

SISTEM REKOMENDASI BERDASARKAN DATA TRANSAKSI PERPUSTAKAAN DAERAH TABANAN DENGAN MENGGUNAKAN K-MEANS CLUSTERING

¹⁾ I Made Agus Widiana Putra, ²⁾ Gede Indrawan, ³⁾ Kadek Yota Ernanda Aryanto

^{1,2,3)} Program Studi Ilmu Komputer, Program Pascasarjana Universitas Pendidikan Ganesha
Singaraja, Indonesia

e-mail: ¹⁾ made.widiana08@gmail.com, ²⁾ gede.indrawan.com, ³⁾ yota.ernanda@gmail.com

Abstrak

Perpustakaan adalah suatu unit kerja yang berupa tempat menyimpan koleksi bahan pustaka yang diatur secara sistematis dengan cara tertentu untuk dipergunakan secara berkesinambungan oleh pemakainya sebagai sumber informasi. Perpustakaan daerah kabupaten Tabanan merupakan salah satu perpustakaan Umum di Bali yang memberikan pelayanan kepada masyarakat terhadap peminjaman buku setiap harinya dari berbagai kalangan. Permasalahan yang dihadapi dalam pelayanan perpustakaan yaitu bagaimana menarik perhatian pengunjung dengan menampilkan rekomendasi kepada peminjam yang berbeda usia. Metode Sistem Rekomendasi merupakan metode yang sesuai dimana metode tersebut menggunakan opini dari suatu komunitas pengguna untuk membantu seorang individu dari komunitas tersebut dalam mencari informasi tertentu. Teknologi clustering data merupakan suatu teknik yang menunjukkan persamaan karakteristik dalam suatu kelompok sehingga akan menghasilkan informasi yang berguna nantinya. K-means merupakan algoritma yang sangat banyak dipergunakan karena efektif dan efisien. Ini dikarenakan K-means sangat mudah dipelajari dan dari segi waktu proses komputasinya relatif singkat. Metode Elbow digunakan untuk menentukan jumlah cluster terbaik untuk penelitian ini. Jumlah cluster yang terbentuk sebanyak 3 cluster dengan menggunakan Metode Elbow. Nilai kemurnian dari cluster yang terbentuk dalam penelitian ini adalah 0.7187017 dengan data yang 930 relevan. Dapat dikatakan hasil dari penelitian ini baik karena nilai purity (kemurnian) mendekati 1.

Kata-kata kunci: K-means, Elbow, Clustering, Rekomendasi, Purity

Abstract

Library is one of the work units that place to save the collection of library materials which are arranged systematically by particular method that used sustainable by the user as the information source. Local library of Tabanan regency is one of the public library in Bali which gives service to the society for borrowing book every day from various circles. The problem that is faced in library service is how to attract the attention of visitor by showing the recommendation to the borrower that different age. Method of recommendation system is a method that suitable where that method using the opinion from user community to help a person from that community in looking for the specific information. Clustering data technology is a technic that show the same characteristic in a group so that produce useful information eventually. K – Means is algorithm that widely used because of the effective and the efficient. This is because of K – Means is very easy learned and relatively short in computation process. Elbow method is used to establish the best cluster amount for this research. The cluster amount that formed is 3 clusters by using elbow method. The value of purity of the cluster formed is 0.7187017 with 930 relevant dat. It can be said the result of this research is good because the purity value approach 1.

Key words: K – Means, Elbow, Clustering, Recommendation, Purity.

I. PENDAHULUAN

Perpustakaan adalah suatu unit kerja yang berupa tempat menyimpan koleksi bahan pustaka yang diatur secara sistematis dengan cara tertentu untuk dipergunakan secara berkesinambungan oleh pemakainya sebagai sumber informasi. Perpustakaan daerah kabupaten Tabanan merupakan salah satu perpustakaan Umum di Bali yang memberikan pelayanan kepada masyarakat terhadap peminjaman buku setiap harinya dari berbagai kalangan. Dengan banyaknya informasi terkadang membuat orang merasa bingung dalam mencari informasi mana yang sesuai dengan apa yang kebutuhannya. Pemasalahan yang dihadapi dalam pelayanan perpustakaan yaitu bagaimana menarik perhatian pengunjung dengan menampilkan rekomendasi kepada peminjam yang berbeda usia. Untuk itu diperlukan sebuah sistem Rekomendasi Pencarian dalam memberikan solusi dari permasalahan tersebut. Metode Sistem Rekomendasi merupakan metode yang sesuai dimana metode tersebut menggunakan opini dari suatu komunitas pengguna untuk membantu seorang individu dari komunitas tersebut dalam mencari informasi tertentu.

Penelitian ini bertujuan untuk pengembangan penggunaan metode data mining dalam membantu proses pengelolaan buku di perpustakaan dengan mengolah data transaksi yang ada. Teknologi clustering data merupakan suatu teknik yang menunjukkan persamaan karakteristik dalam suatu kelompok sehingga akan menghasilkan informasi yang berguna nantinya. Algoritma clustering data sudah banyak dipergunakan diberbagai bidang misalnya untuk proses pengolahan citra, data mining proses pengambilan keputusan, pengenalan pola, maupun dalam bidang bioinformatika. Clustering data akan mengelompokkan objek yang paling dekat dimana terdapat kesamaan dengan objek lain, serta data yang akan dicluster diambil secara acak atau random.

K-Means merupakan algoritma yang sangat banyak dipergunakan karena efektif dan efisien. Ini dikarenakan K-Means sangat mudah dipelajari dan dari segi waktu proses komputasinya relatif singkat. Penentuan nilai

hasil cluster dilihat dari jarak terdekat antar objek data. Sebelum menghitung jarak terdekatnya, harus ditentukan terlebih dahulu jumlah cluster centroid (titik pusat). Selain itu K-Means juga memiliki ketelitian yang cukup tinggi terhadap ukuran objek, sehingga algoritma ini relatif lebih terukur dan efisien untuk pengolahan objek dalam jumlah yang besar.

II. TINJAUAN PUSTAKA

Secara sederhana data mining adalah penambangan atau penemuan informasi baru dengan mencari pola atau aturan tertentu dari sejumlah data yang sangat besar. Data mining juga disebut sebagai serangkaian proses untuk menggali nilai tambah berupa pengetahuan yang selama ini tidak diketahui secara manual dari suatu kumpulan data (Pramudiono, 2007). Data mining, sering juga disebut sebagai knowledge discovery in database (KDD). KDD adalah kegiatan yang meliputi pengumpulan, pemakaian data, historis untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar.

2.1 CLUSTERING

Clustering adalah membagi data ke dalam grup-grup yang mempunyai obyek yang karakteristiknya sama. Garcia-Molina et al. menyatakan clustering adalah mengelompokkan item data ke dalam sejumlah kecil grup sedemikian sehingga masing-masing grup mempunyai sesuatu persamaan yang esensial.

Clustering memegang peranan penting dalam aplikasi data mining, misalnya eksplorasi data ilmu pengetahuan, pengaksesan informasi dan text mining, aplikasi basis data spasial, dan analisis web. Clustering diterapkan dalam mesin pencari di Internet. Web mesin pencari akan mencari ratusan dokumen yang cocok dengan kata kunci yang dimasukkan. Dokumen-dokumen tersebut dikelompokkan dalam cluster-cluster sesuai dengan kata-kata yang digunakan.

2.2 ALGORITMA K-Means

Algoritma K-Means merupakan salah satu algoritma dengan partitional, karena K-Means didasarkan pada penentuan jumlah awal kelompok dengan mendefinisikan nilai centroid awalnya. Algoritma K-Means menggunakan proses secara berulang-ulang untuk mendapatkan basis data cluster.

Dibutuhkan jumlah cluster awal yang diinginkan sebagai masukan dan menghasilkan jumlah cluster akhir sebagai output. Jika algoritma diperlukan untuk menghasilkan cluster K maka akan ada K awal dan K akhir. Metode K-Means akan memilih pola k sebagai titik awal centroid secara acak. Jumlah iterasi untuk mencapai cluster centroid akan dipengaruhi oleh calon cluster centroid awal secara random dimana jika posisi centroid baru tidak berubah. Nilai K yang dipilih menjadi pusat awal, akan dihitung dengan menggunakan rumus Euclidean Distance yaitu mencari jarak terdekat antara titik centroid dengan data/objek. Data yang memiliki jarak pendek atau terdekat dengan centroid akan membentuk sebuah cluster.

2.3 METODE ELBOW

Metode Elbow merupakan suatu metode yang digunakan untuk menghasilkan informasi dalam menentukan jumlah cluster terbaik dengan cara melihat persentase hasil perbandingan antara jumlah cluster yang akan membentuk siku pada suatu titik. Metode ini memberikan ide/gagasan dengan cara memilih nilai cluster dan kemudian menambah nilai cluster tersebut untuk dijadikan model data dalam penentuan cluster terbaik. Dan selain itu persentase perhitungan yang dihasilkan menjadi perbandingan antara jumlah cluster yang ditambah. Hasil persentase yang berbeda dari setiap nilai cluster dapat ditunjukkan dengan menggunakan grafik sebagai sumber informasinya. Jika nilai cluster pertama dengan nilai cluster kedua memberikan sudut dalam grafik atau nilainya mengalami penurunan paling besar maka nilai cluster tersebut yang terbaik.

Untuk mendapatkan perbandingannya adalah dengan menghitung SSE (Sum of Square Error) dari masing-masing nilai cluster. Karena semakin besar jumlah cluster K maka nilai SSE akan semakin kecil. Suum of Square Error merupakan mengukur perbedaan antara data yang diperoleh dengan model perkiraan yang telah dilakukan sebelumnya.

2.4 EVALUASI CLUSTER

Evaluasi algoritma clustering dilihat berdasarkan bagaimana algoritma tersebut dapat

menghasilkan sekumpulan cluster yang mencerminkan pikiran manusia, sebagaimana jika kategorisasi dilakukan secara manual oleh manusia. Terdapat berbagai pendekatan kriteria yang dapat digunakan dalam melakukan evaluasi, salah satunya adalah dengan pendekatan kriteria eksternal, dimana evaluasi kriteria eksternal dilakukan dengan membandingkan hasil cluster yang dilakukan oleh algoritma clustering dengan hasil cluster yang dilakukan secara manual oleh manusia. Purity adalah ukuran “kemurnian” suatu cluster, yaitu seberapa murni solusi clustering yang diperoleh. Nilai kemurnian (purity) dari sebuah cluster berkisaran antara 0 dan 1. Clustering buruk jika nilai purity mendekati 0, dan baik jika nilai purity mendekati 1.

III. METODE PENELITIAN

WEKA berasal dari Selandia Baru (New Zealand) yang dikembangkan oleh Universitas Waikato. Weka adalah tools yang cukup tangguh dan mumpuni. Weka memiliki koleksi berbagai macam jenis algoritma machine learning yang digunakan untuk melakukan generalisasi atau formulasi dari dataset atau sample data.

Meskipun weka cukup unggul di dalam variasi algoritma yang dikumpulkan, namun tetap saja peran user masih menjadi yang utama, Semua karena secanggih apapun weka jika masukan datanya tidak sesuai dengan format, maka weka tidak akan bisa melakukan proses learning. Pada data mining peran data sangatlah penting, sebab keakuratan hasil dari proses ditentukan dari dataset atau inputan yang baik. Inputan bisa dikatakan baik jika sudah tidak memiliki data noise.

Weka mengenal dua format, yaitu CSV dan arff, yang pertama adalah CSV. CSV (Comma Delliimited) "kalau di excel" adalah format yang sering diminati para peneliti karena setiap record data dipisahkan dengan dengan koma (,) atau titik koma (;) yang bisa dibuka menggunakan text editor seperti notepad, wordpad, excel. ARFF adalah format yang dimengerti oleh machine learning weka untuk proses tahap selanjutnya,

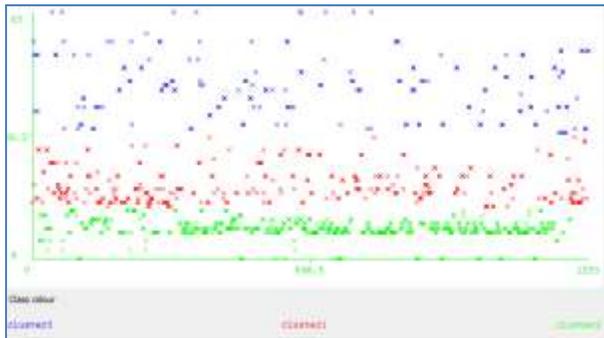
IV. HASIL dan PEMBAHASAN

Penelitian ini dilakukan dengan mengambil data transaksi selama satu tahun terakhir dengan

jumlah data sebanyak 1294 data transaksi pada perpustakaan Daerah Tabanan. Data yang akan diuji cobakan adalah data transaksi peminjaman buku pada perpustakaan Daerah Tabanan dengan atribut data item, nomor anggota, nama anggota, yang dipinjam, pekerjaan dan umur yang melakukan transaksi ditunjukkan pada Tabel 1 dan Gambar 1.

Tabel 1 Hasil Cluster Menggunakan Algoritma K-Means

Attrib	Full Data	Cluster 0	Cluster 1	cluster 2
ute	1294	244	411	639
		19 %	32 %	49 %
Centro id	24.2566	47.7254	24.8759	14.8967
Umur		37-65	20-36	8-19



Gambar 1. Hasil Clustering

Sesuai dengan hasil yang didapat, data transaksi perpustakaan dibagi menjadi 3 cluster. Cluster 0 dengan rentang umur 37 tahun – 65 tahun dengan jumlah anggota 244 dan centroidnya berada di 47.7254. Cluster 1 dengan rentang umur 20 tahun – 36 tahun dengan jumlah anggota 411 dan centroidnya berada di 24.8759. Cluster 2 dengan rentang umur 8 tahun - 19 tahun dengan anggota berjumlah 639 dan centroidnya berada di 14.8967.

Dari hasil tabel 4.2 terdapat buku yang populer disetiap clusternya. Buku populer tersebut akan dijadikan rekomendasi kepada pengunjung yang datang. Disamping peminjam

rekomendasi juga diberikan kepada pihak perpustakaan guna untuk perawatan buku-buku yang ada disana. Buku yang dijadikan rekomendasi merupakan buku yang histori peminjamannya lebih dari sekali. Berikut ini beberapa buku rekomendasi berdasarkan cluster yang ada, ditunjukkan pada Tabel 2.

Tabel 2. Hasil Purity

Jumlah data	Data yang relevan	Purity
1294	930	0.7187017

Nilai kemurnian (purity) dari sebuah cluster berkisaran antara 0 dan 1. Clustering buruk jika nilai purity mendekati 0, dan baik jika nilai purity mendekati 1. Nilai kemurnian dari cluster yang terbentuk dalam penelitian ini adalah 0.7187017 dengan data yang 930 relevan. Dapat dikatakan hasil dari penelitian ini baik karena nilai purity (kemurnian) mendekati 1.

V. PENUTUP

Berdasarkan hasil yang didapat dari penelitian maka dapat disimpulkan sebagai berikut:

- Rekomendasi buku perpustakaan dengan kategori umur dapat dibentuk dengan menggunakan algoritma K-Means.
- Jumlah cluster terbaik yang terbentuk sebanyak 3 cluster dengan menggunakan metode elbow.
- Nilai kemurnian dari cluster yang terbentuk dalam penelitian ini adalah 0.7187017 dengan data yang 930 relevan. Dapat dikatakan hasil dari penelitian ini baik karena nilai purity (kemurnian) mendekati 1.

SARAN

Agar diperoleh hasil yang maksimal, maka beberapa saran untuk pengembangan lebih lanjut adalah sebagai berikut:

Perlu diuji coba menggunakan teknik data mining yang lain sebagai perbandingan teknik yang menghasilkan cluster lebih baik lagi. Misalnya seperti menggunakan teknik klasifikasi dan asosiasi.

DAFTAR PUSTAKA

- Berkhin, Pavel. *Survey on clustering data mining techniques*,
http://www.ee.ucr.edu/~barth/EE242/clustering_survey.pdf
- Deka Dwinavinta Candra Nugraha. 2014. "Klasterisasi Judul Buku dengan Menggunakan Metode K-Means". Yogyakarta
- Eka Merliana, I Putu. 2015. "Analisa Penentuan Jumlah Cluster terbaik pada metode K-means Clustering".
- Garcia-Molina, Hector; Ullman, JD., & Widom, Jennifer. 2002. *Database systems the complete book, International edition*. New Jersey, Prentice Hall.
- Pal, Shankar K & Mitra, Pabitra. 2004. *Pattern Recognition algorithms for data mining*. CRC Press.
- Rismawan, T. 2008. Aplikasi K-Means Untuk Pengelompokan Mahasiswa, 2008(Snati).
- Sri, Andayani. 2007. "Pembentukan Cluster dalam *Knowledge Discovery in Database* dengan Algoritma K-Means".
- Sundika, I Putu. 2013. "Pengelompokan Otomatis *Wall Facebook* dan dampak stemming terhadap kinerjanya".