

ANALISIS DATA MINING UNTUK MENENTUKAN PROFIT PERUSAHAAN MENGUNAKAN METODE K-MEANS

THE DATA MINING ANALYSIS TO DETERMINE COMPANY PROFIT USING K-MEANS METHOD

Yori Adi Atma¹, Suci Montesori

¹Jurusan Teknologi Informasi Politeknik Negeri Padang

*Corresponding Email: yoriadiatma@gmail.com

ABSTRAK

Penelitian ini menerapkan Data Mining dengan menggunakan metode *Clustering* untuk menentukan *profit* perusahaan berdasarkan tingkat penjualan aplikasi di CV. Alele Solutions. Algoritma yang digunakan yaitu *K-Means Clustering*, di mana data dikelompokkan berdasarkan karakteristik yang sama akan dimasukkan ke dalam kelompok yang sama dan set data yang dimasukkan ke dalam kelompok tidak tumpang tindih. Informasi yang ditampilkan berupa kelompok – kelompok data aplikasi berdasarkan tingkat penjualannya, sehingga diketahui *profit* perusahaan pada CV. Alele Solutions yang disesuaikan dengan *cluster* aplikasi dengan tingkat penjualan tinggi, sedang dan rendah. Pengujian dilakukan dengan aplikasi *RapidMiner* 5.3.015, sehingga menghasilkan *cluster-cluster* produk yang dapat dimanfaatkan oleh CV. Alele Solutions dalam menentukan *profit* perusahaan.

Kata Kunci : Data Mining, Clustering, K-Means, Laba, RapidMiner.

ABSTRACT

This research applies data mining using clustering method for determining the level of sales profit corporation based applications in CV. Alele Solutions. The algorithm used is the K-Means Clustering, where the data are grouped by similar characteristics will be included in the same group and a set of data entered into in the group do not overlap. Information displayed in the form of a groups application data based on the level of sales, so the profit of the company known CV. Alele Solutions adapted to cluster applications with a high sales rate, medium and low. Testing is done with the application RapidMiner 5.3.015, resulting in clusters of products that can be used by CV. Alele Solutions in determining the profit of the company.

Keywords: Data Mining, Clustering, K-Means, Profit, RapidMiner.

PENDAHULUAN

Dalam rangka menghadapi persaingan bisnis dan meningkatkan pendapatan perusahaan, pihak terkait dalam perusahaan dituntut untuk dapat mengambil keputusan yang tepat dalam menentukan strategi pemasaran produk yang akan dijualnya. Untuk dapat melaksanakan hal itu, perusahaan memerlukan informasi yang cukup untuk dapat dianalisa lebih lanjut. Ketersediaan data yang melimpah, kebutuhan akan informasi dan pengetahuan sebagai pendukung pengambilan keputusan untuk membuat solusi bisnis, dan dukungan infrastruktur di bidang teknologi informasi merupakan cikal-bakal dari lahirnya teknologi *Data Mining*.

Penelitian ini fokus pada metode *Data Mining* dengan kasus pengelompokan data (*clustering*). Adanya data dalam skala besar memungkinkan metode *Data Mining* dengan teknik *clustering* yang dapat mengelompokkan data ke dalam beberapa kelompok yang diinginkan (Deka, et al, 2014).

Goldie (2012), memaparkan *Knowledge Discovery in Database (KDD)* terdiri dari tiga proses utama, yaitu :

1. *Preprocessing*

Preprocessing dilakukan terhadap data sebelum algoritma *Data Mining* diaplikasikan. Proses ini meliputi data *cleaning*, *integrasi*, *seleksi* dan

transformasi.

2. *Data Mining*

Proses utama dalam KDD adalah proses *Data Mining*, dalam proses ini algoritma *Data Mining* diaplikasikan untuk mendapatkan pengetahuan dari sumber data.

3. *Post processing*

Hasil yang diperoleh dari proses *Data Mining* selanjutnya akan dievaluasi pada *post processing*.

Menurut Tutik (2014) *K-Means* adalah salah satu algoritma yang menggunakan metode partisi. *K-Means* adalah algoritma *clustering* yang membagi masing-masing item data ke dalam satu *cluster*.

K-Means adalah suatu teknik pengelompokan data yang mana keberadaan tiap-tiap titik data dalam suatu *cluster* ditentukan oleh derajat keanggotaan. Teknik ini pertama kali diperkenalkan oleh Jim Bezdek pada tahun 1981 (Deka Dwinavinta *et al*, 2014).

Menurut Tutik (2014), tahapan melakukan *clustering* atau pengelompokan dengan metode *K-Means* adalah sebagai berikut :

1. Menentukan berapa banyak *cluster* yang ingin yang ingin dibentuk, di mana nilai K adalah banyaknya *cluster*/ jumlah *cluster*.
2. Menentukan pusat *cluster (centroid)* awal. *Centroid* awal ditentukan secara

acak dari data yang ada dan jumlah *centroid* awal sama dengan jumlah *cluster*.

3. Setelah menentukan *centroid* awal, maka setiap data akan menemukan *centroid* terdekatnya yaitu dengan menghitung jarak setiap data ke masing-masing *centroid* menggunakan rumus korelasi antar dua obyek yaitu *Euclidean Distance*.
4. Setelah menghitung jarak data ke *centroidnya*, maka langkah berikutnya adalah mengelompokkan data berdasarkan jarak minimumnya. Suatu data akan menjadi anggota dari suatu *cluster* yang memiliki jarak terdekat (terkecil) dari pusat *cluster*-nya.
5. Berdasarkan pengelompokkan tersebut, selanjutnya adalah mencari *centroid* baru berdasarkan membership dari masing-masing *cluster* yaitu dengan menghitung rata-rata dari data masing-masing *cluster*.
6. Kembali ke tahap 3.
7. Perulangan berhenti apabila tidak ada data yang berpindah.

Menurut Tutik (2014) untuk menentukan nilai pusat (*centroid*) pada tahap *iterasi* digunakan rumus sebagai berikut :

$$v_{ij} = \frac{1}{N_i} = \sum_{k=0}^{N_i} x_{ki}$$

Di mana :

\bar{v}_{ij} = *centroid* rata-rata *cluster* ke *i* untuk variable ke *j*

N_i = jumlah anggota *cluster* ke *i*

i, k = *indeks* dari *cluster*

j = *indeks* dari variable

X_{kj} = nilai data ke *k* variable ke *j* dalam *cluster* tersebut

Menurut Afrisawati (2013) untuk menentukan korelasi antar dua obyek yaitu dengan menggunakan rumus *Euclidean Distance* berikut:

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

Di mana :

$d(x,y)$ = jarak data ke *x* ke pusat cluster *y*

x_i = data ke *i* pada atribut data ke *n*

y_i = data ke *j* pada atribut data ke *n*.

Berdasarkan uraian pendahuluan tersebut maka penelitian ini mengangkat tema dengan judul “Analisis Data Mining untuk Menentukan Profit Perusahaan Menggunakan Metode K-Means”.

METODE PENELITIAN

Bagian ini menjelaskan tentang metodologi penelitian yang digunakan dalam penyelesaian masalah penelitian. Metodologi penelitian ini sangat membantu penulis dalam proses penyelesaian masalah. Penelitian ini memiliki beberapa tahapan dalam

pelaksanaan kegiatan yang tertuang pada kerangka kerja penelitian yaitu identifikasi masalah, analisa masalah, menentukan tujuan, mempelajari literatur, mengumpulkan data, analisa Algoritma *K-Means Clustering*, pengolahan data, implementasi, dan pengujian.

Dalam metodologi penelitian ada urutan kerangka kerja yang harus diikuti, urutan kerangka kerja ini merupakan langkah-langkah yang dilakukan dalam penulisan. Adapun uraian langkah kerjanya adalah sebagai berikut :

1. Identifikasi Masalah

Pada tahap identifikasi masalah ini akan dilakukan peninjauan ke sistem yang akan diteliti untuk mengamati serta melakukan eksplorasi lebih dalam dan menggali permasalahan yang ada pada sistem yang berjalan saat ini. Tahap ini adalah langkah awal untuk menentukan rumusan masalah dari penelitian.

2. Analisis Masalah

Analisis masalah ini merupakan langkah yang dilakukan untuk memahami masalah yang telah ditentukan ruang lingkup atau batasannya. Dengan menganalisa masalah yang telah ditentukan tersebut, maka diharapkan masalah dapat dipahami dengan baik.

3. Menentukan Tujuan

Berdasarkan pemahaman dari masalah yang dianalisa, maka ditentukan tujuan yang akan dicapai dari penelitian ini. Pada tujuan ini ditentukan target yang akan dicapai, terutama yang dapat mengatasi masalah-masalah yang ada.

4. Mempelajari Literatur

Literatur-literatur yang dipakai sebagai bahan referensi dalam penelitian ini adalah dari jurnal-jurnal ilmiah dan buku tentang *Data Mining* khususnya *K-Means Clustering* dan bahan bacaan lain yang mendukung penelitian.

5. Mengumpulkan Data

Metode pengumpulan data dilakukan dengan cara melakukan pengamatan langsung di tempat penelitian sehingga dapat diketahui dengan jelas permasalahan yang ada. Selain pengamatan, juga dilakukan wawancara kepada pihak-pihak yang terkait dengan penelitian ini dan juga melakukan pengambilan sampel *database* untuk menunjang penelitian ini. Selain itu, juga dilakukan studi kepustakaan yaitu dengan membaca buku-buku yang menunjang dalam melakukan penganalisaan terhadap data dan informasi yang didapat.

6. Menganalisa Algoritma *K-Means Clustering*

Pada tahap ini akan dilakukan analisa terhadap metode yang digunakan dengan menggunakan literatur-literatur yang telah disiapkan pada tahap studi literatur sebelumnya untuk menentukan jumlah produksi produk. Menganalisa data jumlah penjualan menggunakan algoritma *K-Means Clustering*.

7. Mengimplementasikan *K-Means Clustering*

Pada tahap ini dilakukan implementasi terhadap sistem yang sudah dirancang. Pada penelitian ini penulis mengimplementasikan metode *K-Means Clustering* untuk diterapkan dalam mengelompokkan data penjualan produk dengan menggunakan *software Rapid Miner*.

8. Menguji Hasil Penelitian

Pada tahap ini dilakukan pengujian terhadap sistem yang dikembangkan. Pengujian dilakukan dengan cara menerapkan metode *K-Means Clustering* berdasarkan sampel data yang sudah ada.

9. Menganalisa Hasil Pengujian

Pada tahap ini akan dilakukan pengamatan dan analisa terhadap hasil pengujian metode *K-Means Clustering* dalam menentukan profit perusahaan.

HASIL DAN PEMBAHASAN

1. Analisa Data

Data yang digunakan pada penelitian ini adalah data penjualan aplikasi di CV. Alele Solutions. Cluster yang ingin dibentuk yaitu tingkat penjualan aplikasi berdasarkan Revenue, Cost, dan Profit yang diambil penjualan selama tiga tahun yaitu pada tahun 2013 sampai dengan 2015. Di mana cluster 1 mewakili tingkat penjualan rendah, cluster 2 mewakili tingkat penjualan sedang, dan cluster 3 mewakili tingkat penjualan yang tinggi. Sehingga hasil cluster tersebut dapat menentukan profit perusahaan.

Semua data yang telah didapatkan kemudian diolah untuk dapat dianalisa permasalahan yang terjadi dan kebutuhan yang diharapkan sehingga menghasilkan informasi yang berguna untuk mengatasi permasalahan dan dapat diusulkan suatu perbaikan. Pengolahan data ini menggunakan teori-teori dari Data Mining dengan algoritma *K-Means Clustering*.

Urutan sistem dari algoritma *K-Means Clustering* yaitu mulai dari menentukan jumlah cluster. Pada penelitian ini jumlah cluster yang ditentukan sebanyak tiga cluster yaitu cluster penjualan rendah, cluster penjualan sedang, dan cluster penjualan tinggi. Langkah selanjutnya yaitu menentukan centroid awal, dan menghitung jarak objek ke pusat

(centroid). Setelah jarak objek ke pusat (centroid) telah ditemukan, maka selanjutnya kelompokkan objek berdasarkan jarak minimumnya. Didapatkan pusat cluster baru, dilanjutkan kembali menghitung jarak objek ke pusat (centroid) dan mengelompokkan objek sehingga dapat perbandingan dari hasil perhitungan awal apakah ada objek yang berpindah group. Jika iya, maka proses diulang kembali, dan jika tidak maka proses berakhir dan group yang berakhir akan menjadi hasil dari proses algoritma K-Means Clustering.

Untuk dapat melakukan pengelompokan data-data tersebut menjadi beberapa *cluster* perlu dilakukan beberapa langkah, yaitu :

1. Tentukan jumlah *cluster* yang diinginkan. Dalam penelitian ini data-data yang ada akan dikelompokkan menjadi tiga *cluster*. Yaitu *cluster* 0 merupakan kelompok aplikasi dengan penjualan rendah, *cluster* 1 kelompok aplikasi dengan penjualan sedang, dan *cluster* 2 kelompok aplikasi dengan penjualan tinggi.
2. Tentukan titik pusat awal *cluster* (centroid). Dalam penelitian ini titik pusat awal ditentukan secara random atau acak dan didapat titik pusat dari setiap *cluster* dapat dilihat pada tabel.
3. Setelah menentukan *centroid* awal,

maka setiap data akan menemukan *centroid* terdekatnya yaitu dengan menghitung jarak setiap data ke masing-masing *centroid* menggunakan rumus korelasi antar dua obyek yaitu *Euclidean Distance*. Adapun penghitungan *centroid* awal secara manual hanya menghitung 5 data *sample* saja.

4. Setelah menghitung jarak data ke *centroidnya*, maka langkah berikutnya adalah mengelompokkan data berdasarkan jarak minimumnya.
5. Setelah didapatkan *centroid* baru langkah selanjutnya kembali lagi ke langkah 3, yakni menghitung jarak setiap data ke masing-masing *centroid* menggunakan rumus korelasi antar dua obyek yaitu *Euclidean Distance* berdasarkan *centroid* baru.

Dari pengolahan data tersebut dapat diketahui bahwa kelompok aplikasi yang memiliki penjualan tertinggi dari ketiga *cluster* terdapat pada *cluster* 2. Sehingga dapat disimpulkan bahwa aplikasi pada *cluster* 2 memiliki tingkat *profit* perusahaan yang lebih besar, agar *marketing* lebih ditingkatkan, sedangkan pada *cluster* 1 memiliki penjualan menengah. Dan *profit* perusahaan yang berada pada *cluster* 0 memiliki penjualan rendah, maka *marketing* aplikasi tersebut dikurangi karena dapat mengakibatkan

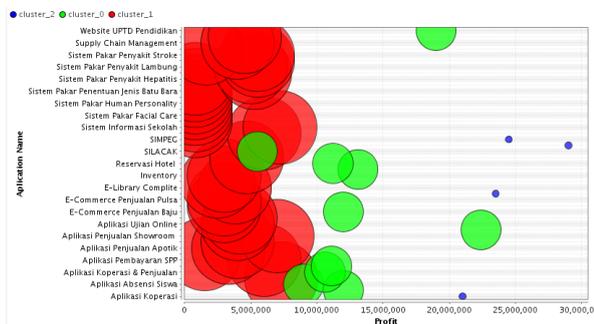
menurunkan *profit* perusahaan.

2. Implementasi dan Pengujian

Pada tahapan ini akan dideskripsikan teknik menggunakan tool RapidMiner dan diselaraskan dengan pembuktian dari analisa metode terhadap permasalahan yang ada pada bab sebelumnya.

Tabel 1. Hasil Clustering

Cluster	Hasil Grafik Pengujian
Cluster 0	<i>Bubble</i> yang berwarna hijau yang berkumpul pada pengelompokan data aplikasidengan tingkat profit sedang
Cluster 1	<i>Bubble</i> yang berwarna merah yang berkumpul pada pengelompokan data aplikasidengan tingkat profit rendah
Cluster 2	<i>Bubble</i> yang berwarna biru yang berkumpul pada pengelompokan data aplikasidengan tingkat profit tinggi



Gambar 1. Tampilan *ExampleSet Plot View*

Dari gambar 1 di atas dapat dijelaskan bahwa grafik menggunakan model *Plotter Bubble* di mana pada sumbu x merupakan batasan data dari profit dan pada sumbu y merupakan batasan data *Application Name*. Bulatan – bulatan atau *Bubble* yang terbentuk pada grafik menunjukkan pengelompokan data, di mana *Bubble* yang berwarna merah

menunjukkan anggota dari cluster 1 yang berjumlah 31 yang menunjukkan tingkat profit perusahaan rendah, *Bubble* yang berwarna hijau menunjukkan anggota dari cluster 0 yang berjumlah 10 yang menunjukkan tingkat profit perusahaan sedang, dan *Bubble* yang berwarna biru menunjukkan anggota dari cluster 2 yang berjumlah 4 yang menunjukkan tingkat profit perusahaan yang lebih besar.

SIMPULAN

Berdasarkan pembahasan dan hasil pengujian manual serta menggunakan *Software RapidMiner*, maka dapat ditarik kesimpulan sebagai berikut :

1. Berdasarkan hasil pengujian, maka didapatkan pengelompokan tingkat *profit* perusahaan di CV. Alele Solutions sebanyak 3 *cluster*. Yaitu *cluster1* yaitu kelompok dengan tingkat *profit* rendah, *cluster0* *cluster* dengan tingkat *profit* sedang dan *cluster2* dengan tingkat *profit* tinggi. Dan aplikasi yang masuk ke dalam masing-masing *cluster* tersebutlah yang menjadi acuan untuk menentukan *profit* perusahaanberikutnya.
2. *Clusteryang* memiliki tingkat *profit* tinggi memiliki penjualan yang tinggi pula. Kemudian *cluster*dengan tingkat *profit* rendah, maka kegiatan marketing penjualan aplikasi untuk

berikutnya dikurangi agar tidak terjadi penambahan *cost* dan penurunan *profit* perusahaan.

Begitupun *cluster* aplikasi dengan tingkat *profit* sedang, jumlah penjualannya disesuaikan dengan tingkat *profit* yang telah dicapai.

3. Baik perhitungan secara manual maupun perhitungan menggunakan *software RapidMiner* didapatkan kesamaan hasil pengujian dengan menggunakan *sample data* (35 *record*).
4. Hasil perhitungan untuk semua data (45 *record*) terdapat perbedaan posisi *cluster* antara pengujian menggunakan *RapidMiner* dengan pengujian secara manual. Yang mana pada *cluster 0* pada *RapidMiner* menunjukkan kelompok aplikasi dengan penjualan sedang, sedangkan pada pengujian manual *cluster 0* menunjukkan tingkat *profit* rendah. Hal yang sama terjadi pada *cluster 1* pada *RapidMiner* menunjukkan kelompok aplikasi dengan *profit* perusahaan rendah, sedangkan pada pengujian manual *cluster 1* menunjukkan tingkat *profit* sedang. Meskipun terdapat perbedaan letak *cluster* pada kedua pengujian tersebut, namun anggota masing-masing *cluster* dengan tingkat *profit* rendah, sedang

dan tinggi adalah sama.

DAFTAR PUSTAKA

- Afrisawati. 2013. *Implementasi Data Mining Pemilihan Pelanggan Potensial Menggunakan Algoritma K-Means*. Ed. Pelita Informatika Budi Darma, Volume : V, No. 3.
- Budanis Dwi Meilani dan Nofi Susanti. 2014, *Aplikasi Data Mining Untuk Menghasilkan Pola Kelulusan Siswa Dengan Metode Naive Bayes*. Ed. Jurnal LINK Vol 21/No.2.
- Deka Dwinavinta, et.al. 2014. *Klasterisasi Judul Buku dengan Menggunakan Metode K-Means*. Ed. Seminar Nasional Aplikasi Teknologi Informasi (SNATI).
- Dr. M.P.S Bhatia and Deepika Khurana. 2013. *Experimental study of Data clustering using k-Means and modified algorithms*. International Journal of Data Mining & Knowledge Management Process (IJDKP). Vol. 3 No.3.
- Fadlina. 2014. *Data Mining untuk Analisa Tingkat Kejahatan Jalanan Dengan Algoritma Association Rule Metode Apriori (Studi Kasus Di Polsekta Medan Sunggal)*. Voumel.III No.1.
- Johan Oscar Ong. 2013. *Implementasi Algoritma K-Means Clustering Untuk Menentukan Strategi Marketing Presiden University*. Ed. Jurnal Ilmiah Teknik Industri, Vol.12, No.1.
- Mujib Ridwan, et.al. 2013. *Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier*. Ed. Jurnal EECCIS Vol.7, No.1.
- Nova Tumoka. 2013. *Analisis Pendapatan Usaha Tani Tomat di Kecamatan Kawangkoan Barat Kabupaten Minahasa*. Jurnal EMBA. Vol.1 No.3
- Rajashree, dkk. 2010. *A hybridized K-means clustering approach for high dimensional dataset*. International Journal of Engineering, Science and Technology Vol. 2, No. 2.
- Ramesh Singh Yadava and P.K.Mishra. 2012. *Performance Analysis of High Performance k-Mean Data Mining Algorithm for Multicore Heterogeneous Compute Cluster*. Vol. 2 No.4.
- Tutik Khotimah. 2014. *Pengelompokan Surat Dalam Al Qur'an Menggunakan Algoritma K-Means*. Ed. Jurnal Simetris, Vol 5 No 1.