

DATA MINING UNTUK PENGELOMPOKAN JENIS USAHA DI RUMAH BUMN BATAM MENGGUNAKAN METODE CLUSTERING

Arif Hernawan¹,
Rahmat Fauzi²

¹Mahasiswa Program Studi Teknik Informatika, Universitas Putera Batam

²Dosen Program Studi Teknik Informatika, Universitas Putera Batam

email: pb180210071@upbatam.ac.id

ABSTRACT

Business actors in Indonesia are generally categorized into large businesses and small and medium enterprises or often known as UKM. There is one State-Owned Enterprise that has the role of gathering and encouraging UKM players to upgrade their classes to be more prosperous, namely Rumah BUMN Batam. The research objective is to classify the types of UKM based on the frequency of sales, so that later the company can carry out further promotions for UKM that get a low number of orders. The k-means clustering algorithm can be used by Rumah BUMN Batam to facilitate the grouping of types of business and the frequency of orders for UKM per year. The author uses the Knowledge Discovery in Database (KDD) process which consists of data cleaning, data integration, data selection, data transformation, data mining, pattern evaluation, and knowledge presentation.

Keywords: Algoritma K-Means, Clustering, Data Mining

PENDAHULUAN

Mayoritas dari 57 juta usaha kecil menengah (UKM) di Indonesia adalah usaha mikro. Bagan ini menunjukkan potensi usaha kecil dan menengah (UKM) sebagai salah satu mesin ekonomi Indonesia untuk meningkatkan kesejahteraan. Telkom Indonesia mengembangkan program Kampung UKM Digital yang memiliki 2 juta UKM terdaftar. seluruh Indonesia. Cikal bakal Rumah BUMN Ekspansi pasar global telah mengubah paradigma bisnis

nasional sehingga usaha kecil dan menengah (UKM) berkontribusi signifikan terhadap kemakmuran perekonomian bangsa melalui penciptaan lapangan kerja, peningkatan kesejahteraan sosial. kesejahteraan, dan pengembangan ide-ide baru.

Dengan Banyaknya pelaku UKM dan berbagai macam jenis produk yang dihasilkan tentu memerlukan sistem yang mengatur tentang data para pelaku UKM agar memudahkan dalam pemasaran produk mereka dan juga memudahkan para pembeli atau costumer dalam

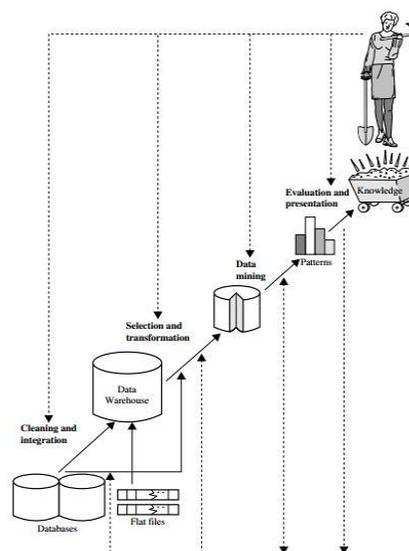
mencari produk yang diinginkan. baik pengelolaan produk maupun pemasaran, serta memudahkan pembeli dalam memenuhi kebutuhannya. Dalam daftar sumber tentang Visi dan Misi Rumah BUMN akan membantu dan mendorong usaha kecil dan menengah (UKM) untuk menjawab tantangan utama pembangunan UKM dengan meningkatkan kompetensi guna mempermudah akses pemasaran dan permodalan.

(Nikmatun & Waspada, 2019) penambangan data adalah bidang multidisiplin yang membahas masalah pengambilan data dari basis data besar menggunakan metode dari pembelajaran mesin, pengenalan pola, statistik, basis data, dan visualisasi. Penambangan data juga dapat digunakan untuk mengekstraksi informasi dari sejumlah besar data untuk membuat prediksi dengan informasi tersebut. Pengelompokan (clustering), regresi (regression), asosiasi (association), dan klasifikasi (classification) hanyalah beberapa dari sekian banyak metode yang digunakan dalam data mining. diambil dari hasil penelitian terdahulu. Selain itu, pendahuluan berisikan tujuan penelitian dan rencana pemecahan masalah pada penelitian.

KAJIAN TEORI

2.1 Knowledge Discovery of Database (KDD)

Knowledge Discovery of Database (KDD) yaitu proses mengolah informasi yang bermanfaat serta belum diketahui sebelumnya dari kumpulan data.



Gambar 2.1 Tahapan Proses Knowledge Discovery Database (KDD)
Sumber: (Gustientiedina et al., 2019)

2.2 Data Mining

Grup Gartner berpendapat Penambangan data merupakan proses penggunaan teknologi pengenalan pola statistik dan matematis untuk menemukan korelasi, pola, dan tren baru yang berarti disimpannya data dalam jumlah besar.

Pendapat (Annur, 2019) ada tiga tehnik yang terkenal pada *data mining*, yaitu:

1. Aturan Asosiatif (*Association rule mining*)

Teknik penambangan yang disebut aturan asosiatif digunakan untuk menemukan aturan asosiatif antara satu set item dengan item yang lain. Untuk menunjukkan bagaimana objek data terkait, aturan asosiatif digunakan. Dua langkah berbeda membuat aturan asosiatif: Temukan

dukungan minimum yang digunakan untuk menentukan keseluruhan pengulangan itemset dalam database terlebih dahulu. Kedua, mengulangi kumpulan item dan menetapkan batas kepercayaan minimum aturan (Aher & Lobo, 2012:1).

2. Klasifikasi (*Classification*)
prosedur menemukan model yang berfungsi mendeskripsikan atau membedakan konsep *class* data untuk memperkirakan kelas dari objek yang tidak memiliki label. Jaringan saraf (*neural network*), pohon keputusan (*decision tree*), atau aturan jika-maka (*if-then*) bisa menjadi model itu sendiri.
3. Pengelompokan (*Clustering*)
Tanpa didasarkan pada kelas data tertentu, clustering mengelompokkan data. Kelas data yang tidak diketahui dapat diberi label dengan bantuan pengelompokan. Tujuan clustering adalah untuk meminimalkan kesamaan antar kelas sekaligus memaksimalkan kesamaan antara anggota satu kelas.

2.3 Metode Data Mining

Metode *clustering* digunakan dalam penelitian ini. Proses pengelompokan sekumpulan objek data menjadi beberapa kelompok atau *cluster* sehingga objek dalam satu cluster memiliki banyak kesamaan dengan yang ada di *cluster* lain tetapi sangat berbeda. Evaluasi kesamaan dan ketidaksamaan bergantung pada jumlah atribut yang menggambarkan objek dan seringkali termasuk perlakuan jarak. Bidang seperti biologi, keamanan, intelijen bisnis, dan pencarian web, dapat memperoleh manfaat dari penggunaan pengelompokan sebagai alat penambangan data. (Priyadi et al., 2019)

2.4 Algoritma K-Means

Cluster awal yang dibuat oleh algoritma K-Means dipusatkan pada sejumlah anggota populasi yang berbeda. Pada titik ini, pusat kluster dipilih secara acak dari sekumpulan data populasi. K-Means menguji data swarm pada langkah berikut untuk mengetahui jarak minimum antara setiap komponen dan setiap pusat cluster yang telah ditentukan sebelumnya. (Wahyudi et al., 2020)

Dibawah ini merupakan langkah-langkah untuk pengelompokan data

1. Menentukan berapa banyak cluster yang dibutuhkan.
2. Kemudian menentukan pusat cluster secara random dan melakukan inialisasi data agar data mudah diolah.
3. Dalam menentukan jarak dari objek satu dengan objek yang lain, posisi dari setiap titik data bisa diletakkan pada cluster yang terdekat. Dalam tahap ini, Euclidean berperan sebagai penghitung jarak dan menentukan perbedaan dan kesamaan data. Dengan rumus sebagai berikut:

Rumus 2.2 Rumus Euclidean

$$d(x, y) = \sqrt{\sum_i^n (x_i - y_i)^2}$$

dimana:

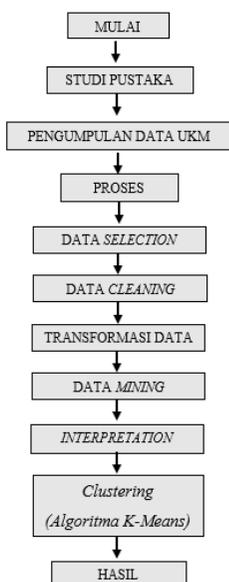
“ $d(x, y)$ = ukuran ketidaksamaan
 $x_i = (x_1, x_2, x_3, \dots, x_n)$ = ukuran ketidaksamaan

$y_i = (y_1, y_2, y_3, \dots, y_n)$ = variabel titik pusat”

4. Rata-rata objek pada cluster digunakan dalam perhitungan pusat cluster. Median bisa juga bisa digunakan dalam menghasilkan perhitungan.
5. Agar proses pengklasteran selesai, maka harus melakukan penghitungan ulang pada jarak diantara objek dan juga pusat cluster sampai dengan hasil cluster sama atau stabil.

METODE PENELITIAN

Metode yang dipilih dan diterapkan dalam penelitian ini menggunakan langkah-langkah dari proses awal penelitian hingga proses akhir penelitian, sehingga ditarik kesimpulan dari seluruh proses penelitian, Adapun langkah-langkah penelitian dapat dilihat pada Gambar 3.1 Berikut:



Gambar 3.1 Diagram Alur Penelitian
Sumber: (Peneliti, 2023)

1. *Data Selection*
Penelitian ini menggunakan tahapan data selection untuk menginisialisasikan Kategori UKM menjadi nominal 1 (Makanan&Minuman), 2 (Jasa), 3 (Craft/Kerajinan Tangan) dan 4 (Fashion/Busana). Tahapan ini harus dilakukan sebelum tahapan penggalian informasi data pada KDD dimulai.
2. *Data Cleaning*
Pada tahapan ini kemudian data UKM yang dilakukan proses pembersihan atau cleaning. Pembersihan data yang dilakukan adalah guna menghapus atribut-atribut yang tidak perlu dalam proses klasifikasi nantinya. Pada penelitian ini atribut yang dihapus seperti nomor telepon, nomor hp, tipe bisnis dan alamat.
3. *Data Transformation*.
Pada proses ini kemudian data yang akan di input dan digunakan terlebih dahulu dilakukan normalisasi. Tujuannya yaitu agar data berada pada range [0-4] sehingga penyebaran data tidak terlalu jauh.
4. *Data Mining*
Tahapan pada proses ini Algoritma K-Means yang merupakan salah satu algoritma dalam metode association rule mining menjadi dasar pengolahan data pada penelitian ini yang dilakukan dengan bantuan software Weka. Oleh karena itu, metode ini menghasilkan aturan dari data yang diproses sebagai keluarannya.
5. *Interpretation*
Tahapan selanjutnya yaitu proses evaluasi, dapat dilihat dari parameter algoritma yang digunakan (support, confidence, lift). Proses evaluasi untuk menganalisa rules yang didapatkan sehingga dihasilkan suatu

pengetahuan baru berdasarkan analisis untuk digunakan dalam pengambilan keputusan nantinya.

HASIL DAN PEMBAHASAN

4.1 Analisis Data

Penelitian ini menggunakan data UKM (Usaha Kecil Menengah) yang termasuk didalamnya yaitu Data Produk, Jenis Produk, Jumlah pemesanan dan Kategori Produk dalam rentang periode Januari 2020 sampai dengan Desember 2021 di Rumah BUMN Batam.

1. Penentuan jumlah cluster

Cluster yang digunakan pada penelitian ini memiliki 4 cluster. Ini mengacu pada kebutuhan yang berguna nantinya bagi Rumah BUMN Batam. Tabel berikut menunjukkan pusat centroid awal dan 4 cluster.

Tabel 4.1 Pusat Centroid Baru

Cluster	JP 2020	JP 2021
C0	436	398
C1	244	320
C2	636	670
C3	29	15

Sumber: (Peneliti, 2023)

2. Menghitung jarak ke Pusat Cluster

Jika titik awal pusat cluster sudah didapatkan, proses selanjutnya yaitu menghitung jarak *Euclidean*, dan dikelompokkan berdasarkan jarak terkecil. Kemudian didapatkan nilai *centroid* baru yang digunakan sebagai

acuan perhitungan dengan nilai *centroid* sebelumnya.

$$d_{1,0} = \sqrt{(350 - 436)^2 + (556 - 398)^2} = 179,8889$$

$$d_{1,1} = \sqrt{(350 - 244)^2 + (556 - 320)^2} = 258,7122$$

$$d_{1,2} = \sqrt{(350 - 636)^2 + (556 - 670)^2} = 307,8831$$

$$d_{1,3} = \sqrt{(350 - 29)^2 + (556 - 15)^2} = 629,0644$$

$$d_{2,0} = \sqrt{(430 - 436)^2 + (237 - 398)^2} = 161,1118$$

$$d_{2,1} = \sqrt{(430 - 244)^2 + (237 - 320)^2} = 203,6787$$

$$d_{2,2} = \sqrt{(430 - 636)^2 + (237 - 670)^2} = 479,505$$

$$d_{2,3} = \sqrt{(430 - 29)^2 + (237 - 15)^2} = 458,3503$$

$$d_{3,0} = \sqrt{(934 - 436)^2 + (230 - 398)^2} = 525,574$$

$$d_{3,1} = \sqrt{(934 - 244)^2 + (230 - 320)^2} = 695,8448$$

$$d_{3,2} = \sqrt{(934 - 636)^2 + (230 - 670)^2} = 531,417$$

$$d_{3,3} = \sqrt{(934 - 29)^2 + (230 - 15)^2} = 930,1882$$

$$d_{4,0} = \sqrt{(440 - 436)^2 + (321 - 398)^2} = 77,10383$$

$$d_{4,1} = \sqrt{(440 - 244)^2 + (321 - 320)^2} = 196,0026$$

$$d_{4,2} = \sqrt{(440 - 636)^2 + (321 - 670)^2} = 400,2712$$

$$d_{4,3} = \sqrt{(440 - 29)^2 + (321 - 15)^2} = 512,4032$$

$$d_{5,0} = \sqrt{(480 - 436)^2 + (326 - 398)^2} = 84,38009$$

$$d_{5,1} = \sqrt{(480 - 244)^2 + (326 - 320)^2} = 236,0763$$

$$d_{5.2} = \sqrt{(480 - 636)^2 + (326 - 670)^2} = 377,7195$$

$$d_{5.3} = \sqrt{(480 - 29)^2 + (326 - 15)^2} = 547,8339$$

$$d_{6.0} = \sqrt{(654 - 436)^2 + (254 - 398)^2} = 261,2661$$

$$d_{6.1} = \sqrt{(654 - 244)^2 + (254 - 320)^2} = 415,2782$$

$$d_{6.2} = \sqrt{(654 - 636)^2 + (254 - 670)^2} = 416,3892$$

$$d_{6.3} = \sqrt{(654 - 29)^2 + (254 - 15)^2} = 669,1383$$

$$d_{7.0} = \sqrt{(231 - 436)^2 + (652 - 398)^2} = 327,9649$$

$$d_{7.1} = \sqrt{(231 - 244)^2 + (652 - 320)^2} = 334,2529$$

$$d_{7.2} = \sqrt{(231 - 636)^2 + (652 - 670)^2} = 405,3159$$

$$d_{7.3} = \sqrt{(231 - 29)^2 + (652 - 15)^2} = 670,1679$$

$$d_{8.0} = \sqrt{(470 - 436)^2 + (321 - 398)^2} = 84,17244$$

$$d_{8.1} = \sqrt{(470 - 244)^2 + (321 - 320)^2} = 226,0022$$

$$d_{8.2} = \sqrt{(470 - 636)^2 + (321 - 670)^2} = 386,4673$$

$$d_{8.3} = \sqrt{(470 - 29)^2 + (321 - 15)^2} = 536,7653$$

$$d_{9.0} = \sqrt{(344 - 436)^2 + (540 - 398)^2} = 169,1981$$

$$d_{9.1} = \sqrt{(344 - 244)^2 + (540 - 320)^2} = 241,6609$$

$$d_{9.2} = \sqrt{(344 - 636)^2 + (540 - 670)^2} = 319,631$$

$$d_{9.3} = \sqrt{(344 - 29)^2 + (540 - 15)^2} = 612,2499$$

$$d_{10.0} = \sqrt{(430 - 436)^2 + (298 - 398)^2} = 100,1798$$

$$d_{10.1} = \sqrt{(430 - 244)^2 + (298 - 320)^2} = 187,2966$$

$$d_{10.2} = \sqrt{(430 - 636)^2 + (298 - 670)^2} = 425,2293$$

$$d_{10.3} = \sqrt{(430 - 29)^2 + (298 - 15)^2} = 490,8055$$

Perhitungan selanjutnya yaitu data ke-11 hingga data ke-87 dilanjutkan hingga hasil perhitungan setiap iterasi dapat digunakan untuk menentukan titik *Centroid* baru.

3. Menentukan Titik *Centroid* Baru

Setelah didapatkan anggota cluster, selanjutnya dilakukan perhitungan untuk menemukan titik pusat cluster baru. Perhitungan dilakukan berdasarkan dari data anggota setiap cluster yang sesuai dengan rumus pusat cluster.

Tabel 4.2 Pusat *Centroid* Baru

CENTROID	JP 2020	JP 2021
C0	459,9375	380,375
C1	248,3181818	297,9090909
C2	819,1666667	897,4
C3	65,89473684	62,73684211

Sumber: (Peneliti, 2023)

Pada tahap pengolahan data, iterasi dilakukan sampai ke iterasi 8, dikarenakan iterasi ke-7 dan iterasi ke-8 sudah tetap atau mempunyai anggota *cluster* yang sama. Maka tahap olah data tidak perlu dilanjutkan atau tidak lagi diperlukan perulangan kembali, karena hasil *clustering* sudah konvergen atau sudah stabil.

Tabel 4.3 Pusat *Centroid* Baru Hasil Iterasi 8

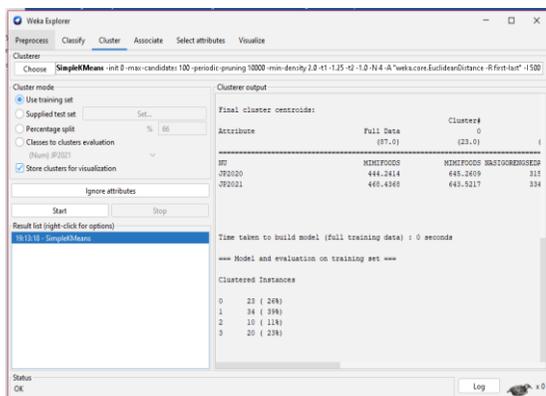
CENTROID	JP 2020	JP 2021
C0	658,9047619	666,2380952
C1	326,2222222	338,1111111
C2	1167,2	1316,1
C3	69,8	71,5

Sumber: (Peneliti, 2023)

Hasil dari jumlah masing-masing cluster dapat ditarik kesimpulan seperti pada berikut ini:

- Cluster 0 : memiliki 21 Anggota yang berarti Banyak Dipesan
- Cluster 1 : memiliki 36 Anggota yang berarti Paling Banyak Dipesan
- Cluster 2 : memiliki 10 Anggota yang berarti Paling Sedikit Dipesan
- Cluster 3 : memiliki 20 Anggota yang berarti Sedikit Dipesan

Hasil perhitungan dari *Software WEKA* 3.9.6 ada pada gambar berikut:



Gambar 4.1 Hasil *Cluster* Akhir
Sumber: (Peneliti, 2023)

Hasil akhir dari jumlah cluster yaitu

- Cluster 0 : 23 (26%)
- Cluster 1 : 34 (39%)
- Cluster 2 : 10 (11%)
- Cluster 3 : 20 (23%)

Pada penelitian ini, ada sedikit perbedaan hasil akhir antara cluster pada hitungan manual *Euclidean Distance* dengan hitungan dari software *WEKA* 3.9.6 ini dikarenakan pada software *WEKA* 3.9.6 ada perpindahan 2 anggota cluster dari cluster 1 ke cluster 0.

Hasil anggota cluster pada hitungan manual *Euclidean Distance* yaitu :
 Cluster 0:21(24%) berarti banyak dipesan
 Cluster 1:36 (41%) berarti paling banyak dipesan
 Cluster 2:10(11%) berarti paling sedikit dipesan
 Cluster 3 :20 (23%) berarti sedikit dipesan

Hasil anggota cluster pada hitungan software *WEKA* 3.9.6 yaitu :
 Cluster 0:23(26%) berarti banyak dipesan
 Cluster 1 : 34 (39%) berarti paling banyak dipesan
 Cluster 2 : 10 (11%) berarti paling sedikit dipesan
 Cluster 3 :20 (23%) berarti sedikit dipesan

Hasil persentase dari kedua perhitungan diatas adalah stabil atau konvergen. Cluster 2 dan 3 harus dilakukan promosi lebih lanjut oleh Rumah BUMN Batam agar frekuensi jumlah pesanan meningkat dan dapat mensejahterakan pelaku UKM tersebut. Jika dilihat dari jenis UKM yang berada di cluster 2 dan 3 yang mendominasi adalah jenis UKM C dan B yaitu Craft/Kerajinan Tangan dan Jasa.

SIMPULAN

Dari hasil pengujian yang dilakukan, maka bisa ditarik kesimpulan seperti berikut :

1. Algoritma k-means clustering dapat dimanfaatkan oleh Rumah BUMN Batam untuk mempermudah pengelompokan jenis usaha dan frekuensi pemesanan UKM pertahun. Untuk UKM yang memiliki jumlah frekuensi pemesanan sedikit dan sangat sedikit akan dibantu promosi lebih lanjut lagi oleh Rumah BUMN.
2. Terlihat pada cluster 2 adalah jenis usaha yang paling banyak menerima pesanan selama periode 2020-2021. Jenis usaha yang termasuk pada cluster 2 bervariasi, dan didominasi oleh kategori 1 yaitu Makanan & Minuman.
3. Untuk cluster 2 dan 3 harus dilakukan promosi lebih lanjut oleh Rumah BUMN Batam agar frekuensi jumlah pesanan meningkat dan dapat mensejahterakan pelaku UKM tersebut. Jika dilihat dari jenis UKM yang berada di cluster 2 dan 3 yang mendominasi adalah jenis UKM C dan B yaitu Craft/Kerajinan Tangan dan Jasa

DAFTAR PUSTAKA

- Annur, H. (2019). Penerapan Data Mining Menentukan Strategi Penjualan Variasi Mobil Menggunakan Metode K-Means Clustering. *Jurnal Informatika Upgris*, 5(1). <https://doi.org/10.26877/jiu.v5i1.3091>
- Gustientiedina, G., Adiya, M. H., & Desnelita, Y. (2019). Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan. *Jurnal Nasional Teknologi Dan Sistem Informasi*, 5(1), 17–24.

<https://doi.org/10.25077/teknosi.v5i1.2019.17-24>

- Nikmatun, I. A., & Waspada, I. (2019). Implementasi Data Mining untuk Klasifikasi Masa Studi Mahasiswa Menggunakan Algoritma K-Nearest Neighbor. *Jurnal SIMETRIS*, 10(2), 421–432.
- Priyadi, I., Santony, J., & Na'am, J. (2019). Data Mining Predictive Modeling for Prediction of Gold Prices Based on Dollar Exchange Rates, Bi Rates and World Crude Oil Prices. *Indonesian Journal of Artificial Intelligence and Data Mining*, 2(2), 93. <https://doi.org/10.24014/ijaidm.v2i2.6864>
- Wahyudi, M., Masitha, M., Saragih, R., Solikhun, S., & Simarmata, J. (2020). *Data Mining: Penerapan Algoritma K-Means Clustering dan K-Medoids Clustering*. Yayasan Kita Menulis.

	<p>Biodata Penulis pertama, Arif Hernawan merupakan mahasiswa Prodi Teknik Informatika Universitas Putera Batam.</p>
	<p>Biodata Penulis kedua, Rahmat Fauzi, S.Kom., M.Kom. merupakan Dosen Prodi Teknik Informatika Universitas Putera Batam.</p>