

Nomor Daftar: 049/S/BD/R/22/VIII/2023

**ANALISIS SEGMENTASI KONSUMEN MENGGUNAKAN ALGORITMA
K-MEANS CLUSTERING BERDASARKAN MODEL RFM SEBAGAI
REKOMENDASI STRATEGI PEMASARAN
(STUDI KASUS DI PT. XYZ)**

SKRIPSI

**Diajukan untuk Memenuhi Sebagian dari Syarat
Memperoleh Gelar Sarjana Program Studi Bisnis Digital**



Oleh

Deni Riswana

1905735

**PROGRAM STUDI BISNIS DIGITAL
KAMPUS TASIKMALAYA
UNIVERSITAS PENDIDIKAN INDONESIA**

2023

**ANALISIS SEGMENTASI KONSUMEN MENGGUNAKAN ALGORITMA
K-MEANS CLUSTERING BERDASARKAN MODEL RFM SEBAGAI
REKOMENDASI STRATEGI PEMASARAN
(STUDI KASUS DI PT. XYZ)**

**Oleh
Deni Riswana**

**Diajukan untuk memenuhi sebagian dari Persyaratan Memperoleh Gelar
Sarjana Bisnis Program Studi Bisnis Digital**

**© Deni Riswana
Universitas Pendidikan Indonesia
Juli 2023**

**Hak Cipta dilindungi undang-undang.
Skripsi ini tidak boleh diperbanyak seluruhnya atau sebagian,
Dengan dicetak ulang, difoto copy, atau cara lainya tanpa izin dari penulis**

DENI RISWANA

**ANALISIS SEGMENTASI KONSUMEN MENGGUNAKAN
ALGORITMA *K-MEANS CLUSTERING* BERDASARKAN MODEL RFM
SEBAGAI REKOMENDASI STRATEGI PEMASARAN
(STUDI KASUS DI PT. XYZ)**

ABSTRAK

Oleh

Deni Riswana

Sejumlah besar perusahaan di dunia telah mengadopsi teknik *data mining*, perusahaan tidak hanya berfokus pada pengembangan produk saja, akan tetapi juga berfokus pada penggalian data konsumen yang dapat diidentifikasi melalui data transaksi konsumen. Berdasarkan pendekatan tersebut, peneliti bertujuan untuk melakukan analisis segmentasi konsumen dengan tujuan dapat mengidentifikasi karakteristik konsumen berdasarkan pola-pola yang terbentuk. Pada penelitian ini, peneliti menggunakan *software* RapidMiner dengan menggunakan algoritma *K-Means clustering* berdasarkan model RFM (*Recency, Frequency, Monetary*). Kemudian penerapan visualisasi data melalui Google Looker Studio bertujuan mempermudah dan mempercepat proses visualisasi. Metode penelitian yang diterapkan adalah *Cross Industry Standard Process for Data Mining* (CRISP-DM). Sumber data pada penelitian ini didapatkan dari *database* internal perusahaan. Hasil penelitian menunjukkan bahwa terdapat 3 kelompok (*cluster*) dan didapatkan nilai *indeks Davies Bouldin* sebesar 0.185 yang menguji validitas model *cluster*. Cluster *Low loyalty* terdiri dari 4222 konsumen, *medium loyalty* terdiri dari 4133 konsumen, dan *high loyalty* terdiri dari 4449 konsumen. Pola dan karakteristik dari setiap segmen konsumen dapat menjadi wawasan dan pengetahuan baru untuk pengembangan strategi pemasaran yang tepat dan terarah.

Kata kunci: Data Mining, K-Means Clustering, Segmentasi Konsumen, Strategi Pemasaran

**ANALYSIS OF CONSUMER SEGMENTATION
USING K-MEANS CLUSTERING ALGORITHM BASED ON
RFM MODEL AS MARKETING STRATEGY RECOMMENDATION
(A CASE STUDY AT PT. XYZ)**

ABSTRACT

By

Deni Riswana

A significant number of companies worldwide have embraced data mining techniques, wherein companies not only develop products and services but also delve into consumer needs identifiable through transactional data and consumer behaviors. Built upon this approach, the researcher aims to conduct consumer segmentation analysis with the goal of identifying consumer characteristics based on emerging patterns. In this study, the researcher employs the RapidMiner software, utilizing the K-Means clustering algorithm based on the RFM (Recency, Frequency, Monetary) model. Subsequently, the application of data visualization through Google Looker Studio aims to simplify and expedite the visualization process. The applied research method adheres to the Cross Industry Standard Process for Data Mining (CRISP-DM). The data source utilized originates from the company's database. The research findings reveal the presence of three distinct clusters, yielding a Davies Bouldin index score of 0.185, validating the cluster model. The "Low loyalty" cluster consists of 4222 consumers, the "medium loyalty" cluster includes 4133 consumers, and the "high loyalty" cluster encompasses 4449 consumers. Patterns and characteristics unique to each consumer segment can provide insights and novel knowledge for the development of marketing strategies tailored to the attributes of each segment.

Keywords: *Data Mining, K-Means Clustering, Consumer Segmentation, Marketing Strategy*

DAFTAR ISI

KATA PENGANTAR.....	V
UCAPAN TERIMA KASIH	VI
ABSTRAK	VIII
ABSTRACT	IX
DAFTAR ISI.....	X
DAFTAR TABEL	XIV
DAFTAR GAMBAR.....	XV
DAFTAR LAMPIRAN.....	XVI
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	4
1.3 Tujuan Penelitian.....	4
1.4 Manfaat Penelitian.....	5
1.4.1 Manfaat Teoritis	5
1.4.2 Manfaat Praktis	5
1.5 Batasan Masalah	5
1.6 Struktur Organisasi Skripsi	6
BAB II KAJIAN PUSTAKA	7
2.1 Kajian Pustaka	7
2.1.1 Segmentasi Konsumen	7
2.1.2 Strategi Pemasaran	8
2.1.3 <i>Data Mining</i>	9
2.1.4 CRISP-DM.....	11

2.1.5	<i>Clustering</i>	13
2.1.6	Algoritma <i>K-Means</i>	13
2.1.6.1	Kekurangan dan Kelebihan Algoritma <i>K-Means</i>	15
2.1.6.2	Metode Pengukuran Jarak pada Algoritma <i>K-Means</i>	16
2.1.7	Model RFM.....	17
2.1.8	Normalisasi <i>Min-Max</i>	18
2.1.9	<i>Davies Bouldin Index</i> (DBI).....	18
2.1.10	<i>Detect Outlier</i>	19
2.1.11	<i>Software</i> RapidMiner Studio.....	19
2.1.12	Google Colab	20
2.1.13	Google Looker Studio.....	20
2.1.14	Microsoft Excel.....	21
2.2	Penelitian Terdahulu.....	21
BAB III METODE PENELITIAN		24
3.1	Objek Penelitian.....	24
3.2	Desain Penelitian	24
3.2.1	<i>Business Understanding</i>	25
3.2.1	<i>Data Understanding</i>	26
3.2.2	<i>Data Preparation</i>	27
3.2.3	<i>Modeling</i>	28
3.2.4	<i>Evaluation</i>	28
3.2.5	<i>Deployment</i>	28
3.3	Metode Penelitian	29
3.3.1	Teknik Pengumpulan Data.....	29

3.3.2	Teknik Analisis Data.....	29
3.4	Alat dan Bahan Penelitian	30
3.4.1	Alat.....	30
3.4.2	Bahan.....	31
BAB IV	HASIL DAN PEMBAHASAN	32
4.1	Pemahaman Bisnis (<i>Business Understanding</i>)	32
4.1.1	Gambaran Umum Perusahaan.....	32
4.1.2	Menentukan Tujuan Bisnis	33
4.1.3	Menilai Situasi Perusahaan	33
4.1.4	Menentukan Tujuan Data Mining.....	34
4.1.5	Menghasilkan Rencana Proyek.....	34
4.2	Pemahaman Data (<i>Data Understanding</i>).....	34
4.2.1	Mengumpulkan Data.....	35
4.2.2	Mendeskripsikan Data.....	35
4.2.3	Mengeksplorasi Data.....	36
4.2.4	Memverifikasi Kualitas Data	36
4.3	Persiapan Data (<i>Data Preparation</i>).....	37
4.3.1	Menyeleksi Data Segmentasi Konsumen.....	37
4.3.2	Integrasi Data Segmentasi Konsumen	38
4.3.3	Membersihkan Data Segmentasi Konsumen	40
4.3.4	Memformat Data Segmentasi Konsumen	42
4.4	Pemodelan (<i>Modeling</i>).....	45
4.4.1	Pemodelan <i>K-Means</i> Menggunakan RapidMiner	45
4.4.2	Pemodelan <i>K-Means</i> Secara Manual	46

4.5 Evaluasi (<i>Evaluation</i>)	48
4.6 Penyebaran (<i>Deployment</i>).....	50
4.6.1 Menyajikan Visualisasi Segmentasi Konsumen	50
4.6.2 Analisis Data Segmentasi Konsumen	51
4.6.2.1 Analisis Karakteristik <i>Low loyalty Cluster</i>	53
4.6.2.2 Analisis Karakteristik <i>Medium loyalty Cluster</i>	54
4.6.2.3 Analisis Karakteristik <i>High loyalty Cluster</i>	54
4.6.2.4 Perbandingan Karakteristik <i>Cluster</i>	55
4.6.3 Rekomendasi Strategi Pemasaran <i>Cluster</i>	55
BAB V SIMPULAN DAN REKOMENDASI.....	57
5.1 Simpulan	57
5.2 Implikasi Manajerial.....	57
5.3 Saran	59
DAFTAR PUSTAKA	60
LAMPIRAN.....	66
RIWAYAT HIDUP	72

DAFTAR PUSTAKA

- Aggarwal, C. C., & Reddy, C. K. (2013). *Data Clustering: Algorithms and Applications* (1st ed.). Chapman & Hall/CRC.
- Alexander, M., Kusleika, R., & Walkenbach, J. (2018). *Excel 2019 Bible* (1st ed.). Wiley Publishing.
- Aprilla Dennis. (2013). Belajar Data Mining dengan RapidMiner. *Innovation and Knowledge Management in Business Globalization: Theory & Practice, Vols 1 and 2*, 5(4), 1–5.
- Arbin, N., Suhaimi, N. S., Mokhtar, N. Z., & Othman, Z. (2016). Comparative analysis between k-means and k-medoids for statistical clustering. *Proceedings - AIMS 2015, 3rd International Conference on Artificial Intelligence, Modelling and Simulation*, 117–121. <https://doi.org/10.1109/AIMS.2015.82>
- Arifin, Z. (2017). *Adopsi Teknologi untuk Keunggulan Daya Saing*.
- Bajracharya, D. (2020). *Comparison between clustering algorithms : k-means and k-medoids using JAVA Program Dinesh Bajracharya*. 8(11), 289–295.
- Carneiro, T., Da Nobrega, R. V. M., Nepomuceno, T., Bian, G. Bin, De Albuquerque, V. H. C., & Filho, P. P. R. (2018). Performance Analysis of Google Colaboratory as a Tool for Accelerating Deep Learning Applications. *IEEE Access*, 6, 61677–61685. <https://doi.org/10.1109/ACCESS.2018.2874767>
- Chandola, V., BANERJEE, A., & KUMAR, V. (2009). Survey of Anomaly Detection. *ACM Computing Survey (CSUR)*, 41(3), 1–72.
- Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Daimlerchrysler, T. R., Shearer, C., & Daimlerchrysler, R. W. (2000). CRISP-DM 1.0. In *SPSS inc* (Vol. 78). DaimlerChrysler.
- Dua, S., & Chowriappa, P. (2012). Data mining for bioinformatics. In *Data Mining for Bioinformatics*. CRC Press. <https://doi.org/10.1201/B13091/DATA->

MINING-BIOINFORMATICS-SUMEET-DUA-PRADEEP-CHOWRIAPPA

- Febrina Wulandari, G. (2020). Segmentasi Pelanggan Menggunakan Algoritma K-Menas Untuk Customer Relationship Management (CRM) Pada Hijab Miulan. *UDINUS Journal*.
- Grubor, A., & Jaksa, O. (2018). Internet Marketing as a Business Necessity. *Interdisciplinary Description of Complex Systems*, 16(2), 265–274. <https://doi.org/10.7906/indecs.16.2.6>
- Guntara, R. G. (2023). Visualisasi Data Laporan Penjualan Toko Online Melalui Pendekatan Data Science Menggunakan Google Colab. 2(6), 2091–2100.
- Guo, F., & Qin, H. (2017). Data Mining Techniques for Customer Relationship Management. *Journal of Physics: Conference Series*, 910(1). <https://doi.org/10.1088/1742-6596/910/1/012021>
- Han, J., Kamber, M., & Pei, J. (2012). Data Mining: Concepts and Techniques. In *Data Mining: Concepts and Techniques* (Third Edit). Morgan Kaufmann. <https://doi.org/10.1016/C2009-0-61819-5>
- Handayani, F. (2022). Aplikasi Data Mining Menggunakan Algoritma K-Means Clustering untuk Mengelompokkan Mahasiswa Berdasarkan Gaya Belajar. *Jurnal Teknologi dan Informasi*, 12(1), 46–63. <https://doi.org/10.34010/jati.v12i1.6733>
- Herdiana, O. (2021). Strategi Pemasaran Produk Industri Kreatif Menggunakan Algoritma K-Means Clustering Berbasis Particle Swarm Optimization. *Jurnal Nuansa Informatika*, 15, 1–13.
- Homburg, C., Jozić, D., & Kuehnl, C. (2015). Customer experience management: Toward implementing an evolving marketing concept. *Journal of the Academy of Marketing Science*, 45. <https://doi.org/10.1007/s11747-015-0460-7>
- Kim, J. (2022). Google Colab 介紹及安裝教學. *Journal of Open Source Education*, Vol. 2, Hal. 51.

- Kohonen, T. (2001). *Self-Organizing neural networks: recent advances and applications*.
- Kotler, P., & Armstrong, G. (2015). Principles of Marketing. In *Economica* (Nomor 19). Pearson. <https://doi.org/10.2307/2548367>
- Kotler, P., & Keller, K. L. (2016a). *Marketing Management (15th ed.)* (Vol. 15E). <https://doi.org/10.1080/08911760903022556>
- Kotler, P., & Keller, K. L. (2016b). *Marketing Management (15th ed.)*. Pearson.
- Kumar, V., & Reinartz, W. (2016). Creating Enduring Customer Value. *Journal of Marketing*, 80(6), 36–68. <http://www.jstor.org/stable/44134973>
- Liao, S., Chu, P., & Hsiao, P. (2012). Data mining techniques and applications – A decade review from 2000 to 2011. *Expert Systems With Applications*, 39(12), 11303–11311. <https://doi.org/10.1016/j.eswa.2012.02.063>
- Manyika, J., Chui, M., Miremadi, M., Bughin, J., George, K., Willmott, P., & Dewhurst, M. (2017). A Future That Works: Automation, Employment, and Productivity. *McKinsey Global Institute, January*, 1–28. [http://njit2.mrooms.net/pluginfile.php/688844/mod_resource/content/1/Executive Summary of McKinsey Report on Automation.pdf](http://njit2.mrooms.net/pluginfile.php/688844/mod_resource/content/1/Executive%20Summary%20of%20McKinsey%20Report%20on%20Automation.pdf)
- Marcoulides, G. A. (2005). Discovering Knowledge in Data: an Introduction to Data Mining. In *Journal of the American Statistical Association* (Vol. 100, Nomor 472). <https://doi.org/10.1198/jasa.2005.s61>
- Muhammad Ali, P. J. (2022). Investigating the Impact of Min-Max Data Normalization on the Regression Performance of K-Nearest Neighbor with Different Similarity Measurements. *ARO-THE SCIENTIFIC JOURNAL OF KOYA UNIVERSITY*, 10(1), 85–91. <https://doi.org/10.14500/aro.10955>
- Muharni, S., & Apri, C. (2022). *Visualisasi Data Menggunakan Data Studio*. Literasi Nusantara Abadi.

- Mulyadi, M. (2013). Riset Desain Dalam Metodologi Penelitian. *Jurnal Studi Komunikasi dan Media*, 16(1), 71. <https://doi.org/10.31445/jskm.2012.160106>
- Nishom, M. (2019). Perbandingan Akurasi Euclidean Distance, Minkowski Distance, dan Manhattan Distance pada Algoritma K-Means Clustering berbasis Chi-Square. *Jurnal Informatika: Jurnal Pengembangan IT*, 4(1), 20–24. <https://doi.org/10.30591/jpit.v4i1.1253>
- North, M. (2012). *Data Mining for the Masses*. A Global Text Project Book. <https://docs.rapidminer.com/downloads/DataMiningForTheMasses.pdf>
- Noviyanto. (2020). Penerapan Data Mining dalam Mengelompokkan Jumlah Kematian. *Jurnal Informatika dan Komputer*, 22(2), 183–188.
- Okta, N. (2022). Analisis Segmentasi Konsumen Menggunakan Metode K-Means Clustering Berdasarkan Model Rfm Dan Pengetahuan Data Lainnya Sebagai Rekomendasi Strategi Pemasaran (Studi Kasus: Yayasan Xyz). In *Digital Library Unila* (Nomor 8.5.2017).
- Putri, Budi R. T. (2022). *Manajemen Pemasaran* (N. Suparta (ed.)). Swasta Nulus.
- Pynam, V., Roje Spanadna, R., & Srikanth, K. (2018). An Extensive Study of Data Analysis Tools (Rapid Miner, Weka, R Tool, Knime, Orange). *SSRG International Journal of Computer Science and Engineering (SSRG-IJCSE)*, 5(9).
- R. C. Blattberg, B.-D. Kim, & S. A. Neslin. (2008). *Database Marketing*. Springer Science+Business Media,.
- Rahmah, L. (2017). PERBANDINGAN HASIL PENGGEROMBOLAN K-MEANS, FUZZY K-MEANS, DAN TWO STEP CLUSTERING. *Jurnal Pendidikan Matematika*, 2(1), 39. <https://doi.org/10.18592/jpm.v2i1.1166>
- Ritonga, H. M. (2018). *Manajemen Pemasaran: Konsep dan Aplikasi*. Universitas Pembangunan Panca Budi.

- Sabzi, A., Farjami, Y., & Zihayat, M. (2011). An improved fuzzy K-medoids clustering algorithm with optimized number of clusters. *Proceedings of the 2011 11th International Conference on Hybrid Intelligent Systems, HIS 2011*, 206–210. <https://doi.org/10.1109/HIS.2011.6122106>
- Saleh, M. Y., & Said, M. (2019). Konsep dan Strategi Pemasaran. In Sobirin (Ed.), *Analisis Standar Pelayanan Minimal Pada Instalasi Rawat Jalan di RSUD Kota Semarang* (Vol. 3). CV Sah Media.
- Saputra, D. B., & Riksakomara, E. (2018). Implementasi Fuzzy C-Means dan Model RFM untuk Segmentasi Pelanggan (Studi Kasus : PT. XYZ). *Jurnal Teknik ITS*, 7(1). <https://doi.org/10.12962/j23373539.v7i1.28230>
- Sawant, K., & Bhogan, S. (2016). Iteration Reduction K Means Clustering Algorithm. *IJISET - International Journal of Innovative Science, Engineering & Technology*, 3(5), 4–6.
- Shafique, U., & Qaiser, H. (2014). A Comparative Study of Data Mining Process Models (KDD , CRISP-DM and SEMMA). *International Journal of Innovation and Scientific Research*, 12(1), 217–222. <http://www.ijisr.issr-journals.org/>
- Sinaga, K. P., & Yang, M. S. (2020). Unsupervised K-means clustering algorithm. *IEEE Access*, 8, 80716–80727. <https://doi.org/10.1109/ACCESS.2020.2988796>
- Situmorang, A., Arifin, A., Rusilpan, I., & Juliane, C. (2022). Analisa dan Penerapan Metode Algoritma K-Means Clustering Untuk Mengidentifikasi Rekomendasi Kategori Baru Pada List Movie IMDb. *JURNAL MEDIA INFORMATIKA BUDIDARMA*, 6(4), 2171–2179. <https://doi.org/10.30865/MIB.V6I4.4729>
- Snipes, G. (2018). Google Data Studio. *Journal of Librarianship and Scholarly Communication*, 6(1). <https://doi.org/10.7710/2162-3309.2214>
- Strecker, S., Kuckertz, A., & Pawlowski, J. M. (2020). Machine Learning and Reasoning Fuzzy Logic Algoritma, Manual, Matlab, & Rapid Miner. In *ICB Research Reports* (Nomor 9). Deepublish.

- Sugiyono. (2019). *Metode Penelitian Kuantitatif, Kualitatif, dan R&D*. Alfabet.
- Supriyati. (2015). *Metodologi Penelitian*. Labkat Press.
- Tidd, J., & Pavitt, K. (2011). *Managing Innovation: Integrating Technological, Market And Organizational Change*.
- Wahyudi, M., & Dkk. (2020). *Data Mining : Penerapan Algoritma K- Means Clustering Dan K- Medoids Clustering*. Kita Menulis.
- Wei, J.-T., Lin, S.-Y., & Wu, H.-H. (2010). A review of the application of RFM model. *African Journal of Business Management*, 4(19), 4199–4206.
- Wei, J., Lin, S., & Wu, H. (2010). A review of the application of RFM model. *African Journal of Business Management*, 4(19), 4199–4206.
- Wickham Grolemond. (2017). *R for Data Science*.
- Widiyanto, A. T., & Witanti, A. (2021). Segmentasi Pelanggan Berdasarkan Analisis RFM Menggunakan Algoritma K-Means Sebagai Dasar Strategi Pemasaran (Studi Kasus PT Coversuper Indonesia Global). *KONSTELASI: Konvergensi Teknologi dan Sistem Informasi*, 1(1), 204–215.
<https://doi.org/10.24002/konstelasi.v1i1.4293>
- Zakariah, M. A., Alfriani, V., & Zakariah, K. M. (2020). Metodologi Penelitian Kualitatif, Kuantitatif, Mix Method, Research and Development. *Madani Media*, 118.
- Zeithaml, V. A., Bitner, M. J., & Gremler, D. D. (n.d.). *Services marketing : integrating customer focus across the firm*. 514.