

Implementasi *Data Mining* Pemilihan Pelanggan Potensial Menggunakan Algoritma *K-Means*

Wahyu Romadhona^{1)*}, Bangkit Indarmawan Nugroho²⁾, Aang Alim Murtopo³⁾

¹⁾³⁾ Program Studi Teknik Informatika STMIK YMI Tegal, Indonesia

²⁾ Program Studi Sistem Informasi STMIK YMI Tegal, Indonesia

¹⁾ wachyurmd@gmail.com, ²⁾ efbeterang@gmail.com ³⁾ aang.alim@gmail.com

Abstrak :

Pelanggan potensial adalah pelanggan yang tertarik dengan produk atau layanan, tetapi tidak selesai mengambil langkah, ditahan oleh masalah yang berbeda-beda. 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. Perusahaan belum dapat mengoptimalkan pelayanan pada pelanggan misalnya jumlah menu yang terbatas serta tempat duduk yang kurang nyaman. Maka diperlukan pemetaan pelanggan sehingga akan menjadi sangat penting untuk menghadapi masalah yang perusahaan alami.

Kata kunci :

Data Mining; Algoritma K-Means; RapidMiner Studio; KKD; K-Means Clustering

PENDAHULUAN

Hubungan antara perusahaan dan pelanggan sangat penting untuk mendukung ekspansi dan keberlanjutan perusahaan. Manajemen perusahaan harus dapat mengidentifikasi dan mempercayai pelanggan potensial dan mengembangkan pemahaman korporasi tentang keinginan pelanggan mereka sebagai individu sehingga mereka tetap setia kepada perusahaan. salah satu hal terpenting tentang manajemen pelanggan adalah bagaimana perusahaan mempertahankan pelanggan mereka. Komitmen terhadap retensi pelanggan ini sangat penting bagi *Exso Coffee & Roastery* mengingat semakin banyaknya bisnis yang sebanding.

Untuk mencegah jenis penyalahgunaan pelanggan ini, Anda ingin memahami basis pelanggan potensial Anda sehingga Anda akan melindungi mereka dengan menawarkan layanan dan hadiah yang paling sederhana. Perusahaan terus memberikan hadiah kepada pelanggan dalam bentuk *souvenir* setahun sekali.

Pelanggan potensial memainkan peran yang sangat penting dalam pengembangan strategi bisnis, dan pelanggan adalah sumber keuntungan bagi perusahaan. Untuk itu, kami ingin mengenal pelanggan kami dengan baik. pemahaman yang baik tentang pembeli memungkinkan perusahaan untuk mengambil posisi di pelanggan potensial. masalah standar adalah bahwa kesulitan menganalisis nilai pelanggan. Banyak pemasar berjuang untuk menemukan pelanggan atau pelanggan yang tepat. Hal ini dapat menyebabkan perusahaan kehilangan calon pelanggan yang tentunya sangat merugikan perusahaan.

Segmentasi pelanggan mungkin merupakan metode umum untuk memilih pelanggan atau klien yang tepat untuk memulai promosi. Dengan mengelompokkan pelanggan Anda berdasarkan perilaku mereka, Anda akan menargetkan perilaku mereka dengan lebih baik. Peluncuran produk yang disesuaikan, tujuan pemasaran, memenuhi harapan pelanggan, dll. Namun, menganalisis sejumlah besar pelanggan atau data pelanggan membutuhkan banyak usaha dan waktu

Penelitian yang akan kami lakukan adalah aplikasi information mining menggunakan algoritma *K-Means*. Sebelumnya, banyak peneliti di bidang teknologi data atau bidang lainnya telah melakukannya.

*Wahyu Romadhona



METODE PENELITIAN

1. *Knowledge discovery In Database*

Seluruh proses penting untuk menemukan pola dalam data dan mengidentifikasi di mana pola yang ditemukan itu valid, berguna dan mudah dipahami (Ndaumanu & Kusri, n.d.). Secara garis besar proses KDD (Nasari & Darma, 2015) dapat digambarkan sebagai berikut:

- a) Pemilihan data
Pemilihan (seleksi) data dari kumpulan data operasional harus dilakukan sebelum KDD memulai tahap ekstraksi informasi. Data terpilih yang digunakan dalam proses data mining disimpan dalam file terpisah dari database operasional.
- b) Pra-perawatan/pembersihan
Sebelum melakukan proses data mining, Anda perlu melakukan proses pembersihan data yang menjadi fokus KDD. Proses pembersihan meliputi menghapus data duplikat, memeriksa data yang tidak konsisten, dan memperbaiki kesalahan data seperti salah ketik.
- c) Konversi
Encoding adalah transformasi data terpilih sehingga cocok untuk proses data mining. Proses coding di KDD adalah proses kreatif dan sangat bergantung pada jenis atau pola informasi yang Anda cari di database.
- d) Data mining
adalah proses mencari pola dan informasi yang menarik pada data terpilih dengan menggunakan teknik dan metode tertentu. Teknik, metode, atau algoritma penambangan data sangat berbeda. Memilih prosedur atau algoritma yang tepat sangat tergantung pada tujuan dan proses KDD secara keseluruhan (Arumawatu et al., 2015). Demikian pula, kami menggunakan algoritma J48 untuk menerapkan data mining untuk memprediksi kinerja siswa berdasarkan status sosial ekonomi orang tua, motivasi, disiplin siswa, dan nilai masa lalu. Kami sedang menyelidiki.
- e) Interpretasi/Evaluasi
Pola informasi yang dihasilkan dari proses data mining harus disajikan dalam format yang mudah dipahami oleh para pemangku kepentingan. Fase ini merupakan bagian dari proses KDD yang disebut interpretasi. Fase ini memeriksa apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesis yang ada

2. *K-Means Clustering*

Ini adalah algoritma pengelompokan berulang. Algoritma *K-Means* dimulai dengan memilih *K* secara acak, dimana *K* adalah jumlah *cluster* yang terbentuk. Nilai *K* kemudian ditetapkan secara acak, dengan nilai sementara ditetapkan ke pusat *cluster*, yang biasa disebut sebagai *centroid*, *mean*, atau 'rata-rata'. Hitung jarak dari semua data ke setiap *centroid* menggunakan rumus Euclidean untuk mencari jarak terdekat dari semua data ke *centroid*. Klasifikasi setiap data berdasarkan kedekatannya dengan *centroid*. Langkah ini dilakukan sampai nilai *centroid* tidak lagi berubah (stabil) (Putra & Wadisman, 2018).

HASIL PENELITIAN DAN DISKUSI

1. Data Uji

Selama proses implementasi dan analisis data, aplikasi *RapidMiner Studio* digunakan untuk menyimpan semua data dari November 2021 hingga April 2022 di *Exso Coffee & Roastery* dalam database dengan sumber data asli menggunakan *Microsoft Access*. Diuji pada 160 sampel data yang dikonversi ke .xlsx. Formatnya di *Microsoft Excel*.

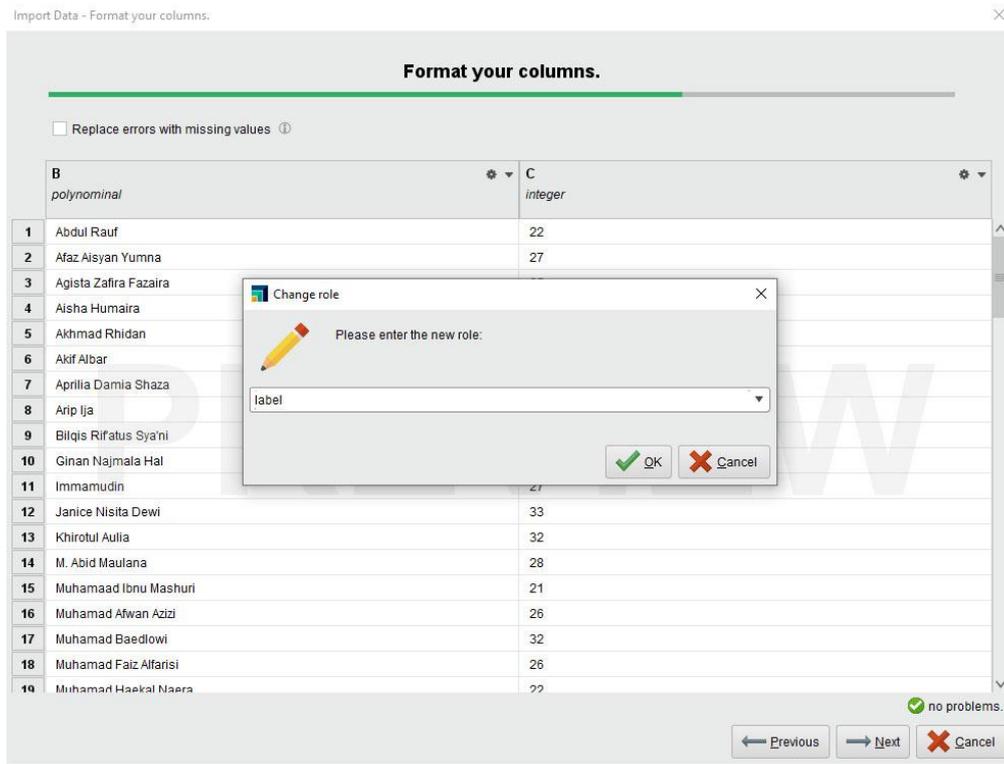
2. Pengujian di *RapidMiner Studio*

Untuk membuktikan kebenarannya, perangkat lunak *RapidMiner Studio* harus digunakan selama fase analisis dan pengujian manual. Berikut adalah langkah-langkah yang harus dilakukan untuk tes ini:

- 1) Impor data dapat dilakukan setelah proses pengumpulan dan pengolahan data survei selesai. Data diproses dalam format .xlsx menggunakan *Microsoft Excel* dan diimpor ke perangkat lunak *RapidMiner Studio*.
- 2) Inisialisasi adalah tindakan menetapkan nilai awal variabel. Dalam penelitian ini, peneliti memvisualisasikan data nama pelanggan sebagai label dan usia pelanggan sebagai bilangan bulat.

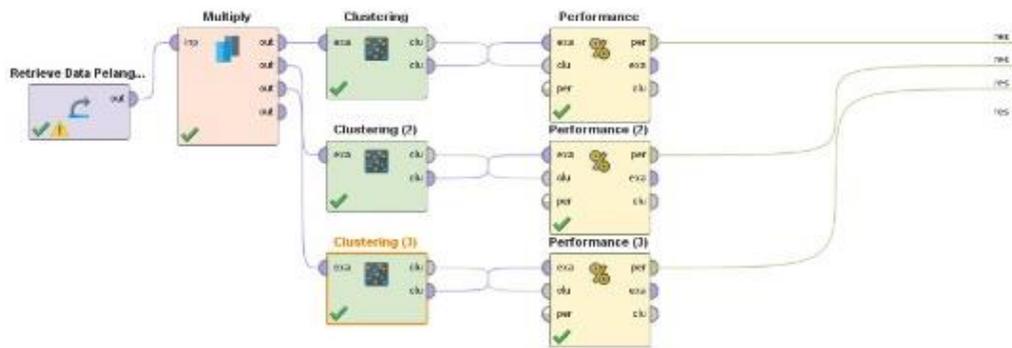
*Wahyu Romadhona





Gambar 1. Tampilan Inisialisasi Data

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



Gambar 2. Tampilan Proses Clustering

- 3) *Clustering* adalah suatu metode pengelompokan data, hal ini merupakan bagian dari *Data Mining* atau Penggalan 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*.
- 4) *Parameters* Selanjutnya yang perlu dilakukan adalah menentukan jumlah *K (parameters)* pada setiap *Cluster*. Dalam hal ini kami memberikan nilai sebagai table berikut:

*Wahyu Romadhona



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.

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

Cluster Model

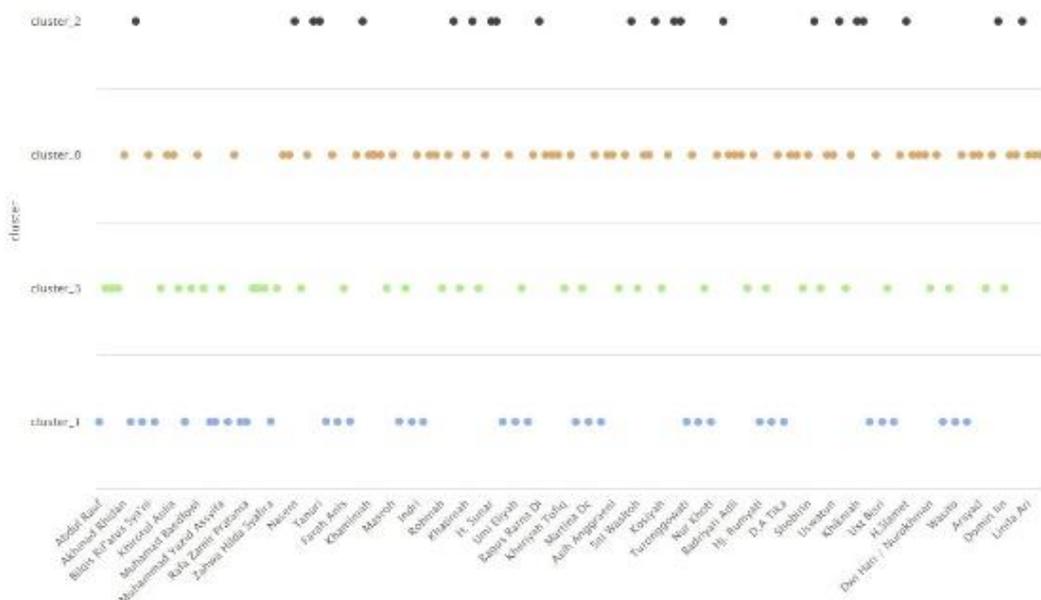
```
Cluster 0: 63 items
Cluster 1: 35 items
Cluster 2: 22 items
Cluster 3: 36 items
Total number of items: 156
```

Gambar 3. Hasil Clustering

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

*Wahyu Romadhona





Gambar 4. Tampilan 3D Hasil Clustering

Dari data diatas bisa disimpulkan 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 *cluster* 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 *cluster* 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.

REFERENSI

Arumawadu, H. I., Rathnayaka, R. M. K. T., & Illangarathne, S. K. (2015). Mining Profitability of Telecommunication Customers Using *K-Means* Clustering. *Journal of Data Analysis and Information Processing*, 03(03), 63–71. <https://doi.org/10.4236/jdaip.2015.33008>

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.

Putra, R. R., & Wadisman, C. (2018). Implementasi Data Mining Pemilihan Pelanggan Potensial Menggunakan Algoritma *K-Means* Implementation of Data Mining for Potential Customer Selection Using *K-Means* Algorithm. *Journal of Information Technology and Computer Science (INTECOMS)*, 1(1), 72–77.

*Wahyu Romadhona



This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.