

Evaluasi Performa Segmentasi Pelanggan Tenant Inkubator Bisnis dengan Menggunakan Model *Consensus Clustering*

Dedy Panji Agustino¹, I Gede Bintang Arya Budaya²

Program Studi Sistem Informasi¹, Program Studi Teknologi Informasi²

Institut Teknologi dan Bisnis STIKOM Bali

Denpasar, Indonesia

e-mail: panji@stikom-bali.ac.id

Abstrak

Dalam dunia bisnis yang semakin kompleks, strategi untuk mempertahankan loyalitas pelanggan dan hubungan jangka panjang dengan mereka menjadi sangat penting. Namun, perusahaan sering menghadapi tantangan dalam memasarkan produk mereka dan mempertahankan pelanggan. Kegagalan dalam merencanakan dan menyusun strategi yang tepat dapat menyebabkan penggunaan sumber daya yang tidak efektif, disinal pengetahuan tentang customer profiling diperlukan. Salah satu teknik dalam customer profiling adalah segmentasi pelanggan menggunakan pendekatan unsupervised learning. Penelitian ini bertujuan untuk mengatasi perbedaan hasil segmentasi yang dihasilkan oleh algoritma K-Means, DBSCAN, dan Affinity Propagation yang telah diterapkan pada penelitian sebelumnya, dengan menggunakan model consensus clustering. Metode consensus clustering dapat mengatasi perbedaan hasil segmentasi yang dihasilkan oleh algoritma clustering yang berbeda, sehingga menghasilkan keputusan yang lebih tepat. Penelitian ini menggunakan dataset transaksi tenant inkubator bisnis yang diolah menjadi data RFM. Hasil penelitian menunjukkan bahwa consensus clustering dengan teknik voting dan averaging merekomendasikan 2 cluster pelanggan dengan sebaran label kelas data poin yang berbeda. Namun hasil consensus clustering teknik voting menghasilkan kelompok segmen yang lebih jelas dengan nilai $DBI = 0.1736$ dan $silhouette\ scores = 0.862$ dengan karakter cluster yang sama dengan penerapan K-Means clustering pada penelitian sebelumnya. Kepastian cluster ini tentu menjadi informasi penting dalam merencanakan strategi bisnis selanjutnya.

Kata kunci: RFM, DBSCAN, Affinity Propagation, K-Means,

Abstract

In today's increasingly complex business world, the strategy to maintain customer loyalty and long-term relationships with them is crucial. However, companies often face challenges in marketing their products and retaining customers. Failure to plan and develop the right strategy can result in ineffective use of resources, hence the need for customer profiling knowledge. One technique used in customer profiling is customer segmentation using unsupervised learning approaches. This study aims to address the differences in segmentation results produced by the K-Means, DBSCAN, and Affinity Propagation algorithms that were applied in previous studies, by using consensus Clustering models. The consensus Clustering method can overcome the differences in segmentation results produced by different Clustering algorithms, resulting in more accurate decisions. The study uses a dataset of business incubator tenant transactions processed into RFM data. The research findings indicate that consensus clustering using voting and averaging techniques recommends 2 customer clusters with different distributions of data point labels. However, the voting-based consensus clustering produces more distinct segment groups with $DBI = 0.1736$ and $silhouette\ scores = 0.862$, showing similar cluster characteristics to the application of K-Means clustering in the previous study. The certainty of these clusters provides essential information for planning future business strategies.

Keywords: RFM, DBSCAN, Affinity Propagation, K-Means

1. Pendahuluan

Dalam dunia bisnis yang semakin kompleks, persaingan antar perusahaan semakin sengit. Oleh karena itu, strategi untuk mempertahankan loyalitas pelanggan dan hubungan jangka panjang dengan mereka menjadi sangat penting. Salah satu cara untuk mencapai hal ini adalah dengan melakukan customer profiling [1], yaitu mengidentifikasi dan memahami pelanggan bisnis secara lebih mendalam. Salah satu

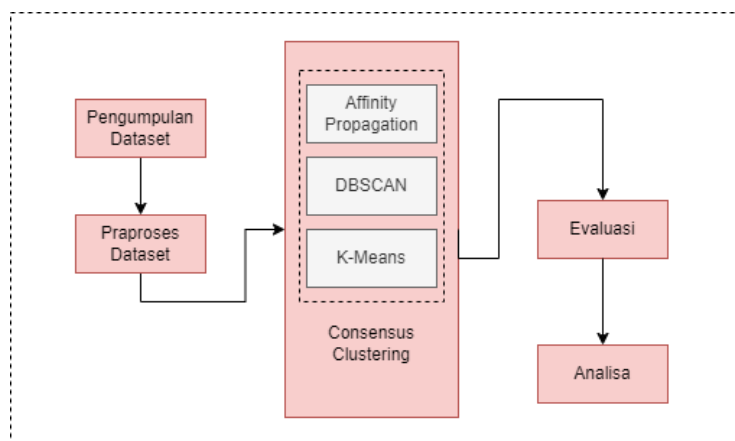
teknik dalam *customer profiling* adalah segmentasi pelanggan menggunakan pendekatan *unsupervised learning*. *Customer profiling* merupakan bagian penting dari *Customer Relationship Management (CRM)* dan dapat membantu perusahaan dalam memilah data pelanggan yang menguntungkan [2], [3]. Penelitian sebelumnya menunjukkan bahwa segmentasi pelanggan dapat meningkatkan penjualan dan keuntungan perusahaan serta memanfaatkan potensi *unsupervised learning* [4], [5].

Inkubator bisnis STIKOM Bali adalah satu dari beberapa inkubator perguruan tinggi yang menyediakan layanan inkubasi bagi perusahaan rintisan. Peran inkubator bisnis adalah untuk membantu pengembangan bisnis dari perusahaan yang bergabung dengan inkubator tersebut [6]–[8]. Namun, berdasarkan data dan hasil wawancara dengan pengelola inkubator, para perusahaan yang bergabung dengan inkubator tersebut menghadapi tantangan dalam memasarkan produk mereka dan mempertahankan pelanggan. Menjaga loyalitas pelanggan merupakan dasar dari pengembangan bisnis dan reputasi bisnis [9], [10]. Kegagalan dalam merencanakan dan menyusun strategi yang tepat dapat menyebabkan penggunaan sumber daya yang tidak efektif. Dalam penelitian sebelumnya, skema segmentasi pelanggan dengan menggunakan algoritma K-Means [11], Fuzzy C Means [12], DBSCAN dan *Affinity Propagation* [13] telah diterapkan. Namun, keempat algoritma tersebut menghasilkan segmen pelanggan yang berbeda. Pada penelitian pertama, hanya terdapat 2 segmen pelanggan yang dihasilkan, sedangkan pada penelitian kedua ditemukan 7 segmen pelanggan, dan pada penelitian ke 3 menghasilkan 4 segmen pelanggan.

Penelitian ini bertujuan untuk mengatasi perbedaan hasil segmentasi yang dihasilkan oleh algoritma K-Means, DBSCAN, dan *Affinity Propagation* yang telah diterapkan pada penelitian sebelumnya, dengan menggunakan model *consensus clustering*. Model ini menggabungkan hasil segmentasi dari beberapa algoritma *clustering* untuk memperoleh hasil yang lebih akurat [14]–[16]. Dalam penelitian ini, akan dilakukan segmentasi pelanggan dengan menggunakan algoritma K-Means, DBSCAN, dan *Affinity Propagation*. Kemudian, hasil segmentasi dari ketiga algoritma *clustering* akan digabungkan menggunakan model *consensus clustering*. Tujuannya adalah untuk memperoleh hasil segmentasi pelanggan yang lebih akurat dan konsisten. Metode *consensus clustering* dapat mengatasi perbedaan hasil segmentasi yang dihasilkan oleh algoritma *clustering* yang berbeda, sehingga menghasilkan keputusan yang lebih tepat.

2. Metode Penelitian

Adapun metode penelitian adalah sebagai berikut, seperti yang terlihat pada Gambar 1.



Gambar 1. Metode penelitian

2.1. Pengumpulan Dataset

Tahap pertama dalam penelitian ini adalah pengumpulan dataset. Pada tahap ini, dilakukan proses pengumpulan data dari salah satu tenant inkubator bisnis STIKOM Bali. Data yang dikumpulkan berupa data transaksi yang terjadi pada tahun 2022 sebanyak 282 baris transaksi dari usaha tersebut. Data transaksi tersebut disimpan dalam dokumen berbentuk excel dengan format .xls. Dataset terdiri dari lima kolom, yaitu email pelanggan, ID transaksi, tanggal transaksi, tipe transaksi, dan jumlah transaksi.

2.2. Praproses Dataset

Tahap kedua dalam penelitian ini adalah tahap praproses dataset. Pada tahap ini, dilakukan proses persiapan data sebelum data tersebut digunakan dalam algoritma segmentasi. Tahap praproses penting

dilakukan untuk memastikan bahwa dataset yang digunakan sudah sesuai dengan kebutuhan dari proses segmentasi yang akan dilakukan. Tahap praproses dataset ini bertujuan untuk membersihkan data dari *noise* atau *outlier*, melakukan identifikasi dan penghapusan data yang duplikat, serta melakukan normalisasi atau standarisasi data apabila diperlukan. Hal ini akan memudahkan proses segmentasi selanjutnya dan meningkatkan kualitas hasil segmentasi yang diperoleh.

2.3. Consensus Clustering

Consensus Clustering adalah metode penggabungan hasil segmentasi dari beberapa algoritma *Clustering* untuk memperoleh hasil yang lebih akurat dan konsisten. Metode ini dapat digunakan untuk mengatasi perbedaan hasil segmentasi yang dihasilkan oleh algoritma *Clustering* yang berbeda, yaitu menggunakan tiga algoritma *Clustering*, yaitu K-Means, *Affinity Propagation*, dan DBSCAN. Dua teknik yang digunakan yaitu dengan *voting* dan *averaging*.

2.4. Evaluasi

Dalam tahap evaluasi, digunakan dua metrik yaitu *Silhouette Scores* dan *Davies-Bouldin Index* (DBI) untuk mengevaluasi kualitas hasil segmentasi. *Silhouette Score* digunakan untuk mengukur sejauh mana objek dalam satu segmen mirip satu sama lain dibandingkan dengan objek di segmen lain. Nilai *Silhouette Score* berkisar antara -1 hingga 1, dengan nilai mendekati 1 menunjukkan bahwa segmen tersebut memiliki kualitas yang baik. Rumus *Silhouette Score* dapat dihitung menggunakan persamaan (1), di mana $a(i)$ 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.

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

DBI digunakan untuk mengukur kualitas dari segmentasi berdasarkan rasio jarak dalam segmen dan jarak antar segmen. Nilai DBI yang lebih rendah menunjukkan kualitas segmen yang lebih baik. Rumus DBI dapat dihitung menggunakan persamaan (2), di mana 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.

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

2.5. Analisa

Tahapan analisis adalah tahap di mana hasil segmentasi pelanggan dievaluasi dan dianalisis untuk mendapatkan pemahaman dari penerapan consensus clustering untuk segmentasi pelanggan, berdasarkan hasil dari evaluasi dengan menggunakan matriks yang telah ditentukan.

3. Hasil dan Pembahasan

Model clustering menerapkan consensus clustering dengan menggunakan tiga algoritma pengelompokan yang berbeda: DBSCAN, *Affinity Propagation*, dan K-Means. Proses consensus clustering melibatkan penggabungan hasil pengelompokan individu dari setiap algoritma dan menghitung baik rata-rata serta skema voting untuk menciptakan himpunan baru dari label cluster. Penentuan cluster akhir diperoleh dengan membulatkan label rata-rata ke bilangan bulat terdekat. Setelah menciptakan label consensus, kode menambahkan kolom baru bernama 'Cluster_Consensus' ke Data Frame untuk menyimpan hasil cluster consensus. Selanjutnya, kode menghitung jumlah label cluster. Terakhir, kode mengevaluasi cluster consensus menggunakan dua metrik: *Davies-Bouldin Index* (DBI) dan *Silhouette Score*. Tabel 1 menampilkan hasil matriks evaluasi dan Gambar 2 menampilkan visualisasi data pelanggan dari 2 teknik consensus Clustering.

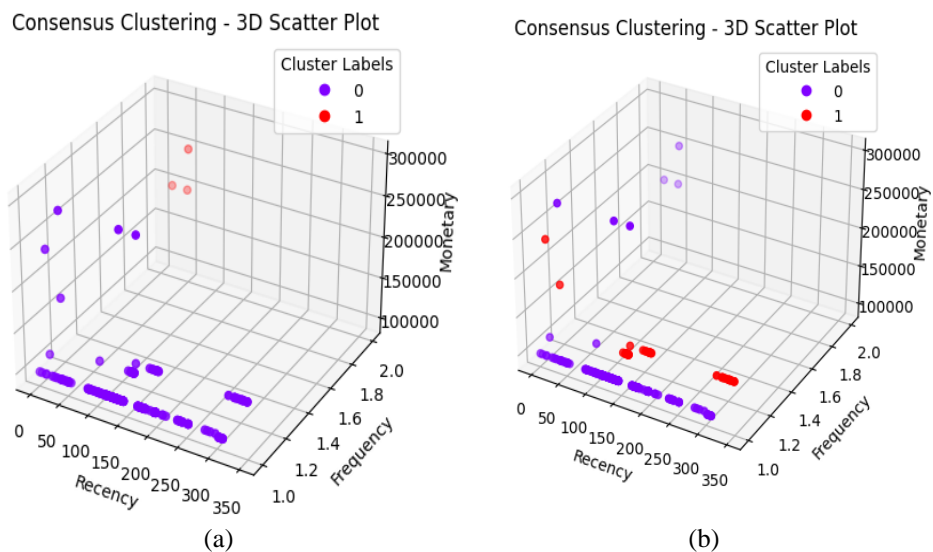
Hasil dari model consensus clustering dengan teknik voting menunjukkan bahwa jumlah optimal label consensus (cluster) adalah 2 dengan 3 data poin terkumpul dalam klaster 2 dan data poin sisanya masuk dalam klaster 1 seperti yang terlihat di Gambar 2 bagian (a), dari label data poin yang dihasilkan memiliki kemiripan dengan hasil clustering dengan menggunakan K-Means dimana pengelompokan ini

dipengaruhi oleh variabel frequency. Hal ini juga berarti bahwa data dapat direpresentasikan dengan baik dengan membaginya menjadi 2 cluster berdasarkan pendekatan consensus clustering teknik voting. *Consensus clustering* dengan teknik *voting* memiliki Nilai DBI sebesar 0.1736 dan *Silhouette Score* sebesar 0.862 menunjukkan bahwa kualitas *clustering* cukup baik. Dalam kasus ini, nilai DBI sebesar 0.1736 sangat rendah, yang menunjukkan bahwa *cluster-cluster*nya terpisah dengan baik dan jelas. Terkait nilai *Silhouette Score* sebesar 0.862 sangat tinggi, yang menunjukkan bahwa objek-objek dalam setiap *cluster* mirip satu sama lain dan berbeda dengan objek-objek dalam *cluster* lainnya. Secara keseluruhan, nilai DBI sebesar 0.1736 dan *Silhouette Score* sebesar 0.862 menunjukkan bahwa kualitas *clustering* cukup baik dan *cluster-cluster*nya terpisah dengan baik dan jelas. Namun, penting untuk diingat bahwa hasil *Clustering* juga harus diinterpretasikan dalam konteks masalah dan domain spesifiknya. Mungkin perlu dilakukan evaluasi lebih lanjut terhadap *cluster-cluster*nya dan karakteristiknya untuk menentukan kegunaan dan relevansinya dalam masalah yang sedang dihadapi.

Hasil dari model consensus clustering dengan teknik averaging juga sama merekomendasikan 2 cluster namun label data poin yang dihasilkan kelas cukup berbeda dimana dipengaruhi oleh variabel frequency dan monetary. Davies-Bouldin Index (DBI) untuk consensus clustering teknik averaging ini adalah adalah 1.111. Nilai DBI yang lebih rendah menunjukkan kualitas *clustering* yang lebih baik. Dalam kasus ini, nilai DBI sebesar 1.111 menunjukkan bahwa kualitas *clustering* tidak optimal, tetapi masih dapat diterima. *Silhouette Score* untuk *consensus clustering* teknik *averaging* adalah 0.452. Skor ini berkisar dari -1 hingga 1, dengan nilai yang lebih tinggi menunjukkan kualitas *clustering* yang lebih baik. Dalam kasus ini, nilai *Silhouette Score* sebesar 0.452 menunjukkan bahwa kualitas clustering masih dapat ditingkatkan. Secara keseluruhan, hasil dari metode consensus clustering teknik averaging menunjukkan bahwa membagi data menjadi 2 cluster adalah pendekatan terbaik, namun tampaknya teknik averaging tidak maksimal dalam menemukan kemiripan data seperti yang terlihat pada Gambar 2 bagian (b) dimana penyebaran data poin yang terlihat belum terpisah secara jelas.

Tabel 1. Hasil evaluasi dengan DBI dan *Silhouette Scores*

No	Algoritma	Klater	DBI	Silhouette Scores
1	K Means [11]	2	0.167	0.867
2	DBSCAN [13]	4	0.525	0.499
3	Affinity Propagation[13]	4	0.429	0.699
4	Consensus Clustering (Voting)	2	0.173	0.862
5	Consensus Clustering (Averaging)	2	1.111	0.452



Gambar 2. Visualisasi data pelanggan setelah pelabelan: (a) *Consensus Clustering-Voting*, (b) *Consensus Clustering-Averaging*

4. Kesimpulan

Penelitian ini bertujuan untuk mengetahui potensi penerapan teknik consensus clustering khususnya dalam analisa segmentasi pelanggan. consensus clustering adalah teknik yang digunakan untuk memperbaiki kualitas clustering dengan menggabungkan beberapa hasil clustering yang diperoleh dari algoritma clustering yang berbeda. Dalam kasus ini menggunakan consensus clustering dengan teknik voting dan averaging dengan 3 algoritma yaitu K-Means, Affinity Propagation dan DBSCAN dengan menggunakan sumber dataset transaksi tenant inkubator bisnis yang diolah menjadi data RFM. Hasil penelitian menunjukkan bahwa consensus clustering dengan teknik voting dan averaging merekomendasikan 2 cluster pelanggan. Namun hasil consensus clustering teknik voting menghasilkan kelompok segmen yang lebih jelas dengan nilai DBI = 0.1736 dan silhouette scores = 0.862 yang dalam kasus ini memiliki kemiripan dengan hasil dari K-Means clustering pada penelitian sebelumnya. Kepastian cluster ini tentu menjadi informasi penting dalam merencanakan strategi bisnis selanjutnya. Kedepan dapat menggunakan dataset transaksi dari sumber yang berbeda dan menambah ukuran dataset untuk mendapatkan [perbandingan](#) yang lebih akurat.

Daftar Pustaka

- [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] D. P. Agustino, I. G. B. A. Budaya, I. G. Harsemadi, I. K. Dharmendra, and I. M. S. A. Pande, "Comparison of the DBSCAN Algorithm and Affinity Propagation on Business Incubator Tenant Customer Segmentation," *Jurnal Sisfokom (Sistem Informasi dan Komputer)*, vol. 12, no. 2, pp. 315–321, 2023.
- [14] Z. Zheng *et al.*, "Subtyping CKD patients by consensus clustering: the chronic renal insufficiency cohort (CRIC) study," *J Am Soc Nephrol*, vol. 32, no. 3, p. 639, 2021.
- [15] A. Onan, "Consensus clustering-based undersampling approach to imbalanced learning," *Sci Program*, vol. 2019, 2019.
- [16] R. Ünlü and P. Xanthopoulos, "Estimating the number of clusters in a dataset via consensus clustering," *Expert Syst Appl*, vol. 125, pp. 33–39, 2019.