Perbandingan Algoritma DBSCAN dan *Affinity Propagation* dalam Segmentasi Pelanggan Tenant Inkubator Bisnis

Dedy Panji Agustino[1]\*, I Gede Bintang Arya Budaya[2], I Gede Harsemadi[3], I Komang Dharmendra[4], I Made Suandana Astika Pande[5]

Prodi Sistem Informasi ITB STIKOM Bali [1], [3], [4], [5], Prodi Teknologi Informasi ITB STIKOM Bali [2]

ITB STIKOM Bali

Denpasar, Indonesia

panji@stikom-bali.ac.id[1]\*

***Abstract*— *The increasingly complex business environment necessitates businesses to design more effective and efficient development strategies, including market expansion. Customer data analysis has become essential in helping businesses understand customer needs and behaviors. One approach for grouping customers is segmentation based on RFM (Recency, Frequency, and Monetary) analysis. This study aims to compare the performance of the DBSCAN and Affinity Propagation (AP) clustering algorithms in providing customer profile segment recommendations using RFM analysis. The transaction dataset used in this research is derived from one of the STIKOM Bali business incubator tenants. The dataset underwent preprocessing stages before being clustered using both the DBSCAN and AP algorithms. The evaluation of the algorithms' performance was conducted using Silhouette Scores and Davies-Bouldin Index (DBI) matrices. The results demonstrate that the AP algorithm outperforms DBSCAN, with AP yielding a Silhouette Score of 0.699 and a DBI of 0.429, and recommending four customer segments. Further analysis of the AP results was conducted using a statistical approach based on the mean values of each segment for RFM variables. The four segments resulting from the AP algorithm, based on the mean RFM variables, exhibit a clear correlation with the concept of customer relationship management. Consequently, this research formulates several strategy recommendations based on the resulting segments.***

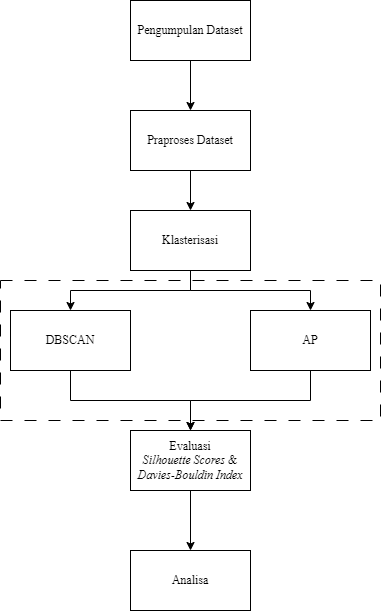
***Keywords—*** ***customer profiling*, *customer relationship manajemen*, *clustering*, *RFM.***

***Abstrak*—Lingkungan bisnis yang semakin kompleks menyebabkan suatu bisnis harus mampu merancang strategi pengembangan perusahaan yang lebih efektif dan efisien, termasuk dalam pengembangan pasar. Analisa data pelanggan menjadi mutlak diperlukan untuk membantu bisnis memahami kebutuhan dan perilaku pelanggan. Salah satu pendekatan yang dapat digunakan untuk mengelompokkan pelanggan adalah segmentasi berdasarkan analisa RFM (*Recency, Frequency, dan Monetary*). Studi ini bertujuan untuk membandingkan kinerja algoritma klasterisasi DBSCAN dan *Affinity Propagation* (AP) dalam memberikan rekomendasi segmen profil pelanggan menggunakan analisa RFM. Dataset transaksi yang digunakan dalam penelitian ini berasal dari salah satu tenant inkubator bisnis STIKOM Bali. Dataset tersebut melalui tahapan praproses dataset dan selanjutnya diklasterisasi menggunakan kedua algoritma DBSCAN dan AP. Tahapan evaluasi kinerja algoritma dilakukan dengan menggunakan matriks *Silhouette Scores* dan Davies-Bouldin Index (DBI). Hasil penelitian menunjukkan bahwa algoritma *AP* memiliki kinerja yang lebih baik ketimbang DBSCAN, dimana AP menghasilkan *Silhouette Scores* dengan nilai 0.699 dan DBI dengan nilai 0.429, serta dengan rekomendasi 4 segmen pelanggan. Analisa lanjutan untuk hasil AP juga dilakukan dengan pendekatan statistik berdasarkan nilai *mean* dari setiap segmen untuk variabel RFM. Hasil keempat segmen dari algoritma AP berdasarkan *mean* variabel RFM menunjukkan korelasi yang jelas dengan konsep *customer relationship management.* Sehingga dalam penelitian ini dapat dirumuskan beberapa rekomendasi strategi berdasarkan segmen yang dihasilkan.**

***Kata Kunci— profiling* pelanggan, *manajemen relasi pelanggan*, klasterisasi*,* RFM**

# Pendahuluan

Dalam lingkungan bisnis yang semakin kompleks, persaingan antar bisnis semakin ketat. Strategi untuk menjaga loyalitas pelanggan dan hubungan jangka panjang antara pelanggan dan pemilik bisnis menjadi salah satu poin yang sangat penting. Salah satu cara untuk mencapai hal tersebut adalah dengan mengenali lebih dalam pelanggan bisnis tersebut disebut juga sebagai *customer profiling*[1]. Salah satu teknik yang dapat digunakan dalam *customer profiling* adalah segmentasi pelanggan dengan menggunakan pendekatan *unsupervised learning. Customer profiling* merupakan bagian dari *Customer Relationship Management* (CRM) dan dapat membantu memilah data pelanggan yang menguntungkan. Beberapa penelitian seperti [2], [3] menggaris bawahi pentingnya mengetahui segmentasi pelanggan untuk meningkatkan penjualan dan keuntungan perusahaan serta potensi pemanfaatan *unsupervised learning*.

Inkubator Bisnis STIKOM Bali adalah salah satu inkubator perguruan tinggi yang memberikan layanan inkubasi bagi perusahaan rintisan. Berdasarkan data dan hasil wawancara dengan pengelola inkubator, tantangan yang dihadapi oleh tenant adalah ketidakefektifan proses pemasaran dan strategi untuk menjaga loyalitas pelanggan. Rencana dan strategi yang tidak tepat dapat menyebabkan penggunaan sumber daya yang tidak efektif. Pada penelitian [4] diterapkan skema segmentasi pelanggan dengan menggunakan algoritma K-Means, dimana dihasilkan 2 klaster pelanggan. Pada penelitian [5] **memanfaatkan algoritma *Fuzzy C Means* dan berdasarkan hasil segmentasi ditemukan 7 klaster pelanggan. Kedua algoritma ini memberikan hasil klaster yang berbeda, meski secara informasi dapat ditemukan alasan objektif untuk pengelompokkan data tersebut.

Dalam proses segmentasi pelanggan tidak ada algoritma yang mampu memiliki kinerja baik pada setiap kasus segmentasi, termasuk dalam segmentasi pelanggan ini. Oleh karena itu peneliti ingin menganalisa pemanfaatan algoritma klasterisasi lainnya yaitu *Density-based Spatial Clustering of Application with Noise* (DBSCAN) dan *Affinity Propagation* (AP) dalam kasus segmentasi data pelanggan untuk menentukan profil pelanggan. DBSCAN mampu mendeteksi klaster dengan bentuk yang tidak linear dan ukuran yang besar, berkat pendekatan berbasis kerapatan (*density*) dan secara eksplisit membedakan *noise* dari *data points* yang merupakan anggota klaster, memungkinkan algoritma untuk mengelompokkan data dengan lebih akurat pada kasus dimana *noise* ada [6], [7].

AP sendiri merupakan algoritma yang cenderung lebih stabil karena algoritma ini mencari solusi yang mengoptimalkan keseluruhan data [8], [9] dan tidak bergantung pada inisialisasi *centroid* awal seperti pada K Means. DBSCAN dan AP sama – sama algoritma yang tidak memerlukan inisiasi awal jumlah klaster. Penelitian ini berfokus pada penerapan DBSCAN dan AP untuk menemukan solusi klaster optimal dalam hal segmentasi data pelanggan. Hasil segmentasi akan dievaluasi dan dibandingkan dengan kinerja pada 2 algortima sebelumnya dan dianalisa terkait hasil yang didapatkan dari DBSCAN dan AP tersebut.

# Metodologi

Adapun metode yang digunakan dalam pelaksanaan penelitian ini, dapat dilihat pada Gambar. 1 berikut.

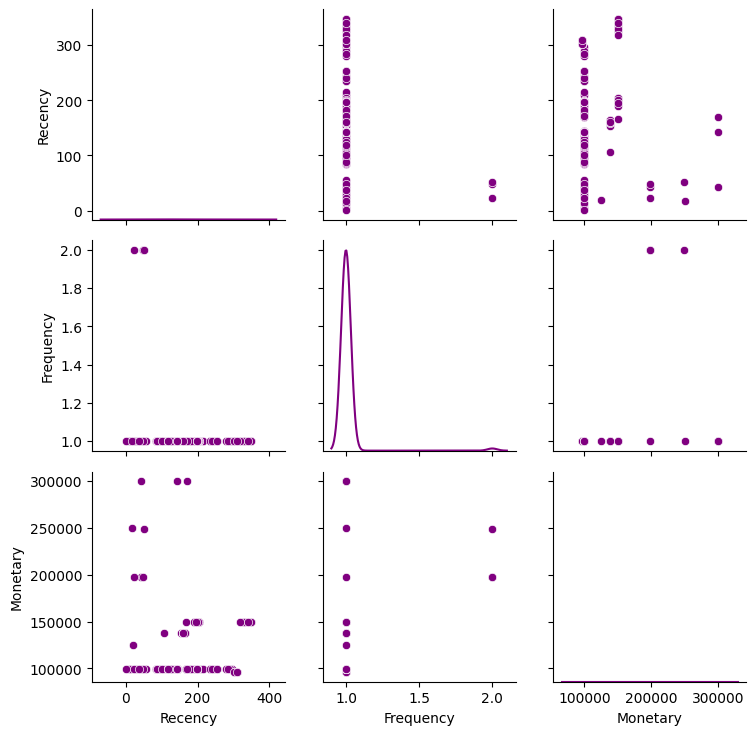
## Pengumpulan Dataset

Pertama adalah tahap pengumpulan dataset, pada tahap ini dilaksanakan proses pengumpulan data yang dalam kasus ini adalah data transaksi dari salah satu tenant inkubator bisnis STIKOM Bali, data merupakan hasil transaksi pada tahun 2022 sebanyak 282 baris transakasi yang merupakan transaksi masuk dari usaha. Selanjutnya berkas transaksi yang didapatkan disimpan dalam dokoumen berbentuk *excel* dalam format *.xls*. Data transkasi sebagai dataset terdiri dari lima kolom, yaitu email pelanggan, ID transaksi, tanggal transaksi, tipe transaksi, dan jumlah transaksi.

## Praproses Dataset

Kedua adalah tahap praproses dari dataset, pada tahap ini dilaksanakan proses persiapan data sebelum digunakan dalam algoritma klasterisasi. Tahapan ini merupakan tahap tang penting dengan tujuan untuk memastikan bahwa dataset yang digunakan sudah sesuai dengan kebutuhan dari proses klastersisasi yang akan dilakukan. Adapun langkah – langkah dalam tahap praproses dalam kasus penelitian ini adalah sebagai berikut.

Gambar. 1 Metode penelitian.

* + 1. Membaca data: Data transaksi yang berupa *file excel* dengan format .xls disimpan menggunakan *pandas* dalam sebuah *Data Frame*.
    2. Pemilihan variabel: dalam data transaksi dipilih variabel yang digunakan dalam pembuatan model klasterisasi. Dalam kasus ini variabel data transaksi yang digunakan yaitu, email pelanggan sebagai unik ID, tanggal transaksi dan jumlah transaksi.
    3. Mengecek nilai *null*, jika ada nilai *null* maka baris data tersebut dihapus dari *Data Frame.* Dalam kasus ini tidak ada nilai *null* yang ditemukan.
    4. Perhitungan RFM (*Recency, Frequency, Monetary*):
       1. *Recency*: Menghitung jumlah hari antara tanggal transaksi terbaru untuk setiap pelanggan dan tanggal referensi (dalam kasus ini, tanggal transaksi terbaru adalah tanggal terbaru dalam dataset + 1 hari).
       2. *Frequency*: Menghitung jumlah transaksi yang dilakukan oleh setiap pelanggan berdasarkan email pelanggan.
       3. *Monetary*: Menghitung total uang yang dihabiskan oleh setiap pelanggan berdasarkan email pelanggan.
    5. Finalisasi data RFM: Setelah menghitung nilai *Recency, Frequency, dan Monetary* hasil perhitungan itu disimpan dalam *Data Frame* RFM. Gambar. 2 menampilkan hasil visualisasi data sebelum dilakukan proses klasterisasi dan TABEL I menampilkan sampel dari isi dataset RFM.
    6. Normalisasi data: Fitur *Recency, Frequency*, dan *Monetary* memiliki rentang nilai yang berbeda oleh karena itu perlu dilakukan normalisasi pada data tersebut. Sehingga algoritma klasterisasi tidak didominasi oleh fitur dengan rentang nilai yang lebih besar dalam kasus ini adalah fitur *monetary*. Dalam kasus ini, digunakan *Standard Scaler*. TABEL II menampilkan sampel dataset setelah dinormalisasi.

Gambar. 2 Visualisasi distribusi data RFM sebelum klasterisasi.

##### TABEL I. Sampel Dataset Setelah Perhitungan RFM

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| No | ***Email*** | ***Recency*** | ***Frequency*** | ***Monetary*** |
| 1 | e1@gmail.com | 41 | 1 | 99000 |
| 3 | e2@gmail.com | 20 | 1 | 125000 |
| 4 | e3@gmail.com | 118 | 1 | 99000 |
| … | …………………… | ………. | ………. | ………. |
| 282 | e282@gmail.com | 143 | 1 | 99000 |

##### TABEL II. Sampel Dataset Setelah Dilakukan Normalisasi

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| No | ***Email*** | ***Recency*** | ***Frequency*** | ***Monetary*** |
| 1 | e1@gmail.com | -1.582898 | -0.103695 | -0.391658 |
| 3 | e2@gmail.com | -1.871522 | -0.103695 | 0.473033 |
| 4 | e3@gmail.com | -0.524611 | -0.103695 | -0.391658 |
| … | …………………… | ………. | ………. | ………. |
| 282 | e282@gmail.com | -0.181011 | -0.103695 | -0.391658 |

## Klasterisasi

Dalam tahap klasterisasi, menggunakan dua algoritma yaitu algoritma DBSCAN dan AP.

* + 1. Algoritma DBSCAN

Algoritma DBSCAN adalah algoritma mampu mendeteksi klaster dengan bentuk yang tidak linear dan ukuran yang besar, berkat pendekatan berbasis kerapatan (*density*) dan secara eksplisit membedakan *noise* dari *data points* yang merupakan anggota klaster, memungkinkan algoritma untuk mengelompokkan data dengan lebih akurat pada kasus dimana *noise* ada [6], [7].

* + 1. Algoritma AP

Algoritma AP sendiri merupakan algoritma yang cenderung lebih stabil karena algoritma ini mencari solusi yang mengoptimalkan keseluruhan data [8], [9] dan tidak bergantung pada inisialisasi *centroid* awal seperti pada K Means.

* 1. *Evaluasi*

Pada tahap evaluasi menggunakan dua matriks, pertama adalah *Silhouette Scores* dan *Davies-Bouldin Index (DBI).*

1. *Silhouette Score* mengukur sejauh mana objek dalam satu klaster mirip satu sama lain dibandingkan dengan objek di klaster lain [10],[11]. Nilai *Silhouette Score* berada pada rentang antara -1 hingga 1, di mana nilai yang mendekati 1 atau yang paling tinggi berarti klaster tersebut memiliki kualitas yang baik.

(1)

dimana *a(i)* adalah rata-rata jarak antara objek *i* dan objek lain dalam klaster yang sama, dan *b(i)* adalah jarak rata – rata minimum antara objek *i* dan objek dalam klaster lain.

1. *Davies-Bouldin Index* mengukur kualitas dari klasterisasi berdasarkan rasio jarak dalam klaster dan jarak antar klaster. Berbeda dengan *Silhouette* Scores, nilai Davies-Bouldin Index yang lebih rendah menunjukkan kualitas klaster yang lebih baik [12], [13].

(2)

Dimana *n* adalah representasi dari jumlah klaster, *s(i)* adalah jarak rata-rata dalam klaster *i*, *s(j)* adalah jarak rata-rata dalam klaster *j*, dan *d(i, j)* adalah jarak antara *centroid* dari klaster i dan j. DBI adalah rata-rata dari rasio maksimum untuk setiap klaster yang ada.

* 1. *Analisa*

Dalam tahapan analisa, untuk lebih memahami karakteristik dari masing-masing klaster yang dihasilkan oleh algoritma DBSCAN dan AP, maka dihitung nilai rata-rata (*mean*) dari setiap variabel dalam klaster, yaitu *Recency, Frequency*, dan *Monetary*. Menghitung nilai rata - rata dari setiap variabel ini akan memberikan gambaran tentang bagaimana pelanggan dalam klaster tersebut berperilaku secara umum, dan rata – rata ini dapat memberikan gambaran lebih jelas untuk membandingkan karakteristik antara klaster.

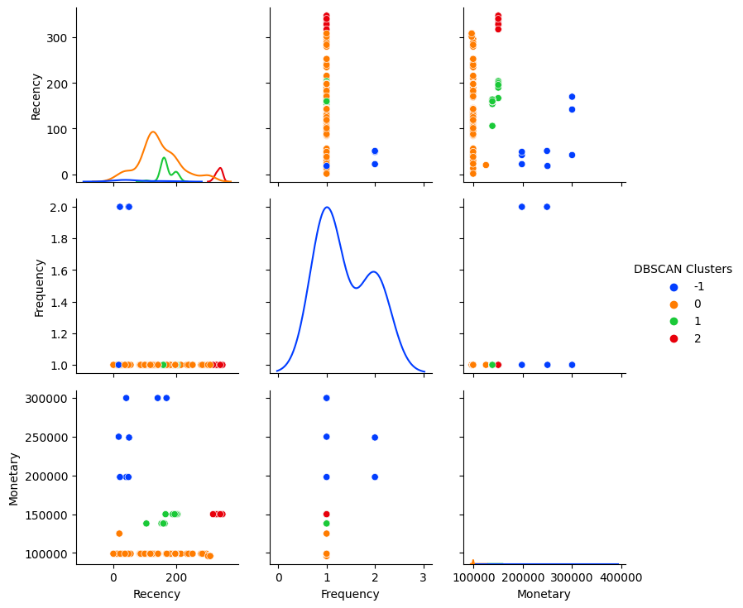
# Hasil Dan Pembahasan

Hasil klasterisasi dengan DBSCAN dan AP adalah sebagai berikut.

##### TABEL III. Hasil Evaluasi Dengan Matriks *Silhouette Scroes* dan DBI

|  |  |  |  |
| --- | --- | --- | --- |
| No | **Algoritma** | ***Silhouette Scores*** | ***DBI*** |
| 1 | DBSCAN | 0.499 | 0.525 |
| 3 | AP | 0.699 | 0.429 |

* 1. *DBSCAN*

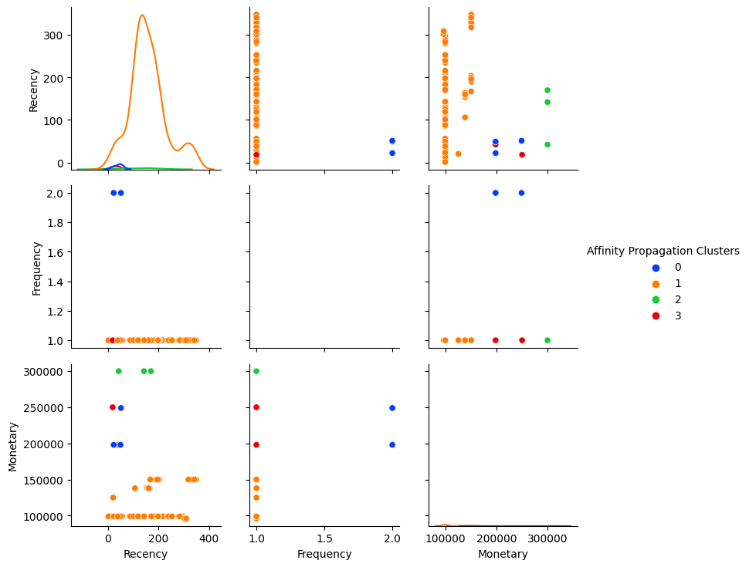
**Pada proses klasterisasi dengan DBSCAN dilakukan beberapa kali percobaan dengan pengaturan parameter yaitu parameter epsilon dan parameter minimal sampel poin. DBSCAN mampu menemukan *noise* pada dataset sehingga proses evaluasi dengan menggunakan *silhouette scores* dan DBI dihitung setelah *noise* dikeluarkan dari proses evaluasi. Berdasarkan hasil evaluasi ditemukan bahwa nilai optimal untuk parameter epsilon adalah 0,9 dan parameter minimal sampel poin adalah 9 dengan hasil seperti yang terlihat pada TABEL III.

Gambar. 3 Visualisasi distribusi data setelah diklasterisasi dengan algoritma DBSCAN.

Jumlah klaster yang direkomendasikan adalah 4 klaster. Gambar. 3 menampilkan visualisasi distribusi data dengan label kelas masing - masing dari hasil klasterisasi dengan algoritma DBSCAN. Klaster pertama adalah *noise* yang diberikan label -1, ada sebanyak 8 data yang termasuk dalam klaster ini. Klaster kedua dengan label klaster 0 memiliki 227 data yang masuk, yang juga menjadi klaster dengan ukuran terbesar. Klaster ketiga dengan label klaster 1 memiliki 34 data, dan yang terakhir klaster keempat dengan label klaster 2 memiliki 13 data.

* 1. *AP*

Pada proses klasterisasi dengan AP dilakukan beberapa kali percobaan dengan pengaturan parameter yaitu parameter *damping*. AP dievaluasi dengan menggunakan *silhouette scores* dan DBI. Berdasarkan hasil evaluasi ditemukan bahwa nilai optimal untuk parameter *damping* adalah 0,9 dan iterasi maksimal 500, dengan hasil seperti yang terlihat pada TABEL III. Jumlah klaster yang direkomendasikan adalah 4 klaster. Gambar. 4 menampilkan visualisasi distribusi data dengan label kelas masing - masing dari hasil klasterisasi dengan menggunakan algoritma AP. Klaster pertama diberikan label 0, ada sebanyak 3 data yang termasuk dalam klaster ini. Klaster kedua dengan label klaster 1 memiliki 274 data yang masuk, yang juga menjadi klaster dengan ukuran terbesar. Klaster ketiga dengan label klaster 2 memiliki 3 data, dan yang terakhir klaster keempat dengan label klaster 3 memiliki 3 data.

**

Gambar. 4 Visualisasi distribusi data setelah diklasterisasi dengan algoritma AP.

* 1. *Analisa Kinerja DBSCAN dan AP*

Berdasarkan hasil evaluasi dengan *silhouette scores* dan DBI dapat diketahui bahwa dalam kasus dataset yang digunakan algoritma AP memiliki kinerja yang lebih baik dari algoritma DBSCAN. Seperti yang juga terlihat pada Gambar. 3 dan Gambar. 4, dimana secara hasil visualisasi pembagian klaster pada algoritma AP lebih jelas perbedaan antar klasternya dari pada algoritma DBSCAN.

Algoritma AP secara otomatis dapat menentukan jumlah klaster yang optimal berdasarkan similaritas antara titik data. Oleh karena itu, AP lebih baik dalam mengenali pola yang muncul dalam data. Sementara itu, jumlah klaster pada DBSCAN tergantung pada pemilihan parameter epsilon dan minimal sampel, yang bisa lebih sulit untuk menemukan kombinasi yang optimal. AP juga tidak menggunakan konsep kepadatan seperti DBSCAN, sehingga lebih mampu mengelompokkan data yang memiliki variasi kepadatan yang besar antar klaster.

Hal ini menjadikan AP lebih *robust* terhadap perbedaan dalam distribusi data yang ada. AP lebih tahan terhadap *noise* dibandingkan dengan DBSCAN. *Noise* adalah titik data yang tidak dapat diatribusikan ke salah satu klaster dan dianggap sebagai *outlier*. DBSCAN mengidentifikasi *noise* sebagai titik data yang tidak termasuk dalam klaster manapun, yang bisa menyebabkan klaster yang kurang jelas jika *noise* cukup signifikan. Sementara itu, AP lebih mampu menangani *noise* dengan mengelompokkannya ke klaster yang paling mirip.

*D. Analisa Segmentasi Pelanggan*

Terkait dengan proses segmentasi pelanggan untuk mengetahui karakteristik pelanggan dilakukan eksplorasi berdasarkan hasil klasterisasi dari masing masing algoritma. Dimana jumlah segmen bagi pelanggan sepadan dengan jumlah klaster rekomendasi dari masing – masing algoritma. Pada TABEL IV dan TABEL V menampilkan hasil perhitungan dari nilai rata – rata variabel RFM dari keempat kelas, pada masing – masing algoritma.

##### TABEL IV. Nilai *Mean* Tiap Kelas Variabel RFM Pada Hasil Klasterisasi Algoritma DBSCAN

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Segmen DBSCAN | Jumlah | *Mean Recency* | *Mean Frequency* | *Mean Monetary* |
| 1 | 8 | 67.00 | 1.37 | 249,125.00 |
| 2 | 227 | 146.69 | 1.00 | 98,995.59 |
| 3 | 34 | 171.38 | 1.00 | 141,882.35 |
| 4 | 13 | 336.77 | 1.00 | 150,000.00 |

##### TABEL V. Nilai *Mean* Tiap Kelas Variabel RFM Pada Hasil Klasterisasi Algoritma AP

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Segmen AP | Jumlah | *Mean Recency* | *Mean Frequency* | *Mean Monetary* |
| 1 | 3 | 40.67 | 2.00 | 215,000.00 |
| 2 | 274 | 158.77 | 1.00 | 106,737.22 |
| 3 | 3 | 118.00 | 1.00 | 300,000.00 |
| 4 | 3 | 30.00 | 1.00 | 224,000.00 |

Pada hasil pembuatan segmen pelanggan dari klasterisasi dengan algoritma DBSCAN, segmen pertama yang dianggap sebagai *noise* memiliki nilai *recency* yang paling kecil dan nilai *frequency,* serta *monetary* yang paling besar. Apabila dilihat secara keseluruhan, data dikelompokkan sebagai *noise* ini memang tidak memiliki keteraturan pola, sehingga segmen ini tidak direkomendasikan menjadi peioritas sasaran bisnis. Berbeda dengan tiga segmen lainnya, dimana segmen kedua yang juga segmen dengan jumlah data terbesar, pengelompokkannya dipengaruhi oleh variabel *monetary*, rata – rata biaya yang dikeluarkan lebih murah dari pada biaya pada segmen lainnya. Pelanggan yang masuk dalam segmen ini adalah pelanggan yang berfokus pada biaya, sehingga pemberian diskon pada setiap layanan bisnis dapat menjadi strategi marketing yang kuat untuk menyasar segmen ini.

Pada segmen ketiga dipengaruhi oleh faktor *recency* dan *monetary*, dimana pelanggan pada segmen ini paling tidak baru berbelanja selama 6 bulan terakhir, dan mau membayar dengan harga yang lebih tinggi. Pelanggan yang masuk dalam segmen ini perlu didekati kembali, selain dengan diskon, pengiriman informasi layanan bisnis dan program secara konsisten melalui email misalnya dapat dilakukan untuk meningkatkan potensi pelanggan berbelanja kembali. Pada segmen ke 4 adalah pelanggan yang waktu belanja terakhir sudah lebih dari 6 bulan, strategi pada segmen sebelumnya dapat dilakukan, namun segmen ini bisa dikategorikan sebagai segmen yang memiliki prioritas lebih rendah.

Pada hasil pembuatan segmen pelanggan dari klasterisasi dengan algoritma AP, secara keseluruhan dapat disimpulkan bahwa segmen yang dihasilkan memiliki karakteristik yang lebih jelas dari pada hasil dari algoritma DBSCAN, seperti yang terlihat pada TABEL V. Hal ini memudahkan dari segi interpretasi dan menarik korelasi dari *customer relationship management* (CRM)berdasarkan *customer value pyramid* yang selanjutnya menjadi dasar dalam perancangan strategi marketing bagi bisnis [14]–[16]. Segmen pertama terdiri dari 3 data dengan pengaruh dari variabel *frequency* dimana hasil klasterisasi ini berdasarkan pelanggan yang sudah melakukan transaksi lebih dari sekali. Pelanggan yang masuk kedalam segmen ini bisa disebut sebagai pelanggan loyal. Pelanggan pada segmen ini perlu diberikan diskon khusus dan kesempatan untuk merekomendasikan produk bisnis ke orang lain melalui skema *referral link* contohnya.

Segmen kedua memiliki jumlah pelanggan terbanyak dan dapat terlihat dengan jelas bahwa mereka yang masuk segmen ini dipengaruhi variabel *monetary* dengan nilai yang paling rendah. Untuk menargetkan pelanggan dari segmen ini pelaku bisnis dapat memberikan diskon secara berkala untuk produk dari bisnisnya. Segmen ketiga adalah pelanggan yang mau membayar dengan biaya yang lebih tinggi ketimbang pelanggan lainnya, dan transaksi terakhir mereka rata – rata diatas 3 bulan, untuk menyasar pelanggan pada segmen ini, pelaku bisnis perlu melakukan kontak kembali dengan pelanggan tersebut, dan dapat memberikan kesempatan kepada pelanggan tersebut untuk menggunakan produk bisnis secara gratis dalam periode tertentu (*trial*). Hal ini bertujuan untuk mengingatkan kembali pelanggan tersebut terkait produk dari bisnis, dan meningkatkan potensi pelanggan tersebut mau bertransaksi kembali. Segmen keempat adalah pelanggan yang secara variabel *recency* atau waktu belanja yang paling dekat. Pelanggan dalam segmen ini masih memiliki ingatan yang jelas terkait pengalaman dalam proses pembelian dan penggunaan produk bisnis. Pelanggan ini perlu diberikan pelayanan khusus, seperti layanan purna jual yang terkait dengan produk bisnis yang telah dibeli dan perlu dipastikan bahwa tidak ada kendala dari produk bisnis yang telah dibeli. Secara garis besar keempat segmen pelanggan hasil klasterisasi dengan algoritma AP memiliki karakteristik yang lebih jelas ketimbang hasil segmen dari algoritma DBSCAN, sehinngga dalam perumusan strategi CRM berdasarkan piramida pelanggan dapat lebih sesuai.

Penelitian sebelumnya [4] dengan menggunakan algoritma *K Means* menghasilkan 2 segmen pelanggan dengan nilai *silhouette scores* yang lebih tinggi dan DBI yang lebih rendah dari pada algoritma AP. Hal ini berarti kinerja algoritma *K Means* masih lebih baik ketimbang algoritma AP dalam hal pembedaan karakteristik. Namun hasil rekomendasi segmen dari algoritma AP dapat diperhitungkan terkait mampu merepresentasikan kompleksitas dari data yang digunakan dengan menghasilkan 4 segmen sebagai profil pelanggan, serta mampu memberikan karakteristik yang jelas dan dapat dijadikan dasar merancang strategi CRM yang lebih mengkhusus sesuai segmen – segmen tersebut.

# Kesimpulan

Penelitian ini bertujuan untuk menganalisa kemampuan algoritma DBSCAN dan AP untuk membuat segmentasi pelanggan berbasis model data RFM. Dataset RFM adalah model dataset yang dibuat dari data catatan transaksi, yang dimana dalam kasus ini adalah data transaksi dari salah satu tenant inkubator bisnis STIKOM Bali. Data tersebut melalui tahapan praproses sebelum digunakan dalam algoritma, dan hasilnya dievaluasi dengan menggunakan *silhouette scores* dan DBI. Berdasarkan hasil evaluasi ditemukan bahwa algoritma DBSCAN dan AP sama – sama merekomendasikan 4 segmen pelanggan, namun AP memiliki kinerja yang lebih baik ketimbang DBSCAN. Setelah dianalisa secara statistik melalui nilai *mean* bagi setiap segmen terkait variabel RFM dari algoritma DBSCAN dan AP, AP memiliki karakteristik yang lebih jelas ketimbang DBSCAN. Sehingga hasil segmen AP lebih mudah diinterpretasikan dan dikorelasikan dengan analisa CRM. Dataset pada penelitian ini berasal dari satu tenant dan transaksi hanya dalam kurun waktu 1 tahun, yang berarti untuk strategi CRM yang dihasilkan tidak bisa digeneralisasi. Kedepan penelitian dapat dilaksanakan dengan menggunakan data transaksi lebih dari satu tahun untuk mengetahui karakteristik pelanggan yang lebih kompleks. Model untuk membuat segmen pelanggan tersebut juga dapat menerapkan algoritma klasterisasi lainnya seperti K Medoids atau menggunakan pendekatan *hybrid* dimana DBSCAN digunakan untuk mendeteksi *outlier* lalu klasterisasi lanjutan menggunakan K Means.

# Ucapan Terimaksih

Kami mengucapkan terimakasih kepada Institut Teknologi dan Bisnis (ITB) STIKOM Bali atas dukungan pelaksanaan penelitian melalui skema program penelitian internal pada tahun 2023 dan terimakasih juga kepada tenant Inkubator Bisnis STIKOM Bali atas kesediannya memberikan informasi dan data usaha.

##### References

[1] S. Hwang and Y. Lee, “Identifying customer priority for new products in target marketing: Using RFM model and TextRank,” *Marketing*, vol. 17, no. 2, pp. 125–136, 2021.

[2] K. Khalili-Damghani, F. Abdi, and S. Abolmakarem, “Hybrid soft computing approach based on clustering, rule mining, and decision tree analysis for customer segmentation problem: Real case of customer-centric industries,” *Applied Soft Computing Journal*, vol. 73, pp. 816–828, Dec. 2018, doi: 10.1016/j.asoc.2018.09.001.

[3] P. Anitha and M. M. Patil, “RFM model for customer purchase behavior using K-Means algorithm,” *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 5, pp. 1785–1792, May 2022, doi: 10.1016/j.jksuci.2019.12.011.

[4] D. P. Agustino, I. G. Harsemadi, and I. G. B. A. Budaya, “Edutech Digital Start-Up Customer Profiling Based on RFM Data Model Using K-Means Clustering,” *Journal of Information Systems and Informatics*, vol. 4, no. 3, pp. 724–736, 2022.

[5] I. G. Harsemadi, D. P. Agustino, and I. G. B. A. Budaya, “Klasterisasi Pelanggan Tenant Inkubator Bisnis STIKOM Bali Untuk Strategi Manajemen Relasi Dengan Menggunakan Fuzzy C-Means,” *JTIM: Jurnal Teknologi Informasi dan Multimedia*, vol. 4, no. 4, pp. 232–243, 2023.

[6] K. Khan, S. U. Rehman, K. Aziz, S. Fong, and S. Sarasvady, “DBSCAN: Past, present and future,” in *The fifth international conference on the applications of digital information and web technologies (ICADIWT 2014)*, IEEE, 2014, pp. 232–238.

[7] E. Schubert, J. Sander, M. Ester, H. P. Kriegel, and X. Xu, “DBSCAN revisited, revisited: why and how you should (still) use DBSCAN,” *ACM Transactions on Database Systems (TODS)*, vol. 42, no. 3, pp. 1–21, 2017.

[8] P. Kalpana, A. M. Nivetha, R. Madhumitha, P. Tamijeselvy, and S. S. Devi, “Improvisation of spectral clustering through affinity propagation,” in *AIP Conference Proceedings*, AIP Publishing LLC, 2021, p. 030004.

[9] H. Keller, H. Möllering, T. Schneider, and H. Yalame, “Balancing quality and efficiency in private clustering with affinity propagation,” *Cryptology ePrint Archive*, 2021.

[10] D.-T. Dinh, T. Fujinami, and V.-N. Huynh, “Estimating the optimal number of clusters in categorical data clustering by silhouette coefficient,” in *International Symposium on Knowledge and Systems Sciences*, Springer, 2019, pp. 1–17.

[11] K. R. Shahapure and C. Nicholas, “Cluster quality analysis using silhouette score,” in *2020 IEEE 7th international conference on data science and advanced analytics (DSAA)*, IEEE, 2020, pp. 747–748.

[12] I. M. K. Karo, K. MaulanaAdhinugraha, and A. F. Huda, “A cluster validity for spatial clustering based on davies bouldin index and Polygon Dissimilarity function,” in *2017 second international conference on informatics and computing (ICIC)*, IEEE, 2017, pp. 1–6.

[13] C. Legány, S. Juhász, and A. Babos, “Cluster validity measurement techniques,” in *Proceedings of the 5th WSEAS international conference on artificial intelligence, knowledge engineering and data bases*, World Scientific and Engineering Academy and Society (WSEAS) Stevens Point …, 2006, pp. 388–393.

[14] A. Curry and J. Curry, *The customer marketing method: how to implement and profit from customer relationship management*. Simon and Schuster, 2002.

[15] T. BELHADJ, “Customer Value Analysis Using Weighted RFM model: Empirical Case Study”.

[16] R. Srivastava, “Identification of customer clusters using RFM model: a case of diverse purchaser classification,” *International Journal of Information, Business and Management*, vol. 9, no. 4, p. 201, 2017.