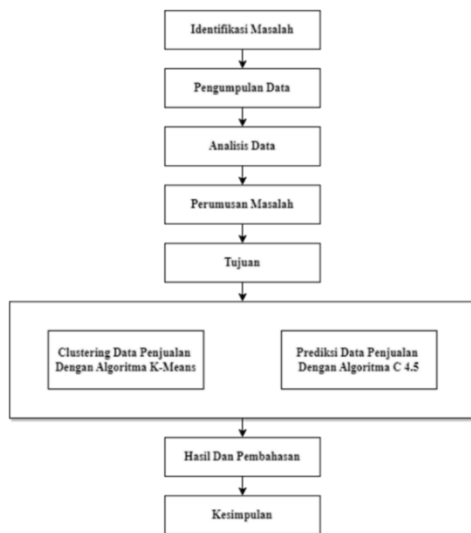
dari sistem baik atau buruk memerlukan elemen-elemen kunci diantaranya

- a. Akurasi Prediksi, yang menentukan seberapa akurat suatu model dalam memprediksi keluaran.
- b. Kecepatan, yang menunjukkan seberapa cepat suatu model dalam memproses data masukan.
- c. Robustness, menggambarkan kemampuan suatu model melakukan prediksi yang akurat walau dalam kondisi ekstrim dan banyak gangguan terjadi
- d. Skalabilitas, kemampuan suatu model memproses data baik dalam ukuran yang lebih besar maupun data dari bidang lain yang berbeda.
- e. Interpretability, menggambarkan kemudahan suatu model untuk dipahami dan diinterpretasikan.
- f. Kesederhanaan, merupakan sifat yang cenderung dipilih untuk menyelesaikan suatu permasalahan.

Umumnya, pengukuran kinerja klasifikasi dilakukan dengan *confusion matrix*. *Confusion matrix* merupakan alat pengukuran yang dapat digunakan untuk menghitung kinerja atau tingkat kebenaran proses klasifikasi. Dengan *confusion matrix* dapat dianalisa seberapa baik *classifier* dapat mengenali *record* dari kelas-kelas yang berbeda.

3 METODE PENELITIAN

Penelitian ini menggunakan metode atau pendekatan studi kasus (*Case Study*). Penelitian ini menggunakan jenis data kuantitatif. Sumber data diperoleh dari Toko CTR Outdoor Data tersebut diminta langsung oleh peneliti kepada pemilik Toko CTR Outdoor Data yang digunakan dalam penelitian ini adalah data penjualan 6 bulan terakhir.



Gambar 2. Skema Penelitian

4 HASIL DAN PEMBAHASAN

Dari data 6 bulan terakhir yang telah dikumpulkan didapat 10 jenis barang maka akan diolah menggunakan K-Means Clustering dan Prediksi menggunakan Algoritma C 4.5. Langkah awal dilakukan dengan menentukan banyak kluster dan nilai centroid awal jumlah kluster ditentukan berdasarkan variable pengelompokan data penjualan Tidak Laku, Kurang Laku, Laku, Sangat Laku, selanjutnya menentukan centroid awal dapat dilakukan secara acak.

Tabel 1. Data Penjualan Toko CTR Outdoor

No	Nama Produk	Juni	Juli	Agustus	September	Oktober
1	Sepatu Cayman Lite	24	29	20	20	25
2	Sepatu Gaea Riding	12	2	16	18	17
3	Sepatu Sandal Valor	21	12	15	13	23
4	Sandal Caldera Pinch	15	18	14	21	17
5	Jas Hujan Riding Rexon	42	15	32	8	11
6	Jaket Kanawa	22	24	30	21	18
7	Waist Bag Wall Run	26	34	30	24	4
8	Tas Bruni	25	21	37	41	23
9	Selempang X-Handlepack	6	26	4	40	18
10	Waist Bag Sender	15	15	37	46	22

Tabel 2. Nilai Awal Centroid

Centroid	Juni	Juli	Agustus	September	Oktober
C1	12	2	16	18	17
C2	15	18	14	21	17
C3	22	24	30	21	18
C4	25	21	37	41	23

Langkah selanjutnya, dari empat nilai centroid diatas, lalu kita hitung jarak data dengan centroid dengan rumus Euclidean

Distance dan langsung kita kelompokkan data berdasarkan kedekatan (similaritas)-nya, sehingga didapat data sebagai berikut (Tabel 3).

Maka jika dilihat dari jarak terpendek maka C1 berisikan 1 data, C2 berisikan 2 data, C3 berisikan 4 data, C4 berisikan 2 data yang dimana C1 adalah tidak laku, C2 kurang laku, C3 laku, C4 sangat laku. Selanjutnya akan di lakukan Prediksi menggunakan algoritma C 4.5.

Tabel 3. Jarak Terpendek Dan Similaritas Centroid Pada Iterasi 1

Literasi Pertama						
Data ke-i	C1	C2	C3	C4	Jarak Terpendek	Cluster
1	30.94	17.41	13.38	28.27	13.38	C3
2	0.00	16.67	28.11	39.19	0.00	C1
3	15.59	13.15	21.42	36.95	13.15	C2
4	16.67	0.00	18.49	32.77	0.00	C2
5	38.22	35.59	26.51	39.79	26.51	C3
6	28.11	18.49	0.00	22.18	0.00	C3
7	40.26	28.48	17.92	29.48	17.92	C3
8	39.19	32.77	22.18	0.00	0.00	C4
9	35.23	24.64	36.01	38.74	24.64	C2
10	37.79	34.47	28.64	12.73	12.73	C4

Tabel 4. Format Data Prediksi

No	Nama Barang	Warna	Size	Status
1	Sepatu Cayman Lite	Hitam	42	Laku
2	Sepatu Cayman Lite	Hitam	42	Laku
3	Sepatu Cayman Lite	Hitam	43	Laku
4	Sepatu Cayman Lite	Coklat	41	Laku
5	Sepatu Cayman Lite	Coklat	42	Kurang Laku
6	Sepatu Cayman Lite	Coklat	43	Kurang Laku
7	Sepatu Cayman Lite	Hitam	41	Laku
8	Sepatu Cayman Lite	Hitam	42	Laku
9	Sepatu Cayman Lite	Hitam	43	Kurang Laku
10	Sepatu Cayman Lite	Coklat	41	Kurang Laku
11	Sepatu Cayman Lite	Coklat	42	Laku
12	Sepatu Cayman Lite	Coklat	43	Kurang Laku
13	Sepatu Cayman Lite	Hitam	41	Laku
14	Sepatu Cayman Lite	Hitam	42	Laku
15	Sepatu Cayman Lite	Hitam	43	Laku
16	Sepatu Cayman Lite	Coklat	41	Laku
17	Sepatu Cayman Lite	Coklat	42	Kurang Laku
18	Sepatu Cayman Lite	Coklat	43	Kurang Laku
19	Sepatu Cayman Lite	Hitam	41	Laku
20	Sepatu Cayman Lite	Hitam	41	Kurang Laku
21	Sepatu Cayman Lite	Hitam	42	Kurang Laku
22	Sepatu Cayman Lite	Hitam	43	Laku
23	Sepatu Cayman Lite	Coklat	41	Kurang Laku
24	Sepatu Cayman Lite	Coklat	42	Laku
25	Sepatu Cayman Lite	Coklat	43	Laku
26	Sepatu Cayman Lite	Hitam	41	Laku
27	Sepatu Cayman Lite	Hitam	42	Laku
28	Sepatu Cayman Lite	Hitam	43	Kurang Laku

Berdasarkan format data pada tabel 4 diatas, akan dilakukan klasifikasi dengan algoritma C 4.5 dengan membuat pohon keputusan. Untuk memilih atribut sebagai akar, didasarkan pada nilai Gain tertinggi dari atribut-atribut yang ada. Selanjutnya dilakukan perhitungan Gain dan Entropy sebagai berikut (Tabel 5).

Tabel 5. Nilai Gain Dan Entropy

Atribut		Jumlah	Laku	Kurang laku	Entropy	Gain
Total		28	17	11	0.966619	
Nama Barang	Sepatu Cayman Lite	28	17	11	0.966619	0
Warna	Hitam	16	13	4	0.743393	0.541823
	Coklat	12	5	7	0.979869	
Size	41	9	6	3	0.918296	0.671452
	42	10	7	3	0.881291	
	43	9	4	5	0.991076	

Tahap ini menentukan atribut yang akan digunakan sebagai akar dengan cara menentukan Entropy terlebih dahulu lalu menentukan gain dan memilih gain tertinggi untuk digunakan sebagai akar. Pada tabel diatas Gain tertinggi adalah size yaitu sebesar 0.671452. Dengan demikian, size dapat di jadikan node akar. Ada 3 atribut pada size yaitu 41, 42, 43, sehingga perlu dilakukan perhitungan lagi.

Tabel 6. Nilai Atribut Size 41

Atribut		Jumlah	Laku	Kurang laku	Entropy	Gain
Total		28	17	11	0.966619	
Nama Barang	Sepatu Cayman Lite	28	17	11	0.966619	0
Warna	Hitam	5	4	1	0.721928	0.837703
	Coklat	4	2	2	1	

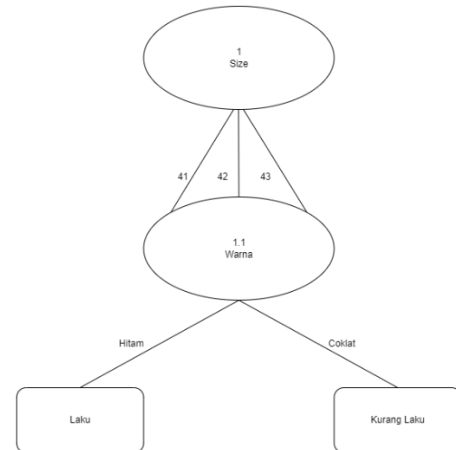
Tabel 7. Nilai Atribut Size 42

Atribut		Jumlah	Laku	Kurang laku	Entropy	Gain
Total		28	17	11	0.966619	
Nama Barang	Sepatu Cayman Lite	28	17	11	0.966619	0
Warna	Hitam	6	5	1	0.650022	0.827328
	Coklat	4	2	2	1	

Tabel 8. Nilai Atribut Size 43

Atribut		Jumlah	Laku	Kurang laku	Entropy	Gain
Total		28	17	11	0.966619	
Nama Barang	Sepatu Cayman Lite	28	17	11	0.966619	0
Warna	Hitam	5	3	2	0.970951	0.793235
	Coklat	4	1	2	1	

Berdasarkan 3 tabel diatas maka warna hitam termasuk kategori Laku dan warna coklat termasuk kategori Kurang Laku.



Gambar 3. Decision Tree

Berdasarkan gambar diatas maka dapat diambil keputusan bahwa warna hitam diminati konsumen, jika warna coklat kurang diminati konsumen

5 SIMPULAN

Berdasarkan hasil penelitian yang dilakukan, maka dapat di simpulkan bahwa penerapan K-Means dalam pengelompokan data penjualan dapat menghasilkan rekomendasi barang yang sangat laku, laku, kurang laku, dan tidak laku. Serta ditambah prediksi menggunakan C 4.5 dimana didapatkan bahwa faktor tertinggi yang mempengaruhi penjualan sepatu adalah warna sepatu, faktor size tidak mempengaruhi penjualan karena dengan size yang berbeda beda sepatu yang dijual masih diminati oleh konsumen. Sehingga dapat dijadikan bahan rujukan oleh toko CTR Outdoor dalam mengelola stok barang.

DAFTAR PUSTAKA

Arikunto, S. (2013). *Prosedur Penelitian*. Jakarta: Rineka Cipta.
 Hadi, F., & Diana, Y. (2020). Pengklusteran Penjualan Bahan Bangunan Menggunakan Algoritma K-Means. *JOISIE (Journal Inf. Syst. Informatics Eng*, 4(1), 22.

- Han, J. K. (2012). *Data Mining Concept and Techniques*. Amsterdam: Morgan Kaufmann-Elsevier.
- Indriyani, F., & Irfiani, E. (2019). Clustering Data Penjualan pada Toko Perlengkapan Outdoor Menggunakan Metode K-Means. *JUITA: Jurnal Informatika*, 7(2), 109-113.
- Mardalius, M. (2018). Pemanfaatan Rapid Miner Studio 8.2 Untuk Pengelompokan Data Penjualan Aksesoris Menggunakan Algoritma K-Means. *JURTEKSI (Jurnal Teknologi dan Sistem Informasi)*, 4(2), 123-132.
- Mudzakkir, B. D. (2018). Pengelompokan Data Penjualan Produk Pada Pt Advanta Seeds Indonesia Menggunakan Metode K-Means. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 2(2), 34-40.
- Normah, N., Nurajizah, S., & Salbinda, A. (2021). Penerapan data mining metode k-means clustering untuk analisa penjualan pada toko fashion hijab Banten. *Jurnal Khatulistiwa Informatika*, 7(2), 158-163.
- Pambudi, W. T., & Witanti, A. (2022). Penerapan Algoritma K-Means Untuk Menganalisis Data Penjualan Pada Toko Ayu Collection Berbasis Web. *Jurnal Sistem Informasi dan Bisnis Cerdas*, 15(1), 1-9.
- Putra, P. P., & Chan, A. S. (2018). Pengembangan Aplikasi Perhitungan Prediksi Stock Motor Menggunakan Algoritma C 4.5 Sebagai Bagian dari Sistem Pengambilan Keputusan (Studi Kasus di Saudara Motor). *Jurnal Inovtek Polbeng Seri Informatika*, 3(1), 24-33.
- Rahmah, S. A. (2020). Klasterisasi Pola Penjualan Pestisida Menggunakan Metode K-Means Clustering (Studi Kasus Di Toko Juanda Tani Kecamatan Hutabayu Raja). *Djtechno: Jurnal Teknologi Informasi*, 1(1), 1-5.
- Santosa, B. (2007). *Teknik Pemanfaatan Data untuk Keperluan Bisnis*. Yogyakarta: Graha Ilmu.
- Silalahi, M. (2018). Analisis Clustering Menggunakan Algoritma K-Means Terhadap Penjualan Produk Padapt Batamas Niaga Jaya. *Computer Based Information System Journal*, 6(2), 20-35.
- Siregar, M. H. (2018). Data Mining Klasterisasi Penjualan Alat-Alat Bangunan Menggunakan Metode K-Means (Studi Kasus Di Toko Adi Bangunan). *Jurnal Teknologi Dan Open Source*, 1(2), 83-91.
- Suyanto, D. (2017). Data Mining untuk klasifikasi dan klasterisasi data. *Bandung: Informatika Bandung*.
- Tamba, S. P., & Kesuma, F. T. (2019). Penerapan Data Mining Untuk Menentukan Penjualan Sparepart Toyota Dengan Metode K-Means Clustering: data mining; k-means-clustering. *Jurnal Sistem Informasi dan Ilmu Komputer Prima (JUSIKOM PRIMA)*, 2(2), 67-72.
- Tan, P. S. (2006). *Introduction to Data Mining*. Boston: Pearson Education.
- Vulandari, R. T. (2017). *Data Mining*. Yogyakarta: Gava Media.
- Witten, E. A. (2016). *Data Mining Practical Machine Learning Tools and Technique*. San Francisco: Morgan Kaufmann.
- Xu, R., & Wunsch II, D. C. (2009). *Clustering*, A John Wiley & Sons. Inc., Publication.