

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

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

Prodi Sistem Informasi ITB STIKOM Bali <sup>[1], [3], [4], [5]</sup>, Prodi Teknologi Informasi ITB STIKOM Bali <sup>[2]</sup>

ITB STIKOM Bali

Denpasar, Indonesia

panji@stikom-bali.ac.id<sup>[1]\*</sup>, bintang@stikom-bali.ac.id<sup>[2]</sup>, harsemadi@stikom-bali.ac.id<sup>[3]</sup>, dharmendra@stikom-bali.ac.id<sup>[4]</sup>, suandana@stikom-bali.ac.id<sup>[5]</sup>

**Abstract**— The increasingly complex business environment necessitates businesses to design more effective and efficient strategies for company development, including market expansion. To understand customer behaviors, customer data analysis becomes crucial. One common approach used to group customers is segmentation based on RFM analysis (Recency, Frequency, and Monetary). This study aims to compare the performance of two clustering algorithms, namely DBSCAN and Affinity Propagation (AP), in providing customer profile segment recommendations using RFM analysis. DBSCAN algorithm is employed due to its ability to identify arbitrarily shaped clusters and handle data noise. On the other hand, Affinity Propagation (AP) algorithm is chosen for its capability to discover cluster centers without requiring a pre-defined number of clusters. The transaction dataset used in this research is obtained from one of the business incubator tenants at STIKOM Bali. The dataset undergoes preprocessing steps before being segmented using both DBSCAN and AP algorithms. Performance evaluation of the algorithms is conducted using the Silhouette Scores and Davies-Bouldin Index (DBI) matrices. The research findings indicate that the AP algorithm outperforms DBSCAN in this customer segmentation case. The AP algorithm yields Silhouette Scores of 0.699 and DBI of 0.429, along with recommendations for 4 customer segments. Furthermore, further analysis is performed on the AP results using a statistical approach based on the mean values of each segment for the RFM variables. The four customer segments generated by the AP algorithm, based on the mean values of the RFM variables, can be associated with the concept of customer relationship management.

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

**Abstrak**— Lingkungan bisnis yang semakin kompleks membutuhkan bisnis untuk merancang strategi pengembangan perusahaan yang lebih efektif dan efisien, termasuk dalam pengembangan pasar. Untuk memahami perilaku pelanggan, analisis data pelanggan menjadi sangat penting. Salah satu pendekatan yang umum digunakan untuk mengelompokkan pelanggan adalah segmentasi berdasarkan analisis RFM (*Recency, Frequency, dan Monetary*). Penelitian ini bertujuan untuk membandingkan kinerja dua algoritma segmentasi, yaitu

DBSCAN dan Affinity Propagation (AP), dalam memberikan rekomendasi segmen profil pelanggan menggunakan analisis RFM. Algoritma DBSCAN digunakan karena kemampuannya dalam mengidentifikasi segmen berbentuk bebas dan dapat mengatasi *noise* data. Sementara itu, algoritma *Affinity Propagation* (AP) dipilih karena kemampuannya dalam menemukan pusat-pusat segmen tanpa memerlukan jumlah segmen yang ditentukan sebelumnya. Dataset transaksi yang digunakan dalam penelitian ini diperoleh dari salah satu tenant inkubator bisnis STIKOM Bali. Dataset tersebut melewati tahap pra-proses sebelum dilakukan segmentasi menggunakan kedua algoritma DBSCAN dan AP. 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 dibandingkan DBSCAN dalam kasus segmentasi pelanggan ini. Algoritma AP menghasilkan *Silhouette Scores* dengan nilai 0.699 dan DBI dengan nilai 0.429, serta memberikan rekomendasi 4 segmen pelanggan. Selain itu, analisis lanjutan dilakukan pada hasil AP dengan menggunakan pendekatan statistik berdasarkan nilai rata-rata dari setiap segmen untuk variabel RFM. Keempat segmen pelanggan yang dihasilkan oleh algoritma AP berdasarkan nilai rata-rata variabel RFM dapat dikaitkan dengan konsep manajemen relasi pelanggan.

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

## I. 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 [2], [3]. Beberapa penelitian seperti penelitian [4], [5] 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, peran inkubator bisnis adalah untuk membantu pengembangan bisnis dari tenant yang didampingi [6]–[8]. Berdasarkan data dan hasil wawancara dengan pengelola inkubator, tantangan yang dihadapi oleh tenant adalah ketidakefektifan proses pemasaran dan strategi untuk menjaga loyalitas pelanggan, padahal menjaga loyalitas pelanggan adalah salah satu dasar pengembangan bisnis dan reputasi bisnis [9], [10]. Rencana dan strategi yang tidak tepat dapat menyebabkan penggunaan sumber daya yang tidak efektif. Pada penelitian [11] diterapkan skema segmentasi pelanggan dengan menggunakan algoritma *K-Means*, dimana dihasilkan 2 segmen pelanggan. Pada penelitian [12] memanfaatkan algoritma *Fuzzy C Means* dan berdasarkan hasil segmentasi ditemukan 7 segmen pelanggan. Kedua algoritma ini memberikan hasil segmen yang berbeda.

Pada penelitian [11] hanya variabel frekuensi yang menjadi faktor kunci yang membedakan kedua segmen pelanggan. Hal ini mengindikasikan wawasan yang didapatkan relatif minim untuk memahami pelanggan lebih dalam. Pada penelitian [12] menghasilkan 7 segmen pelanggan yang dimana jumlah ini cukup besar, namun karakteristik dari tiap segmen berdasarkan analisa RFM cenderung memiliki kemiripan, sehingga dalam merumuskan karakteristik pelanggan relatif sulit. Oleh karena itu untuk mendapatkan pemahaman komprehensif dari potensi profil pelanggan dengan kompleksitas yang tepat pada kasus segmentasi pelanggan tenant ini, peneliti ingin menganalisa pemanfaatan algoritma segmentasi lainnya yaitu *Density-based Spatial Clustering of Application with Noise (DBSCAN)* dan *Affinity Propagation (AP)* serta membandingkan performa keduanya. DBSCAN mampu mendeteksi segmen 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 segmen, memungkinkan algoritma untuk mengelompokkan data dengan lebih akurat pada kasus dimana *noise* ada [13], [14].

AP sendiri merupakan algoritma yang cenderung lebih stabil karena algoritma ini mencari solusi yang mengoptimalkan keseluruhan data [15], [16] dan tidak bergantung pada inisialisasi *centroid* awal seperti pada *K Means*. DBSCAN dan AP sama-sama algoritma yang tidak memerlukan inisiasi awal jumlah segmen. Penelitian ini berfokus pada penerapan DBSCAN dan AP untuk menemukan solusi optimal dalam hal segmentasi data pelanggan. Data pelanggan yang dimaksud adalah model data *Recency, Frequency, Monetary (RFM)* [17], [18]. Hasil segmentasi akan dievaluasi. Performa Algoritma DBSCAN dan AP selanjutnya dibandingkan, untuk mengetahui algoritma yang memiliki hasil yang baik dan cocok pada kasus segmentasi ini. Hasil segmen yang didapatkan dari DBSCAN dan AP tersebut juga dianalisa karakteristiknya dalam kaitannya dengan manajemen relasi pelanggan.

## II. METODOLOGI

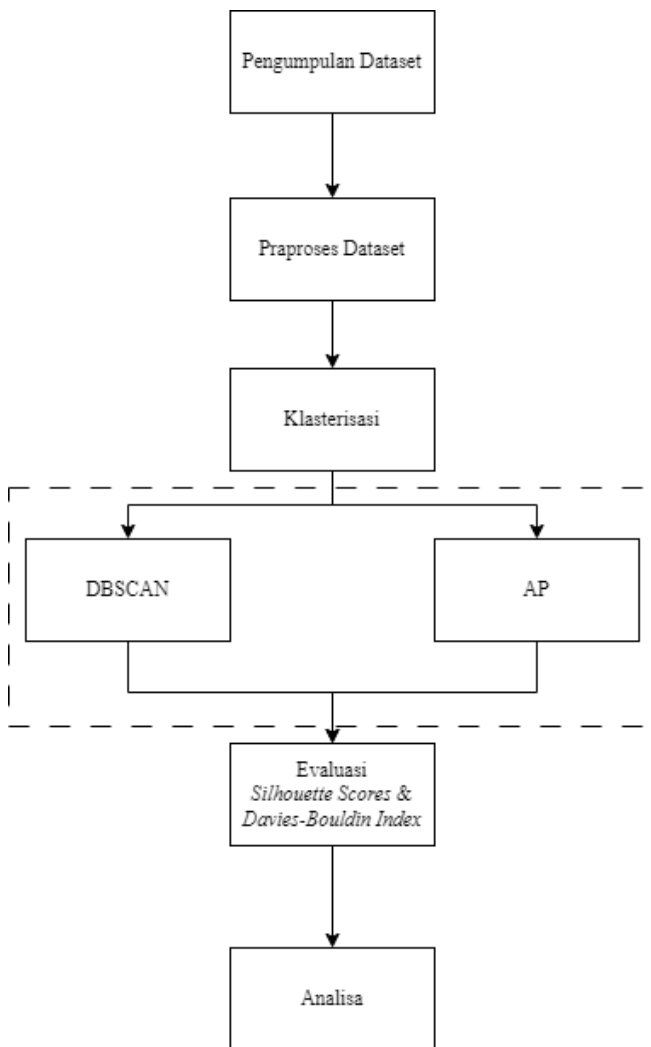
Adapun metode yang digunakan dalam pelaksanaan penelitian ini, dapat dilihat pada **Error! Reference source not found.** berikut.

### A. 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 transaksi yang merupakan transaksi masuk dari usaha. Selanjutnya berkas transaksi yang didapatkan disimpan dalam dokumen berbentuk Excel dalam format .xls. Data transaksi sebagai dataset terdiri dari lima kolom, yaitu *email* pelanggan, ID transaksi, tanggal transaksi, tipe transaksi, dan jumlah transaksi.

### B. Praproses Dataset

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

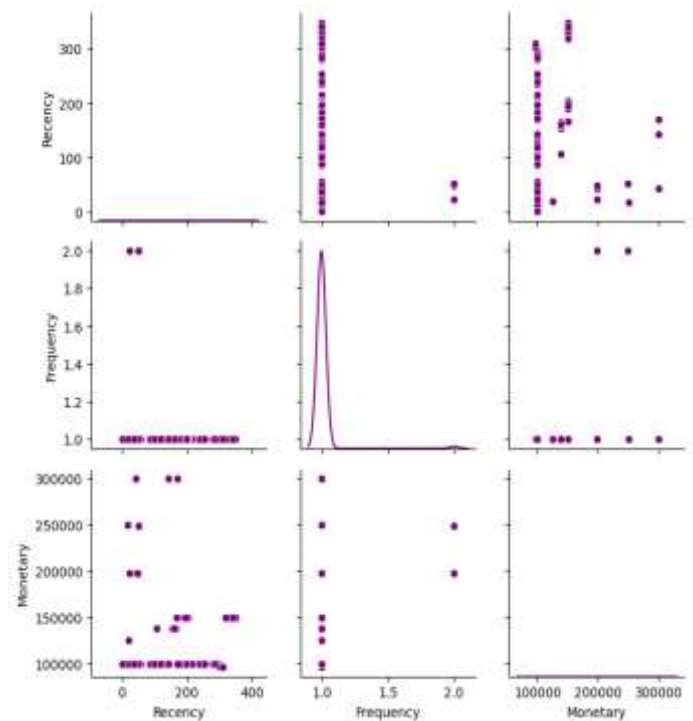


Gambar. 1 Metode penelitian.

disimpan dalam *Data Frame* RFM. Gambar. 2 menampilkan hasil visualisasi data sebelum dilakukan proses segmentasi dan

6) TABEL 1 menampilkan sampel dari isi dataset RFM.

7) Normalisasi data: Fitur *Recency*, *Frequency*, dan *Monetary* memiliki rentang nilai yang berbeda oleh karena itu perlu dilakukan normalisasi pada data tersebut. Sehingga algoritma segmentasi tidak didominasi oleh fitur dengan rentang nilai yang lebih besar dalam kasus ini adalah fitur *monetary*. Normalisasi dalam kasus ini menggunakan *Standard Scaler*.



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 segmentasi. 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*):

a) *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).

b) *Frequency*: Menghitung jumlah transaksi yang dilakukan oleh setiap pelanggan berdasarkan email pelanggan.

c) *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

8) TABEL II menampilkan sampel dataset setelah dinormalisasi.

Gambar. 2 Visualisasi distribusi data RFM sebelum segmentasi.

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 RFM 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

C. Segmentasi

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

1) Algoritma DBSCAN

Algoritma DBSCAN adalah Algoritma yang mampu mendeteksi segmen 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 segmen, memungkinkan algoritma untuk mengelompokkan data dengan lebih akurat pada kasus dimana *noise* ada [13], [14].

2) Algoritma AP

Algoritma AP sendiri merupakan algoritma yang cenderung lebih stabil karena algoritma ini mencari solusi yang mengoptimalkan keseluruhan data [15], [16] dan tidak bergantung pada inisialisasi *centroid* awal seperti pada *K Means*.

D. 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 segmen mirip satu sama lain dibandingkan dengan objek di segmen lain [19],[20]. Nilai *Silhouette Score* berada pada rentang antara -1 hingga 1, di mana nilai yang mendekati 1 atau yang paling tinggi berarti segmen tersebut memiliki kualitas yang baik.

$$silhouette(i) = \frac{b(i)-a(i)}{\max(a(i),b(i))} \tag{1}$$

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

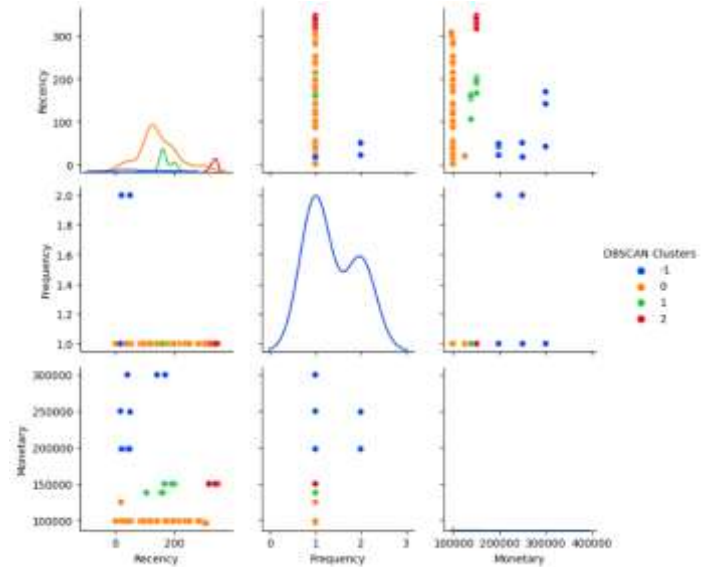
2) *Davies-Bouldin Index* mengukur kualitas dari segmentasi berdasarkan rasio jarak dalam segmen dan jarak antar segmen. Berbeda dengan *Silhouette Scores*, nilai *Davies-Bouldin Index* yang lebih rendah menunjukkan kualitas segmen yang lebih baik [21], [22].

$$DBI = \frac{1}{n} \sum_{i=1}^n \max\left(\frac{s(i)+s(j)}{d(i,j)}\right) \tag{2}$$

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

E. Analisa

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



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

membandingkan karakteristik antara segmen.

III. HASIL DAN PEMBAHASAN

Hasil segmentasi dengan DBSCAN dan AP adalah sebagai berikut.

A. Hasil Segmentasi DBSCAN

Pada proses segmentasi 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.

Jumlah segmen yang direkomendasikan adalah 4 segmen, sudah termasuk dengan 1 segmen yang dikategorikan sebagai *outlier*.

menampilkan visualisasi distribusi data dengan label kelas masing - masing dari hasil segmentasi dengan algoritma

DBSCAN. Segmen pertama adalah *noise* yang diberikan label -1, ada sebanyak 8 data yang termasuk dalam segmen ini. Segmen kedua dengan label segmen 0 memiliki 227 data yang masuk, yang juga menjadi segmen dengan ukuran terbesar. Segmen ketiga dengan label segmen 1 memiliki 34 data, dan yang terakhir segmen keempat dengan label segmen 2 memiliki 13 data.

**B. Hasil Segmentasi AP**

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

**C. Analisa Segmentasi Pelanggan**

Terkait dengan proses segmentasi pelanggan untuk mengetahui karakteristik pelanggan dilakukan eksplorasi berdasarkan hasil segmentasi dari masing masing algoritma. Pada TABEL III dan

TABEL III. Nilai *Mean* Tiap Kelas Variabel RFM Pada Hasil Segmentasi Algoritma DBSCAN menampilkan hasil perhitungan dari nilai rata – rata variabel RFM dari keempat kelas, pada masing – masing algoritma.

TABEL III. NILAI *MEAN* TIAP KELAS VARIABEL RFM PADA HASIL SEGMENTASI 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 IV. NILAI *MEAN* TIAP KELAS VARIABEL RFM PADA HASIL SEGMENTASI ALGORITMA AP

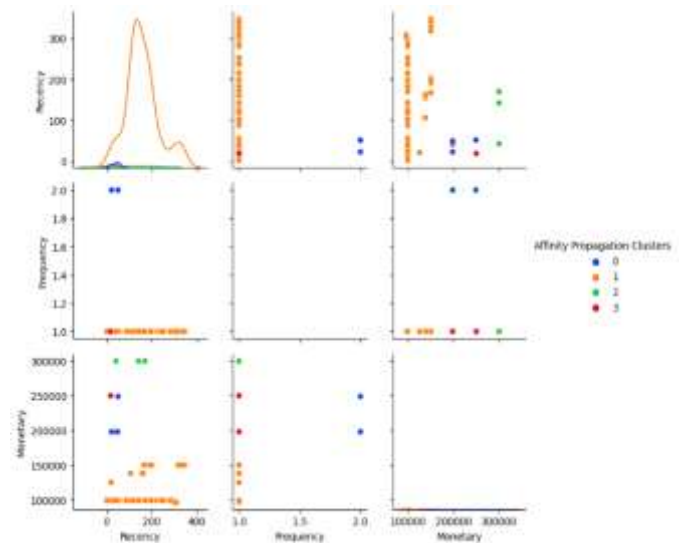
Segmen AP	Jumlah	Mean Recency	Mean Frequency	Mean Monetary
1	3	40.67	2.00	215,000.00
2	274	158.77	1.00	106,737.22
3	3	118.00	1.00	300,000.00
4	3	30.00	1.00	224,000.00

Pada hasil pembuatan segmen pelanggan dari segmentasi 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, pengelompokannya 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 [23], [24]. 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.

Selanjutnya pada hasil pembuatan segmen pelanggan dengan algoritma AP, secara keseluruhan dapat disimpulkan bahwa segmen yang dihasilkan memiliki karakteristik yang lebih jelas dari pada hasil dari algoritma DBSCAN. Hal ini memudahkan dari sisi interpretasi dan menarik keterkaitan dengan konsep *customer relationship management* (CRM) yang selanjutnya menjadi dasar dalam perancangan strategi marketing bagi bisnis [25]–[27]. Terkait penjelasan detail dari segmen yang dihasilkan dengan algoritma AP, segmen pertama terdiri dari 3 data dengan pengaruh dari variabel *frequency* dimana hasil segmentasi 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* dan memberikan *reward* [10].

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

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 mendekati 3 bulan, untuk menysasar pelanggan pada segmen ini, pelaku bisnis perlu melakukan kontak kembali dengan pelanggan tersebut [28], dan dapat memberikan kesempatan kepada pelanggan tersebut untuk menggunakan produk bisnis secara gratis dalam periode tertentu (*trial*) [29]. Hal ini bertujuan untuk mengingatkan kembali pelanggan tersebut terkait produk dari bisnis, dan meningkatkan potensi pelanggan tersebut mau bertransaksi kembali. Terakhir adalah segmen keempat dimana segmen pelanggan yang secara variabel dipengaruhi 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 [30].

D. Perbandingan 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. 5 dan Gambar. 6, dimana secara hasil visualisasi pembagian segmen pada algoritma AP lebih jelas perbedaan antar segmennya dari pada algoritma DBSCAN. Berdasarkan hasil evaluasi dengan *Silhouette Scores* dan DBI juga menunjukkan bahwa AP memiliki performa yang lebih baik daripada DBSCAN seperti yang terlihat pada TABEL V.

TABEL V. HASIL EVALUASI DENGAN MATRIKS *SILHOUETTE SCROES* DAN DBI

No	Algoritma	<i>Silhouette Scores</i>	DBI
1	DBSCAN	0.499	0.525
3	AP	0.699	0.429

Algoritma AP secara otomatis dapat menentukan jumlah segmen yang optimal berdasarkan similaritas antara titik data. Oleh karena itu, AP lebih baik dalam mengenali pola yang muncul dalam data. Sementara itu, jumlah segmen 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 segmen.

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 segmen dan dianggap sebagai *outlier*. DBSCAN mengidentifikasi *noise* sebagai titik data yang tidak termasuk dalam segmen manapun, yang bisa menyebabkan segmen yang kurang jelas jika *noise* cukup signifikan. Sementara itu, AP lebih mampu menangani *noise* dengan mengelompokkannya ke segmen yang paling mirip.

Secara garis besar keempat segmen pelanggan hasil segmentasi dengan algoritma AP memiliki karakteristik yang lebih jelas ketimbang hasil segmen dari algoritma DBSCAN, sehingga dalam perumusan strategi CRM berdasarkan piramida pelanggan dapat lebih sesuai.

Penelitian sebelumnya [11] dengan menggunakan algoritma *K-Means* menghasilkan 2 segmen pelanggan dengan nilai *silhouette scores* yang lebih tinggi dan DBI yang lebih rendah daripada algoritma AP. Hal ini berarti kinerja algoritma *K-Means* masih lebih baik ketimbang algoritma AP dalam hal perbedaan 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.

IV. KESIMPULAN

Penelitian ini bertujuan untuk menganalisa dan membandingkan 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, meski DBSCAN dapat juga dikaitkan dengan CRM secara umum, namun 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 segmentasi lainnya seperti *K Medoids* atau menggunakan pendekatan *hybrid* dimana DBSCAN digunakan untuk mendeteksi *outlier* lalu segmentasi lanjutan menggunakan *K-Means*.



UCAPAN TERIMA KASIH

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] R. W. Palmatier, J. A. Petersen, and F. Germann, *Marketing Analytics: Based on First Principles*. Bloomsbury Publishing, 2022.
- [3] V. Kumar and W. Reinartz, *Customer relationship management*. Springer, 2018.
- [4] 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.
- [5] 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.
- [6] B. D. Komara and H. C. B. Setiawan, "Inkubator Bisnis Sebagai Pendorong Tumbuhnya Wirausaha Muda: Studi Tentang Sukses Kewirausahaan Mahasiswa Universitas Muhammadiyah Gresik," *Jurnal Riset Entrepreneurship*, vol. 3, no. 1, pp. 33–39, 2020.
- [7] N. Lutfiani, U. Rahardja, and I. S. P. Manik, "Peran Inkubator Bisnis dalam Membangun Startup pada Perguruan Tinggi," *Jurnal Penelitian Ekonomi dan Bisnis*, vol. 5, no. 1, pp. 77–89, 2020.
- [8] W. Gunadi, "Pengembangan Kewirausahaan Usaha Mikro, Kecil Dan Menengah Melalui Inkubator Bisnis," *Jurnal Ilmiah M-Progress*, vol. 8, no. 1, 2018.
- [9] M. M. D. Alam, R. Al Karim, and W. Habiba, "The relationship between CRM and customer loyalty: The moderating role of customer trust," *International Journal of Bank Marketing*, vol. 39, no. 7, pp. 1248–1272, 2021.
- [10] R. U. Khan, Y. Salamzadeh, Q. Iqbal, and S. Yang, "The impact of customer relationship management and company reputation on customer loyalty: The mediating role of customer satisfaction," *Journal of Relationship Marketing*, vol. 21, no. 1, pp. 1–26, 2022.
- [11] 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.
- [12] 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.
- [13] M. Hahsler, M. Piekenbrock, and D. Doran, "dbscan: Fast density-based clustering with R," *J Stat Softw*, vol. 91, pp. 1–30, 2019.
- [14] W. Lai, M. Zhou, F. Hu, K. Bian, and Q. Song, "A new DBSCAN parameters determination method based on improved MVO," *Ieee Access*, vol. 7, pp. 104085–104095, 2019.
- [15] P. Kalpana, A. M. Nivetha, R. Madhumitha, P. Tamijeselv, and S. S. Devi, "Improvisation of spectral clustering through affinity propagation," in *AIP Conference Proceedings*, AIP Publishing LLC, 2021, p. 030004.
- [16] H. Keller, H. Möllering, T. Schneider, and H. Yalame, "Balancing quality and efficiency in private clustering with affinity propagation," *Cryptology ePrint Archive*, 2021.
- [17] M. Tavakoli, M. Molavi, V. Masoumi, M. Mobini, S. Etemad, and R. Rahmani, "Customer segmentation and strategy development based on user behavior analysis, RFM model and data mining techniques: a case study," in *2018 IEEE 15th International Conference on e-Business Engineering (ICEBE)*, IEEE, 2018, pp. 119–126.
- [18] Q. Zhang, H. Yamashita, K. Mikawa, and M. Goto, "Analysis of purchase history data based on a new latent class model for RFM analysis," *Industrial Engineering & Management Systems*, vol. 19, no. 2, pp. 476–483, 2020.
- [19] 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.
- [20] 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.
- [21] Y. A. Wijaya, D. A. Kurniady, E. Setyanto, W. S. Tarihoran, D. Rusmana, and R. Rahim, "Davies bouldin index algorithm for optimizing clustering case studies mapping school facilities," *TEM J*, vol. 10, no. 3, pp. 1099–1103, 2021.
- [22] A. K. Singh, S. Mittal, P. Malhotra, and Y. V. Srivastava, "Clustering Evaluation by Davies-Bouldin Index (DBI) in Cereal data using K-Means," in *2020 Fourth international conference on computing methodologies and communication (ICCMC)*, IEEE, 2020, pp. 306–310.
- [23] J. S. Thomas, C. Chen, and D. Iacobucci, "Email Marketing as a Tool for Strategic Persuasion," *Journal of Interactive Marketing*, vol. 57, no. 3, pp. 377–392, May 2022, doi: 10.1177/10949968221095552.
- [24] C. E. Khedkar and A. E. Khedkar, "Email Marketing: A Cost-Effective Marketing Method," *Vidyabharati International Interdisciplinary Research Journal 13 (1)*, 2021.
- [25] İ. SABUNCU, E. TÜRKAN, and H. POLAT, "Customer segmentation and profiling with RFM analysis," *Turkish Journal of Marketing*, vol. 5, no. 1, pp. 22–36, 2020.
- [26] T.-H. Chou and S.-C. Chang, "The RFM Model Analysis for VIP Customer: A case study of golf clothing brand," *International Journal of Knowledge Management (IJKM)*, vol. 18, no. 1, pp. 1–18, 2022.
- [27] A. T. Widiyanto and A. Witanti, "Segmentasi Pelanggan Berdasarkan Analisis RFM Menggunakan Algoritma K-Means Sebagai Dasar Strategi Pemasaran (Studi Kasus PT Coversuper Indonesia Global)," *KONSTELASI: Konvergensi Teknologi dan Sistem Informasi*, vol. 1, no. 1, pp. 204–215, 2021.
- [28] H. Rangriz and Z. Bayrami Shahrivar, "The impact of E-CRM on customer loyalty using data mining techniques," *BI Management Studies*, vol. 7, no. 27, pp. 175–205, 2019.
- [29] I. G. Juanamasta *et al.*, "The role of customer service through customer relationship management (Crm) to increase customer loyalty and good image," *International Journal of Scientific and Technology Research*, vol. 8, no. 10, pp. 2004–2007, 2019.
- [30] C. M. Durugbo, "After-sales services and aftermarket support: a systematic review, theory and future research directions," *Int J Prod Res*, vol. 58, no. 6, pp. 1857–1892, 2020.