

Implementasi K-Means Untuk Pengelompokan Produk Terbaik PT. Koko Pelli

Muhammad Dahria, Rudi Gunawan, Zulkifli Lubis

Program Studi Informasi, STMIK Triguna Dharma, Medan, Indonesia
Email: ¹m_dahria@trigunadharma.ac.id, ²rudi_gunawan8899@yahoo.com, ³zlubis0968@gmail.com

Abstrak

PT. Koko Pelli adalah sebuah perusahaan yang bergerak dibidang retail. Dimana pada prosesnya masih memiliki kendala yang belum mampu menentukan produk terbaik untuk promosi kedepannya. Akibat dari hal tersebut sering terjadi kekurangan stok produk yang digemari konsumen. Untuk mengatasi masalah tersebut diperlukan sebuah sistem yang mampu menentukan produk terbaik yang tidak berpatokan pada satu produk saja melainkan kelompok produk terbaik. Pengelompokan produk terbaik akan dihitung menggunakan metode K-Means yang mampu menghitung secara tepat dan konsisten. Aplikasi yang dibangun dapat mengelompokkan produk terbaik dengan cepat yang akan membantu mengatasi masalah yang ada pada PT. Koko Pelli.

Kata Kunci: Retail, pengelompokan produk, k-means, clustering, promosi produk

1. PENDAHULUAN

PT. Koko Pelli adalah sebuah perusahaan yang bergerak dibidang retail. Dalam menjalankan usaha ini, setiap transaksi penjualan sudah menggunakan sistem komputer. Data-data penjualan tersebut selalu dicatat dan tersimpan dalam sebuah basis data (database). Pendokumentasian setiap transaksi ini sangat berguna untuk keperluan menganalisa penjualan.

Dalam proses promosi yang terjadi masih memiliki kendala, dimana dalam penentuannya produk terbaik hanya dapat diperoleh melalui data penjualan. Namun pada kenyataannya produk terbaik bukan hanya dinilai melalui penjualan tetapi juga kualitas dan harga. promosi produk biasanya berpatokan pada satu jenis produk yang berakibat banyak produk lain yang ketinggalan atau tidak dikenal oleh konsumen. Pengelompokan produk terbaik adalah solusi terbaik untuk mengatasi masalah yang terjadi pada PT. Koko Pelli.

Pengelompokan produk terbaik ini diharapkan dapat meningkatkan penjualan yang secara otomatis meningkatkan keuntungan. Maka dari itu diperlukan suatu algoritma atau metode dalam menentukan produk terbaik, sehingga tujuan yang diharapkan tersebut dapat tercapai.

Ketersediaan data yang cukup banyak, kebutuhan akan informasi (pengetahuan) sebagai pendukung pengambilan keputusan untuk membuat solusi bisnis serta dukungan infrastruktur dibidang teknologi informasi merupakan lahirnya suatu teknologi Data Mining. Data mining yang dimaksud untuk memberikan solusi nyata dalam pengambilan keputusan.

Salah satu metode untuk tujuan diatas adalah menggunakan teknik pengelompokan data dengan metode K-Means. K-Means merupakan salah satu algoritma clustering. Dimana algoritma K-Means sangat terkenal karena kemudahan dan kemampuannya untuk mengklarifikasi data besar dan outlier dengan sangat cepat.

Data penjualan yang sudah ada akan diolah dan dianalisa untuk mengetahui tingkat kecenderungan konsumen (pelanggan) dalam membeli produk terbaik yang mereka sukai. Dari pengolahan data tersebut diperoleh suatu pengetahuan yang dapat digunakan sebagai pendukung keputusan dalam membuat pengelompokan produk terbaik.

2. METODOLOGI PENELITIAN

Data mining adalah penambangan atau penemuan informasi baru dengan mencari pola atau aturan tertentu dari sejumlah data yang sangat besar. Data mining juga disebut sebagai serangkaian proses untuk menggali nilai tambah berupa pengetahuan yang selama ini tidak diketahui secara manual dari suatu kumpulan data.[1] [4] [9]

Istilah Data Mining dan Knowledge Discovery In Database(KDD) sering kali digunakan secara bergantian untuk menjelaskan proses penggalian informasi tersembunyi dalam suatu basis data yang besar. Sebenarnya kedua istilah tersebut memiliki konsep yang berbeda akan tetapi akan berkaitan satu sama lain. Knowledge Discovery In Database (KDD) adalah kegiatan yang meliputi pengumpulan pemakaian data historis untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar [2] [3] [5]

K-Means merupakan salah satu algoritma clustering. Metode pengelompokan data nonhierarki (sekatan) yang berusaha mempartisi data yang ada kedalam bentuk dua atau lebih kelompok. Metode ini mempartisi data ke dalam kelompok sehingga data yang berkarakteristik berbeda dikelompokkan ke dalam kelompok yang lain K-Means termasuk dalam *partitioning clustering* yaitu setiap data harus masuk dalam cluster tertentu dan memungkinkan bagi setiap data yang termasuk dalam cluster tertentu pada suatu tahapan proses, pada tahapan berikutnya berpindah ke *cluster* yang lain. Algoritma K-Means sangat terkenal karena kemudahannya dan kemampuannya untuk mengklasifikasikan data besar dan *outlier* dengan sangat cepat. Berikut adalah langkah-langkah algoritma K-Means:[7] [8]

1. Tentukan k sebagai jumlah *cluster* yang dibentuk.
2. Tentukan pusat (*centroid*) *cluster* awal.

Dalam menentukan buah pusat cluster awal dilakukan pembangkitan bilangan random yang mempresentasikan urutan data input.



Pusat awal *cluster* didapatkan dari data sendiri bukan dengan menentukan titik baru, yaitu dengan mengacak (random) pusat awal dari data. Kemudian untuk menghitung *centroid cluster* ke-I berikutnya digunakan rumus sebagai berikut :

$$v = \frac{\sum_{i=1}^n x_i}{n}; i = 1,2,3, \dots n \quad (1)$$

Keterangan :

v : *centroid* pada *cluster*

xi : objek ke-j

n : banyaknya objek/jumlah objek yang menjadi anggota *cluster*

3. Perhitungan jarak dengan pusat *cluster*

Untuk mengukur jarak antara data dengan pusat *cluster* digunakan *Euclidian Distance*. Jarak hasil perhitungan akan dilakukan perbandingan dan dipilih jarak terdekat antara data dengan pusat *cluster*, jarak ini menunjukkan bahwa data tersebut berada dalam suatu kelompok dengan pusat *cluster* terdekat.

Rumusnya adalah :

$$d(x,y) = ||x - y|| = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}; i = 1,2,3, \dots n \quad (2)$$

Keterangan :

xi : pusat *cluster*

yi : data

d(x,y) : *Euclidian Distance* yaitu jarak antara data pada titik x dan titik y menggunakan kalkulasi matematika.

4. Pengelompokan data

Setelah sejumlah populasi data tersebut menemukan kedekatan dengan salah satu *centroid* yang ada maka secara otomatis populasi data tersebut masuk kedalam kelas yang memiliki *centroid* yang bersangkutan.

5. Lakukan iterasi, kemudian posisi *centroid* baru dengan menggunakan persamaan (1)

6. Ulangi langkah 3 apabila masih ada data yang berpindah kelompok, atau ada perubahan nilai *centroid* diatas nilai ambang yang ditentukan, atau juga apabila perubahan nilai ambang yang ditentukan, atau juga apabila perubahan nilai pada fungsi objektif yang digunakan masih di atas nilai ambang yang ditentukan.

3. ANALISA DAN PEMBAHASAN.

3.1 Analisa Permasalahan

Produk merupakan segala sesuatu yang dapat ditawarkan produsen untuk diperhatikan, diminta, dicari, dibeli, digunakan atau dikonsumsi pasar sebagai pemenuhan kebutuhan atau keinginan pasar yang bersangkutan. Secara konseptual produk adalah pemahaman subyektif dari produsen atas sesuatu yang bisa ditawarkan, sebagai usaha untuk mencapai tujuan organisasi melalui pemenuhan kebutuhan dan keinginan konsumen, sesuai dengan kompetensi dan kapasitas organisasi serta daya beli pasar

Iterasi 1 adalah langkah awal yang digunakan dalam menyelesaikan suatu permasalahan. Dimana iterasi akan dilakukan berulang sampai masalah dapat dipecahkan.

1. Penentuan pusat (*centroid*) awal *cluster*

Untuk menentukan pusat (*centroid*) awal ditentukan dengan mengacak (*random*) dari data nilai yang sudah ada. Pada kasus ini pusat (*centroid*) awalnya adalah :

Tabel 1. Pusat (*Centroid*) Awal Cluster

Centroid 1	3	2	2
Centroid 2	4	4	4
Centroid 3	4	4	3

2. Perhitungan jarak dengan pusat *cluster* dengan rumus :

$$d(x,y) = ||x - y|| = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}; i = 1,2,3, \dots n$$

Perhitungan jarak dari data ke - 1 terhadap pusat *cluster* adalah

$$d(1,1) = \sqrt{(4 - 3)^2 + (4 - 2)^2 + (4 - 2)^2} = 3$$

$$d(1,2) = \sqrt{(4 - 4)^2 + (4 - 4)^2 + (4 - 4)^2} = 0$$

$$d(1,3) = \sqrt{(4 - 4)^2 + (4 - 4)^2 + (4 - 3)^2} = 1$$

Dan Seterusnya dilakukan jarak untuk data ke - 2 sampai data ke - 60. Kemudian akan didapatkan hasil perhitungan jarak setiap data terhadap pusat *cluster* baru sebagai berikut :

Tabel 2. Hasil Perhitungan Jarak data terhadap *cluster* pada iterasi ke - 1

KD	C1	C2	C3	Jarak Terdekat	JCT	JCT^2
A1	3,000	0,000	1,000	C2	0,000	0,000
A2	0,000	3,000	2,449	C1	0,000	0,000

A3	2,236	2,000	2,236	C2	2,000	4,000
----	-------	-------	-------	----	-------	-------

3. Menghitung Besar Rasio

Rasio Besaran Antara BCV (Betwen Cluster Variation) dengan WCV (Within Cluster Variation)

Karena Centroid $M1=(3,2,2)$, $M2=(4,4,4)$, $M3=(4,4,3)$

$$d(m1, m2) = \sqrt{(3 - 4)^2 + (2 - 4)^2 + (2 - 4)^2} = 3$$

$$d(m1, m3) = \sqrt{(3 - 4)^2 + (2 - 4)^2 + (2 - 3)^2} = 2,449$$

$$d(m2, m3) = \sqrt{(4 - 4)^2 + (4 - 4)^2 + (4 - 3)^2} = 1$$

$$BCV = d(m1, m2) + d(m1, m3) + d(m2, m3) = 3 + 2,449 + 1 = 6,449$$

$$\text{Besar Rasio} = 6,44 / 72 = 0,09$$

Tabel 3. Kesimpulan Akhir Pengelompokan Produk Terbaik

KD	Nama	Kelompok / Jarak Terdekat
A1	Andyra 1	C2
A2	Jovin	C1
A3	Narada 1	C1
A4	Narada 2	C3
A5	Gwine	C2
A6	Neomi 1	C2
A7	Shanel 8	C2
A8	Neomi 2	C2
A9	Victory 2	C2
A10	Chista 1	C2
A11	Snowbell 6	C2
A12	Chista 2	C2
A13	Denita	C3
A14	Feisol	C3

3.2 Hasil

Implementasi merupakan langkah yang digunakan untuk mengoperasikan sistem yang dibangun. Dalam bab ini akan dijelaskan bagaimana menjalankan sistem tersebut. Dibawah ini merupakan tampilan dari implemenatsi Data Mining Pengelompokan Produk Terbaik Pada PT. Koko Pelli Menggunakan Metode K-Means



Gambar 1. Form Produk



Gambar 2. Form Penilaian



Gambar 3. Form Proses

4. KESIMPULAN

Dari hasil pembahasan bab I sampai bab IV mengenai Data Mining Pengelompokan Produk Terbaik Pada PT. Koko Pelli Menggunakan Metode

K-Means dapat diambil beberapa kesimpulan sebagai berikut:

1. Untuk menyelesaikan permasalahan terkait pengelompokan produk terbaik pada PT. Koko Pelli dibutuhkan sebuah sistem data mining dengan metode K-Means.
2. Untuk merancang sistem untuk mengelompokkan pola penjualan produk terbaik pada PT. Koko Pelli dibutuhkan sebuah aplikasi yang dapat membangun sebuah sistem yaitu Microsoft Visual Studio 2010.
3. Untuk menerapkan metode K-Means dalam menganalisa pengelompokan produk terbaik pada PT. Koko Pelli dilakukan dengan menentukan kriteria penilaian dan data produk yang akan dinilai

REFERENCES

- [1] Wu, X., Kumar, V., Ross, Q. J., Ghosh, J., Yang, Q., Motoda, H., ... Steinberg, D. (2008). Top 10 algorithms in data mining. Knowledge and Information Systems. <https://doi.org/10.1007/s10115-007-0114-2>
- [2] Huang, Z. (1997). A Fast Clustering Algorithm to Cluster Very Large Categorical Data Sets in Data Mining. Research Issues on Data Mining and Knowledge Discovery. <https://doi.org/10.1.1.6.4718>
- [3] Bienkowski, M., Feng, M., & Means, B. (2014). Enhancing teaching and learning through educational data mining and learning analytics: An issue brief. In Educational Improvement Through Data Mining and Analytics.
- [4] Münz, G., Li, S., & Carle, G. (2007). Traffic Anomaly Detection Using K-means Clustering. Leistungs-, Zuverl• assigkeits- Und Verl• asslichkeitsbewertung von Kommunikationsnetzen Und Verteilten Systemen.
- [5] Huang, J. Z., Ng, M. K., Rong, H., & Li, Z. (2005). Automated variable weighting in k-means type clustering. IEEE Transactions on Pattern Analysis and Machine Intelligence. <https://doi.org/10.1109/TPAMI.2005.95>
- [6] Zhang, B., Hsu, M., & Dayal, U. (1999). K-Harmonic means - A data clustering algorithm. HP Laboratories Technical Report.
- [7] Shi, N., Liu, X., & Guan, Y. (2010). Research on k-means clustering algorithm: An improved k-means clustering algorithm. 3rd International Symposium on Intelligent Information Technology and Security Informatics, IITSI 2010. <https://doi.org/10.1109/IITSI.2010.74>
- [8] Sumathi, S., & Esakkirajan, S. (2007). Data mining and data warehousing. Studies in Computational Intelligence. https://doi.org/10.1007/978-3-540-48399-1_10
- [9] Singh, A., Yadav, A., & Rana, A. (2013). K-means with Three different Distance Metrics. International Journal of Computer Applications. <https://doi.org/10.5120/11430-6785>