

Penerapan Algoritma K-Means dan Cure Dalam Menganalisa Pola Perubahan Belanja Dari Retail ke E-Commerce

Muhammad Miftakhul Syaikhuddin¹⁾ Prihandoko²⁾

¹⁾Magister Teknik Informatika, Universitas Amikom, Yogyakarta

²⁾Fakultas Ilmu Komputer dan Teknologi Informasi, Universitas Gunadharma, Depok

Terima Naskah : 10 September 2017

Terima Revisi : 25 September 2017

ABSTRAK

Pola-pola perubahan dunia pastilah terjadi, tidak menutup kemungkinan terjadinya sebuah pola perubahan masyarakat dalam melakukan proses transaksi jual beli. Seiring perkembangan zaman dan didorong dengan munculnya perkembangan teknologi yang menyebar ke seluruh lapisan masyarakat, maka telah terjadi perubahan pola belanja masyarakat yang sebelumnya melakukan belanja hanya di ritel atau toko kelontong, dengan adanya teknologi dan website e-commerce maka masyarakat melakukan belanja dengan menggunakan e-commerce. Terjadi pola perubahan tersebut akan disertai dengan beberapa faktor-faktor pendukung yang mengubahnya, baik itu secara social, ekonomi, psikologi, ataupun perilaku masyarakat itu sendiri. Faktor-faktor tersebut perlu dilakukan sebuah pengelompokan agar dapat diketahui faktor apa yang merupakan pendorong terjadinya perubahan tersebut, sehingga didapatkan sebuah informasi guna menentukan strategi penjualan yang lebih baik. Proses pengelompokan tersebut dapat menggunakan teknik dari data mining yakni dengan algoritma K-means dan Clustering Representative (CURE). Dengan algoritma k-means yang mudah dan memiliki tingkat akurasi yang tinggi, ditunjang dengan algoritma CURE yang memiliki akurasi dalam pengolahan data yang besar maka, diharapkan memberikan sebuah keluaran yang akurat dan sesuai dengan tujuan penelitian ini.

Kata Kunci : E-commerce, Ritel, Algoritma K-means, Algoritma CURE.

ABSTRACT

Patterns of change in the world must have happened, did not cover the possibility of the occurrence of a change in the pattern of community in the process of buying and selling. As the development of the times and is driven by the emergence of technological development that spread to all walks of life, then it's been going shopping patterns change communities that previously did shopping only at retail or grocery store, with technology and e-commerce website then do some shopping community by using e-commerce. The change pattern occurs will be accompanied by some of the factors supporting that change it, whether it is social, economics, psychology, or behavior of the community itself. These factors need to be done a grouping to be known what factors is driving these changes, so that the occurrence of acquired information to determine a strategy for better sales. The grouping process can use the techniques of data mining: with K-means algorithm for Clustering and Representative (CURE). With the k-means algorithm that is easy and has a high degree of accuracy, supported by the CURE algorithm and like accuracy in large data processing then, is expected to give an output that is accurate and in accordance with the objectives of this research .

Keywords: E-commerce, Retail, K-means Algorithm, CURE Algorithm

PENDAHULUAN

Perkembangan teknologi mengubah beberapa aspek kehidupan masyarakat dalam melakukan aktivitasnya. Perubahan aktivitas tersebut didukung dengan perkembangan aplikasi-aplikasi yang dikembangkan oleh beberapa perusahaan startup. Setiap aktivitas masyarakat tak lepas dari

gawainya yang digunakan untuk mendukung kemudahan dalam melakukan aktivitasnya. Berdasarkan laporan Akamai.com, penggunaan gawai untuk akses internet untuk menggunakan aplikasi online memiliki peningkatan 70% pada trafik internet dan aksesnya[1].

Berbanding terbalik dengan peningkatan akses internet dan aplikasi mobile, di sisi lain terdapat

sebuah penurunan penjualan di toko ritel terutama di Indonesia. Penjualan ritel sampai saat ini sebesar 3.6%-3.8% yang mengalami penurunan dari semester II tahun 2016 yakni sebesar 7-8% [2]. Hal ini berbanding terbalik dengan pertumbuhan ekonomi sebesar 5,01% yang diatas pertumbuhan ekonomi dunia. Permasalahan tersebut merupakan sebuah fenomena yang perlu diadakannya sebuah penelitian mengenai hal tersebut.

Pola-pola perubahan dalam dunia pasti terjadi, beberapa penelitian mengungkapkan adanya suatu pola perubahan belanja mulai dari pola perubahan belanja dari toko dekat rumah, berubah ke mall atau ritel. Pola perubahan tersebut juga dapat terjadi dengan perubahan pemilihan suatu produk lain atau brand lainnya. Menurut [3] dalam penelitiannya mengungkapkan bawasannya pola perubahan belanja masyarakat terjadi sejak tahun 1990 dan perubahan tersebut terjadi juga pada tahun 2001, 2013, dan 2014 seiring dengan perkembangan pusat perdagangan modern atau ritel.

Berdasarkan laporan penelitian yang telah dilakukan diseleuruh dunia oleh Nielsen diperoleh sebuah fakta yang aman seperempat responden melakukan pemesanan sebuah produk secara online dan lebih dari setengahnya atau 55% diantaranya akan melakukannya. Penggunaan kupon dan mobile dan list daftara belanja merupakan hal yang sering dilakukan saat ini [4]. Hal ini bisa menjadi sebuah indicator berkembangnya dunia *E-commerce* baik di Indonesia maupun di Dunia.

Penggunaan algoritma k-means dapat dilakukan untuk melakukan segmentasi pelanggan guna membuat strategi dalam menarget pelayanan pelanggan. Penelitian yang mendukung penggunaan k-means ini dilakukan oleh [5] dalam penelitiannya didapati kemurnian dalam data yang diolah mencapai 0.95 dengan indikasi 95% keakuratan segmentasi pelanggan.

Penelitian lain yang menggunakan algoritma CURE dalam proses pengelolaan data agar data yang didapatkan efisien dan akurat. Penelitian performa algoritma CURE dalam proses analisa kumpulan data kecil dapat dilakukan dalam satu atau dua detik dan menunjukkan peningkatan rata-rata keakuratan data dalam setiap meningkatnya ukuran data [6].

Dengan adanya fenomena dan didukung dengan penelitian yang dilakukan sebelumnya, peneliti berkeinginan mengetahui fenomena

penurunan penjualan ritel yang terjadi berkorelasi atau tidak dengan perkembangan teknologi khususnya perkembangan pembelian masyarakat melalui marketplace atau *E-commerce*. Perubahan dengan fenomena tersebut dapat dilakukan dengan melakukan penelitian. Penelitian tersebut dengan cara melakukan observasi dan penghimpunan data melalui survey atau penyebaran kuesioner yang akan diisi oleh masyarakat dan data yang didapatkan dari survey tersebut selanjutnya diolah dengan metode clustering dan menggunakan 2 teknik pengolahan data yakni algoritma K-means dan CURE. Penelitian ini dilakukan guna mendapatkan jawaban mengenai fenomena yang terjadi dimasyarakat Indonesia, agar dapat dijadikan sebagai salah satu pertimbangan dalam menentukan strategi peningkatan ekonomi masyarakat Indonesia dan dalam dunia akademis, penelitian ini diharapkan memperkaya dan menyempurnakan teknik pengolahan data mining terutama terhadap algoritma K-means

Belanja

Menurut [7] dalam bukunya, menjelaskan aktifitas berbelanja adalah suatu aktifitas yang membutuhkan suatu pertimbangan dalam pembelian suatu produk dan jasa, mencari toko yang menyediakan produk dan jasa, mencari produk dan jasa dalam toko tersebut, sebelum membuat keputusan untuk membelinya.

Data Mining

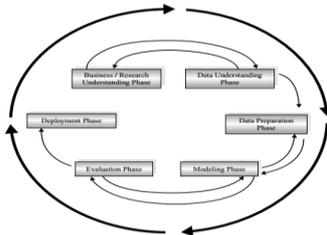
Data Mining adalah proses menemukan pola menarik dan pengetahuan dari sejumlah besar sumber data. Data bisa berupa database, data gudang, Web, repositori informasi lain, atau data yang dialirkan ke sistem secara dinamis[8].

Data mining [9] adalah sebuah proses menemukan korelasi baru yang berate, pola dan trend dengan memilah sejumlah data dalam jumlah besar yang tersimpan dalam sebuah repository dengan menggunakan teknologi pengenalan pola serta statistic dan teknik matematika. Proses pencarian informasi dengan menggunakan data dalam jumlah besar dapat dilakukan dengan metode yang ada dalam data mining dengan melalui sebuah proses tahapan pencarian pengetahuan. Proses tersebut dapat dilakukan secara *iterative*.

Adapun tahapan dalam data mining yang bernama The Cross-Industry Standard Process for Data Mining (CRISP-DM) yang dikembangkan oleh analisis DaimlerChrysler, SPSS, dan NCR.

CRISP-DM memiliki enam fase dalam melakukan data mining diantaranya[9]:

- Business understanding phase*, Proses merupakan proses dimana menentukan tujuan, syarat batasan, sasaran dalam perumusan masalah dalam data mining serta menyiapkan strategi dalam pemecahan masalah agar tercapainya tujuan.
- Data understanding phase*, yakni proses mengumpulkan data, menganalisa data untuk menemukan wawasan awal, mengevaluasi data dan memilih subset yang mengandung pola yang dapat ditindaklanjuti.
- Data preparation phase*, menyiapkan data mentah dari fase sebelumnya yang akan digunakan sampai akhir, memilih kasus, memilih variable yang akan dianalisa, melakukan transformasi variable dan membersihkan data mentah agar siap dimodelkan.
- Modeling phase*, Fase untuk menentukan teknik permodelan yang sesuai dan mengatur model dengan mengkalibrasi untuk menghasilkan data yang optimal.
- Evaluation phase*, Melakukan proses evaluasi dari model dari fase sebelumnya dan mengevaluasi model tersebut apakah sesuai dengan tujuan yang akan dicapai pada fase pertama.
- Deployment phase*. Memanfaatkan model yang telah dibuat dengan membuat sebuah laporan dengan berdasarkan model yang telah dibuat.



Gambar 1. Proses Tahapan CRISP-DM

Algoritma K-means

Algoritma k-means merupakan algoritma yang mudah dan efektif untuk menemukan kluster dalam data. Adapun tahapan dalam proses clustering dengan k-means antara lain[9].

- Menentukan berapa banyak data yang akan dipartisi.
- Secara acak menetapkan k record menjadi lokasi cluster center awal.
- Menemukan pusat cluster terdekat. Jadi, dalam arti tertentu, masing-masing cluster center "memiliki" subset dari catatan, sehingga mewakili

sebuah partisi dari kumpulan data Oleh karena itu kami memiliki k cluster, C1, C2, . . . , Ck.

- Menemukan centroid kluster dan memperbarui lokasi masing-masing cluster ke dalam nilai centroid yang baru pada masing-masing k cluster
- Mengulang langka 3 sampai 5 sampai konvergensi atau terhenti.

Kriteria terdekat pada langkah ketiga,biasanya jarang Euclidean, walaupun kriteria lainnya dapat diterapkan. Cluster centroid pada langkah 4 ditemukan sebagai berikut. Misalkan terdapat nilai n titik data (a1, b1, c1), (a2, b2, c2) ,..., (an,bn,cn), centroid dari masing-masing titik ini adalah proses gravitasi titik dan terletak di titik ($\sum a_i/n$, $\sum b_i/n$, $\sum c_i/n$).

Algoritma k-means akan berakhir apabila centroid tidak lagi berubah. Maksudnya, algoritma berakhir ketika semua kluster C1, C2, C3 ,..., Ck. Semua caatanyang dimiliki oleh masing-masing cluster centermain di kluster tersebut. Dan juga algoritma k-means akan berakhir ketika beberapa kriteria konvergensi dipenuhi, seperti tidak ada penyusutan signifikan dalam jumlah kesalahan kuadrat.

$$SSE = \sum_{i=1}^k \sum_{p \in C_i} d(p, m_i)^2$$

Dimana $p \in C_i$ mewakili setiap titik data pada cluster i dan m_i mewakili centroid dari cluster i.

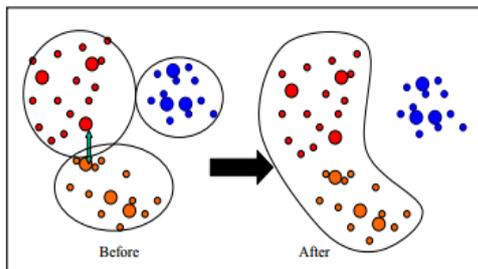
Algoritma Clustering Representative (CURE)

Algoritma CURE, pertama kali diperkenalkan oleh Guha dkk, yakni sebuah algoritma clustering agglomerative hirarki. Algoritma ini membutuhkan sebuah perawatan khusus pada outlier dan pada tahap pemberian label. Algoritma ini menggunakan 2 perangkat untuk mencapai skalabilitas[10]. Yang pertama, data sampling (bagian Skalabilitas dan Eksistensi VLDB), perangkat kedua adalah partisi data dari partisi p, sehingga cluster granularity halus dibangun dipartisi terlebih dahulu. Fitur utama dari

CURE adalah bahwa hal itu mewakili sebuah cluster dengan jumlah angka c yang terpencah di sekitarnya. Itu jarak antara dua kelompok yang digunakan dalam proses aglomerasi sama dengan minimum jarak antara dua perwakilan yang tersebar. Oleh karena itu, CURE mengambil pendekatan middleground antara metode graph (all-points) dan geometris (one centroid) metode. Keterkaitan tautan tunggal dan rata-rata digantikan oleh agregat perwakilan kedekatan. Memilih perwakilan yang tersebar di sekitar sebuah cluster

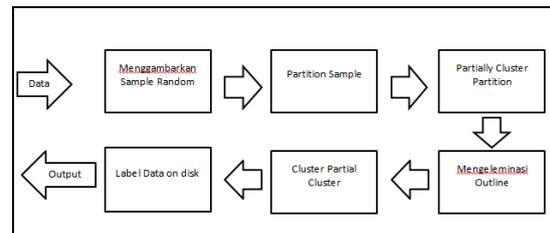
memungkinkan untuk menutupi bentuk non-bola. Seperti sebelumnya, aglomerasi berlanjut sampai permintaan nomor k dari cluster tercapai CURE menggunakan satu perangkat tambahan: awalnya dipilih berserakan poin menyusut ke pusat geometris cluster oleh faktor yang ditentukan pengguna α . Penyusutan menekan pengaruh outlier sejak outlier terjadi lebih jauh dari centroid cluster daripada perwakilan lainnya yang tersebar. CURE mampu menemukan kelompok dengan berbagai bentuk dan ukuran, dan ini tidak sensitif terhadap outlier. Sejak CURE menggunakan sampling, estimasi kompleksitasnya tidak langsung. Untuk dimensi rendah penulis data memberikan perkiraan kompleksitas $O(N^2_{Sample})$ yang didefinisikan dalam hal ukuran sampel[10].

Batas yang lebih tepat tergantung pada parameter masukan: faktor mengecilkan α , jumlah perwakilan titik c , jumlah partisi, dan ukuran sampel. Gambar 2 menggambarkan aglomerasi di CURE. Tiga kelompok, masing-masing dengan tiga perwakilan, diperlihatkan sebelum dan sesudah bergabung dan menyusut. Dua perwakilan terdekat dihubungkan dengan panah. Sedangkan algoritma CURE bekerja dengan atribut numerik (terutama dimensi rendah data spasial)



Gambar 2. Angglomerasi dalam CURE

Menurut [11] CURE adalah algoritma clustering hirarkis aglomeratif yang menciptakan keseimbangan antara pendekatan centroid dan all point. Pada dasarnya CURE adalah clustering hirarkis algoritma yang menggunakan partisi dataset. Kombinasi random sampling dan partisi digunakan disini agar database besar bisa ditangani. Dalam proses ini secara acak sampel yang diambil dari dataset dipartisi terlebih dahulu dan kemudian setiap partisi sebagian berkerumun. Cluster parsial kemudian dikelompokkan lagi pada jalur kedua untuk menghasilkan cluster yang diinginkan



Gambar 3. Gambaran Algoritma CURE

METODE

Penelitian mengenai pola perubahan belanja ritel ke e-commrce ini dilakukan pada data masyarakat yang sudah pernah melakukan transaksi baik secara ritel ataupun melalui e-commrce. Adapun rencana pengumpulan data dalam penelitian ini, diantaranya sebagai berikut:

1. Sumber Data Prime yang berupa Kuesioner dan hasil Observasi.
2. Sumber Data Sekunder yang merupakan studi dokumentasi studi kepustakaan.

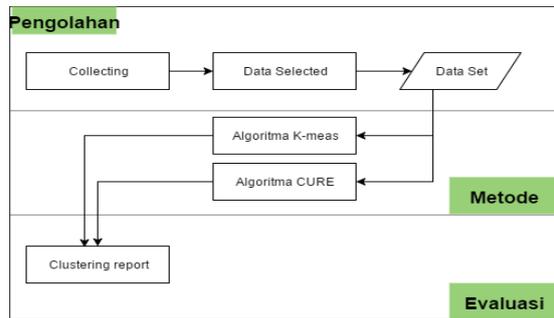
Analisa Data

Data yang digunakan dalam penelitian ini berseumber dari angket yang diberikan kepada responden yang selanjutnya dilakukan analisa data dengan menggunakan metode clustering dan teknik data yang digunakan adalah algoritma K-means dan CURE dan analisa datanya menggunakan aplikasi WEKA atau aplikasi alternative lainnya yang sesuai dengan algoritma K-means dan CURE.

Alur Penelitian

Alur penelitian dari penelitian ini dimulai dengan mengelola data yang sudah didapatkan dari kuesioner (pengolahan data awal), proses evaluasi dan terakhir yakni hasil penelitian.

Pengolahan data dalam penelitian ini menggunakan metode clustering dan teknik yang digunakan adalah algoritma K-means dan CURE. Pada proses perhitungan menggunakan algoritma k-means dengan menentukan nilai k sebagai langkah awal. Setelah hasil pengolahan data dengan menggunakan algoritma K-means dan CURE maka akan didapatkan sebuah hasil cluster data. Data tersebut kemudian dibandingkan antara data yang dicluster dengan algoritma K-means dan CURE mempunyai tingkat akurasi yang tinggi. Alur tahapan penelitian dengan metode tersebut dapat dilihat pada gambar 4.



Gambar 4. Diagram Alur Penelitian

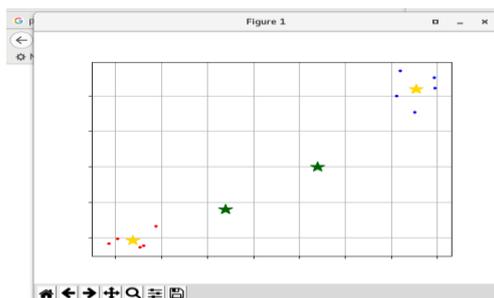
Model penelitian ini menggunakan algoritma K-means dan CURE.

HASIL DAN PEMBAHASAN

Hasil yang akan dicapai dalam penelitian ini adalah sebuah model yang didapatkan dari proses modeling dari algoritma k-means dan CURE yang berupa report cluster yang kemudian diinterpretasikan untuk mendapatkan sebuah informasi yang diinginkan dalam rumusan masalah dan tujuan dari penelitian ini. Yakni faktor yang mempengaruhi terjadi perubahan pola belanja masyarakat dari belanja secara ritel beralih belanja menggunakan aplikasi e-commerce.

Hasil dari data sampel produk yang dibeli masyarakat yang dijalankan dengan menggunakan aplikasi pyclustering, penggunaan algoritma k-means, ditemukan bawasannya tingkat akurasi dari metode k-means dengan data sejumlah data sample yang ada, didapati tingkat akurasi sebesar 85,1%. Dengan jumlah cluster sebanyak 2 cluster.

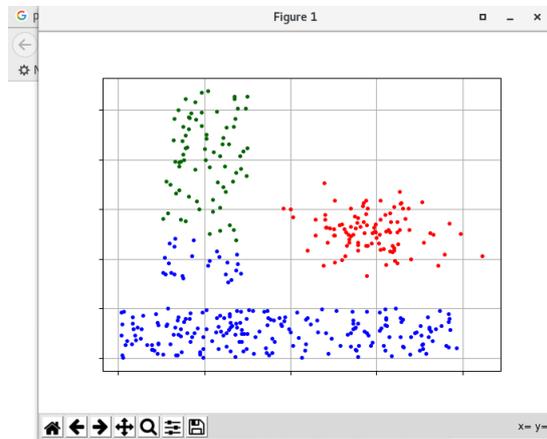
Sebaran dari data cluster tersebut dapat dilihat pada gambar berikut.



Gambar 5. Sebaran data Cluster dari Algoritma K-means

Penggunaan algoritma CURE, difungsikan sebagai algoritma pembanding yang diperuntukan untuk data yang lebih besar sehingga dicapai

tingkat akurasi yang lebih baik. Hal ini dapat dilihat pada gambar 6 Sebaran cluster dari data CURE.



Gambar 6 Sebaran Data Cluster Algoritma CURE

Gambaran sebaran algoritma k-means dan algoritma CURE memperlihatkan adanya sebuah perbedaan dalam sebarannya yakni dengan data yang besar sebaran dari cure cenderung lebih baik dibandingkan dengan k-means.

SIMPULAN

Dari hasil penelitian penulis dapat memberikan gambaran faktor-faktor yang mempengaruhi perubahan pola belanja masyarakat dan memberikan informasi mengenai tingkat akurasi dari kedua algoritma yang digunakan dalam penelitian ini.

DAFTAR PUSTAKA

- [1] Akamai, "Q1 2016 State of the Internet - Connectivity Report | Akamai," 2016.
- [2] Y. Fauzi, "Penjualan Lebaran Anjlok, Bisnis Ritel Semester I Tumbuh 3,8%," *CNN Indonesia*, 24-Sep-2017.
- [3] A. P. Usandy, G. Yudana, and E. F. Rini, "PERUBAHAN POLA PERGERAKAN BELANJA MASYARAKAT PASCA DIBANGUN PUSAT PERDAGANGAN MODERN DI SOLO BARU," *Arsitektura*, vol. 15, no. 1, pp. 10–17, 2017.
- [4] Nielsen, "The Future of Grocery," 2015.
- [5] C. P. Ezenkwu, S. Ozuomba, and C. Kalu, "Application of K-Means Algorithm for Efficient Customer Segmentation: A Strategy for Targeted Customer Services," *Int. J. Adv. Res. Artif. Intell.*, vol. 4, no. 10, pp. 40–44,

- 2015.
- [6] S. Sagunthaladevi and B. R. V. R. Raju, "Performance Analysis of Cure and Rock Algorithms on Constructing a New Attribute with Mixed Datasets," *Int. J. Innov. Res. Sci. Eng. Technol.*, vol. 5, no. 1, pp. 127–134, 2017.
 - [7] P. Huddleston and S. Minahan, *Consumer Behavior: Women and Shopping*. New York: Business Expert Press, 2011.
 - [8] J. Han, M. Kamber, and J. Pei, *Data Mining : Concept and Techniques*, vol. 1. 2012.
 - [9] D. T. Larose, *Discovering Knowledge in Data*. Hoboken: John Wiley & Sons, Inc., 2005.
 - [10] Pavel Berkhin, "Survey Of Clustering Data Mining Techniques," pp. 1–56, 2002.
 - [11] Y. Rani and H. Rohil, "A Study of Hierarchical Clustering Algorithm," *Int. J. Inf. Comput. Technol.*, vol. 3, no. 10, pp. 1115–1122, 2013.