

## SISTEM APLIKASI BERBASIS OPTIMASI METODE *ELBOW* UNTUK PENENTUAN *CLUSTERING* PELANGGAN

Elly Muningsih<sup>1</sup>, Sri Kiswati<sup>2</sup>  
AMIK BSI Yogyakarta

**Abstract:** *Customer is a very important asset for the company. Having customers who are loyal to the company is an absolute and important for the progress of the company. This study aims to help companies, especially in the online shop to create a better customer management by identifying and grouping customers into several clusters or groups to know the characteristics of their loyalty to the company. The method used in this research is K-Means method which is one of the best and most popular method in clustering algorithm. To overcome the weakness of the K-Means method in determining the number of clusters, we use the Elbow method where this method gets the comparison of the number of clusters added by calculating the SSE (Sum of Square Error) of each cluster value. This research starts from collecting the necessary data and will be processed. From total transaction data 478 then done cleaning of data and result 73 data. Then the data processed with RapidMiner software from Cluster 2 up to 10 to search the data center of each cluster. From the calculated SSE value found that the best number of clusters is 3. The end result of the research is a Visual Basic based application program that is expected to provide ease in grouping or clustering customers. Software development method using Waterfall method.*

**Keywords:** *customer, Elbow, value Sum of Square Error, Waterfall*

### I. Pendahuluan

Perkembangan teknologi internet yang sangat pesat saat ini dapat dilihat terutama dalam dunia bisnis, yaitu munculnya *electronic commerce (e-commerce)* (Ustadiyanto, 2001). *E-Commerce* dapat memberikan manfaat yang luar biasa bagi perusahaan sehingga akan menjadi keunggulan kompetitif bagi perusahaan yang mengaplikasikannya. (Rudy, Wahyudiarti, Megaputri, & Wihardini, 2008). Lembaga atau perusahaan yang mengaplikasikan *E-Commerce* dalam kegiatan penjualan atau perdagangannya biasa dikenal dengan nama Toko Online atau *Online Shop*. Sebuah *online shop* melakukan transaksi tanpa berinteraksi langsung dengan konsumennya, karena semua dilakukan secara *online* menggunakan *Internet*. Di tengah persaingan bisnis dewasa ini, *online shop* harus menyadari bahwa hubungan dengan konsumen sangat penting karena menunjang perkembangan dan kelangsungan *online shop*. Dengan meningkatkan pemahaman *online shop* akan kebutuhan konsumen secara individu, pihak manajemen *online shop* harus bisa mengenali konsumen terbaiknya sehingga dapat mempertahankan loyalitas konsumennya terhadap perusahaan.

Dalam pengelolaan pelanggannya satu hal yang penting adalah bagaimana suatu perusahaan dapat mempertahankan pelanggan yang dimilikinya. Perpindahan konsumen menjadi hal yang wajar karena persaingan *online shop* yang semakin ketat. Untuk mencegah terjadinya perpindahan pelanggan ini, maka perlu diketahui kelompok pelanggan yang loyal sehingga *online shop* bisa melindungi pelanggan loyal tersebut dengan cara memberikan pelayanan prima, memberikan potongan harga atau hadiah. Permasalahan yang terjadi saat ini disebabkan karena *online shop* mengalami kesulitan dalam menentukan pelanggan yang loyal menjadi beberapa kelompok berdasarkan transaksi

penjualan. Penentuan yang dilakukan selama ini adalah secara manual dan hasil yang di dapat tidak akurat. Untuk dapat mengatasi permasalahan yang terjadi, maka *online shop* membutuhkan suatu metode untuk mengidentifikasi dan mengelompokkan pelanggannya menjadi beberapa *cluster* atau kelompok.

Penentuan kelompok pelanggan yang kurang akurat karena harus berdasarkan pengetahuan dari jumlah data transaksi penjualan yang besar (Setiawan, 2011). Karena hal itu untuk mendapatkan pengetahuan tersebut maka diperlukan suatu proses pengolahan data historis transaksi besar diperlukan suatu teknik *data mining*. Teknik data mining yang akan digunakan pada penelitian ini adalah metode *K-Means* dengan optimasi penentuan jumlah cluster menggunakan metode *Elbow* dan software pendukung yang digunakan untuk pengolahan data adalah *RapidMiner*. Penelitian diharapkan dapat menghasilkan sebuah program aplikasi yang dibangun menggunakan bahasa pemrograman Visual Basic dengan metode pengembangan perangkat lunak metode *Waterfall* untuk memudahkan pengguna dalam menentukan *cluster* atau kelompok pelanggan loyal dengan lebih cepat sehingga dapat memaksimalkan strategi pemasaran yang lebih baik yang pada akhirnya akan meningkatkan penjualan produk. Hal ini tentunya akan memberikan efek positif bagi perkembangan perusahaan.

### 1. Data Mining

*Data mining* merupakan penyelesaian masalah dengan menganalisa data yang ada pada database, dimana data tersimpan secara elektronik dan pencariannya dilakukan otomatis seperti pada komputer. *Data mining* merupakan sebuah inti dari proses *Knowledge Discovery in Database (KDD)*, meliputi dugaan algoritma yang mengeksplor data,

membangun model dan menemukan pola yang belum diketahui. KDD bersifat otomatis, dapat didefinisikan sebagai pengorganisasian proses untuk pengidentifikasian yang benar, berguna dan penemuan pola dari kumpulan data yang besar dan kompleks. (Maimon & Rokach, 2010). Tahapan pada proses KDD pada database digambarkan sebagai berikut (Maimon & Rokach, 2010):

Ada banyak metode atau fungsi *data mining* yang bisa digunakan untuk menemukan, menggali dan menambang pengetahuan, namun menurut (Larose, 2005), (Moertini, 2002) dan (Susanto & Suryadi, 2010) ada enam fungsi utama *data mining*, yaitu: *Description* (deskripsi), *Estimation* (estimasi), *Prediction* (prediksi), *Classification* (klasifikasi), *Clustering* (pengelompokan) dan *Association* (asosiasi).

## 2. Metode K-Means

Metode *K-Means* adalah salah satu metode dalam fungsi *clustering* atau pengelompokan. Menurut (Larose, 2005) *clustering* mengacu pada pengelompokan data, observasi atau kasus berdasar kemiripan objek yang diteliti. Sebuah *cluster* adalah suatu kumpulan data yang mirip dengan lainnya atau ketidakmiripan data pada kelompok lain. Sedangkan (Xu & Wunsch II, 2009) menjelaskan bahwa *clustering* adalah membagi objek data (bentuk, entitas, contoh, ketaatan, unit) ke dalam beberapa jumlah kelompok (grup, bagian atau kategori). Tujuan proses *clustering* adalah meminimalkan terjadinya *objective function* yang diset dalam proses *clustering*, yang pada umumnya digunakan untuk meminimalisasikan variasi dalam suatu *cluster* dan memaksimalkan variasi antar *cluster* (Agusta, 2007).

Ada dua jenis data *clustering* yang sering digunakan dalam proses pengelompokan data yaitu *hierarchical* (hirarki) dan *non-hierarchical* (non hirarki) data *clustering*. *K-Means* adalah salah satu metode data clustering non hirarki yang mempartisi data ke dalam bentuk satu atau lebih cluster/kelompok, sehingga data yang memiliki karakteristik yang sama dikelompokkan dalam satu cluster yang sama dan data yang memiliki karakteristik berbeda dikelompokkan ke dalam kelompok lain (Agusta, 2007).

Algoritma *K-Means* merupakan sebuah metode sederhana untuk membagi suatu kumpulan data dalam suatu angka spesifik dari *cluster*, yaitu *k*. Algoritma atau metode *K-Means* ditemukan oleh beberapa peneliti dengan disiplin ilmu berbeda-beda yaitu oleh Lloyd (1957, 1982), Forgery (1965), Friedman dan Rubin (1967), dan terakhir adalah McQueen (1967) dalam (Wu, et al., 2008). Disebutkan bahwa metode *K-Means* adalah metode yang cepat dan efisien yang dapat digunakan dalam clustering data (Larose, 2005). Algoritma dasar *clustering* data menggunakan metode *K-Means* dapat dilakukan dengan cara (Agusta, 2007), (Suprihatin, 2011):

1. Tentukan jumlah *cluster*

2. Inialisai *k* sebagai pusat *cluster* (beri nilai-nilai random)
3. Alokasikan setiap data atau obyek ke *cluster* terdekat. Kedekatan dua obyek ditentukan berdasarkan jarak antar kedua obyek tersebut. Jarak paling dekat antara satu data dengan satu *cluster* tertentu akan menentukan suatu data masuk ke dalam *cluster* yang mana.
4. Hitung kembali pusat *cluster* dengan anggota *cluster* yang sekarang. Pusat *cluster* adalah rata-rata semua data atau obyek dalam *cluster*.
5. Tugaskan lagi setiap obyek memakai pusat *cluster* yang baru. Jika pusat *cluster* sudah tidak berubah lagi, maka proses peng-*cluster-an* selesai.
6. Kembali ke langkah 3 sampai pusat *cluster* tidak berubah lagi

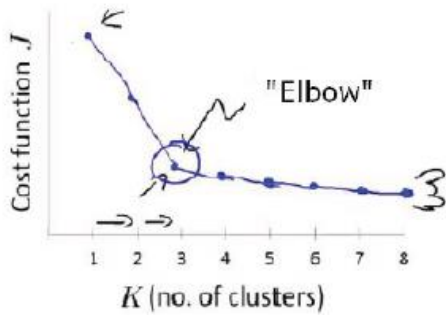
## 3. Metode Elbow

Menurut Madhulatha (2012) dalam Merliana, Ernawati & Santoso (2015) disebutkan bahwa 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 *Elbow* ini memberikan ide atau 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 pembanding 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. Rumus SSE pada *K-Means* adalah (Irwanto, et al 2012):

$$SSE = \sum_{K=1}^K \sum_{x_j \in S_K} \|x_j - c_k\|_2^2 \quad \dots\dots\dots 1)$$

Setelah dilihat akan ada beberapa nilai *K* yang mengalami penurunan paling besar dan selanjutnya hasil dari nilai *K* akan turun secara perlahan-lahan sampai hasil dari nilai *K* tersebut stabil. Misalnya nilai *cluster* *K*=2 ke *K*=3, kemudian dari *K*=3 ke *K*=4, terlihat penurunan drastis membentuk siku pada titik *K*=3 maka nilai *cluster* *k* yang ideal adalah *K*=3.



Gambar 1. Grafik Nilai K

Algoritma Metode *Elbow* dalam menentukan nilai K pada *K-Means*

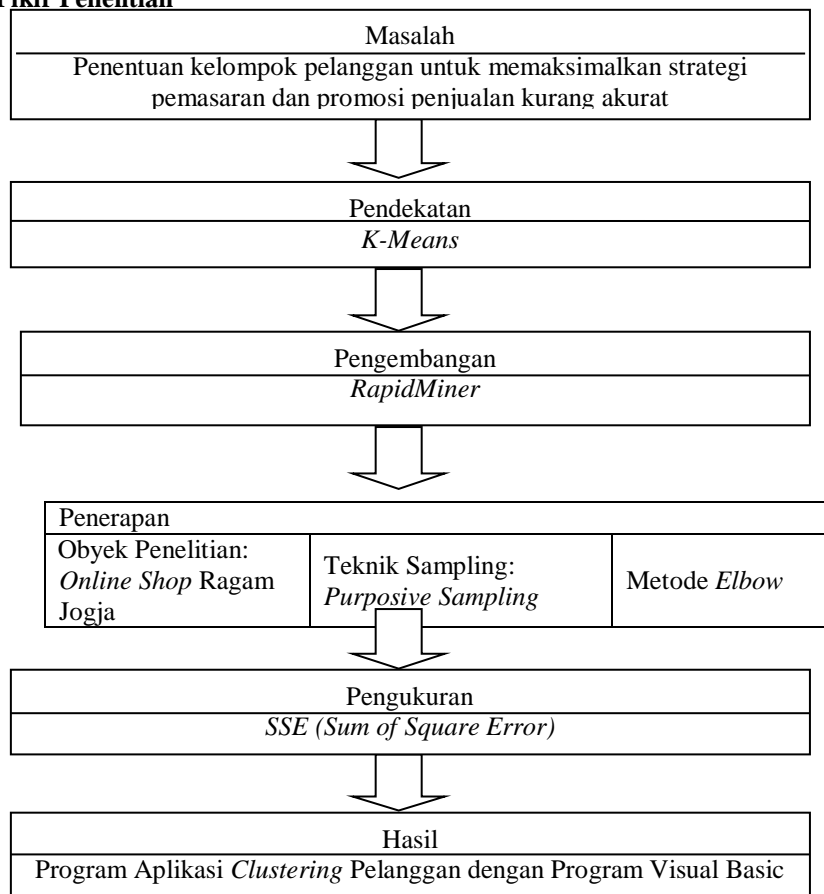
1. Mulai
2. Inisialisasi awal nilai K
3. Naikkan nilai K
4. Hitung hasil *sum of square error* dari tiap nilai K

5. Melihat hasil *sum of square error* dari nilai K yang turun secara drastis
6. Tetapkan nilai K yang beberbentuk siku
7. Selesai

**4. Framework RapidMiner**

*RapidMiner* dibuat oleh Dr. Markus Hofmann dari Institute of Technology Blanchardstown dan Ralf Klittenberg dari rapid-i.com (Firmansyah, 2011). *RapidMiner* merupakan sebuah *framework* berbasis *Graphical User Interface* (GUI) dan bersifat *open source* untuk *data mining* yang menawarkan banyak operator yang bisa diisikan pada suatu proses, dimana proses yang dimaksud adalah menganalisa data yang diterima pada awal proses (Jungermann, 2011). *RapidMiner* menyediakan banyak metode mulai dari klasifikasi, klusterung, asosiasi dan lain-lain.

**5. Kerangka Pikir Penelitian**



Gambar 2. Kerangka Pikir Penelitian

**II. Metode Penelitian**

Jenis penelitian yang digunakan dalam penelitian ini adalah jenis penelitian kuantitatif, yaitu melakukan analisis berdasarkan pada data-data numerik (angka) dari transaksi penjualan. Data penelitian diambil dari tempat penelitian yaitu salah satu *online shop* yang berdomisili di Jogja.

**1. Populasi dan Teknik Pengambilan Sampel**

Suatu populasi adalah kumpulan dari individu-individu dengan kualitas serta ciri-ciri yang sudah ditetapkan, sedangkan sampel adalah bagian dari populasi (Nazir, 2014: 240). Populasi target dalam penelitian ini adalah data historis transaksi penjualan dari bulan Januari s/d Desember 2015 dengan total 478 transaksi. Mengingat jenis

sampel yang diambil tidak dipilih secara acak dan unsur populasi yang terpilih menjadi sampel disebabkan karena sudah direncanakan oleh peneliti, teknik pengambilan sampel menggunakan teknik *Purposive Sampling*, dimana dalam penarikan sampel peneliti mengambil data yang digunakan adalah data pelanggan yang bertransaksi minimal 2 kali dalam rentang waktu data yang digunakan dalam penelitian.

**2. Metode Pengumpulan Data**

Berdasarkan sumber data yang digunakan pada penelitian ini, maka metode pengumpulan data yang digunakan adalah sebagai Wawancara, Dokumentasi dan Studi Pustaka.

**3. Metode Analisis**

Ada beberapa tahap dalam melakukan eksperimen ini. Peneliti menggunakan metode KDD yang terdiri dari sembilan langkah (Maimon & Rokach, 2010) yaitu:

- a. Tahap *Domain Understanding* dan *KDD Goals*  
Berdasarkan pengamatan, selama ini penentuan pelanggan loyal dilakukan secara manual sehingga tidak akurat dan mengakibatkan strategi pemasaran yang dijalankan tidak maksimal
- b. Tahap Selection dan Addition  
Data historis diambil dari data transaksi penjualan dari bulan Januari – Desember 2015. Penelitian ini difokuskan pada pelanggan yang melakukan transaksi *online* saja dimana data yang diolah adalah pelanggan yang melakukan transaksi minimal 2 kali.
- c. Tahap *Preprocessing* dan *Data Cleansing*.  
*Preprocessing* data pada tahap ini adalah mengambil data transaksi yang dengan cara *Purposive Sampling* setelah terlebih dahulu melakukan pembersihan data yang dianggap tidak mewakili. Dari 478 data, diambil 73 data yang digunakan dalam penelitian.
- d. Tahap Transformation.  
Proses transformasi data dengan cara mengganti atribut yang ada dengan atribut yang akan digunakan dalam pengolahan data yaitu Kode Cust, Jumlah Transaksi, Jumlah Item dan Total Bayar (dalam ribuan rupiah).

Untuk mengetahui wilayah atau daerah pelanggan dengan tingkat loyalitasnya masing-masing, maka wilayah pengiriman peneliti buat dalam Tabel 1 berikut:

Tabel 1. Wilayah dan Kota Pengiriman

No	Wilayah	Kota Pengiriman
1	Aceh	Banda Aceh
2	Sumatera Utara	Medan
3	Sumatera Selatan	Palembang
4	Riau	Pekanbaru
5	Bengkulu	Bengkulu
.....	.....	.....

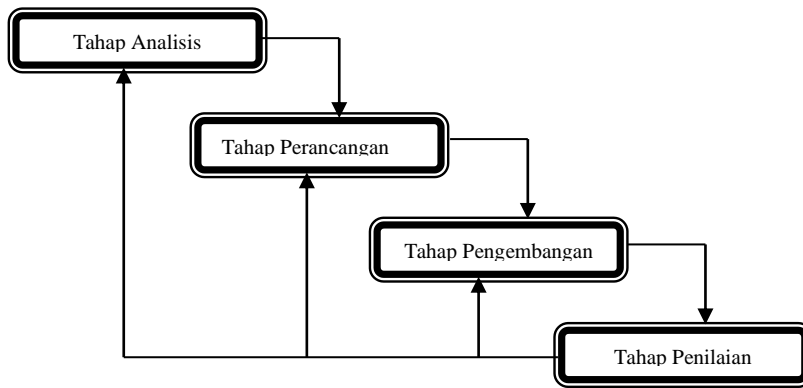
Tipe Data Mining yang digunakan dalam penelitian ini adalah *clustering* untuk pengelompokan pelanggan loyal.

- f. Tahap Data Mining memilih algoritma data mining.  
Metode yang digunakan dalam penelitian adalah metode *K-Means* yang mengkluster pelanggan menjadi beberapa kelompok untuk mengetahui karakteristik pelanggan. Dalam penelitian ini untuk optimasi *cluster* yang akan dihitung adalah cluster 2 s/d 10
- g. Tahap *Data Mining* dalam penggunaan algoritma *data mining*.  
Pada tahap ini dilakukan implementasi dari algoritma *data mining* yang telah ditentukan pada tahap sebelumnya. Untuk mengetahui data *centroid* atau pusat *cluster* pada pengolahan metode *K-Means* digunakan *software RapidMiner*. Pada Tabel dibawah ini ditampilkan data pusat *cluster* untuk *cluster 2*
- h. Tahap *Evaluation*.  
Pada tahap ini dilakukan evaluasi terhadap kelompok atau *cluster* yang ada dengan cara melakukan optimasi jumlah *cluster*. Dari tahap ini akan diketahui berapa jumlah *cluster* terbaik dengan menggunakan metode *Elbow* yang mendapatkan perbandingan jumlah *cluster* dengan cara menghitung SSE (*Sum of Square Error*) dari masing-masing nilai *cluster*.
- i. Tahap *Discovered Knowledge ( Visualization dan Integration)*  
Penggunaan pengetahuan yang diperoleh. Pada tahap ini, penerapan metode *K-Means* dengan optimasi jumlah *cluster* metode *Elbow* yang mendapatkan perbandingan jumlah *cluster* dengan cara menghitung SSE (*Sum of Square Error*) dari masing-masing nilai *cluster*.

**4. Metode Pengembangan Perangkat Lunak**

Model pengembangan perangkat lunak dalam penelitian ini menggunakan model proses *Waterfall*, yang digambarkan di Gambar 3.

- e. Tahap Data Mining memilih tipe data mining yang cocok.



Gambar 3. Metode Wartefall

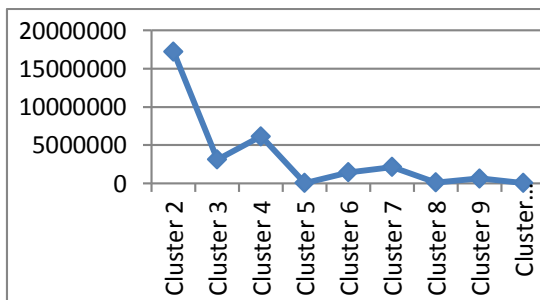
III. Hasil dan Pembahasan

1. Hasil Clustering

Dari pengolahan data yang sudah dilakukan, dengan menghitung nilai SSE cluster 2 s/d 10 dihasilkan perbandingan nilai SSE tiap cluster yang ditampilkan pada tabel 2 di bawah ini:

Tabel 2. Nilai SSE cluster 2 s/d 10

Cluster	Hasil Sum Square Error	Selisih
K2	17230921,28	17230921,28
K3	14106361,99	3124559,296
K4	7942555,06	6163806,926
K5	7904037,699	38517,36116
K6	6471172,149	1432865,55
K7	4323339,378	2147832,771
K8	4196989,481	126349,8968
K9	3566702,705	630286,7761
K10	3524587,387	42115,31824



Gambar 4. Grafik Sum of Square Error

Dari nilai SSE yang dihitung didapatkan bahwa jumlah cluster terbaik adalah 3, dimana jumlah anggota cluster 1 adalah 47 pelanggan dengan nilai rata-rata jumlah transaksi 2 – 3 kali, jumlah item yang dibeli 7 dan jumlah bayar yang dilakukan 416.000. Cluster 2 ada 18 pelanggan dengan nilai rata-rata jumlah transaksi 5, jumlah item yang dibeli 21 dan jumlah bayar yang dilakukan 1.204.000. Dan cluster 3 memiliki 8 pelanggan dengan nilai rata-rata jumlah transaksi 5 - 6 kali; jumlah item yang dibeli sekitar 48 dan jumlah bayar yang dilakukan 2.121.000.

2. Halaman Antarmuka Sistem

Beberapa tampilan utama pada aplikasi program pengelompokan atau clustering pelanggan ditampilkan dibawah ini.

a. Tampilan Halaman Login

Halaman ini berfungsi untuk menjadi pintu masuk ke program aplikasi jika username dan password yang dimasukkan diterima (benar).



Gambar 5. Tampilan Halaman Login

b. Tampilan Halaman Input Pelanggan

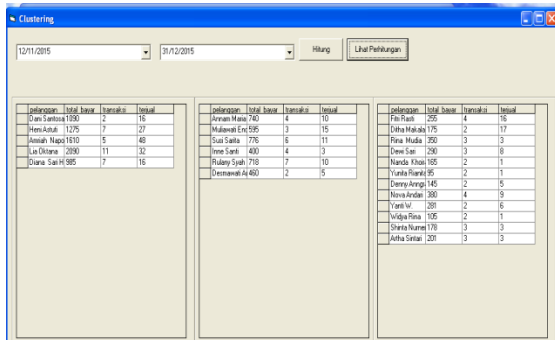
Halaman ini berfungsi untuk menambahkan data pelanggan.



Gambar 6. Tampilan Halaman Input Pelanggan

c. Tampilan Halaman Proses Pengelompokan Pelanggan

Halaman ini berfungsi untuk menampilkan hasil pengelompokan pelanggan yang dibedakan menjadi 3 kelompok.



Gambar 7. Tampilan Halaman Proses Pengelompokan Pelanggan

3. Pengujian Unit

Pengujian terhadap program yang dibuat menggunakan *Blacbox Testing* yang fokus pada proses masukan dan keluaran program

a. Pengujian terhadap form Login

Tabel 3. Pengujian Form Login

No	Skenario	Test Case	Hasil yang diharapkan	Hasil Pengujian	Kesimpulan
1.	Username otomatis terisi admin, jika password tidak diisi kemudian klik Login	Username: (admin) Password: (kosong)	Sistem akan menolak akses user dan menampilkan “Username atau Password salah”	Sesuai Harapan	Valid
2.	Mengetikkan Username dan atau Password yang salah	Username: admin (benar) Password: 123 (salah)	Sistem akan menolak akses user dan menampilkan “Username atau Password salah”	Sesuai Harapan	Valid
3.	Mengetikkan Username dan Password yang benar	Username: admin (benar) Password: admin (benar)	Sistem akan menerima akses Login dan langsung menampilkan Menu Utama	Sesuai Harapan	Valid

2. Pengujian terhadap form Proses Pengelompokan Pelanggan

Tabel 4. Pengujian Form Proses Pengelompokan Pelanggan

No	Skenario	Test Case	Hasil yang diharapkan	Hasil Pengujian	Kesimpulan
1.	Pilih tanggal awal kemudian pilih tanggal akhir kemudian klik Hitung	Tanggal: 1/1/2015 s/d 31/12/2015	Sistem akan menampilkan tiga kelompok pelanggan	Sesuai Harapan	Valid
2.	Klik tombol Lihat Perhitungan	Klik Tombol	Sistem menampilkan hasil perhitungan secara manual	Sesuai Harapan	Valid

IV. Kesimpulan

Berdasarkan hasil yang dicapai terkait dengan penelitian clustering atau pengelompokan pelanggan maka dapat disimpulkan bahwa:

1. Berdasarkan pengolahan data yang dilakukan menggunakan metode clustering K-Means dan optimasi metode Elbow dengan mengetahui nilai SSE (*Sum of Square Error*) dihasilkan 3 kelompok pelanggan yang memiliki nilai maksimal atau terbaik.
2. Dengan adanya program aplikasi, penelitian ini bisa dimanfaatkan juga untuk pelaku bisnis lainnya untuk pengelompokan pelanggannya secara lebih cepat dan akurat.

REFERENSI

[1] Agusta, Y. (2007). K-Means - Penerapan, Permasalahan dan Metode Terkait. *Jurnal Sistem dan Informatika*, 3, 47-60.

[2] Firmansyah. (2011). *Penerapan Algoritma Klasifikasi C4.5 Untuk Penentuan Kelayakan Pemberian Kredit Koperasi*. Jakarta: STMIK Nusa Mandiri.

[3] Jungermann, F. (2011). *Documentation of the Information Extraction Plugin for RapidMiner*. Dortmund, Jerman: Department of Computer Science - Artificial Intelligent Group.

[4] Larose, D. T. (2005). *Discovering Knowledge in Data: An Introduction to Data Mining*. Hoboken, New Jersey: John Wiley & Sons, Inc.

- [5] Maimon, O., & Rokach, L. (2010). *Data Mining and Knowledge Discovery Handbook*. London: Springer Science+Business Media.
- [6] Merliana., N.P.E., Ernawatii & Santoso., A.J. (2015). Analisa Penentuan Jumlah Cluster Terbaik pada Metode K-Means Clustering, ISBN: 978-979-3649-81-8.
- [7] Rudy, Wahyudiarti, R., Megaputri, V., & Wihardini, R. (2008). Analisis Dan Perancangan E-Commerce (Studi Kasus: Roemah Soetera Batik Dan Bordir). *Seminar Nasional Informatika 2008 (semnasIF 2008) UPN "Veteran" Yogyakarta*. Yogyakarta.
- [8] Setiawan, F. H. (2011). *Penerapan Fuzzy C-Means Dan Apriori Untuk Rekomendasi Promosi Produk Berdasarkan Segmentasi Konsumen*. Semarang.
- [9] Suprihatin. (2011, Februari). Klatering K-Means untuk Penentuan Nilai Ujian. *JUSI Vol. I, No. I*, 53-62.
- [10] Susanto, S., & Suryadi, D. (2010). *Pengantar Data Mining: Menggali Pengetahuan dari Bongkahan Data*. Yogyakarta: Penerbit ANDI.
- [11] Ustadiyanto, R. (2001). *Framework e-Commerce*. Yogyakarta: Penerbit Andi.
- [12] Widiarina., & Wahono, R.S. (2015). Algoritma Cluster Dinamik untuk Optimasi Cluster pada Algoritma K-Means dalam Pemetaan Nasabah Potensial. *Journal of Intelligent System Vol. 1, No. 1*, 32-35.
- [13] Witten, I. H., Frank, E., & Hall, M. A. (2011). *Data mining: Practical Machine Learning Tools and Techniques 3rd Edition*. Elseiver.
- [14] Xu, R., & Wunsch II, D. C. (2009). *Clustering*. Kanada: IEEE Press.

Halaman Sengaja Dikосongkan