



PEMANFAATAN DATA MINING DENGAN METODE CLUSTERING UNTUK EVALUASI BIAYA DOKUMEN EKSPOR DI PT WINSTAR BATAM

Yuli Siyamto

Program Studi Sistem Informasi, Fakultas Teknik, Universitas Putera Batam - Batam
Jln. Raden Patah Nagoya - Batam

ABSTRAK

Data Mining merupakan teknologi yang sudah ada dan berkembang cukup lama. Banyak perusahaan yang kurang memanfaatkan data-data lama yang tersimpan di dalam database, untuk dimanfaatkan semaksimal mungkin. Penelitian ini mencoba menerapkan teknologi Data Mining dengan menggunakan metode clustering untuk mendapatkan pengetahuan baru, yang diharapkan bermanfaat bagi kelangsungan dan perkembangan perusahaan. Algoritma yang digunakan yaitu K-Means Clustering yaitu proses pengelompokan sejumlah data atau objek ke dalam cluster (group) sehingga setiap cluster tersebut akan berisi data yang semirip mungkin dan berbeda dengan objek dalam cluster yang lainnya. Pengujian dilakukan dengan aplikasi Rapid Miner 5.3. Hasilnya adalah cluster-cluster pengelompokan data pelanggan yang disarankan untuk di evaluasi biaya pengurusan dokumen eksportnya.

Kata Kunci : Data Mining, K-Means Clustering, Evaluasi Biaya Dokumen Ekspor.

ABSTRACT

Data Mining is a technology that already exists and grows long enough. Many companies are less use of old data stored in the database, to be used as much as possible. This research is trying to apply Data Mining technology by using the clustering method to get new knowledge, which is expected to be useful for the continuity and development of the company. The algorithm used is K-Means Clustering which is the process of grouping a number of data or objects into a cluster (group) so that each cluster will contain data that is as similar as possible and different from other cluster objects. Tests were performed with the Rapid Miner 5.3 application. The result is clustered clustering customer data suggested for the evaluation of the cost of processing the export documents.

Keywords: Data Mining, K-Means Clustering, Evaluate the Cost of Export Documents

I. PENDAHULUAN

Dalam rangka menghadapi ketatnya persaingan bisnis dan menjaga loyalitas para pelanggan, maka setiap perusahaan dituntut untuk dapat mengambil keputusan yang tepat dalam menentukan strategi bisnis. Kota Batam dengan lokasi strategis dalam jalur perdagangan, terdapat banyak perusahaan yang bergerak di bidang jasa pengurusan dokumen ekspor. Dalam hal ini, strategi dalam penetapan harga atau biaya pengurusan dokumen perlu mendapatkan perhatian dan evaluasi berkala dari perusahaan. Falsafah customer satisfaction dan reward kepada pelanggan setia, merupakan prioritas dalam menjaga relasi bisnis perusahaan yang telah lama terjalin.

Untuk memenuhi kebutuhan strategi bisnis perusahaan, dapat dilakukan dengan pemanfaatan data-data yang telah ada di dalam database perusahaan dengan menggunakan Data Mining. Hasil pengolahan dengan Data Mining berupa sebuah pengetahuan (knowledge) yang baru yaitu pengelompokan data pelanggan, dapat dimanfaatkan perusahaan sebagai informasi untuk mengevaluasi biaya dokumen para pelanggan. Sehingga perusahaan mampu mempertahankan loyalitas pelanggan dan menjadikan perusahaan lebih kompetitif serta meningkatkan pendapatannya.

II. TEORITIS

A. Data Mining

Data Mining adalah teknik analisa data secara otomatis untuk membuka atau membongkar hubungan dari banyak data yang tidak diketahui sebelumnya. *Data Mining* sering dihubungkan dengan analisa simpanan data di dalam sebuah *warehouse*. Secara umum ada tiga teknik utama *Data Mining* yaitu regresi, klasifikasi dan clustering.

B. Clustering

Clustering merupakan bagian dari ilmu *Data Mining* yang bersifat tanpa arahan (*unsupervised*). *Clustering* adalah proses pembagian data ke dalam kelas atau *cluster* berdasarkan tingkat kesamaannya. *Clustering* merupakan pekerjaan yang memisahkan data atau vektor ke dalam sejumlah kelompok atau *cluster* menurut karakteristiknya masing-masing. Data-data yang memiliki kemiripan karakteristik akan berkumpul dalam kelompok atau *cluster* yang sama. Data-data yang memiliki perbedaan karakteristik, akan berkumpul dalam kelompok atau *cluster* yang berbeda. Tujuan utama dari metode *clustering* adalah pengelompokan sejumlah data atau obyek ke dalam *cluster (group)* sehingga dalam setiap *cluster* akan berisi data yang semirip mungkin.

C. Algoritma K-Means

K-means merupakan salah satu teknik pengelompokan yang bekerja berdasarkan *partitioned clustering*. Prinsip kerja dari pengelompokan *hierarchical clustering* dilakukan



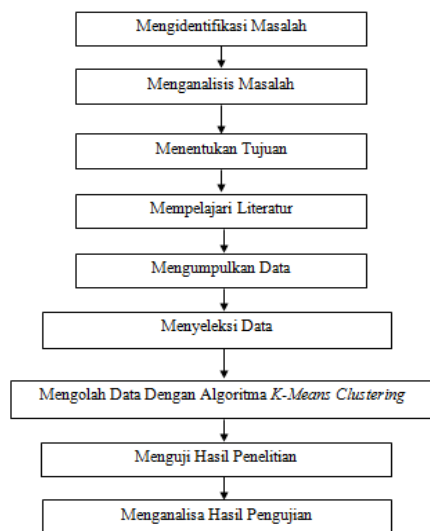
secara bertahap. Dan disetiap iterasi dari pengelompokan *hierarchical clustering* hanya ada satu pemilihan penggabungan suatu item terhadap item lainnya. Sedangkan prinsip kerja dari pengelompokan *partitioned clustering* adalah mengelompokkan item secara acak karena dipengaruhi *centroid*. Dan disetiap iterasi dari pengelompokan *partitioned clustering* dapat memungkinkan untuk terjadinya lebih dari satu pemilihan item yang akan digabungkan.

Langkah-langkah pada proses *clustering* dengan menggunakan algoritma *K-Means*, adalah sebagai berikut:

1. Menentukan Jumlah *Cluster*
Jumlah cluster yang ditentukan untuk mengelompokkan data pada penelitian ini sebanyak 4 *cluster*.
2. Menentukan *Centroid*
Pusat awal *cluster (centroid)* ditentukan secara *random* atau acak.
3. Menghitung Jarak dari *Centroid*
Menghitung jarak antara titik *centroid* dengan titik tiap objek dengan menggunakan *Euclidian Distance*.
4. Alokasikan masing-masing objek ke *centroid* terdekat. Untuk mengalokasikan objek ke dalam masing-masing *cluster* dengan cara mengelompokkan berdasarkan jarak minimum objek ke pusat *cluster*.

III. ANALISA dan PEMBAHASAN

Dalam metodologi penelitian terdapat urutan kerangka kerja yang harus diikuti. Kerangka kerja ini merupakan langkah langkah yang dilakukan dalam penelitian. Adapun kerangka kerja yang digunakan dalam penelitian ini adalah seperti terlihat pada gambar 1.



Gambar 1. Kerangka Kerja Penelitian

Metode pengumpulan data yang digunakan adalah dengan mengambil data transaksi dalam *database*. Data yang digunakan pada penelitian ini adalah data transaksi perusahaan selama periode

Januari 2015 sampai Desember 2015. Sepanjang tahun 2015, terdapat 5537 transaksi dokumen. Data transaksi dokumen tersebut merupakan data mentah yang harus diolah terlebih dahulu melalui proses seleksi data, *pre-processing*, pembersihan data dan *transformation*, sebelum di proses dalam *Data Mining*. Data transaksi dokumen ekspor impor direkapitulasi berdasarkan nama pelanggan, sehingga menghasilkan 223 data pelanggan. Variabel yang digunakan adalah jumlah frekuensi transaksi pelanggan, total kontainer yang digunakan dan total barang pelanggan pada tahun 2015. Hasil adalah data rekapitulasi yang digunakan dapat dilihat pada tabel 1.

Tabel 1. Sampel Data Awal

NO	CUSTOMER	FREK	JUM CON	JUM BRG
15	PT.DELIMAX INDONESIA	219	61	432
16	PT.DUTA MULTIKARYA	1	1	9
17	PT.ENGNGLAP PLASTRONIC	30	20	61
18	PT.GHIMLI INDONESIA	61	52	206
19	PT.GLOBAL KONSTRINDO ENGINEERING	7	1	15
20	PT.GREEN FLEXIBLE INDUSTRIES	171	225	413
21	PT.HASSANI CAN PACKAGING	1	4	14
22	PT.KARIMUN ANUGRAH SEJATI	9	0	38
23	PT.KARYA SEMPURNA ABADI	3	0	6
24	PT.MADEIRA INDONESIA	1	0	10
25	PT.METAL MULTI VARIASI COATING	2	0	10
26	PT.MILLIONBUILT	19	0	191
27	PT.MULTIPLASTINDO UTAMA	2	2	4
28	PT.PETROTECAIR POWER	1	0	1
29	PT.POHKIN INDONESIA	11	11	149
30	PT.PRIORITAS SOLUSI	5	0	6

Hasil pengelompokan data pelanggan dengan *Rapid Miner* :

1. *Data view*: merupakan *sheet* di *Result Perspective* untuk menampilkan data yang telah diolah secara keseluruhan lengkap dengan klasternya dari *example set (read excel)*. *Data view* dapat lihat pada gambar 2.

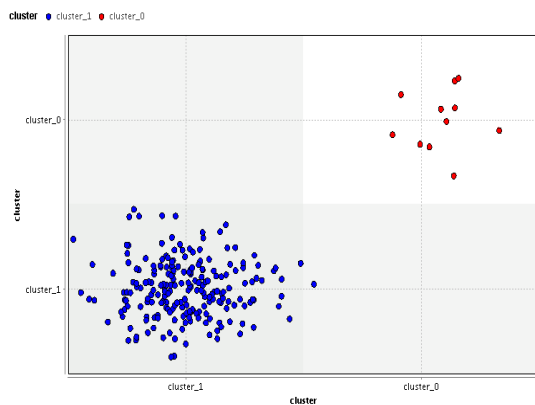
Row No.	A	cluster
66	PT. MCDERMOTT INDONESIA	cluster_1
67	PT. M-I INDONESIA	cluster_1
68	PT. NSP TECHNOLOGY BATAM	cluster_1
69	PT. PIPAMAS PUTIH	cluster_1
70	PT. PROFAB INDONESIA	cluster_0
71	PT. PROSERV	cluster_1
72	PT. REMOTE OILFIELD SERVICES	cluster_1
73	PT. SUPRA TEKNOLOGI PLASTIK	cluster_1
74	PT. TOYO KANETSU INDONESIA	cluster_1
75	PT. TRANSALINDO EKA PERSADA	cluster_1
76	PT. UNION PLASTIC INDUSTRIES	cluster_1
77	PT. VME PROCESS	cluster_1
78	PT. WASCO ENGINEERING INDONESIA (BT)	cluster_0
79	PT. WEICHAI UTAMA MANDIRI	cluster_1
80	PT. WORLD WIDE EQUIPMENT SOUTH EAST ASIA	cluster_1
81	STAR ENERGY (KAKAP) LTD	cluster_1
82	PT. DOK & PERKAPALAN KODJA BAHARI	cluster_1
83	PT. EXPRO INDONESIA	cluster_1
84	PT. M-I INDONESIA	cluster_1
85	PT. PEMBANGUNAN PERUMAHAN (PERSERO) TBK	cluster_1
86	PT. TRAKINDO UTAMA	cluster_1

Gambar 2. Tampilan *Data View*

2. *Flot view* : merupakan *sheet* di *Result Perspective* untuk menampilkan data yang telah diolah secara keseluruhan lengkap dengan klasternya dari

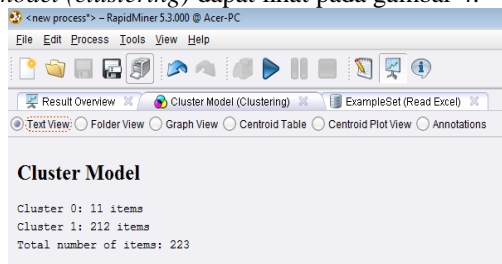


example set (read excel) dalam bentuk diagram Scatter. Tampilannya dapat dilihat di gambar 3.



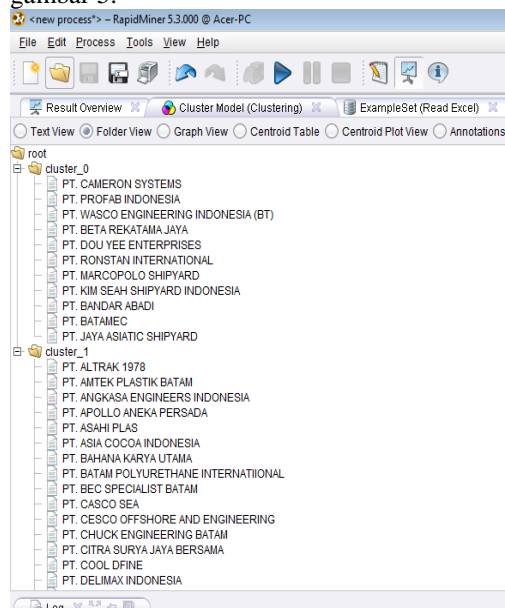
Gambar 3. Tampilan Flot View

3. *Text view* : merupakan *sheet* untuk menampilkan *database* yang telah diolah secara keseluruhan lengkap dengan *clusternya*. Dilihat dari *cluster model (clustering)* dapat dilihat pada gambar 4.



Gambar 4. Tampilan Text View

4. *Folder view* : merupakan *sheet* untuk menampilkan *database* yang telah diolah secara keseluruhan lengkap dengan *clusternya*. dilihat dari *cluster model (clustering)* dapat dilihat pada gambar 5.



Gambar 5. Tampilan Folder View

IV. KESIMPULAN

Kesimpulan yang didapatkan dari penelitian ini adalah pengelompokan data pelanggan adalah

sebanyak 2 *cluster*. Yaitu *cluster 0* adalah kelompok data pelanggan yang biaya dokumen ekspornya disarankan untuk dievaluasi, *cluster 1* yaitu kelompok data pelanggan yang biaya dokumen ekspornya disarankan untuk tidak perlu dievaluasi. Pelanggan yang biaya dokumen ekspornya disarankan untuk dievaluasi hanya sebanyak 11 pelanggan. Evaluasi dimaksudkan bahwa biaya pengurusan dokumen ekspornya ditinjau kembali, diberikan sejumlah diskon atau *reward*, sehingga relasi dan kelangsungan bisnis antara perusahaan dan pelanggan semakin terjaga.

DAFTAR REFERENSI

- Alfina, T., Santosa, B. dan Barakbah, A.R. (2012). "Analisa Perbandingan Metode *Hierarchical Clustering*, *K-Means* Dan Gabungan Keduanya Dalam *Cluster Data* (studi kasus : problem kerja praktek jurusan teknik industri ITS)". *Jurnal Teknik ITS*.1. 521-525.
- Pudjiantoro, T. H., Renaldi, F. dan Teogunadi, A. (2011). "Penerapan *Data Mining* Untuk Menganalisis Kemungkinan Pengunduran Diri Calon Mahasiswa Baru". *Konferensi Nasional Sistem dan Informatika*. 51-60.
- Lindawati. (2008). "Data Mining Dengan Teknik *Clustering* Dalam Pengklasifikasian Data Mahasiswa Studi Kasus Prediksi Lama Studi Mahasiswa Universitas Bina Nusantara". *Seminar Nasional Informatika*. 174-180.
- Ediyanto, Mara, M.N. dan Satyahadewi, N. (2013). "Pengklasifikasian karakteristik Dengan Metode *K-Means Cluster Analysis*". *Buletin Ilmiah Mat.Stat, dan Penerapannya (Bimaster)*.2 (2). 133-136.
- Rismawan, T. dan Kusumadewi, S. (2008). "Aplikasi *K-Means* Untuk Pengelompokan Mahasiswa Berdasarkan Nilai Body Mass Index (BMI) & Ukuran Kerangka". *Seminar Nasional Aplikasi Teknologi Informatika*. E43-E48.
- Handoyo, R., Rumani, R. dan Nasution, S.M. (2014). "Perbandingan Metode *Clustering* Menggunakan Metode *Single Linkage* Dan *K-Means* Pada Pengelompokan Dokumen". *JSM STMIK Mikroskil*. 15 (2). 73-82.
- Suprihatin. (2011). "Klastering *K-Means* Untuk Penentuan Nilai Ujian". *Jusi*. 1 (1). 53-62.
- Ikhwan, A., Nofriansyah, D. dan Sriani. (2015). "Penerapan *Data Mining* Dengan Algoritma *Fp- Growth* Untuk Mendukung Strategi Promosi Pendidikan (Studi Kasus Kampus STMIK Triguna Dharma)". *Jurnal Ilmiah SAINTIKOM*. 14 (3). 211-226.
- Aravind, H., C. Rajgopal. dan K.P. Soman. (2010). "A Simple Approach to Clustering in Excel". *International Journal of Computer Applications*. 11 (7). 19-25.
- Prasetyo, E. (2014). "Data Mining Mengolah Data Menjadi Informasi Menggunakan *Matlab*". Yogyakarta : CV Andi Offset.
- Ong, J.O. (2013). "Implementasi Algoritma *K-Means Clustering* Untuk Menentukan Strategi *Marketing* Di President University". *Jurnal Ilmiah Teknik Industri*. 12 (1). 10-20.
- Tayal, M.A. dan Raghuvanshi, M.M. (2010). "Review on Various Clustering Methods for the Image Data". *Journal of Emerging Trends in Computing and Information Sciences*. 2. 34-38.
- Ridwan, M., Suyono, H. dan Sarosa, M. (2013). "Penerapan *Data Mining* Untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma *Naive Bayes Classifier*". *Jurnal EECIS*. 7 (1). 59-64.
- Sharma, N., Bajpai, A. and Litoriya, R. (2012). "Comparison the various clustering algorithms of Weka Tools". *International Journal of Emerging Technology and Advanced Engineering*. 2 73-80.
- Sidhu, N.K. and Kaur, R. (2013). "Clustering In Data Mining". *International Journal of Computer Trends and Technology (IJCTT)*. 4. 710-7014.
- Krotha, R.S. and Merugula, S. (2013). "A Brief Survey On Document Clustering Techniques Using *MATLAB*". *International Journal of Computer & Organization Trends*. 3. 1-6.



Hal 28-31

- Sutrisno, Afriyudi, Widiyanto. (2013). "Penerapan *Data Mining* Pada Penjualan Menggunakan Metode *Clustering* Study Kasus PT. Indomarco Palembang". *Jurnal Ilmiah Teknik Informatika Ilmu Komputer*.10 (10). 1-11.
- Oyelade, O. J., Oladipupo, O.O. and Obagbuwa, I.C. (2010)." *Application of k-Means Clustering algorithm for prediction of Students' Academic Performance*". (*IJCSIS*) *International Journal of Computer Science and Information Security*. 7 (1). 292-295.
- Sirait, T.H. dan Ong, J.O. (2011). "Analisis Keberhasilan Mahasiswa dengan Metode *Cluster ing K-Means*". *SNASTIA*. 1-6.
- Ali, H. dan Wangdra, T.(2010). "Technopreneurship dalam perspektif bisnis *online*". 1. Batam : Baduose Media.
- Mardiani. (2014)." Perbandingan Algoritma *K-Means* dan EM untuk Clusterisasi Nilai Mahasiswa Berdasarkan Asal Sekolah". *Citec Journal*. 1 (4). 316-325.