Vol. 11, No. 3, May 2021, pp. 67 – 71
P-ISSN: 2089 – 0256, e-ISSN: 2598 – 3016

PENERAPAN *DATA MINING* MENGGUNAKAN ALGORITMA *K-MEANS* PENGELOMPOKAN PELANGGAN BERDASARKAN KUBIKASI AIR TERJUAL MENGGUNAKAN WEKA

Ayu Pangestu[1], dan Taufik Ridwan[2]
[1,2] Program Studi Pendidikan Sistem dan Teknologi Informasi, Universitas Pendidikan Indonesia
ayupangestu@upi.edu[1], taufikridwan@upi.edu[2]

Abstrak

Air merupakan komponen penting bagi pemenuhan kebutuhan sehari-hari yang bersumber dari sungai, bendungan, sumur buatan, PDAM, dan sebagainya. PAM Kerta Raharja merupakan salah satu perusahaan air yang dikelola oleh pemerintahan desa. Permasalahan yang muncul pada penggunaan air yaitu adanya ketidaklancaran air yang mengalir pada rumah warga yang berada di dataran tinggi. Beberapa penelitian yang berkaitan dengan pengolahan data penggunaan air pada perusahaan pengairan menggunakan *data mining* metode *clustering*. Penelitian-penelitian tersebut menggunakan algortma yang sama yaitu K-Means tetapi dengan alat yang berbeda. Pada penelitian ini akan melakukan pengelompokan penggunaan air berdasarkan kubikasi. Tujuan dari penelitian ini yaitu memberikan informasi mengenai kelompok pengguna air berdasarkan data penjualan air di PAM Kerta Raharja. Metode yang digunakna pada penelitian ini yaitu menggunakan *K-Means*. Data yang telah diolah menggunakan aplikasi Weka akan mengelompokkan data dengan kategori hemat, sedang, dan boros. Hasil penelitian ini diharapkan dapat membantu PAM Kerta Raharja untuk menindaklanjuti penggunaan pelanggan yang boros dan mengatasi kelancaran air yang mengalir bagi para warga.
Kata Kunci: *Pemakaian air, K-Means, Weka*

Abstract

Water is an important component for meeting daily needs sourced from rivers, dams, artificial wells, PDAMs, and so on. PAM Kerta Raharja is a water company managed by the village government. The problem that arises in the non-smoothness of the water flowing in the houses of residents in the highlands. Several studies related to water use data processing in irrigation companies use data mining clustering methods. These studies use the same algorithm, namely K-Means but with different tools, in this study, water use will be grouped based on cubication. The purpose of this study is to provide information about water user groups based on water sales data at PAM Kerta Raharja. The method used is this research is using K-Means. The data that has been processed using the Weka application will group the data into economical, medium, and wasteful categories. The result of this study are expected to help PAM Kerta Raharja to follow up on the use of wasteful customers and overcome the smooth running of water for residents.
Keywords: *Water consumption, K-Means, Weka*

1. Pendahuluan

Air merupakan salah satu unsur yang ada di bumi yang sangat penting untuk keberlangsungan kehidupan. Ketersediaan air di bumi berhubungan erat dengan waktu yang

p-ISSN 2089-0265
e-ISSN 2598-3016

tersedia, jumlah, lokasi, kualitas, dan ilmu lain yang dipelajari dalam sumber daya air. Saat ini kebutuhan air bersih bagi kebutuhan rumah tangga dapat berasal dari sumber air buatan seperti PDAM atau PAM yang dikelola oleh suatu lembaga. Namun, dalam penggunaan air masih terdapat pengguna yang menggunakan secara boros, sedang, dan hemat. Hal ini menjadi topik yang dapat digunakan oleh suatu lembaga yang mengelola perairan untuk menindaklanjuti terhadap pemakaian air.

PAM Kerta Raharja merupakan salah satu pengairan yang ada di Desa Kertawirama, Kecamatan Nusaherang, Kabupaten Kuningan, Jawa Barat. Pengairan ini didistribusikan kepada para warga untuk memenuhi kebutuhan rumah tangga seperti minum, mencuci pakaian, mandi, dan kebutuhan lainnya. Penggunaan air yang beragam di setiap kepala keluarga menjadi data yang setiap bulan diperoleh untuk diarsipkan. Adapun kriteria pengguna yang dapat dikelompokkan hemat, sedang, dan boros. Perbedaan penggunaan setiap warga dan jarak menjadi faktor penghambat air mengalir pada rumah warga yang berada di daerah yang lebih tinggi. Tidak jarang di waktu-waktu tertentu air tidak mengalir karena banyak warga yang tinggal di daerah dataran rendah banyak yang menggunakan air.

Strategi untuk mengambil keputusan pada pengelola air dapat menggunakan cara yang modern saat ini. Data-data yang dimiliki dapat menjadi modal utama dalam mengambil keputusan untuk perbaikan suatu perusahaan kedepannya. Proses pengolahan data yang dapat digunakan salah satunya adalah data mining. Data mining merupakan sebuah langkah yang digunakan untuk analisis data dalam jumlah besar untuk mengetahui hubungan antar data yang disajikan dalam bentuk yang mudah dipahami (Han, 2006). Jika dilihat dari berbagai sudut pandang, data mining merupakan suatu pengetahuan yang dapat dibagi menjadi karakteristik, diskriminasi, asosiasi, klasifikasi, *clustering*, *trend*, dan *outlier*.

Riset sejenis yang membahas mengenai penggunaan *Data Mining* pernah dilakukan dalam pengelompokan data pada penjualan air. Seperti yang pernah dilakukan oleh Siska (Siska, 2016). Pada riset tersebut menghasilkan pengelompokkan data berdasarkan jumlah kubikasi air yang dipakai dalam satu bulan. Objek penelitian yaitu PAM Kab. 50 dengan aplikasi yang digunakan dalam pengolahan data penelitian yaitu RapidMiner.

2. Tinjaun Pustaka

Data Mining

Data mining sebuah langkah yang digunakan untuk analisis data dalam jumlah besar untuk mengetahui hubungan antar data dan disajikan dalam bentuk yang mudah dipahami (Han, 2006). Dari hasil pemrosesan data tersebut diperoleh sebuah pola. Dari pengertian *data mining* di atas, tujuan dari data mining yaitu menemukan pengetahuan dari data yang ada. Salah satu metode yang digunakan dalam *data mining* adalah *Knowledge Discovery in Database Process* (KDD). Proses yang ada pada KDD yaitu seleksi data, pemrosesan data untuk memilih yang lebih penting, integrasi data, data mining, menghasilkan model, dan pengujian.

Jika dilihat dari berbagai sudut pandang, data mining merupakan suatu pengetahuan yang dapat dibagi menjadi karakteristik, diskriminasi, asosiasi, klasifikasi, *clustering*, *trend*, dan *outlier*. Teknik yang terdapat pada *data mining*, diantaranya *database*, *machine learning*, statistik, dan visualisasi. Data yang bisa digunakan yaitu berupa relasi, transaksi, multimedia, *web*, dan *text*.

Clustering

Clustering adalah teknik pengelompokan objek berdasarkan kesamaan antar objek. *Clustering* dapat mengelompokan data menjadi satu *cluster* dengan mempertimbangkan nilai kemiripan maksimum dan minimum (Tan, 2006). Nilai pada metode ini difokuskan untuk data bertipe numerik. Pada *clustering* terdapat istilah anomali. Anomali merupakan data yang telah di uji pada *clustering* dan tidak menemukan kelas yang cocok.

Metode clustering memiliki konsep apabila kedua objek yang memiliki nilai kemiripan yang tinggi maka akan menghasilkan nilai kesamaan yang tinggi (Irwansyah & Faisal, 2015). Pada dasarnya kualitas hasil dari data *clustering* bergantung

p-ISSN 2089-0265
e-ISSN 2598-3016

pada metode yang digunakan. Ada banyak cara untuk melakukan perhitungan *clustering* salah satunya *Euclidean Distance*. *Euclidean distance* merupakan rumus yang digunakan dalam menghitung jarak pada dua buah objek dengan menggunakan nilai dari masing-masing objek. Berikut adalah rumus perhitungan dan pengukuran pada *Euclidean distance*:

Perhitungan

$$Distance\ (a,b)\ = (\sum_{k}^{n} \mu k|ak - bk|r)\ 1/r \quad (1)$$

Pengukuran

$$j\ (v_1,\ v_2) = \sqrt{\sum_{k=1}^{n}(v_1(k) - v_2(k))^{\ 2}} \quad (2)$$

3. Metode Penelitian

Penelitian ini menggunakan algoritma *K-Means* untuk mengelompokan pengguna hemat, sedang, dan boros. Penentuan kelompok pengguna air dapat menjadi acuan sebagai pengambilan keputusan PAM Kerta Raharja dalam menangani pengguna yang boros.

K-Means

K-Means adalah metode yang ada pada *data mining* dengan proses awal mengambil sebagian komponen populasi untuk dijadikan titik pusat di *cluster* awal. Penetapan kelompok dalam satu *cluster* dapat dilakukan dengan menghitung jarak setiap objek dengan titik pusat. Algoritma *clustering K-Means* dapat membagi data berdasarkan jarak antar data pada kelompok yang telah ditetapkan. Algoritma ini bergantung pada fungsi untuk mengukur data yang mempunyai ciri khas sama. Jarak itu sendiri dihitung menggunakan fungsi *Euclidean*. Kemudian data dimasukkan dalam kelompok yang mempunyai jarak terdekat (Wahyudi dkk, 2020).

Algoritma ini memiliki keunggulan, yaitu dapat memproses data dengan jumlah yang banyak dengan waktu yang cepat. Adapun aturan yang ada di dalam algoritma *K-Means*, diantaranya

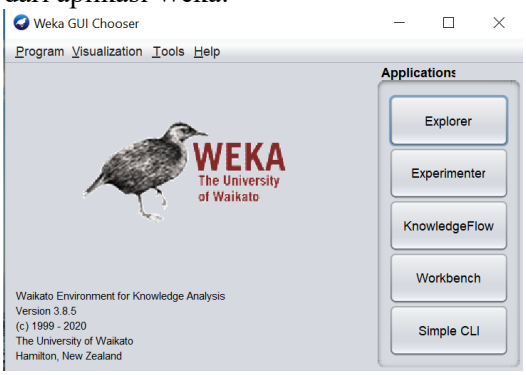
1) Jumlah *cluster* perlu dimasukan.
2) Hanya memiliki atribut bertipe numerik.

Berikut adalah langkah-langkah yang dapat digunakan pada algoritma *K-Means*:

1) Tentukan jumlah *cluster*.
2) Inisialisasi awal dan pusat cluster dilakukan secara acak.
3) Terdapat jarak antara objek dengan outsalt cluster. Pada tahap ini jarak dihitung dengan menentukan kemiripan dan ketidakmiripan data dengan metode jarak *Euclidean distance*.
4) Hitung pusat *cluster* yang baru dengan keanggotaan yang baru dengan cara menghitung rata-rata objek pada cluster.
5) hitung kembali jarak tiap objek dengan pusat *cluster* yang baru, hingga *cluster* tidak berubah.

Perangkat lunak yang digunakan dalam penelitian ini adalah Weka. Weka merupakan salah satu sistem yang digunakan untuk melakukan mining data. Aplikasi ini berlisensi *GNU General Public License* sehingga dapat digunakan secara gratis. Weka menggunakan *framework Java* sehingga dapat digunakan di berbagai macam sistem operasi. Penggunaan aplikasi ini bertujuan untuk memudahkan pengguna dalam memproses data dari mulai pemrosesan awal hingga pemodelan data (Adinugroho & Sari, 2018). Pada halaman pertama di Weka terdapat menu aplikasi berupa *Explorer*, *Experimenter*, *KnowledgeFlow*, *Workbench*, dan *Simple CU*.

Berikut adalah tampilan halaman awal dari aplikasi Weka:



Gambar 1. Tampilan Awal Aplikasi Weka

4. Hasil dan Pembahasan

Data Uji

p-ISSN 2089-0265
e-ISSN 2598-3016

Data uji yang digunakan terdiri dari beberapa kolom yang masing-masing memiliki komponen sebagai berikut:

1) Memiliki 4 atribut yaitu nama, jumlah pemakaian awal, jumlah pemakian akhir, dan pemakaian sekarang.
2) Jumlah data pengguna sebanyak 70 warga.

Proses Algoritma K-Means

Adapun proses dari algoritma ini adalah sebagai berikut:

1) Menentukan jumlah *cluster*. Penenutuan jumlah *cluster* dibuat menjadi 3 *cluster* yaitu hemat, sedang, dan boros. Atribut yang digunakan sebanyak 4 atribut yaitu nama, jumlah pemakaian air awal, jumlah pemakaian air akhir, dan pemakaian air sekaranng.
2) Menentukan titik pusat pada masing-masing *cluster*.

TABEL 1
CENTROID

Centroid	Quantity
C1	46,6
C2	13,6
C3	24,5

3) Menghitung jarak setiap data ke pusat *cluster*. Penentuan nilai pusat dari setiap *cluster* berguna untuk menjadi acuan dalam melakukan perhitungan pada setiap table data uji. Contohnya penentuan jarak objek ke titik pusat dilakukan pada nama pengguna Dian yang menggunakan air sebanyak 10 M^3. Perhitungan ini mengacu pada rumus *Euclidean* yang telah disederhanakan menjadi:

Jarak = *centroid* y – pemakaian pengguna (3)

Sehingga diperoleh
Jarak 0 = (46,6 - 10) = 36,6
Jarak 1 = (13,6 - 10) = 3,6
Jarak 2 = (24,5 - 10) = 14,5

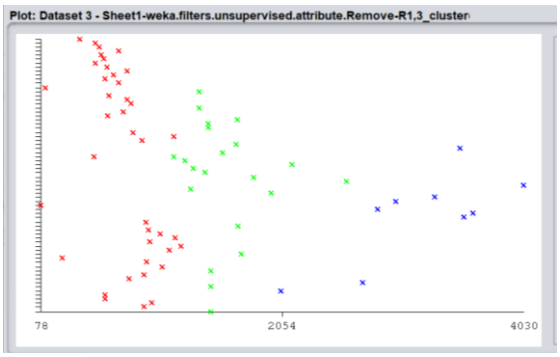
Dari hasil perhitungan didapatkan hasil jarak 1 = 3,6 yang memiliki nilai mendekati *cluster* 1, sehingga dapat diketahui bahwa pengguna bernama Dian termasuk ke dalam kelompok pengguna yang sedang.

4) Pengelompokan objek pada data dilihat berdasarkan data minimum yang diperoleh dari perhitungan langkah 2. Hasil perhitungan jarak akan digunakan untuk menentukan *clustering*.

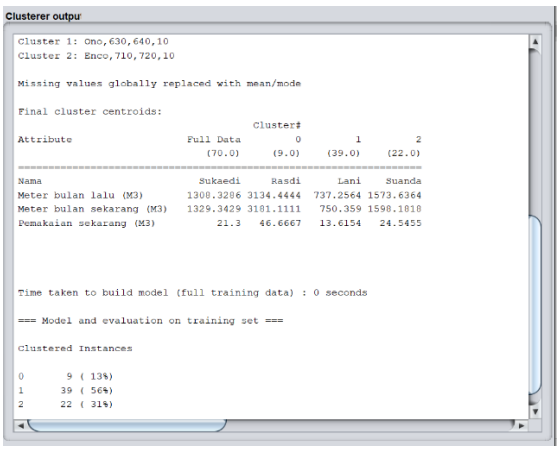
Pengujian Aplikasi Weka

Pengujian data dengan aplikasi Weka menghasilkan data berupa

1) Nilai *cluster centroids* dan cluster instances.
2) Grafik *clustering* posisi setiap pengguna air pada *cluster* masing-masing.



Gambar 2. Plot grafik *clustering* pada Weka



Gambar 2. Hasil cluster centroid dan cluster instances pada Weka

Hasil Pengujian

Dari data yang telah diuji, diperoleh 3 kelompok dengan hasil sebagai berikut:

1) Pelanggan dengan pemakaian air = 46,6 pada cluster 0, sebanyak 9 orang dari 70 pelanggan (12%).

p-ISSN 2089-0265
e-ISSN 2598-3016

2) Pelanggan dengan pemakaian air = 13,6 pada cluster 1, sebanyak 39 orang dari 70 pelanggan (55%).
3) Pelanggan dengan pemakaian air = 24,5 pada cluster 2, sebanyak 22 orang dari 70 pelanggan (31%).

Maka cluster 0 dengan pemakaian air paling tinggi dapat ditindak lanjuti mengenai penggunaan air yang boros.

5. Kesimpulan

Dari hasil perhitungan menggunakan algoritma *K-Means* didapatkan nilai *centroid* 0 (46,6), *centroid* 1 (13,6), dan *centroid* 2 (25,4). Kelompok *cluster* 0 merupakan *cluster* boros yaitu sebanyak 9 orang, cluster 1 meupakan *cluster* sedang, dan *cluster* 2 merupakan pelanggan yang hemat.

Daftar Pustaka

Adinugroho, S., Sari, Y.A. (2018). *Implementasi Data Mining Menggunakan Weka*. Malang: UB Press.

Irwansyah, E. & Faisal, M. (2015). *Teori dan Aplikasi*. Yogyakarta: Deepublish.

Prasetyowati, W. (2017). *Data Mining Pengelompokan Data Untuk Informasi dan Evaluasi*. Makassar: Duta Media Publishing.

Triatmadja, R. (2019). *Teknik Penyediaan Air Minum Perpipaan*, Yogyakarta: Gajah Mada University Press.

Wahyudi, M., dkk. (2020). *Data Mining Penerapan Algoritma K-Means Clustering dan K-Medoids Clustering*. Medan: Yayasan Kita Menulis.

Werdiningsih, I., Nuqoba, B. & Muhammadun. (2020). *Data Mining Menggunakan Android, Weka, dan SPSS*. Surabaya: Airlangga University Press.

Asroni., Andrian, R. 2015. *Penerapan Metode K-Means Untuk Clustering Mahasiswa Berdasarkan Nilai Akademik Dengan Weka Interface Strudi Kasus Pada Jurusan Teknik Informatika UMM Magelang*. Jurnal Ilmiah Semesta Teknika, 18*(1), 76-82.*

Siska, S.T. 2016. Analisa dan Penerapan Data Mining Untuk Menentukan Kubikasi Air Terjual Berdasarkan Pengelompokan Pelanggan Menggunakan Algoritma K-Means Clustering. Jurnal Ekonomi dan Studi Pembangunan, 1(15), 78-96. 299-304.

Sulastri, H. & Gufroni., A.I. 2017. *Penerapan Data Mining Dalam Pengelompokan Penderita Thalassaemia*. Jurnal Nasional Teknologi dan Informasi dan Sistem Informasi, 3(2),

Fahlevi, A (2021, September 30). Retrived from Andi Fahlevi Daring: https://sis.binus.ac.id/2021/09/30/proses-data-mining-kdd/