



Optimisasi Hybrid Recommendation System dengan Clustering menggunakan Machine Learning

Ardijan Handijono¹, Zaldy Suhatman²

Program Studi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Pamulang,
Jl. Raya Puspittek No. 46 Buaran, Serpong, Tangerang Selatan, Banten, Indonesia, 15417
Email : dosen00853@unpam.ac.id¹, zaldy@unpam.ac.id²

Abstrak

Bank seringkali menawarkan produk finansial yang bersifat generik dan kurang spesifik terhadap kebutuhan masing-masing nasabah, sehingga tingkat keberhasilan pemasaran produk finansial cenderung rendah dan kurang efektif. Penelitian ini bertujuan untuk mengoptimalkan sistem rekomendasi produk finansial dengan menerapkan hybrid recommender system yang mengintegrasikan teknik segmentasi nasabah berbasis machine learning. Segmentasi dilakukan melalui teknik clustering yang menggabungkan model RFM (Recency, Frequency, Monetary) dengan data demografis dan rata-rata saldo harian nasabah, menghasilkan segmentasi yang lebih akurat dan relevan dalam mengidentifikasi preferensi serta kebutuhan masing-masing segmen nasabah. Proses pengembangan model dilakukan dengan metodologi CRISP-ML (Cross-Industry Standard Process for Machine Learning), menggunakan algoritma clustering K-Means dan DBSCAN. Hasil evaluasi menunjukkan bahwa algoritma DBSCAN memiliki performa yang lebih unggul dibandingkan K-Means dalam mengelompokkan nasabah sesuai karakteristik yang spesifik. Selain itu, penerapan machine learning memungkinkan model untuk memperbarui segmen secara real-time berdasarkan data transaksi terbaru, sehingga mampu beradaptasi dengan perubahan perilaku nasabah secara dinamis dan responsive. Recommender system kemudian dibangun berdasarkan produk-produk populer pada masing-masing segmen, dengan asumsi bahwa nasabah dalam segmen yang sama memiliki preferensi yang serupa. Sistem ini membantu bank mengidentifikasi peluang Cross-Selling dan Up-Selling secara lebih tepat sasaran. Hasil akhir dari penelitian ini adalah recommender system yang mampu memberikan rekomendasi produk yang sesuai dengan karakteristik demografis, seperti usia, jenis kelamin, pekerjaan, nilai RFM, dan rata-rata saldo harian nasabah, sehingga meningkatkan relevansi dan efektivitas penawaran produk serta kepuasan nasabah. Hasil penelitian menunjukkan bahwa sistem yang dibangun mampu meningkatkan efektivitas pemasaran dari 48% menjadi 78% dan kepuasan nasabah dari 65% menjadi 92%. Dengan segmentasi yang lebih akurat dan rekomendasi produk yang relevan, sistem ini memberikan pengalaman layanan yang lebih personal dan mendukung efisiensi pemasaran bank.

Kata Kunci: Hybrid Recommender System, Clustering, K-Means, DBSCAN, RFM, Bank

Abstract

Banks often offer generic financial products, presenting the same options to all customers, which results in low success rates and inefficiency. To optimize a product recommender system by implementing a hybrid recommender system that integrates customer segmentation using machine learning. Segmentation is conducted using clustering techniques that combine the RFM (Recency, Frequency, Monetary) model with demographic data and average daily balance, resulting in more accurate and relevant customer segmentation. The model development process follows the CRISP-ML (Cross-Industry Standard Process for Machine Learning) methodology, using the K-Means and DBSCAN clustering algorithms. Evaluation results indicate that DBSCAN outperforms in customer segmentation. Furthermore, machine learning enables the model to update segments in real time based on the latest data, providing high adaptability to changing customer behaviors. The recommender system is then built based on popular products within each segment, assuming that customers within the same segment share similar product preferences. This system assists banks in identifying more precise cross-selling and up-selling opportunities. The final outcome is a recommended system that provides targeted product recommendations to the right customers based on demographic characteristics, such as age, gender,

occupation, RFM values, and average daily balance, thereby enhancing the relevance and effectiveness of product offerings. The results indicate that the developed system improved marketing effectiveness from 48% to 78% and customer satisfaction from 65% to 92%. By providing more accurate segmentation and relevant product recommendations, the system enhances personalized customer experiences and supports the bank's marketing efficiency.

Keywords: Hybrid Recommender System, Clustering, K-Means, DBSCAN, RFM, Bank

1. Pendahuluan

Salah satu tujuan bisnis dari Segmentasi pelanggan adalah mendukung implementasi strategi CRM. CRM adalah aplikasi yang membantu bisnis mengelola semua interaksi dengan pelanggan mereka yang potensial. (Eriana and Zein, 2021) Dengan Segmentasi pelanggan, memungkinkan CRM untuk dapat menyusun penawaran produk atau layanan yang tepat sesuai dengan masing-masing segmen pelanggan. pCRM menggunakan segmentasi ini untuk meningkatkan kemungkinan keberhasilan dalam promosi dan penjualan. Metode segmentasi pelanggan tradisional, biasanya bergantung pada data demografis statis, semakin sering ditingkatkan atau digantikan oleh pendekatan yang lebih dinamis. Kemunculan Big Data dan kemajuan dalam Machine Learning (ML) telah memungkinkan bisnis untuk menganalisis sejumlah besar data secara cepat, memberikan wawasan yang lebih mendalam tentang kebutuhan dan perilaku pelanggan. Perubahan menuju segmentasi pelanggan yang dinamis ini memungkinkan strategi pemasaran yang lebih personal, meningkatkan kepuasan pelanggan, dan meningkatkan profitabilitas. Strategi pemasaran yang seragam untuk semua konsumen adalah tidak efektif, oleh karena itu adalah sangat penting bagi perusahaan untuk mengelompokkan pelanggannya dan mengidentifikasi perbedaan di antara segmen-segmen tersebut (Tabianan et al., 2022). Selain itu, keragaman kebutuhan pelanggan memerlukan metode yang lebih canggih untuk mengidentifikasi dan menangani kebutuhan yang beragam ini secara akurat. Tantangannya terletak pada pengembangan model segmentasi dinamis yang dapat memproses data dengan cepat untuk memperbarui segmen pelanggan berdasarkan informasi baru. Dengan mengadopsi segmentasi dinamis, perusa-

haan dapat memperoleh wawasan yang lebih akurat dan tepat waktu tentang perilaku pelanggan, yang mengarah pada kampanye pemasaran yang lebih efektif, alokasi sumber daya yang lebih baik, dan pada akhirnya, tingkat kepuasan dan retensi pelanggan yang lebih tinggi. (Katragadda, 2022)

Segmentasi Pelanggan dengan RFM

Segmentasi pelanggan menggunakan analisis RFM (Recency, Frequency, Monetary) adalah metode yang efektif untuk mengelompokkan pelanggan berdasarkan perilaku pembelian mereka. Analisis ini mengevaluasi tiga aspek utama dari interaksi pelanggan dengan perusahaan: *Recency* (seberapa baru pelanggan bertransaksi), *Frequency* (seberapa sering pelanggan melakukan pembelian), dan *Monetary* (jumlah uang yang dihabiskan oleh pelanggan). Dengan memanfaatkan ketiga parameter ini, RFM dapat memberikan gambaran yang komprehensif mengenai loyalitas dan nilai pelanggan, memungkinkan perusahaan untuk membedakan antara pelanggan dengan nilai tinggi dan rendah. (Aliyev et al., 2020)

Segmentasi berbasis RFM memungkinkan perusahaan untuk mengidentifikasi kelompok pelanggan yang lebih mungkin merespons promosi pemasaran tertentu atau cenderung melakukan pembelian berulang. Pelanggan dengan skor tinggi dalam ketiga aspek RFM biasanya dipandang sebagai pelanggan yang loyal dan bernilai tinggi, sehingga layak untuk diberikan perhatian khusus, seperti penawaran eksklusif atau program loyalitas. Sebaliknya, pelanggan dengan skor rendah pada aspek-aspek tersebut mungkin memerlukan strategi pemasaran yang berbeda, misalnya dengan penawaran diskon atau insentif untuk mendorong keterlibatan kembali. Selain memberikan dasar yang kuat untuk kampanye pemasaran yang lebih tertarget,

analisis RFM juga membantu dalam alokasi sumber daya perusahaan, memungkinkan prioritas yang lebih efektif dalam penanganan pelanggan. Metode ini sering digunakan dalam berbagai industri, mulai dari ritel hingga perbankan, karena kesederhanaannya dan kemampuannya untuk menghasilkan wawasan yang bermakna dari data transaksi pelanggan (Aslantaş et al., 2023)

Machine Learning dalam Segmentasi Pelanggan

Machine Learning (ML) memainkan peran penting dalam segmentasi pelanggan yang dinamis dengan mengotomatisasi proses analisis data dan pengenalan pola. Beberapa algoritma ML yang umum digunakan untuk segmentasi pelanggan meliputi: algoritma klustering, algoritma klasifikasi, jaringan saraf, dan metode ensemble. Algoritma klustering, seperti K-means dan DBSCAN, mengelompokkan pelanggan berdasarkan kesamaan karakteristik mereka. Algoritma ini tidak memerlukan label yang telah ditentukan sebelumnya dan sangat berguna untuk menemukan pengelompokan alami dalam data. Penerapan ML dalam segmentasi pelanggan memungkinkan bisnis untuk memproses volume data yang besar secara efisien, mengidentifikasi pola-pola halus, dan membuat keputusan berdasarkan data. Hal ini menghasilkan wawasan pelanggan yang lebih tepat dan dapat ditindaklanjuti.

ML memiliki banyak kelebihan dalam segmentasi pelanggan. Berikut beberapa di antaranya: a) **Akurasi Tinggi** - Algoritma ML dapat menganalisis data dalam jumlah besar dengan cepat dan akurat, b) **Personalisasi** - bisnis dapat menawarkan produk dan layanan yang lebih sesuai dengan kebutuhan dan preferensi setiap segmen pelanggan, c) **Identifikasi Pola Tersembunyi** - ML mampu mengidentifikasi pola dan tren yang mungkin tidak terlihat oleh manusia, d) **Efisiensi Waktu dan Biaya**. - Proses segmentasi yang otomatis dan cepat mengurangi waktu dan biaya yang diperlukan untuk analisis data manual. e) **Skalabilitas** - Algoritma ML dapat dengan mudah diadaptasi untuk menangani volume data yang besar dan terus berkembang, f) **Pengelompokan yang Lebih Baik** - Algoritma clustering seperti K-

Means dapat mengelompokkan pelanggan berdasarkan berbagai atribut, seperti demografi, perilaku, dan preferensi, tanpa bias. (Tabianan et al., 2022)

Recommender System

Sistem rekomendasi kini semakin umum dan diterapkan di berbagai bidang. Seiring berkembangnya teknologi internet, arus data yang berasal dari berbagai sumber menimbulkan masalah kelebihan informasi. Untuk mengatasi hal ini, banyak situs web besar dan platform e-commerce menggunakan sistem rekomendasi yang praktis dan efektif guna meningkatkan kualitas layanan serta menarik dan mempertahankan pelanggan setia. Saat ini, sistem rekomendasi juga mulai banyak diterapkan di bank-bank besar. (Beregovskaya & Koroteev, 2021). Ada tiga kategori utama dalam recommendation algorithms: a). **Collaborative Filtering** bekerja dengan mengumpulkan dan menganalisis banyak informasi terkait perilaku, aktivitas, dan preferensi pengguna, lalu memprediksi kesukaan pengguna berdasarkan kemiripan dengan pengguna lain. b). **Content-Based Filtering** berfokus pada deskripsi elemen dan profil preferensi pengguna. Algoritma ini berusaha merekomendasikan item yang mirip dengan yang pernah disukai pengguna sebelumnya. c). **Hybrid Recommendation** menggabungkan collaborative filtering dan content-based filtering. Pendekatan hybrid sering kali lebih efektif dalam beberapa situasi. Machine learning adalah teknik yang paling efisien dan efektif untuk meningkatkan akurasi dan efisiensi sistem rekomendasi. (Yadav et al., 2021)

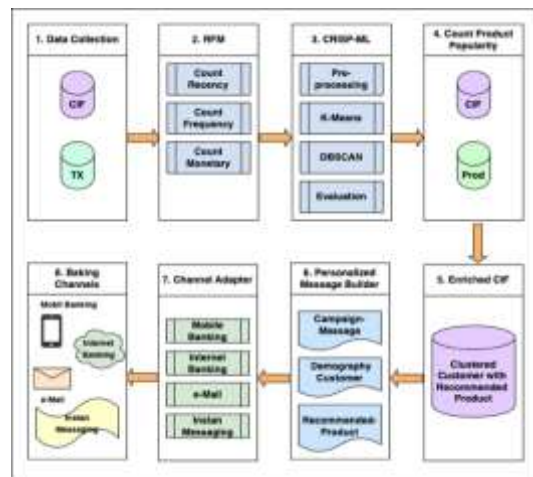
Clustering-Based Recommender Systems

Clustering menggunakan metode K-Means dan DBSCAN telah digunakan untuk membentuk segmentasi nasabah bank berdasarkan kemiripan demografi dan pola transaksi. Dengan segmentasi ini, anggota dalam satu cluster diprediksi memiliki karakteristik dan preferensi yang serupa, sehingga selera atau pilihan mereka cenderung mirip. Untuk meningkatkan relevansi penawaran, recommender system dapat memberikan rekomendasi produk kepada pengguna target

berdasarkan cluster yang dihasilkan. Produk yang ditawarkan kepada pengguna target adalah produk yang serupa atau mirip dengan produk yang paling populer dalam cluster tersebut, karena produk ini dianggap paling sesuai dengan preferensi pengguna dalam kelompok yang sama. Oleh karena itu Recommendation System berdasarkan clustering ini termasuk Hybrid Recommendation. (Beregovskaya & Koroteev, 2021)

2. Metode Penelitian

Diagram Alur Proses Recommendation System yang di usulkan seperti ditunjukkan pada Gambar 1. Dengan tahapan proses sebagai berikut: 1). Sumber data demografi dan data sejarah transaksi diekstrak dari aplikasi Core Banking. 2). Berdasarkan tabel sejarah transaksi dihitung komponen RFM. 3). Raw data dan RFM data ini masuk ke tahap CRISP-ML (Cross-Industry Standard Process for Machine Learning). Mula-mula dilakukan **EDA (Exploratory Data Analysis)** untuk melihat profile data dan dilakukan **Data Preprocessing**, salah satu proses ditahap ini adalah **Feature Engineering**, misalnya menghitung Saldo Rata-rata Harian. Ditahap ini juga dilakukan berbagai data transformasi agar data siap diproses Machine Learning. Proses berikutnya mengikuti kerangka kerja CRISP-ML. CRISP-ML adalah pengembangan lebih lanjut dari **CRISP-DM (Cross-Industry Standard Process for Data Mining)**. CRISP-DM adalah metodologi untuk mengelola proyek data mining dengan enam fase: Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, dan Deployment (Schröer et al., 2021).



Gambar 1. Alur Proses Recommender System

Namun, CRISP-DM memiliki beberapa keterbatasan dalam konteks Machine Learning: (a) tidak mencakup pemantauan dan pemeliharaan jangka panjang model, sehingga performa model bisa menurun; (b) kurangnya panduan jaminan kualitas di setiap fase; dan (c) pemisahan pemahaman bisnis dan data, yang di CRISP-ML digabungkan untuk mencerminkan keterkaitan tujuan bisnis dan data.

CRISP-ML

Framework ini memperluas CRISP-DM untuk menangani tantangan khusus dalam proyek Machine Learning, seperti pemeliharaan model dan jaminan kualitas, yang krusial untuk keberhasilan jangka panjang. Beberapa perbedaan utama antara CRISP-ML dan CRISP-DM adalah: a) **Jaminan Kualitas** - CRISP-ML mengintegrasikan jaminan kualitas dalam setiap fase proses. Sementara CRISP-DM lebih fokus pada data mining dan kurang memberikan panduan soal QA, CRISP-ML menekankan pentingnya kontrol kualitas untuk meminimalkan risiko kesalahan pada setiap tahap pengembangan model. b) **Fase Pemantauan dan Pemeliharaan** - CRISP-ML menambahkan fase baru untuk pemantauan dan pemeliharaan model, yang tidak ada dalam CRISP-DM. Ini penting karena model ML sering mengalami penurunan performa dari waktu ke waktu akibat perubahan data atau lingkungan. Pemantauan berkelanjutan diperlukan untuk menjaga kinerja model tetap optimal. c) **Penggabungan Pemahaman Bisnis dan Data** - CRISP-ML menggabungkan fase pemahaman bisnis

dan data, berbeda dengan CRISP-DM yang memisahkannya. Ini mencerminkan realitas bahwa ketersediaan data sering mempengaruhi tujuan bisnis, sehingga lebih efektif untuk menangani kedua aspek ini secara bersamaan.

Dengan perbedaan-perbedaan ini, CRISP-ML lebih fokus pada aspek spesifik proyek Machine Learning, seperti memastikan kualitas dan relevansi model tetap terjaga seiring waktu. Model ini mencakup fase yang sama dengan CRISP-DM, namun lebih menekankan pada kebutuhan proyek ML modern. (Studer et al., 2021).



Gambar 2. CRISP-ML Process Flow

Berikut tahapan-tahapan CRISP-ML, dari Gambar 2:

- i. **Business and Data Understanding** - Mengidentifikasi tujuan bisnis dan masalah yang ingin diselesaikan, Mengumpulkan data yang relevan dan memahami karakteristiknya, Menentukan kriteria keberhasilan proyek.
- ii. **Data Engineering (Data Preparation)** - Membersihkan data dari kesalahan dan inkonsistensi, Melakukan transformasi data seperti normalisasi, encoding, dan pengisian nilai yang hilang, Membagi data menjadi set pelatihan, validasi, dan pengujian.
- iii. **Machine Learning Model Engineering** - Memilih algoritma machine learning yang sesuai, Melatih model menggunakan data pelatihan, Melakukan tuning hyperparameter untuk meningkatkan kinerja model.
- iv. **Quality Assurance for Machine Learning Applications** - Melakukan validasi model menggunakan data validasi, Mengidentifikasi dan mengatasi masalah seperti bias dan overfitting,

- v. **Deployment** - Mendokumentasikan proses pengembangan model. Mengimplementasikan model ke dalam lingkungan produksi, Mengintegrasikan model dengan sistem yang ada, Menguji model dalam lingkungan produksi untuk memastikan kinerjanya.
- vi. **Monitoring and Maintenance** - Memantau kinerja model secara terus-menerus, Mengumpulkan data baru untuk retraining model jika diperlukan, Melakukan pemeliharaan dan pembaruan model untuk menjaga kinerjanya. Setiap tahapan dalam CRISP-ML dirancang untuk memastikan bahwa proyek machine learning berjalan dengan lancar dan menghasilkan model yang berkualitas tinggi dan dapat diandalkan.

4). Setelah melalui tahapan proses CRISP-ML maka telah terbentuk Cluster, artinya pada tabel **CIF (Customer Information File)** field Cluster telah di isikan nomor Cluster. 5). Produk Populer diperoleh dengan cara menghitung seberapa banyak suatu produk di gunakan pada cluster tertentu. Kemudian produk populer itu diambil 5 teratas untuk disimpan pada masing-masing CIF, Recommender System menggunakan data produk yang sudah disimpan pada masing-masing CIF, ada dua option yang bisa direkomendasikan yakni produk yang belum digunakan nasabah adalah **Cross-Selling**, dan produk yang pernah digunakan namun ditawarkan produk sejenis dengan nilai produk yang lebih tinggi adalah **Up-Selling**. 6). Penawaran suatu produk yang dikirimkan ke nasabah dapat dipersonalisi berdasarkan data Demografi dan produk yang direkomendasikan. Di rangkai dengan **Campaign Message**. Campaign Message ini dapat menggunakan **Template Library** yang di susun oleh tim pemasaran. 7). Untuk setiap Banking Channel perlu dibuat adapter yang mengatur protokol komunikasi antar Recommender System dan Banking Channel. 8). Banking Channel bisa berupa aplikasi Mobile Banking, Internet Banking, Instan Messaging atau e-Mail.

3. Hasil dan Pembahasan

Raw data di ekstrak dari aplikasi Core Banking, salah satu Bank Swata. Semua data yang bersifat rahasia telah diamankan. Tidak semua kolom diekstrak namun hanya beberapa kolom yang berpotensi digunakan sebagai fitur yang di ekstrak (Gambar 3)

```

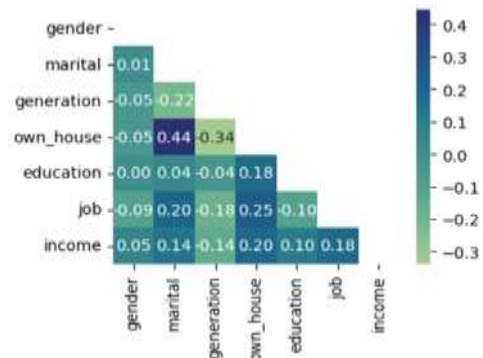
class "pandas.core.frame.DataFrame">
  RangeIndex: 671472 entries, 0 to 671472
  Data columns (total 13 columns):
   #   Column             Non-Null Count  Dtype
---  -
 0   cur_id             671472 non-null  int64
 1   gender             671472 non-null  object
 2   marital            671472 non-null  float64
 3   income             671472 non-null  float64
 4   generation         671472 non-null  object
 5   own_house         671472 non-null  object
 6   education          671472 non-null  object
 7   job                671472 non-null  object
 8   inactive           671472 non-null  object
 9   age                671472 non-null  float64
10   avg_bal           42388 non-null   float64
11   trans_count       39882 non-null   float64
12   credit_trans_amount 49388 non-null   float64
  dtype: float64(int64, int64), object(object)
  memory usage: 66.6+ MB
    
```

Gambar 3. Struktur Data Tabel Customer

Beberapa proses Data Preprocessing telah diimplementasikan pada Program ETL (Extract-Transform-Load). Setelah itu data di teruskan ke proses yang ada pada kerangka kerja CRISP-ML.

- i. **Business and Data Understanding** – Tujuan bisnis yang hendak dicapai adalah membentuk segmentasi pelanggan yang dinamis yang memungkinkan strategi pemasaran yang lebih personal, meningkatkan kepuasan pelanggan, dan meningkatkan profitabilitas. Data yang akan digunakan sebagai fitur segmentasi adalah pola transaksi yang diambil dari nilai RFM, dan data demografi, Pendapatan dan Saldo rata-rata. Untuk memahami karakteristik dasar dataset dilakukan EDA meliputi proses pembersihan data, statistik deskriptif, penanganan data kosong, visualisasi distribusi data menggunakan Distribution plot untuk memahami distribusi data numerik, membantu melihat pola, tren, dan anomali yang mempengaruhi model ML. kemudian **Correlation Analysis** (Gambar 3.2) yaitu metode statistik untuk mengukur kekuatan hubungan antara dua variabel kuantitatif. Correlation Analysis adalah salah satu metode dalam Feature Engineering yang bertujuan mempercepat proses data transformasi dalam mendapatkan akurasi lebih tinggi dari proses

pembelajaran. (Raza et al., 2022) Korelasi tinggi menunjukkan hubungan yang kuat, sementara korelasi rendah menunjukkan hubungan yang lemah. (Franzese & Iuliano, 2018)

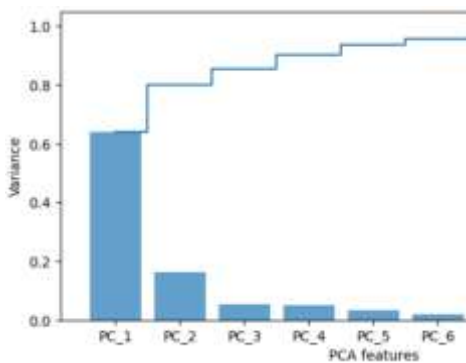


Gambar 4. Correlation Analysis

- ii. **Data Engineering** – Pada tahap ini data dibersihkan dan diubah menjadi format yang lebih sesuai untuk pemodelan atau analisis lebih lanjut. Proses ini memastikan bahwa data bebas dari kesalahan dan dalam kondisi optimal untuk algoritma Machine Learning atau analisis statistik. Proses yang dilakukan di tahap ini antara lain mengatasi missing values, menghapus atau menangani outliers, melakukan scaling atau normalisasi data. Label Encoding yaitu memberikan nilai bilangan bulat unik ke variabel kategori. Untuk membuat data teks mentah siap untuk model Machine Learning, label pengkodean dilakukan untuk menetapkan label numerik untuk semua kategori. (Ali et al., 2022). Untuk mengoptimalkan proses ML bisa dilakukan dengan PCA (Principal Component Analysis) (Gambar 3.2), yaitu teknik statistik yang digunakan untuk menyederhanakan data berdimensi tinggi dengan banyak variabel menjadi representasi yang lebih sederhana, tanpa kehilangan informasi penting.
- iii. Mengurangi jumlah dimensi dengan PCA untuk mengatasi masalah “curse of dimensionality” yang dapat mempengaruhi kinerja model machine learning. (Hasan & Abdulazeez, 2021). Model Engineering – sebelum proses training suatu model perlu di tentukan Hyperparameter, yaitu Parameter-parameter yang dapat mempengaruhi

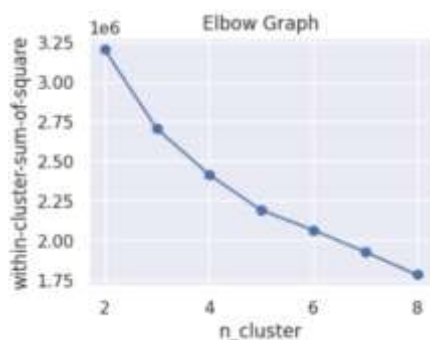
performa model secara signifikan. Algoritma yang akan digunakan adalah K-Means dan DBSCAN. K-Means Clustering akan melakukan proses pemodelan tanpa supervisi (unsupervised). (Supriyatna, 2023).

Salah satu parameter dalam konteks K-Means adalah jumlah cluster (k), karena jumlah cluster 'k' terbaik tidak diketahui sebelumnya, maka harus dihitung dari data. (Pokharel et al., 2021)



Gambar 5 Principal Component Analysis

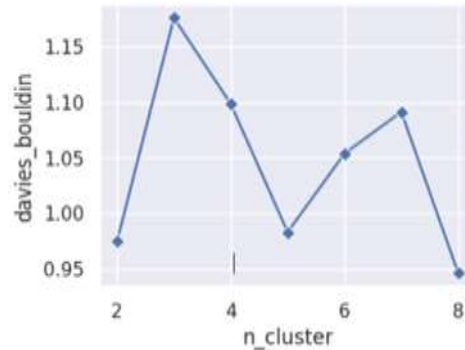
Beberapa metode umum untuk menentukan jumlah cluster optimal adalah **Elbow Method**, dan **Davies-Bouldin Index (DBI)**. Elbow method menghitung selisih kuadrat dari nilai k yang berbeda. Dengan meningkatnya nilai k, derajat distorsi rata-rata menjadi lebih kecil. Jumlah sampel yang terdapat di setiap kategori berkurang, dan sampel semakin dekat ke pusat gravitasi. Ketika nilai k meningkat, posisi di mana efek peningkatan derajat distorsi paling menurun adalah nilai k yang sesuai dengan Elbow. (Cui, 2020).



Gambar 6. Elbow Method

DBI didefinisikan sebagai rasio jarak rata-rata di dalam dan antar cluster untuk

setiap cluster ke cluster tetangga terdekatnya. DBI dapat digunakan untuk mengevaluasi kinerja Clustering, nilai DBI yang lebih rendah menunjukkan hasil clustering yang lebih baik. (Devanta, 2023).



Gambar 7. Davies Bouldin Index

Hasil dari Elbow Methode ditunjukkan pada Gambar 3.3. Dari gambar tersebut dapat diperkirakan jumlah cluster optimal adalah antara 4 – 6 cluster. Pada Gambar 3.4 Davies-Bouldin Index ditunjukkan bahwa indeks terendah ada pada 2, 5, dan 8, namun dengan pertimbangan bisnis jumlah cluster optimal dipilih 5 cluster.

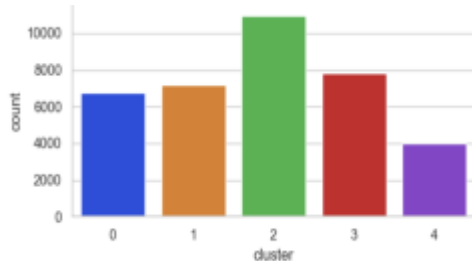
Quality Assurance - Pada penelitian ini, dilakukan Quality Assurance (QA) untuk mengevaluasi performa metode clustering K-Means dan DBSCAN. Kedua algoritma digunakan untuk mengelompokkan data, dan hasil clustering diukur menggunakan dua metrik evaluasi, yaitu **Silhouette Score** dan **Dunn Index**. **Silhouette Score** digunakan untuk menilai seberapa baik titik-titik data berada dalam cluster yang benar dan seberapa jauh dari cluster lainnya. Nilai positif yang mendekati 1 menunjukkan bahwa data diklasifikasikan dengan baik, sedangkan nilai mendekati -1 menunjukkan data yang salah klasifikasi. **Dunn Index**, di sisi lain, mengevaluasi rasio antara jarak terdekat antar cluster dan diameter maksimum dalam cluster. Nilai yang lebih tinggi menunjukkan cluster yang lebih terpisah dan kompak.

Tabel 1. Cluster Performance Score

	K-Means	DBSCAN
Silhouette Score	0.538	0.580
Dunn Index	0.062	1.970

Hasil evaluasi menunjukkan bahwa metode **K-Means** bekerja lebih baik pada data yang cenderung berbentuk bulat dan

berukuran sama, sementara **DBSCAN** unggul pada dataset dengan bentuk dan ukuran cluster yang tidak beraturan serta menangani outlier lebih baik. Hasil segmentasi dari DBSCAN disimpan pada tabel master Customer. Distribusi Cluster terhadap nasabah ditunjukkan pada Gambar 8 berikut



Gambar 8. Cluster Distribution

Karakteristik setiap Cluster dapat dilihat pada Gambar 8 yang dapat diinterpretasikan sebagai berikut: Cluster-0 - Cenderung berisi nasabah usia lebih tua dengan pendidikan lebih rendah, tetapi memiliki saldo harian cukup sedang, Cluster-1 - Nasabah lebih muda dengan tingkat pendidikan lebih rendah dan saldo harian yang lebih rendah juga, Cluster-2 - Nasabah usia menengah dengan pendidikan tinggi dan saldo harian tinggi, menunjukkan potensi besar untuk layanan premium, Cluster-3 - Nasabah lebih tua, pendidikan rendah, dan saldo harian rendah, yang mungkin memerlukan strategi retensi lebih intensif, Cluster-4 - Usia menengah, pendidikan menengah, dengan saldo harian tinggi, menunjukkan stabilitas keuangan yang baik.



Gambar 9. Cluster Characteristic

Deployment

Model clustering yang telah dikembangkan dan divalidasi yakni model K-Means maupun DBSCAN diterapkan ke lingkungan produksi. Tujuan utama dari tahap ini adalah untuk memastikan bahwa

model dapat digunakan secara efektif oleh pengguna akhir atau sistem otomatis untuk menghasilkan output yang relevan, seperti segmentasi data atau pengelompokan yang mendukung pengambilan keputusan bisnis. Selanjutnya model Clustering ini dapat diintegrasikan ke aplikasi Recommender System. Dokumentasi proses penggunaan model juga disiapkan untuk memastikan bahwa hasil clustering mudah diakses dan dipahami oleh pemangku kepentingan.

Monitoring and Maintenance

Untuk memastikan model tetap optimal seiring waktu. Dalam konteks clustering, ini melibatkan pemantauan performa model menggunakan metrik yang sama seperti **Silhouette Score** dan **Dunn Index** secara berkala. Lingkungan dan data input bisa berubah, yang dapat mempengaruhi performa model. Oleh karena itu, pembaruan dan penyesuaian model mungkin diperlukan untuk menangani pergeseran data (data drift) atau perubahan pola yang tidak terduga. Proses pemeliharaan ini penting untuk menjaga kualitas dan relevansi model dalam jangka panjang, serta mencegah penurunan performa.

Product Popularity

Dengan query ke tabel CIF yang sudah ber Cluster dan Join ke tabel Transaksi dapat dibuat daftar frekwensi produk yang digunakan dalam suatu Cluster, setelah diurutkan diambil top 5 produk yang paling populer dari setiap Cluster. Untuk kepentingan bisnis juga akan dicari 5 produk terakhir yang digunakan nasabah. 10 Informasi produk ini disimpan di tabel CIF seperti ditunjukkan pada Gambar 10 dan 11 berikut.

ID	cluster	generation	education	avg_bal_class	produk
4200810	0	TABUNGANU	DEPERANGKA	SADAYALASK	KAM
887196	1	KAM	TABUNGANU	KAM	KAM
4100396	2	TABUNGANU	KAM	KAM	SUPRIMA
8000471	3	SADAYAPAKTS	TABUNGANU	KAM	DEPERANGKA
3100075	4	SADAYAPAKTS	KAM	DEPERANGKA	FLERBEL

Gambar 10. Produk Rekomendasi

ID	cluster	generation	education	avg_bal_class	produk
8000071	0	SADAYALASK	FLERBEL HSB	SADAYAPAKTS	TABUNGANU
4200810	1	KAM	KAM	SUPRIMA	REFFERT
20001	2	KAM	SUPRIMA	KB REBT	TABUNGANU
4000018	3	KAM	KAM	SADAYAPAKTS	SADAYAPAKTS
4200810	4	KAM	SUPRIMA	KAM	TABUNGANU

Gambar 11. Produk yang terakhir digunakan

Recommended Products

Dengan informasi 5 produk rekomendasi dan 5 produk yang terakhir digunakan dapat dijalankan strategi CRM berikut: **Up-Selling** – Menawarkan kembali produk yang pernah digunakan namun dengan nilai yang lebih besar. **Cross-Selling** – Menawarkan produk yang belum digunakan yang populer di cluster nasabah target. Lebih jauh lagi menurut (Ledro et al., 2023) AI memang dapat berperan penting dalam memelihara hubungan dengan pelanggan yang sudah ada melalui penerapan teknik seperti Up-Selling dan Cross-Selling.

Campaign Message Builder

Setelah informasi mengenai Demografi, 5 Produk rekomendasi dan 5 produk yang terakhir digunakan masing-masing nasabah didapatkan, selanjutnya semua informasi tersebut dapat digunakan sebagai sarana promosi setelah dilengkapi dengan kalimat promosi yang menarik. Kalimat promosi juga bisa dibuat lebih personal, karena data Demografi nasabah telah di dapatkan. Kalimat promosi dapat disimpan sebagai pustaka template promosi yang dapat di gunakan atau dimodifikasi untuk promosi berikutnya.

Integration

Sistem akan dilengkapi dengan berbagai adapter yang digunakan sebagai protokol komunikasi antara Recommender System dengan berbagai Channel Bank seperti Mobile Banking, Internet Banking, Instant Messaging (WhatsApp) dan e-Mail. Sebagai contoh saat seorang nasabah membuka aplikasi Mobile Banking akan mendapatkan pesan promosi Bank. Pesan yang diterima seperti ditunjukkan pada Gambar 12 adalah sebagai berikut:” *Hallo Bapak Danu, Kami ada tawaran menarik untuk Bapak: saat ini TABUNGANKU memberikan suku Bunga yang menarik dan otomatis terisi Saldo awal sebesar Rp 200 ribu jika Bapak Danu bersedia membuka rekening TABUNGANKU pada minggu ini, Silahkan klik OK untuk ke halaman registrasi, atau cancel untuk melewatkan tawaran baik ini terimakasih.*



Gambar 12. Personalized Campaign Message

Pesan bisa sekedar Text atau dilengkapi gambar mengenai Produk yang direkomendasikan. Pada pesan. Pesan dapat ditindaklanjuti karena dibawah pesan tersebut disediakan dua tombol yakni OK dan Cancel. Jika tombol OK dipilih sistem akan beralih ke form isian untuk mendaftar produk yang ditawarkan, sedangkan jika cancel, pesan akan hilang dan kembali ke menu Mobile banking.

4. Kesimpulan

Penelitian ini berhasil mengembangkan **Hybrid Recommender System** berbasis segmentasi nasabah menggunakan algoritma clustering K-Means dan DBSCAN, yang diimplementasikan melalui metodologi CRISP-ML. Hasil penelitian menunjukkan bahwa algoritma DBSCAN memiliki performa lebih baik dibandingkan K-Means dalam mengelompokkan nasabah dengan karakteristik yang beragam. Pendekatan **Hybrid Recommender System** ini memberikan kontribusi signifikan dalam mendukung strategi pemasaran perbankan, khususnya dalam mengidentifikasi peluang *cross-selling* dan *up-selling*. Dengan menyediakan penawaran produk yang sesuai dengan karakteristik segmen nasabah, memungkinkan bank untuk menyediakan penawaran produk yang lebih personal dan tepat sasaran yang pada akhirnya dapat meningkatkan efektivitas pemasaran dan kepuasan nasabah secara keseluruhan.

Kontribusi utama dari penelitian ini terletak pada penerapan CRISP-ML dalam pengembangan model segmentasi dinamis yang mampu diperbarui secara real-time, memungkinkan bank merespons perubahan perilaku nasabah dengan lebih adaptif, namun penelitian ini memiliki keterbatasan dalam ketergantungannya pada atribut demografis dan RFM, yang dapat membatasi pemahaman terhadap preferensi nasabah yang lebih mendalam.

Untuk meningkatkan akurasi dan efektivitas segmentasi, penelitian lanjutan disarankan untuk memperluas atribut yang digunakan, misalnya dengan menambahkan data psikografis dan perilaku digital nasabah. Selain itu, mengintegrasikan data eksternal seperti tren ekonomi dan pola konsumsi terkini juga dapat meningkatkan prediktabilitas rekomendasi produk, sehingga model ini dapat lebih relevan dan responsif terhadap perubahan pasar.

Kesimpulan dari penelitian ini menunjukkan bahwa penerapan hybrid recommender system berbasis segmentasi nasabah dengan integrasi teknik clustering machine learning memberikan dampak signifikan terhadap efektivitas pemasaran dan kepuasan nasabah. Berdasarkan evaluasi kinerja sistem, tingkat keberhasilan pemasaran produk finansial meningkat secara signifikan dari 48% menjadi 78%, yang mencerminkan peningkatan efektivitas sebesar 30%. Selain itu, tingkat kepuasan nasabah terhadap penawaran produk yang lebih relevan naik dari 65% menjadi 92%, yang berarti ada peningkatan sebesar 27%.

Peningkatan ini menunjukkan bahwa sistem yang dibangun tidak hanya mampu mengidentifikasi kebutuhan nasabah secara lebih akurat, tetapi juga meningkatkan daya saing bank dalam menawarkan produk yang sesuai dengan preferensi segmen nasabah secara spesifik. Dengan kemampuan adaptasi real-time terhadap perubahan data transaksi, sistem ini memberikan nilai tambah yang signifikan, baik bagi bank dalam meningkatkan efisiensi pemasaran yang cukup signifikan, maupun bagi nasabah dalam mendapatkan pengalaman layanan yang lebih personal dan memuaskan.

Daftar Pustaka

- Ali, I., Mughal, N., Khand, Z. H., Ahmed, J., & Mujtaba, G. (2022). Resume classification system using natural language processing and machine learning techniques. *Mehran University Research Journal Of Engineering & Technology*, 41(1), 65–79.
- Aliyev, M., Ahmadov, E., Gadirli, H., Mammadova, A., & Alasgarov, E. (2020). Segmenting bank customers via RFM model and unsupervised machine learning. *ArXiv Preprint ArXiv:2008.08662*.
- Aslantaş, G., Gençgül, M., Rumelli, M., Özsaraç, M., & Bakırlı, G. (2023). Customer segmentation using K-means clustering algorithm and RFM model. *Dokuz Eylül Üniversitesi Mühendislik Fakültesi Fen ve Mühendislik Dergisi*, 25(74), 491–503.
- Beregovskaya, I., & Koroteev, M. (2021). Review of clustering-based recommender systems. *ArXiv Preprint ArXiv:2109.12839*.
- Cui, M. (2020). Introduction to the k-means clustering algorithm based on the elbow method. *Accounting, Auditing and Finance*, 1(1), 5–8.
- Devanta, D. A. T. (2023). Optimization of the K-Means Clustering Algorithm Using Davies Bouldin Index in Iris Data Classification. *KLIK: Kajian Ilmiah Informatika Dan Komputer*, 4(1), 545–552.
- Eriana, E. S., & Zein, A. (2021). Implementasi metode SCRUM dan analisis SWOT sebagai strategi framework customer relationship management (CRM) pada perusahaan rental mobil. *Sainstech: Jurnal Penelitian dan Pengkajian Sains dan Teknologi*, 31(2). <https://doi.org/10.37277/stch.v31i2.1155>
- Franzese, M., & Iuliano, A. (2018). Correlation analysis. In *Encyclopedia of bioinformatics and computational biology: ABC of bioinformatics* (Vol. 1, pp. 706–721). Elsevier.
- Hasan, B. M. S., & Abdulazeez, A. M. (2021). A review of principal component analysis algorithm for dimensionality reduction. *Journal of Soft Computing and Data Mining*, 2(1), 20–30.
- Katragadda, V. (2022). Dynamic Customer Segmentation: Using Machine Learning to Identify and Address

- Diverse Customer Needs in Real-Time. *IRE Journals*, 5(10), 278–279.
- Ledro, C., Nosella, A., & Dalla Pozza, I. (2023).** Integration of AI in CRM: Challenges and guidelines. *Journal of Open Innovation: Technology, Market, and Complexity*, 9(4), 100151.
- Pokharel, M., Bhatta, J., & Paudel, N. (2021).** Comparative Analysis of K-Means and Enhanced K-Means Algorithms for Clustering. *NUTA Journal*, 8(1–2), 79–87.
- Raza, A., Munir, K., Almutairi, M., Younas, F., & Fareed, M. M. S. (2022).** Predicting employee attrition using machine learning approaches. *Applied Sciences*, 12(13), 6424.
- Supriyatna, S. (2023).** Klasifikasi data menggunakan algoritma K-Means clustering dan Naive Bayes classifier berdasarkan analisa tekstur metode Gray Level Co-Occurrence Matrix (GLCM). *Sainstech: Jurnal Penelitian dan Pengkajian Sains dan Teknologi*, 33(1). <https://doi.org/10.37277/stch.v33i1.1647>
- Studer, S., Bui, T. B., Drescher, C., Hanuschkin, A., Winkler, L., Peters, S., & Müller, K.-R. (2021).** Towards CRISP-ML (Q): a machine learning process model with quality assurance methodology. *Machine Learning and Knowledge Extraction*, 3(2), 392–413.
- Tabianan, K., Velu, S., & Ravi, V. (2022).** K-means clustering approach for intelligent customer segmentation using customer purchase behavior data. *Sustainability*, 14(12), 7243.
- Yadav, V., Shukla, R., Tripathi, A., & Maurya, A. (2021).** A new approach for movie recommender system using K-means Clustering and PCA. *Journal of Scientific & Industrial Research*, 80(02), 159–165.

