

**IMPLEMENTASI K-MEANS UNTUK SEGMENTASI DATA
PELANGGAN DENGAN MODEL LRFM**

Skripsi



oleh

NOVIA TONAPA BASSY

22104995

PROGRAM STUDI INFORMATIKA FAKULTAS TEKNOLOGI INFORMASI
UNIVERSITAS KRISTEN DUTA WACANA

2018

IMPLEMENTASI K-MEANS UNTUK SEGMENTASI DATA PELANGGAN DENGAN MODEL LRFM

Skripsi



Diajukan kepada Program Studi Informatika Fakultas Teknologi Informasi
Universitas Kristen Duta Wacana
Sebagai Salah Satu Syarat dalam Memperoleh Gelar
Sarjana Komputer

Disusun oleh

NOVIA TONAPA BASSY

22104995

**PROGRAM STUDI INFORMATIKA FAKULTAS TEKNOLOGI INFORMASI
UNIVERSITAS KRISTEN DUTA WACANA**

2018

PERNYATAAN KEASLIAN SKRIPSI

Saya menyatakan dengan sesungguhnya bahwa skripsi dengan judul:

IMPLEMENTASI K-MEANS UNTUK SEGMENTASI DATA PELANGGAN DENGAN MODEL LRFM

yang saya kerjakan untuk melengkapi sebagian persyaratan menjadi Sarjana Komputer pada pendidikan Sarjana Program Studi Informatika Fakultas Teknologi Informasi Universitas Kristen Duta Wacana, bukan merupakan tiruan atau duplikasi dari skripsi kesarjanaan di lingkungan Universitas Kristen Duta Wacana maupun di Perguruan Tinggi atau instansi manapun, kecuali bagian yang sumber informasinya dicantumkan sebagaimana mestinya.

Jika dikemudian hari didapati bahwa hasil skripsi ini adalah hasil plagiasi atau tiruan dari skripsi lain, saya bersedia dikenai sanksi yakni pencabutan gelar kesarjanaan saya.

Yogyakarta, 9 Januari 2018



NOVIA TONAPA BASSY
22104995

HALAMAN PERSETUJUAN

Judul Skripsi : IMPLEMENTASI K-MEANS UNTUK
SEGMENTASI DATA PELANGGAN DENGAN
MODEL LRFM

Nama Mahasiswa : NOVIA TONAPA BASSY

N I M : 22104995

Matakuliah : Skripsi (Tugas Akhir)

Kode : TIW276

Semester : Gasal

Tahun Akademik : 2017/2018

Telah diperiksa dan disetujui di
Yogyakarta,
Pada tanggal 10 Januari 2018

Dosen Pembimbing I



Gloria Virginia, S.Kom., MAI, Ph.D.

Dosen Pembimbing II



R. Gunawan Santosa, Drs. M.Si.

HALAMAN PENGESAHAN

IMPLEMENTASI K-MEANS UNTUK SEGMENTASI DATA PELANGGAN DENGAN MODEL LRFM

Oleh: NOVIA TONAPA BASSY / 22104995

Dipertahankan di depan Dewan Penguji Skripsi
Program Studi Informatika Fakultas Teknologi Informasi
Universitas Kristen Duta Wacana - Yogyakarta
Dan dinyatakan diterima untuk memenuhi salah satu syarat memperoleh gelar
Sarjana Komputer
pada tanggal 18 Desember 2017

Yogyakarta, 10 Januari 2018
Mengesahkan,

Dewan Penguji:

1. Gloria Virginia, S.Kom., MAI, Ph.D.
2. R. Gunawan Santosa, Drs. M.Si.
3. Budi Susanto, SKom., M.T.
4. Sri Suwarno, Dr. Ir. M.Eng.

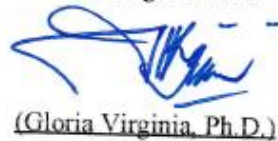


Dekan



(Budi Susanto, S.Kom., M.T.)

Ketua Program Studi



(Gloria Virginia, Ph.D.)

UCAPAN TERIMAKASIH

Segala pujian, hormat dan kemuliaan kepada Tuhan Yesus Kristus atas kemurahan-Nya yang besar, penghiburan dan kekuatan yang melimpah dianugerahkan kepada penulis sehingga penulis dapat menyelesaikan penelitian yang berjudul “Implementasi k-means untuk segmentasi data pelanggan dengan model LRFM”.

Selama penulisan penelitian ini penulis banyak mendapatkan dukungan dan bimbingan sehingga penulis ingin mengucapkan terimakasih kepada:

1. Ibu Gloria Virginia, S.Kom., MAI, Ph.D. selaku dosen pembimbing I yang dengan sabar membimbing memberi masukan dan koreksi dalam penelitian ini.
2. Bapak R. Gunawan Santosa, Drs. M.Si. selaku dosen pembimbing II yang mengajari teori - teori yang penulis belum pahami.
3. Bapak Budi Susanto, S.Kom., M.T. selaku dosen penguji yang memberi masukan dan saran untuk perbaikan penelitian.
4. Bapak Sri Suwarno, Dr. Ir. M.Eng. selaku dosen penguji yang memberi masukan dan saran untuk perbaikan penelitian.
5. Ibu Roselina Tonapa selaku orang tua dan seluruh keluarga yang memberi dukungan dalam doa, semangat dan materi kepada penulis selama ini.
6. Sahabat penulis Ila, Inge, Grefti, Flo, Yeni, Venny, Ria dan keluarga Body Voice yang memberi dukungan kepada penulis sehingga sangat menguatkan.

Akhir kata kiranya penelitian ini dapat berguna bagi penelitian selanjutnya dan pengembangan dalam bidang bisnis dan teknologi.

Yogyakarta, 7 Januari 2018

Penulis

INTISARI

Informasi yang diperoleh dari pemrosesan data transaksi pada sebuah penjualan pengecer secara *online* dapat digunakan untuk meningkatkan keuntungan dalam sebuah perusahaan karena perusahaan dapat mengetahui cara yang tepat untuk memberikan pelayanan kepada pelanggannya berdasarkan aktifitas pelanggan dalam bertransaksi. Pelanggan yang telah dikelompokkan berdasarkan kedekatan dari setiap atribut yang dimiliki dapat diberikan pelayanan khusus sesuai kebutuhan pelanggan tersebut. Salah satu algoritma yang dapat mengelompokkan pelanggan berdasarkan kedekatan dari setiap atribut yang dimiliki adalah algoritma k-means.

Pada penelitian ini penulis mengimplementasikan algoritma k-means untuk proses klusterisasi data transaksi tiap pelanggan yang akan ditransformasikan ke dalam model data LRFM (*Length, Recency, Frequency, Monetary*) dari data transaksi yang didapatkan pada sebuah perusahaan pengecer secara *online*. LRFM merupakan model yang sederhana dan baik untuk memprediksi *customer lifetime value*. Setelah dilakukan proses klusterisasi hasil segmentasi akan dipetakan berdasarkan *customer value matrix* dan *customer loyalty matrix*. Sehingga manfaat dari mengetahui karakteristik tiap segmen pelanggan adalah perusahaan dapat menentukan strategi marketing pada tiap segmen.

Keyword : Clustering, K-means, LRFM

DAFTAR ISI

HALAMAN SAMPUL	Error! Bookmark not defined.
HALAMAN JUDUL.....	Error! Bookmark not defined.
PERNYATAAN KEASLIAN SKRIPSI.....	iii
HALAMAN PERSETUJUAN.....	iv
HALAMAN PENGESAHAN.....	v
UCAPAN TERIMA KASIH.....	vi
INTISARI.....	vii
DAFTAR ISI.....	viii
DAFTAR GAMBAR	x
DAFTAR TABEL.....	xi
BAB 1 PENDAHULUAN.....	1
1.1 Latar Belakang Masalah.....	1
1.2 Perumusan Masalah.....	2
1.3 Batasan Sistem	2
1.4 Tujuan Penelitian.....	2
1.5 Manfaat Penelitian.....	2
1.6 Metodologi Penelitian	3
1.7 Sistematika Penulisan.....	3
BAB 2 LANDASAN TEORI	5
2.1 Tinjauan Pustaka	5
2.2 Landasan Teori	6
2.2.1 Metode Pengelompokan (<i>Clustering</i>).....	6
2.2.2 Algoritma K-means.....	7
2.2.3 Davies-Bouldin Index	10
2.2.4 <i>Sum of the Squared Error (SSE)</i>	12

2.2.5	<i>Length, Recency, Frequency, Monetary (LRFM)</i>	12
BAB 3	ANALISIS DAN PERANCANGAN SISTEM	14
3.1	Spesifikasi Kebutuhan Sistem	14
3.2	Penggunaan Bahasa Pemrograman	14
3.3	Perangkat Keras	14
3.4	Perangkat Lunak	14
3.5	Alur Penelitian	15
3.6	Praproses Data	16
3.6.1	Set Data	17
3.6.2	Segmen Pelanggan	18
3.6.3	Data model LRFM	19
3.7	Algoritma dan <i>Flowchart</i>	20
3.7.1	Algoritma K-Means	20
BAB 4	IMPLEMENTASI DAN ANALISIS SISTEM	22
4.1	Implementasi Sistem	22
4.1.1	Praproses Data	22
4.2	Analisis Kluster	27
BAB 5	KESIMPULAN DAN SARAN	34
5.1	Kesimpulan	34
5.2	Saran	34
	DAFTAR PUSTAKA	35
	LAMPIRAN	37

DAFTAR GAMBAR

Gambar 2.1 <i>customer value matrix</i> (Wu, Lin, & Chih-Wei, 2014).....	13
Gambar 2.2 <i>customer loyalty matrix</i>	13
Gambar 3.1 Alur penelitian.....	15
Gambar 3.2 Set data <i>online retail</i>	17
Gambar 3.3 <i>flowchart</i> algoritma K-means.....	21
Gambar 4.1 Tabel atribut yang dipilih.....	22
Gambar 4.2 Baris yang memiliki <i>CustomerID</i> kosong.....	23
Gambar 4.3 Baris yang memiliki <i>Quantity</i> dibawah nol dan <i>InvoiceNo</i> kosong.....	24
Gambar 4.4 Baris yang memiliki <i>UnitPrice</i> dibawah nol atau sama dengan nol.....	24
Gambar 4.5 Tabel data dalam model LRFM.....	27
Gambar 4.6 Nilai rata-rata L terhadap kluster.....	29
Gambar 4.7 Nilai rata-rata R terhadap kluster.....	29
Gambar 4.8 Nilai rata-rata F terhadap kluster.....	30
Gambar 4.9 Nilai rata-rata M terhadap kluster.....	30

DAFTAR TABEL

Tabel 2.1 jarak titik data ke titik pusat iterasi ke-1.....	8
Tabel 2.2 jarak titik data ke titik pusat iterasi ke-2.....	9
Tabel 2.3 jarak titik data ke titik pusat iterasi ke-3.....	9
Tabel 3.1 Deskripsi variabel pada data transaksi <i>online retail</i>	17
Tabel 3.2 Grup pelanggan berdasarkan <i>Customer value matrix</i> dan <i>customer loyalty matrix</i>	18
Tabel 3.3 Deskripsi variabel pada data <i>online retail</i> yang akan ditransformasi ke dalam LRFM.....	19
Tabel 4.1 Query pembersihan data uji.....	23
Tabel 4.2 Query untuk transformasi data kedalam bentuk LRFM.....	24
Tabel 4.3 Deskripsi dari <i>Length, Recency, Frequency, Monetary</i> berdasarkan data uji.....	27
Tabel 4.4 hasil klasterisasi algoritma k-means dengan 16 jumlah klaster.....	28
Tabel 4.5 Pemetaan pelanggan berdasarkan <i>customer value matrix</i> dan <i>customer loyalty matrix</i>	31

INTISARI

Informasi yang diperoleh dari pemrosesan data transaksi pada sebuah penjualan pengecer secara *online* dapat digunakan untuk meningkatkan keuntungan dalam sebuah perusahaan karena perusahaan dapat mengetahui cara yang tepat untuk memberikan pelayanan kepada pelanggannya berdasarkan aktifitas pelanggan dalam bertransaksi. Pelanggan yang telah dikelompokkan berdasarkan kedekatan dari setiap atribut yang dimiliki dapat diberikan pelayanan khusus sesuai kebutuhan pelanggan tersebut. Salah satu algoritma yang dapat mengelompokkan pelanggan berdasarkan kedekatan dari setiap atribut yang dimiliki adalah algoritma k-means.

Pada penelitian ini penulis mengimplementasikan algoritma k-means untuk proses klusterisasi data transaksi tiap pelanggan yang akan ditransformasikan ke dalam model data LRFM (*Length, Recency, Frequency, Monetary*) dari data transaksi yang didapatkan pada sebuah perusahaan pengecer secara *online*. LRFM merupakan model yang sederhana dan baik untuk memprediksi *customer lifetime value*. Setelah dilakukan proses klusterisasi hasil segmentasi akan dipetakan berdasarkan *customer value matrix* dan *customer loyalty matrix*. Sehingga manfaat dari mengetahui karakteristik tiap segmen pelanggan adalah perusahaan dapat menentukan strategi marketing pada tiap segmen.

Keyword : Clustering, K-means, LRFM

BAB 1

PENDAHULUAN

1.1 Latar Belakang Masalah

Penerapan data mining pada bidang bisnis diperlukan oleh perusahaan agar perusahaan dapat memahami dan mengelola pelanggannya sesuai dengan karakteristik pelanggan. Segmentasi pelanggan dapat dilakukan dengan analisis klusterisasi untuk memperoleh informasi mengenai pengelompokan tingkat kesetiaan pelanggan terhadap perusahaan berdasarkan data transaksi, sehingga sebuah perusahaan dapat melakukan perlakuan khusus bagi pelanggan dengan tingkat potensial yang berbeda-beda berdasarkan data. Model yang digunakan untuk mengolah data pelanggan sebelum masuk dalam proses klusterisasi adalah model LRFM (*Length, Recency, Frequency, Monetary*).

Model LRFM diterapkan pada data transaksi pelanggan sebagai salah satu cara untuk mengetahui *Customer Lifetime Value (CLV)* pada sebuah perusahaan pengecer secara *online* setelah data ditransformasikan kedalam bentuk LRFM data kemudian disegmentasi dengan menggunakan metode pengelompokan.

Metode pengelompokan pada data mining yang sering digunakan untuk mengolah data yang belum diketahui kelas-kelasnya adalah metode klusterisasi. Salah satu algoritma dari metode klusterisasi adalah algoritma k-means. Algoritma k-Means mengambil titik pusat atau *centroid* dari data yang ada untuk dijadikan acuan oleh data-data lain yang akan diukur jarak dan kemiripan data dengan titik pusat. Jika jarak dari data yang satu dengan *centroid* lebih dekat atau sifat dari data dengan *centroid* homogen maka data merupakan anggota dari kluster tersebut (Hidayat & Martiana, 2008).

1.2 Perumusan Masalah

Berdasarkan uraian permasalahan di atas, perumusan masalah adalah sebagai berikut:

1. Bagaimana menerapkan algoritma k-means pada proses klasterisasi data LRFM ?
2. Bagaimana mengambil informasi dari hasil klasterisasi data dengan model LRFM ?

1.3 Batasan Sistem

Batasan masalah pada sistem adalah sebagai berikut :

1. Algoritma yang digunakan untuk klasterisasi data LRFM adalah algoritma k-means menggunakan data *Online retail*. Set data *online retail* adalah set data yang dapat digunakan untuk klasterisasi yang diperoleh dari situs <https://archive.ics.uci.edu/ml/datasets> dengan jumlah *instance* 541909.
2. Atribut yang digunakan adalah *Stock Code*, *Invoice Date*, *Unit Price*, *Quantity*, *CustomerId*.
3. Kualitas klaster diukur menggunakan *Sum of Squared Error* (SSE).

1.4 Tujuan Penelitian

Mempelajari dan menganalisis hasil implementasi algoritma k-means pada data transaksi *online retail* yang ditransformasikan ke dalam bentuk LRFM.

1.5 Manfaat Penelitian

Dari hasil penelitian yang dilakukan diharapkan dapat memberikan manfaat sebagai berikut:

1. Mengetahui cara mengimplementasikan algoritma k-means untuk klasterisasi.
2. Mengetahui hasil kinerja algoritma k-means pada klasterisasi data transaksi dengan model LRFM.

3. Memperoleh informasi dari hasil klasterisasi data LRFM.

1.6 Metodologi Penelitian

1. Studi Literatur dan Pustaka

Mempelajari teori mengenai bagaimana algoritma k-means diterapkan dalam proses klasterisasi dan studi yang mengembangkan model-model algoritma dalam klasterisasi melalui sumber-sumber pustaka yang relevan.

2. Pemahaman Data dan Praproses Data

Data yang telah diperoleh dilakukan praproses agar dapat digunakan dan diterapkan kedalam model LRFM.

3. Perancangan Sistem Klasterisasi

Perancangan sistem meliputi analisa kebutuhan sistem, pembuatan *flowchart*. Perancangan berdasarkan algoritma yang akan digunakan.

4. Implementasi dan Testing

Sistem klasterisasi yang telah dibangun menggunakan algoritma yang diteliti, akan diuji coba dengan data-data yang telah disediakan.

5. Evaluasi

Keluaran dari sistem klasterisasi dievaluasi untuk hasil pengelompokan dari algoritma k-means. Kemudian, disajikan dalam bentuk tabel, diagram atau penyajian informasi lain yang dibutuhkan.

6. Konsultasi

Bimbingan oleh dosen pembimbing untuk keseluruhan penelitian yang dilakukan.

1.7 Sistematika Penulisan

Bab 1 merupakan pendahuluan yang berisi gambaran tentang penelitian yang dilakukan dan dijelaskan dalam latar belakang masalah, perumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, metodologi penelitian, dan sistematika penulisan.

Bab 2 merupakan tinjauan pustaka yang berisi uraian teori penunjang penelitian yang dilakukan. Teori diperoleh dari berbagai sumber. Kemudian analisis dan perancangan sistem akan dijabarkan pada

Bab 3 yang berisi langkah-langkah perancangan sistem dan analisis kebutuhan sistem, alur penelitian, *flowchart*.

Bab 4 berisi implementasi dan analisis sistem yang menjabarkan hasil dari implementasi sistem dan pembahasan analisa sistem pada penelitian berdasarkan data-data yang diperoleh.

Bab 5, berisi kesimpulan dari penelitian yang telah dilakukan dan saran untuk perbaikan, serta pengembangan sistem selanjutnya.

©UKDW

BAB 5

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Dari hasil pengembangan algoritma k-means pada data model LRFM diperoleh kesimpulan sebagai berikut:

1. Pada studi kasus di perusahaan pengecer secara *online* pelanggan dapat dibagi dalam empat kelompok besar yaitu *core customer*, *lost customer*, *cosuming resource customer*.
2. Berdasarkan hasil pemetaan pelanggan dengan karakteristik *core customer* berada pada klaster 3, 4, 5, 7, 8, 9, 10, 11, 12, 15, 16 dan *core customer* dengan nilai *monetary* rendah berada pada klaster 1 dan 14.
3. Berdasarkan set data yang digunakan dengan model LRFM nilai *Davies-Bouldin index* adalah sebesar 0.49247025 dan nilai SSE sebesar 1017298.621.

5.2 Saran

1. Pada pengembangan selanjutnya sistem dapat dikembangkan dengan menambahkan perhitungan untuk mengetahui kemurnian klaster.
2. Menambahkan metode pada k-means untuk mendapatkan titik pusat awal yang optimal melihat kelemahan dari algoritma k-means pada penentuan titik pusat awal.

DAFTAR PUSTAKA

- Al-Shboul, B., & Myaeng, S.-H. (2009). Initializing K-Means using Genetic Algorithms. *International conference on Computational Intelligence and Cognitive Informatics*, 114-118.
- B M, A. S., & S, K. H. (2012). Dynamic Clustering of data with Modified K-means. *ICICN 2012*, 221-225.
- Chang, D.-X., Zhang, X.-D., & Zheng, C.-W. (2008). A Genetic Algorithm with Gene Rearrangement for K-means Clustering. *Pattern Recognition*, 1210--1222.
- Chittu V., & Sumathi, N. (2011). A Modified Genetic Algorithm Initializing K-Means Clustering. *Global Journal of Computer Science and Technology*, 56-62.
- daoud, R. A., Amine, A., Bouikhalene, B., & Lbibb, R. (2015). Customer Segmentation Model in E-commerce Using Clustering Techniques and LRFM Model: The Case of Online Store in Morocco. *International Journal of Computer, Electrical, Automation, Control and Information Engineering Vol:9*.
- Hermawati, F. A. (2013). *Data Mining*. Yogyakarta: ANDI.
- Hidayat, A., & Martiana, E. (2008). Deteksi Mahasiswa Berprestasi dan Bermasalah dengan metode K-Means Klastering yang Dioptimasi dengan Algoritma Genetika. *Seminar Nasional Industri dan Teknologi [SNIT]2008*, 24-28.
- Kantardzic, M. (2003). *DATA MINING: Concepts, Models, Method, and Algorithms*. United States of America: Wiley Interscience.
- Liu, B. (2011). *Web Data Mining*. New York: Springer-Verlag.
- Muhammad, A. F. (2015). KLASTERISASI PROSES SELEKSI PEMAIN MENGGUNAKAN ALGORITMA K-MEANS (Study Kasus : Tim Hockey Kabupaten Kendal). 1-5.

- Parvaneh, A., Abbasimehr, H., & Tarokh, M. J. (2012). Data mining application in retailer segmentation based on LRFM variables: case study. *AWERProcedia Information Technology & Computer Science*, 128-133.
- Prasetyo, E. (2012). *DATA MINING - Konsep dan Aplikasi Menggunakan Matlab*. Yogyakarta: Andi.
- Rahman, M., & Islam, M. (2014). A hybrid clustering technique combining a novel genetic algorithm with K-means . *Knowledge-Based Systems*.
- ROIHA, N. U., Suprpto, Y. K., & Wibawa, A. D. (2017). *Web users segmentation using Genetic K-Means Algorithm*. Surabaya: Masters thesis, Institut Teknologi Sepuluh Nopember.
- Susanty, S. E., Hendrawan, R. A., & Anggraeni, W. (n.d.). Segementasi Pelanggan Menggunakan Two Stage Clustering dan LRFM Model pada Divisi Marketing PT.XYZ untuk mendukung Strategi Pengelolaan Pelanggan. 1-8.
- Wardhani, F. K., Suryani, E., & Mukhlason, A. (2012). Penerapan metode GA-Kmeans untuk pengelompokan pengguna pada Bapersip. *JURNAL TEKNIK ITS Vol. 1*, 545-550.
- Wu, H.-H., Lin, S.-Y., & Chih-Wei, L. (2014). Analyzing Patients' Values by Applying Cluster Analysis and. *The Scientific World Journal*, 2-7.
- Zacharski, R. (n.d.). *A Programmer's Guide to Data Mining: The Ancient Art of the Numerati*.