

IMPLEMENTASI DATA MINING PEMILIHAN PELANGGAN POTENSIAL
MENGUNAKAN ALGORITMA K-MEANS
IMPLEMENTATION OF DATA MINING FOR POTENTIAL CUSTOMER SELECTION
USING K-MEANS ALGORITHM

Wahyu Romadhona*, Bangkit Indarmawan Nugroho, Aang Alim Murtopo

^{1,3} Program Studi Teknik Informatika STMIK YMI Tegal

² Program Studi Sistem Informasi STMIK YMI Tegal

E-mail Korespondensi : 18165027@mhs.stmik-tegal.ac.id

History Artikel

Diterima : 18 Juli 2022 Disetujui : 26 September 2022 Dipublikasikan : 14 Oktober 2022

Abstract

Potential customers are customers who are interested in a product or service, but do not finish taking action, are held back by different issues (eg price, place, convenience, opinions of other customers, etc.). But this can be overcome by identifying and assessing the problems experienced by the company naturally. The purpose of this research is to determine potential customers and study data mining using the K-Means algorithm implemented at Exso Coffee & Roastery when selecting prospective customers. From the results of data collection conducted through observations and interviews with the owners and customers of the company concerned, as well as by conducting data analysis to determine what is needed in the process of identifying potential customers. Based on customer data from November 2021 to April 2022 at Exso Coffee & Roastery, 160 data samples have been obtained. The company has not been able to optimize service to customers, for example the limited number of menus and uncomfortable seats. So it is important to attract the attention of these customers, it will be very important to deal with the problems that the company is experiencing so that problem solving will be easier, more focused and precise. Through experiments using the K-Means algorithm and RapidMiner Studio software, the results are that potential customers are customers aged between 35 and 40 years. which will then be declared as cluster 0.

Keywords: Data Mining, K-Means Algorithm, RapidMiner Studio

Abstrak

Pelanggan potensial adalah pelanggan yang tertarik dengan produk atau layanan, tetapi tidak selesai mengambil langkah, ditahan oleh masalah yang berbeda-beda (misalnya harga, tempat, kenyamanan, pendapat pelanggan lain, dll.). Tetapi hal ini bisa diatasi dengan cara mengidentifikasi serta mengkaji masalah-masalah yang perusahaan alami. Tujuan dari penelitian ini adalah untuk menentukan pelanggan potensial serta mempelajari data *mining* menggunakan algoritma *K-Means* yang diimplementasikan di Exso Coffee & Roastery saat memilih calon pelanggan. Berdasarkan dari hasil pengumpulan data yang dilakukan melalui observasi dan wawancara dengan pemilik dan pelanggan perusahaan yang bersangkutan, juga dengan melakukan analisa data untuk menentukan apa saja yang dibutuhkan dalam proses mengidentifikasi pelanggan potensial. Berdasarkan data pelanggan pada bulan November 2021 hingga April 2022 di Exso Coffee & Roastery yang telah didapatkan adalah sebanyak 160 *sample* data. Perusahaan belum dapat mengoptimalkan pelayanan pada pelanggan misalnya jumlah menu yang terbatas serta tempat duduk yang kurang nyaman. Maka diperlukan pemetaan pelanggan sehingga dengan adanya pemetaan pelanggan ini, akan menjadi sangat penting untuk menghadapi masalah yang perusahaan alami sehingga penyelesaian masalah pun akan lebih mudah, terarah dan tepat. Melalui eksperimen menggunakan algoritma *K-Means* dan perangkat lunak RapidMiner Studio, memberikan hasil pelanggan potensial adalah pelanggan berusia antara 35 hingga 40 tahun. yang selanjutnya akan dideklarasikan sebagai *cluster* 0.

Kata Kunci: : Data *Mining*, Algoritma *K-Means*, RapidMiner Studio

How to Cite: W. Romadhona (2021). Implementasi Data *Mining* Pemilihan Pelanggan Potensial Menggunakan Algoritma *K-Means*. KOMPUTEK : Jurnal Teknik Universitas Muhammadiyah Ponorogo Vol 6 (2): Halaman 76-83

© 2022 Universitas Muhammadiyah Ponorogo. All rights reserved

ISSN 2614-0985 (Print)
ISSN 2614-0977 (Online)

PENDAHULUAN

Hubungan antara perusahaan dengan pelanggannya sangat penting untuk mendukung pertumbuhan dan kelangsungan perusahaan. Manajemen perusahaan harus dapat mengidentifikasi pelanggan potensial dan mempercayai mereka dengan meningkatkan pemahaman perusahaan tentang kebutuhan mereka sebagai individu sehingga mereka dapat tetap loyal kepada perusahaan. Salah satu hal terpenting dalam manajemen pelanggan adalah bagaimana perusahaan mempertahankan pelanggan. Upaya mempertahankan pelanggan ini penting bagi Exso Coffee & Roastery mengingat semakin banyaknya bisnis serupa.

Untuk mencegah terjadinya penyimpangan pelanggan seperti ini, perlu dipahami basis pelanggan potensial sehingga perusahaan dapat melindungi pelanggan potensial tersebut dengan memberikan pelayanan prima dan memberikan hadiah. Perusahaan memberikan hadiah berupa cinderamata kepada pelanggannya, yang terus berlanjut setiap tahun.

Pelanggan Potensial menduduki posisi penting dalam pengembangan strategi bisnis, pelanggan juga merupakan salah satu sumber keuntungan dalam perusahaan. Untuk itu diperlukan suatu pemahaman yang baik tentang pelanggan. Pemahaman yang baik terhadap pelanggan dapat digunakan perusahaan untuk berinvestasi terhadap pelanggan yang potensial. Masalah yang

sering dihadapi adalah kesulitan dalam menganalisa nilai pelanggan. Banyak pemasar mengalami kesulitan untuk mengidentifikasi pelanggan atau nasabah yang tepat hal tersebut dapat mengakibatkan perusahaan kehilangan pelanggan potensial dan tentunya akan sangat merugikan perusahaan.

Segmentasi pelanggan adalah metode populer yang digunakan untuk memilih pelanggan atau nasabah yang tepat untuk memulai promosi. Dengan segmentasi nasabah berdasarkan prilakunya, kita dapat menargetkan tindakan mereka dengan lebih baik. Seperti peluncuran produk yang disesuaikan, target pemasaran dan untuk memenuhi harapan pelanggan. Namun untuk menganalisa data pelanggan atau nasabah dalam jumlah besar memerlukan tenaga dan waktu yang banyak (DIKRIANSYAH, 2018)

Penelitian yang akan kami lakukan adalah pengaplikasi data *mining* menggunakan algoritma K-Means. Yang sebelumnya sudah banyak dilakukan oleh banyak peneliti di bidang teknologi informasi atau bidang lainnya.

METODE PENELITIAN

1. *Knowledge discovery In Database*

adalah keseluruhan proses non-trivial untuk mencari dan mengidentifikasi pola (*pattern*) dalam data, dimana pola yang ditemukan bersifat sah, baru dapat bermanfaat dan dapat dimengerti (Ndaumanu & Kusri, n.d.). Proses KDD secara garis besar (Nasari

& Darma, 2015) dapat dijelaskan sebagai berikut:

a) *Data Selection*

Pemilihan (seleksi) data dari sekumpulan data operasional perlu dilakukan sebelum tahap penggalian informasi dalam KDD dimulai. Data hasil seleksi yang akan digunakan untuk proses *data mining* disimpan dalam suatu berkas, terpisah dari basis data operasional.

b) *Pre-processing / Cleaning*

Sebelum proses *data mining* dapat dilaksanakan, perlu dilakukan proses pembersihan pada data yang menjadi fokus KDD. Proses pembersihan mencakup antara lain membuang duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data, seperti kesalahan cetak (tipografi).

c) *Transformation*

Coding adalah transformasi pada data yang telah dipilih, sehingga data tersebut sesuai untuk proses Data mining. Proses coding dalam KDD merupakan proses kreatif dan sangat tergantung pada jenis atau pola informasi yang akan dicari dalam basis data.

d) *Data mining*

Data mining adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan

teknik atau metode tertentu. Teknik, metode, atau algoritma dalam Data mining sangat bervariasi. Pemilihan metode atau algoritma yang tepat sangat bergantung pada tujuan dan proses KDD secara keseluruhan. Seperti halnya (Susanto & Sudiyatno, 2014) meneliti tentang penerapan *data mining* untuk memprediksi prestasi belajar siswa berdasarkan status sosial ekonomi orang tua, motivasi, kedisiplinan siswa dan prestasi masa lalu dengan algoritma J48.

e) *Interpretation / Evaluation*

Pola informasi yang dihasilkan dari proses *data mining* perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan. Tahap ini merupakan bagian dari proses KDD yang disebut *interpretation*. Tahap ini mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesis yang ada sebelumnya

2. *K-Means Clustering*

Merupakan algoritma *clustering* yang berulang-ulang. Algoritma K-Means dimulai dengan pemilihan secara acak K, K disini merupakan banyaknya *cluster* yang ingin dibentuk. Kemudian ditetapkan nilai-nilai K secara *random*, untuk sementara nilai tersebut menjadi pusat dari *cluster*

atau biasa disebut dengan *centroid*, *mean* atau “*means*”. Adalah hitungan jarak setiap data yang terhadap pada masing-masing *centroid* menggunakan rumus *Euclidian* hingga ditemukan jarak yang paling dekat dari setiap data dengan *centroid*. Pengklasifikasian setiap data berdasarkan kedekatannya dengan *centroid* Langkah ini dilakukan hingga nilai *centroid* tidak dapat lagi berubah (stabil) (Andi Cuhwanto & R, 2021).

HASIL DAN PEMBAHASAN

A. Data Uji

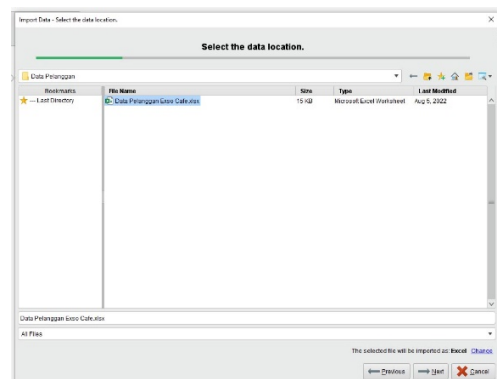
Proses implementasi dan analisis data menggunakan aplikasi RapidMiner Studio, dengan melakukan uji pada semua data yang di mulai dari bulan November 2021 sampai dengan April 2022 di Exso *Coffee & Roastery* dengan menggunakan 160 sampel data, di antaranya sumber data asli disimpan ke dalam database menggunakan *Microsoft Access* dan dikonversikan ke format *.xlsx* melalui *Microsoft Excel*.

B. Pengujian pada *Rapid Miner Studio*

Untuk membuktikan kebenaran Selama fase analisis dan pengujian secara manual, diperlukan penggunaan perangkat lunak RapidMiner Studio, langkah-langkah yang harus diambil dalam pengujian ini adalah:

1. Impor data

Impor data dapat dilakukan setelah proses pengumpulan dan pengolahan data penelitian selesai, data diolah dalam format *.xlsx* menggunakan *Microsoft Excel*, untuk selanjutnya di import ke dalam *Software RapidMiner Studio*.



Gambar 1. Mengimpor Data Pelanggan

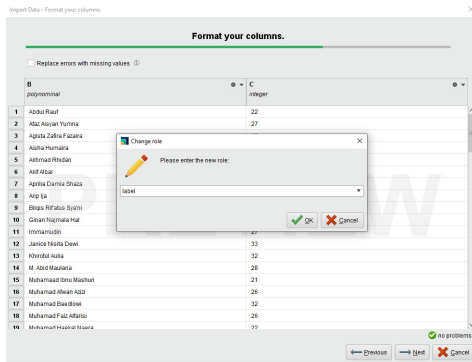
Untuk melihat isi datanya Kita bisa masuk dengan menu import lalu pilih data yang akan digunakan, setelah itu klik *Next* untuk melanjutkan, seperti yang ditunjukkan pada gambar.

2. Seleksi Data

Selanjutnya Perlu kita lakukan adalah penyeleksian data yang akan digunakan.

3. Inisialisasi

Adalah proses mengeset nilai awal suatu variabel. Dalam penelitian ini peneliti mencontohkan data nama pelanggan sebagai *label*, dan usia pelanggan sebagai *interger*.

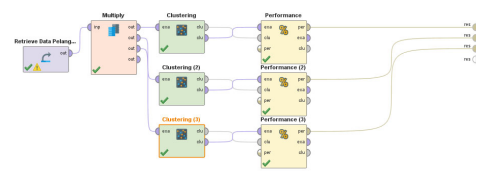


Gambar 2. Tampilan Inisialisasi Data

Setelah data berhasil di *import* selanjutnya yang perlu dilakukan adalah klusterisasi.

4. *Clustering*

adalah suatu metode pengelompokan data, hal ini merupakan bagian dari Data *Mining* atau Penggalian data, yaitu ekstraksi pola yang menarik dari data dalam jumlah besar. dalam hal ini peneliti menggunakan tiga operator atau *tools* yang terdapat pada Aplikasi RapidMiner Studio yaitu: *multiply* digunakan untuk menghubungkan banyak operator agar bisa dijalankan secara bersamaan, *clustering* digunakan untuk mengelompokkan jumlah data dan *cluster distance performance* untuk melihat nilai terbaik dari masing-masing *cluster*.



Gambar 3. Tampilan Proses Clustering

5. *Parameters*

Selanjutnya yang perlu dilakukan adalah menentukan jumlah *K* (*parameters*) pada setiap *Cluster*. Dalam hal ini kami memberikan nilai sebagai table berikut:

Tabel 1. Parameter Pengelompokan

No.	Cluster	K
1	0	2
2	1	3
3	2	4

Atau nilai 2 pada *cluster* 0, 3 pada *cluster* 1 dan 4 pada *cluster* 2. Sehingga menghasilkan nilai (*Bouldin*) sebagai di tabel berikut:

Tabel 2. Nilai Bouldin

No.	Cluster	K	Bouldin
1	0	2	-0.481
2	1	3	-0.531
3	2	4	-0.470

Nilai *bouldin* terkecil adalah nilai yang terbaik, abaikan (-) karena nilai ini bersifat *absolute* sehingga (+) ataupun (-) tidak berpengaruh.

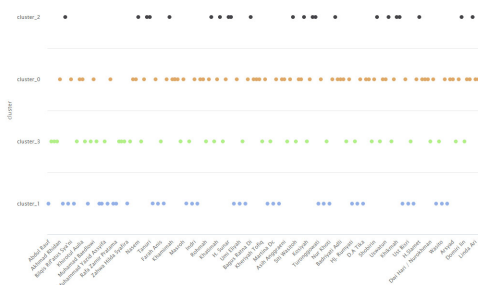
6. Hasil Berdasarkan data nilai *table Bouldin* di atas bisa disimpulkan bahwa nilai terbaik adalah *Cluster 0* dengan pengelompokan *K* berjumlah 4 (*parameters*).

Cluster Model

```
Cluster 0: 63 items
Cluster 1: 35 items
Cluster 2: 22 items
Cluster 3: 36 items
Total number of items: 156
```

Gambar 4. Hasil *Clustering*

Dalam penelitian ini performance terbaiknya adalah 4 *K* sehingga menghasilkan, data seperti yang ditunjukkan pada gambar hasil *clustering*.



Gambar 5. Tampilan 3D Hasil *Clustering*

Dari data diatas bisa disimpul bahwa data yang memiliki *centroid* terbesar adalah yang termasuk dalam pelanggan paling potensial yaitu yang

berusia antara 35 hingga 40 tahun, atau yang dideklarasikan sebagai *cuslter 0*.

KESIMPULAN

Dari penelitian yang ada, maka dapat ditarik kesimpulan seperti di bawah ini:

1. Paket data

Algoritma K-Means digunakan untuk menentukan jumlah *cluster*, dengan cara menghitung jarak terdekat ke pusat gugus. Data terdekat di deklarasikan sebagai anggota cluster. Perhitungan dilakukan secara berulang sampai data tidak dapat lagi dipindahkan ke *cluster* lain.

2. Data Utama

Data utama didapatkan setelah perhitungan Algoritma K-Means selesai. Data yang memiliki *centroid* terbesar adalah yang termasuk dalam pelanggan paling potensial adalah pelanggan yang berusia antara 35 hingga 40 tahun. Atau yang dideklarasikan sebagai *cuslter 0*.

3. Data Prospek

Data prospek diperoleh dengan menggunakan metode *interview* kepada pelanggan, yang kemudian dicatat dengan menggunakan microsoft excel dan selanjutnya data tersebut diolah menggunakan aplikasi RapidMiner Studio. Data yang diproses meliputi data nama pelanggan, dan usia pelanggan.

DAFTAR PUSTAKA

- Andi Cuhwanto, Y. N., & R, D. A. (2021). Implementasi Data Mining Pemilihan Pelanggan Potensial Menggunakan Algoritma K-Means. *Petir*, 15(1), 48–56.
<https://doi.org/10.33322/petir.v15i1.1358>
- DIKRIANSYAH, F. (2018). No
主観的健康感を中心とした在宅高齢者における
健康関連指標に関する共分散構造分析
Title. *Biomass Chem Eng*, 3(2),
تفتتفتت. Retrieved from
http://journal.stainkudus.ac.id/index.php/equilibrium/article/view/1268/1127%0Ahttp://publicacoes.cardiol.br/portal/ijcs/portugues/2018/v3103/pdf/3103009.pdf%0Ahttp://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S0121-75772018000200067&lng=en&tlng=
- Nasari, F., & Darma, S. (2015). *Seminar Nasional Teknologi Informasi dan Multimedia 2015 PENERAPAN K-MEANS CLUSTERING PADA DATA PENERIMAAN MAHASISWA BARU (STUDI KASUS : UNIVERSITAS POTENSI UTAMA)*. 6–8.
- Ndaumanu, R. I., & Kusriani, M. R. (n.d.). *A.,. 2014, Analisis Prediksi Tingkat Pengunduran Diri Mahasiswa dengan Metode K-Nearest Neighbor*. Jatisi.
- Susanto, H., & Sudiyatno, S. (2014). Data mining untuk memprediksi prestasi siswa berdasarkan sosial ekonomi, motivasi, kedisiplinan dan prestasi masa lalu. *Jurnal Pendidikan Vokasi*, 4(2), 222–231.
<https://doi.org/10.21831/jpv.v4i2.2547>