

Penerapan K-means Clustering pada Penjualan dan Pajak BBM di DKI Jakarta

Pramudya Aziz Wisnuadi^{1*}, Bagus Irfanzah Arda Nugraha², Hiskia Kus Setiawan³

^{1,2}Teknik Informatika

Universitas Duta Bangsa Surakarta

^{1*}202020588@mhs.udb.ac.id, ²202021094@mhs.udb.ac.id, ³202020630@mhs.udb.ac.id

Abstrak— Bahan Bakar Minyak (BBM) adalah bagian dari barang dagangan utama yang berasal dari hasil penyulingan minyak bumi dan gas bumi, baik melalui pengambilan langsung maupun pengolahan minyak mentah. Pajak merupakan kewajiban yang harus dibayar oleh orang pribadi atau badan kepada negara berdasarkan paksaan undang-undang. Pada penelitian ini metode yang digunakan adalah K-Means Clustering data mining yang berfungsi mengelompokkan data yang telah terkumpul menjadi beberapa bagian, mekanisme K-Means Clustering diimplementasikan dengan tool RapidMiner. Sumber data yang terekam di website Jakarta Open Data, khususnya data penjualan BBM dan perpajakan DKI Jakarta digunakan dalam penelitian ini. Kriteria yang digunakan adalah penentuan titik berat secara acak. Proses iterasi dilakukan sebanyak 7 kali, sehingga terbentuk cluster pertama dengan jumlah elemen terbanyak yaitu 195 item dan cluster kedua dengan 9 item.

Kata kunci— Data mining, K-means, clustering, Rapid Miner.

Abstract— Fuel oil (BBM) is part of the main merchandise originating from refined petroleum and natural gas, either through direct extraction or processing of crude oil. Tax is an obligation that must be paid by an individual or company to the state based on coercion by law. In this study, the method used is K-Means Clustering data mining, which functions to group data that has been collected into several parts, the K-Means Clustering mechanism is implemented with the RapidMiner tool. Sources of data recorded on the Jakarta Open Data website, especially data on fuel sales and taxation for DKI Jakarta are used in this study. The criterion used is the determination of the center of gravity randomly. The iteration process is carried out 7 times, so that the first cluster with the highest number of elements is formed, including 195 items and the second with 9 items.

Keywords— Data mining, K-means, clustering, Rapid Miner.

I. PENDAHULUAN

Bahan bakar minyak (BBM) merupakan sebagian komoditas yang berasal dari sumber daya alam minyak dan gas bumi yang dapat ditemukan atau diolah dari minyak bumi. Minyak mentah merupakan produk alami yang terbentuk melalui proses hidrokarbon pada tekanan dan suhu tertentu, dan berbentuk cair atau padat. Minyak bumi merupakan SDA sangat penting yang tidak bisa direformasi karena dikuasi oleh negara. Keberadaannya memiliki mendasar dalam hajat hidup semua orang.

Pajak adalah suatu keharusan yang wajib dipatuhi oleh masyarakat dan badan hukum oleh Negara yang sesuai dengan aturan perundang-undangan, tanpa menimbulkan ketidakseimbangan secara langsung, dan dimaksudkan guna memenuhi kebutuhan Negara yaitu upaya mencapai kesejahteraan rakyatnya. Pemenuhan pajak merupakan wujud tanggung jawab negara dan partisipasi aktif wajib pajak dalam pemenuhan kewajiban perpajakan untuk mendukung pembiayaan negara serta ekspansi nasional.

Berdasarkan asas hukum perpajakan, penunaian pajak bukan hanya suatu kewajiban, tetapi juga hak setiap warganya guna ikut serta melalui bentuk kontribusi dalam pemenuhan negara serta ekspansi nasional. Dewasa ini kendaraan pajak BBM meningkat yang membuat semakin data masuk ke dalam sistem penjualan dan pajak BBM di Jakarta. Data yang menumpuk disebabkan oleh banyaknya data yang masuk. Sehingga dapat ditemukan informasi yang tersembunyi dengan mengelola data yang tersimpan selama bertahun-tahun, tentunya pihak pemerintah sangat membutuhkan informasi tersebut.

Metode yang diterapkan untuk mengatasi penumpukan volume data penjualan BBM dan pajak di DKI Jakarta adalah melalui implementasi data mining. Teknik data mining berfungsi untuk mengolah data dalam jumlah besar menghasilkan informasi yang relevan yang sering disebut sebagai knowledge discovery in databases (KDD). Dalam rangka pengelompokan serta penjualan BBM dan pajak di DKI Jakarta, metode yang digunakan yaitu metode clustering. Menurut Narwati (2010), Clustering (pengelompokan) merupakan suatu pendekatan guna mengidentifikasi kelompok-

kelompok objek yang mempunyai kesamaan mengungkapkan model distribusi dan keterkaitan terhadap sekumpulan data yang besar. Pada bagian rangkaian pengelompokan, hal yang perlu diperhatikan yaitu mengelompokkan model pada kelompok yang sesuai guna mengidentifikasi kesamaan atau perbedaan yang mengarah pada hasil akhir yang sangat penting.

Metode clustering memanfaatkan algoritma kmeans pada saat pengelompokan data terhadap penjualan dan pajak BBM di DKI Jakarta.

Metode cluster k-means analisis dapat dijadikan solusi guna mengklasifikasikan ciri-ciri dari suatu objek. Algoritma k-means mempunyai akurasi teliti yang relatif tinggi mengenai bentuk objek, dengan demikian algoritma tersebut lebih efisien dan terukur untuk mengolah suatu objek dalam jumlah yang tidak sedikit serta tidak dipengaruhi oleh rangkaian objek (Aranda, 2016).

Berdasarkan latar belakang tersebut, permasalahan yang perlu diselesaikan pada studi ini yaitu untuk mengetahui ketimpangan jumlah data penjualan dan pajak BBM di DKI Jakarta, kemudian menerapkan metode clustering melalui penggunaan algoritma k-means pada pengelompokan data penjualan dan pajak BBM di DKI Jakarta.

Studi ini memiliki sejumlah tujuan antara lain: a. Menerapkan metode clustering memanfaatkan algoritma k-means pada pengelompokan data penjualan dan pajak BBM di DKI Jakarta, b. Mengetahui pengelompokan data penjualan dan pajak BBM di DKI Jakarta menggunakan algoritma k-means.

Menurut Fayyad pada buku (Kusrini, 2009), data mining dan knowledge discovery in database (KDD) banyak digunakan silih berganti guna mengidentifikasi proses pencarian informasi yang tidak terlihat pada basis data yang besar. Pada umumnya, istilah diatas memiliki konsep yang tidak sama, namun saling berkaitan satu dengan yang lain. tidak terlepas yaitu salah satu langkah pada keseluruhan proses KDD yaitu data mining. Proses KDD secara garis besar (Narwati, 2010):

1. Data selection

Penentuan data yang berasal dari sejumlah data operasional dapat dijalankan dari sebelum

proses pencarian informasi pada KDD dilakukan.

2. Cleaning

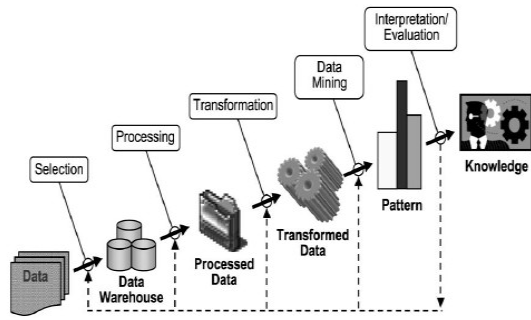
Sebelum tahap data mining bisa dilakukan (Pre-processing), melakukan cleaning pada data yang membentuk pusat KDD. Pada tahapan ini meliputi pembuangan penggandaan data, menguji data yang berubah-ubah, serta menyempurnakan kesalahan dalam data seperti kesalahan dalam mencetak. Kemudian melakukan proses enrichment, merupakan mekanisme memperbanyak data yang sudah terkumpul atau informasi yang jelas guna keperluan KDD (data atau informasi eksternal).

3. Transformation Coding

Ialah perubahan terhadap data yang telah terpilih guna tahapan data mining. Mekanisme coding pada KDD adalah proses kreatif dan memiliki ketergantungan terhadap corak dan ragam keterangan yang akan dicari pada basis data.

4. Data mining, ialah cara mendapatkan model dan informasi yang jelas serta menarik pada data yang dipilih dengan Teknik tertentu. Metode algoritma pada data mining memiliki banyak variasi. Proses memilih metode algoritma yang sesuai tergantung pada tujuan dan mekanisme KDD secara menyeluruh.

5. Interpretation (Evaluation Pola) merupakan model informasi yang didapatkan dari tahapan data mining yang butuh diperlihatkan ke bentuk yang sederhana dan dapat dipahami oleh bagian yang memiliki kepentingan. Tahapan ini adalah salah satu tahap KDD disebut dengan interpretation, meliputi pengujian apakah model maupun informasi yang didapatkan tidak sama menggunakan keadaan dan hipotesis sebelumnya.



Gambar 8. tahapan data mining

Menurut Eko Prasetyo (2012), clustering merupakan kategorisasi beberapa data ke dalam kelompok sesuai dengan karakteristik dan kesamaan tiap data pada kelompok yang tersedia. Pengklasifikasian data dibedakan berdasarkan susunan, kedudukan data, serta solidaritas pada pada kelompok.

Pada susunan, pengklasifikasian data dibagi menjadi partitioning (membagi tiap data cuma memiliki peserta satu kelompok) dan hierarki (data tunggal dianggap suatu kelompok berjumlah dua atau lebih). Pada kedudukan, pengklasifikasian data dibagi menjadi eksklusif (suatu data yang pasti cuma berubah jadi peserta satu kelompok dan bukan dengan kelompok lainnya) dan tumpang tindih (proses mengelompokkan data dimana hanya diperoleh satu data menjadi peserta lebih dari satu kelompok). Pada kategori solidaritas, pengklasifikasian data dibagi menjadi komplet dan parsial. Apabila seluruh data dapat bergabung menjadi satu maka dapat diartikan seluruh data kompak menjadi kelompok. Namun jika ada satu atau lebih data yang tidak bergabung pada kelompok mayoritas, maka diartikan data tersebut memiliki karakteristik menyimpan (Outlier/noise/uninterestedbackground).

K-means ialah salah satu proses pembagian data non-hierarki yang membatasi data yang tersedia dalam bentuk dua atau lebih kelompok, dengan demikian dapat memiliki karakteristik sejenis kemudian digabungkan ke satu kelompok yang serupa, data yang memiliki karakter berbeda dimasukkan ke dalam kelompok lain. Guna pengelompokkan data antara lain untuk meminimalisir manfaat objektif yang sudah diatur dalam kelompok metode ini berupaya untuk mengklasifikasikan data yang tersedia ke beberapa

kelompok yang memiliki karakteristik berbeda dengan data yang dimiliki kelompok lain.

Pokok algoritma antara lain:

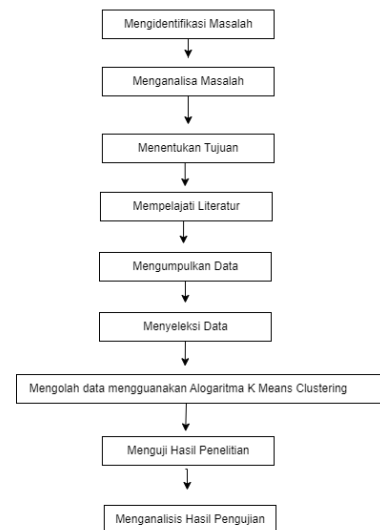
1. Menentukan skor K sebagai kuantitas kluster yang ingin diciptakan.
2. Inisial K merupakan centroid yang didirikan secara acak.
3. Menghitung jarak tiap data ke setiap centroid dengan persamaan Euclidean Distance:

$$\text{dist}(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (1)$$

4. Mengelompokkan tiap data berdasarkan jarak paling dekat antara centroid dengan datanya.
5. Menentukan letak centroid baru (k)
6. Mengulang Langkah ketiga jika letak centroid baru dan lama berbeda.

II. METODOLOGI PENELITIAN

Bagian ini memaparkan secara detail mengenai tahap-tahap studi yang dijalani. Untuk kerangka kerja yang digunakan pada studi ini adalah sebagai berikut :



Gambar 9. kerangka kerja penelitian

III. HASIL DAN PEMBAHASAN

A. Sumber data

Pada penelitian ini, data didapatkan dari Open Data Jakarta yang mana data tersebut ialah data penjualan BBM dan penerimaan BBM dari bulan 1 sampai bulan ke 12. Berikut data penjualan dan pajak BBM di DKI Jakarta

Tabel 1. Dataset

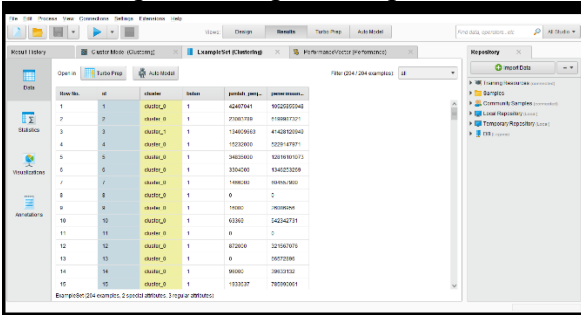
bulan	jumlah penjualan	penerimaan pajak
1	42407041	10525855948

D. Pengaplikasian pada aplikasi RapidMiner

Keluaran dari clustering data penjualan dan pajak BBM di DKI Jakarta menggunakan Rapid Miner

1. Data view

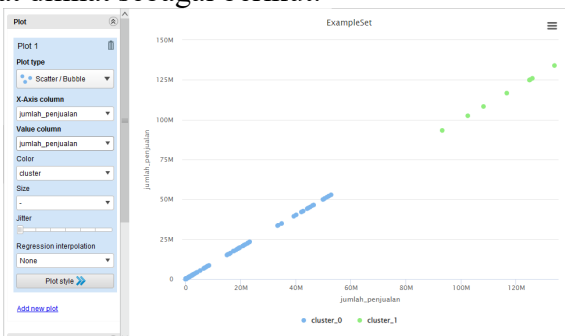
Data view: ialah menu yang berada di Result Perspective untuk memperlihatkan data yang sudah diproses dengan menyeluruh komplit dengan clusternya dari example set (read csv). Tampilan Data View dapat dilihat pada tampilan berikut:



Gambar 10 Data View

2. Flot view

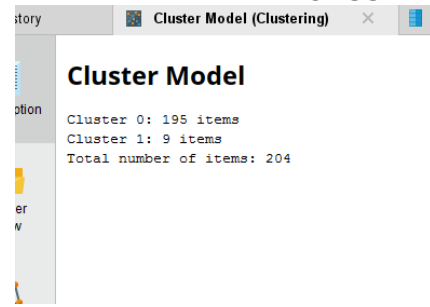
Flot view, ialah menu yang berada di Result Perspective guna memperlihatkan semua data yang selesai diolah dengan klasternya dari example set (read csv) dengan susunan diagram scatter. Tampilan dapat dilihat sebagai berikut:



Gambar 11 Flot View

3. Text view

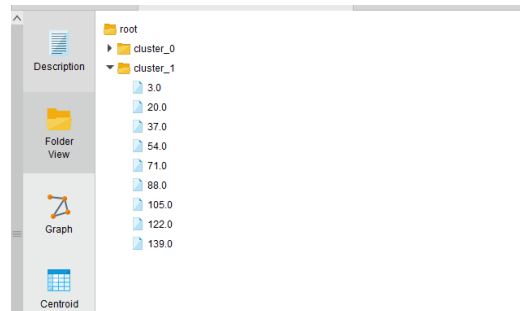
Text view:ialah menu untuk memperlihatkan dataset yang sudah diproses dengan menyeluruh komplit beserta clusternya. Tampilan bisa dilihat sebagai berikut:



Gambar 12 Text View

4. Folder view

Folder view ialah sheet guna memperlihatkan dataset yang selesai diolah secara menyeluruh komplit beserta clusternya. Berikut tampilan dari folder view.



Gambar 13 Folder View

IV. KESIMPULAN

Dalam penelitian ini, penulis dapat menyimpulkan bahwa percobaan yang dilaksanakan menghasilkan clustering pada penjualan dan pajak BBM terjadi sebanyak 7 kali iterasi dengan rata rata centroid distance 16359244117552204000.000. Pengolahan data di atas menggunakan aplikasi Rapid Miner guna menentukan nilai centroid pada kedua cluster. Cluster 1 dengan jumlah item paling tinggi yaitu 195 item dan cluster 2 terdapat 9 item.

REFERENSI

[1] C. J. M. S. Fina Nasari, "Penerapan Algoritma K-Means Clustering Untuk Pengelompokan Penyebaran Diare Di Kabupaten Langkat," pp. 108–119..

[2] S. Mulyati, "Penerapan Data Mining Dengan Metode Clustering Untuk Pengelompokan Data Pengiriman Burung," vol. 1, no. Senatkom, 2015..

[3] R. Hidayat, R. Wasono, and M. Y. Darsyah, "Pengelompokan Kabupaten / Kota Di Jawa Tengah," pp. 240–250, 2017.. <https://pajak.go.id/id/>

[4] https://ppsdmmigas.esdm.go.id/id/Landing/lihat_berita/6FtsKXqpM.

[5] Fatmawati, K., & Windarto, A. P. (2018). DATA MINING: PENERAPAN RAPIDMINER DENGAN K-MEANS CLUSTER PADA DAERAH TERJANGKIT DEMAM BERDARAH DENGUE (DBD) BERDASARKAN PROVINSI (Vol. 3, Issue 2)