

PENGELOMPOKAN DAN ANALISIS PELANGGAN DENGAN MENGGUNAKAN FUZZY C-MEANS CLUSTERING

Benedictus Andrian Henry Threstanto¹, Dhinta Darmantoro², Kiki Maulana³

¹Teknik Informatika, Fakultas Teknik Informatika, Universitas Telkom

Abstrak

Pada saat ini volume informasi pelanggan yang dimiliki oleh perusahaan semakin meningkat. Data mining dapat digunakan untuk menggali informasi berharga yang sebelumnya tidak diketahui dari suatu database pelanggan untuk mengetahui karakteristik pelanggan. Informasi tersebut dapat digunakan oleh perusahaan untuk melakukan manajemen pelanggan secara lebih efektif. Tugas akhir ini mengimplementasikan salah satu teknik data mining yaitu clustering untuk melakukan pengelompokan dan analisis pelanggan pada suatu perusahaan Telekomunikasi. Metode clustering yang digunakan adalah Fuzzy C-Means Clustering (FCM). Fuzzy C-Means Clustering adalah suatu teknik pengklasteran fuzzy dimana keberadaan tiap-tiap titik data dalam suatu klaster ditentukan oleh derajat keanggotaan. Karena metode clustering bersifat unsupervised learning maka digunakan cluster validity index untuk menganalisis kualitas hasil segmentasi. Dari hasil yang didapat menunjukkan bahwa Fuzzy C-Means Clustering dapat digunakan untuk melakukan segmentasi pelanggan. Namun dalam penentuan validasi klaster menggunakan cluster validity index untuk fuzzy clustering, perlu menggunakan variasi parameter masukan yang berbeda-beda untuk menentukan hasil segmentasi yang optimal.

Kata Kunci : data mining, fuzzy c-means, clustering, pengelompokan pelanggan,

Abstract

Nowadays, the amount of customer information that companies have, has increased significantly. Data mining can be used to retrieve valuable informations that are previously unknown to retrieve the customer characteristics. These informations can be used by the company to effectively carry out customer management. This final project implements one of the techniques in data mining, which is clustering to execute customer segmentation and analysis. The clustering method that is used is Fuzzy C-Means Clustering. Fuzzy C-Means Clustering is a fuzzy clustering technique where each data membership in a cluster is determined by a membership degree. Because clustering is an unsupervised method, cluster validity index is used to analyze the quality of the segmentation results. The testing results shows that Fuzzy C-Means Clustering can be used for customer segmentation. However in defining cluster validation using cluster validity index for fuzzy clustering, it is necessary to use a variety of input parameter for defining the optimum segmentation results.

Keywords : data mining, fuzzy c-means, clustering, customer segmentation, cluster

1. Pendahuluan

1.1 Latar belakang

Pada saat ini volume informasi pelanggan yang dimiliki oleh perusahaan semakin meningkat, oleh karena itu ada tuntutan untuk adanya metode yang dapat dipakai untuk menggali informasi yang sebanyak mungkin mengenai minat, potensi, dan karakteristik pelanggan. *Data mining*, sebagai salah satu ilmu di bidang teknologi informasi, dapat digunakan untuk mengekstraksi informasi berharga yang sebelumnya tidak diketahui dari suatu *database*. Informasi tersebut dapat digunakan oleh perusahaan untuk melakukan manajemen pelanggan secara lebih efektif.

Salah satu informasi yang dibutuhkan oleh perusahaan adalah segmentasi/pengelompokan pelanggan, hal ini dilakukan untuk mendukung strategi manajemen yang bisa jadi berbeda untuk tiap segmen/kelompok. Pada saat ini segmentasi pelanggan di PT.Telkom masih dilakukan secara manual berdasarkan pada *revenue*/pemasukan rata-rata per bulan yang dievaluasi per tahun. Tugas Akhir ini bertujuan untuk memanfaatkan adanya teknik *data mining* seperti *clustering*, untuk memberikan prediksi pengelompokan dan analisis pelanggan berdasarkan dari basis data pelanggan yang dimiliki.

Salah satu metode *clustering* yang dapat digunakan untuk segmentasi data adalah *fuzzy c-means clustering*. *Fuzzy C-Means Clustering* adalah suatu teknik pengklasteran *fuzzy* dimana keberadaan tiap-tiap titik data dalam suatu kluster ditentukan oleh derajat keanggotaan. Dengan metode ini, data pelanggan akan dikelompokkan ke dalam kelompok-kelompok yang berbeda beserta persentase keanggotaannya di masing-masing kelompok. Selain itu akan dihitung juga *cluster validity index* untuk mengetahui apakah segmentasi yang dilakukan sudah optimal. Kemudian dilakukan analisa dari hasil segmentasi untuk mengetahui segmentasi yang optimal. Analisis yang dihasilkan diharapkan dapat membantu efektifitas perusahaan dalam melakukan manajemen pelanggan.

1.2 Perumusan masalah

Berdasarkan latar belakang masalah di atas, maka masalah-masalah yang akan diangkat dalam Tugas Akhir ini adalah:

- a. Bagaimana mengimplementasikan metode Fuzzy C-Means Clustering untuk menentukan prediksi segmentasi pelanggan
- b. Bagaimana menganalisis hasil segmentasi dengan metode *cluster validity index* untuk memperoleh hasil segmentasi yang optimal

Dalam Tugas Akhir ini ada beberapa batasan masalah yaitu :

- a. Data yang akan digunakan sebagai studi kasus adalah data tagihan pelanggan layanan PT.Telkom
- b. Data yang digunakan telah mengalami *pre-processing* terlebih dahulu, sehingga telah siap untuk di-mining
- c. Segmentasi yang akan dilakukan adalah segmentasi berdasarkan data tagihan pelanggan, faktor-faktor lain tidak diperhitungkan

1.3 Tujuan

Secara umum tujuan penulisan yang ingin dicapai dalam tugas akhir ini adalah:

- a. Menerapkan metode *data mining*, *fuzzy c-means clustering* untuk melakukan segmentasi dan analisis pelanggan
- b. Memberikan prediksi segmentasi pelanggan dengan metode *fuzzy c-means clustering* dan analisis segmentasi pelanggan yang optimal dengan *cluster validity index*.

1.4 Metodologi penyelesaian masalah

Metodologi penyelesaian masalah yang dilakukan dalam tugas akhir ini mencakup hal-hal berikut:

- a. Mengumpulkan bahan-bahan literatur dan jurnal-jurnal yang relevan dengan permasalahan ini, yang meliputi: *Data Mining*, *Clustering*, *Fuzzy C-Means Clustering*, *Cluster Validity Index*, *Fuzzy Logic*, dan *Business Intelligence Systems*
- b. Studi literatur tentang *data mining*, *business intelligence*, *fuzzy c-means clustering*, *fuzzy logic*, *cluster validity index*, dan hal-hal lain yang mendukung pendalaman materi
- c. Melakukan pencarian data-data pelanggan dari suatu perusahaan dan melakukan survei ke perusahaan untuk menganalisis manajemen pelanggannya.
- d. Merancang sistem aplikasi untuk melakukan segmentasi dan analisa pelanggan
- e. Menyusun algoritma yang akan digunakan dalam *fuzzy c-means clustering* dan *cluster validity index*
- f. Mengimplementasikan rancangan sistem ke dalam program
- g. Melakukan pengujian sistem menggunakan data-data pelanggan yang diperoleh
- h. Melakukan analisis hasil implementasi segmentasi pelanggan
- i. Menentukan kesimpulan dari hasil analisis dan implementasi
- j. Penyusunan laporan tugas akhir

1.5 Sistematika Penulisan

Penulisan tugas akhir ini dibagi dalam lima bab, yaitu terdiri atas :

BAB 1 Pendahuluan

Berisi latar belakang, perumusan masalah, batasan masalah, tujuan, metodologi penyelesaian masalah dan sistematika penulisan.

- BAB 2 Landasan Teori**
Bab ini dikemukakan berbagai teori dasar yang mendukung terlaksananya tugas akhir ini, antara lain mengenai *data mining*, *clustering*, *distance space*, *fuzzy c-means clustering*, dan *cluster validity index*.
- BAB 3 Analisis dan Perancangan Sistem**
Membahas tentang analisis dan perancangan awal sistem
- BAB 4 Implementasi dan Pengujian**
Bab ini membahas kebutuhan perangkat lunak dan perangkat keras yang digunakan untuk merealisasikan sistem. Selain itu, pada bab ini akan dibahas mengenai scenario pengujian perangkat lunak/simulasi, hasil uji coba dan analisa dari hasil yang didapatkan.
- BAB 5 Kesimpulan dan Saran**
Berisi kesimpulan akhir dan saran terhadap pengembangan dari penelitian tugas akhir ini selanjutnya.



5. Kesimpulan dan Saran

5.1 Kesimpulan

Kesimpulan yang dapat diambil dari Tugas Akhir ini adalah :

- 1) Metode Fuzzy C-Means Clustering dapat digunakan untuk melakukan segmentasi data pelanggan
- 2) Performansi Fuzzy C-Means Clustering pada dataset Iris lebih baik bila dibandingkan dengan K-Means Clustering yaitu 92.66667% berbanding 88.66667%. Rata-rata kesalahan clustering oleh Fuzzy C-Means Clustering pada dataset Iris yaitu 6.62156%.
- 3) Metode Fuzzy C-Means dapat dipakai untuk menentukan label kedua suatu data yang dinilai berdasarkan derajat keanggotaannya.
- 4) Untuk menentukan segmentasi yang optimal dengan Fuzzy C-Means clustering pada suatu dataset yang belum diketahui kelas-kelas dan jumlah klasternya, dapat dilakukan dengan menggunakan *cluster validity index*.
- 5) Untuk pengujian hasil segmentasi dengan *cluster validity index* harus dilakukan dengan parameter masukan yang berbeda-beda untuk menentukan hasil segmentasi yang optimal.

5.2 Saran

Saran terhadap pengembangan yang akan dilakukan terhadap TA ini adalah :

- 1) Menggunakan metode clustering lain untuk melakukan segmentasi pelanggan
- 2) Menentukan hasil segmentasi yang optimal dengan metode validasi klaster yang lain

Daftar Pustaka

- [1] Agusta, Yudi, PhD, *K-Means – Penerapan, Permasalahan dan Metode Terkait*, 2007, Jurnal Sistem dan Informatika Vol. 3 (Februari 2007).
- [2] *Clementine 9.0 Algorithm Guide*, 2004, Integral Solutions Limited, USA.
- [3] Duo, Chen, Li Xue, Cui Du Wu, 2007, *An Adaptive Cluster Validity Index for the Fuzzy C-means*, IJCSNS International Journal of Computer Science and Network Security.
- [4] Han, Jiawei, Micheline Kamber, 2000, *Data Mining: Concepts and Techniques*, Morgan Kaufmann Publishers.
- [5] Halkidi, Maria, Y. Batistakis, M. Vazirgiannis, 2005, *Clustering Validity Checking Methods Part II*, Dept of Informatics Athens Univ. of Economics and Business.
- [6] Kusumadewi, Sri, Hari Purnomo, 2004, *Aplikasi Fuzzy Logic untuk Pengambil Keputusan*, Yogyakarta: Graha Ilmu.
- [7] Modenesi, V.M., Myrian Costa, Alexandre Evsukoff, Nelson Ebecken, 2006, *Parallel Fuzzy C-Means Cluster Analysis*, Rio de Janeiro, Brazil.
- [8] Pramudiono, Iko, *Pengantar Data Mining: Menambang Permata Pengetahuan di Gunung Data*, 2003, IlmuKomputer.Com.
- [9] Simha, B. Jay, S.S Iyengar, 2005, *Customer Value Analysis with Fuzzy Data Mining*, Dept. of Computer Science Louisianan State University, USA.
- [10] Tan, Pang Ning, Michael Steinbach, Vipin Kumar, 2000, *Introduction to Data Mining*.
- [11] Turban, Efraim, Jay E. Aronson, Ting Peng Liang, 2005, *Decision Support Systems and Intelligent Systems*, Yogyakarta: Andi.