KLUSTERISASI PENGGUNA APLIKASI MPSTORE DAN ANALISIS POLA PEMBELIAN MENGGUNAKAN *K-MEANS* DAN *APRIORI* (STUDI KASUS PT. MITRA PEDAGANG INDONESIA)

Khoirul Anam, Wahyudi Setiawan, Eka Mala Sari Rochman
Sistem Informasi, Universitas Trunojoyo Madura
Jalan Raya Telang Kamal, Bangkalan, Indonesia
190441100129@student.trunojoyo.ac.id

ABSTRAK

Di era digital ini, aplikasi telah berkembang pesat dan menjadi bagian penting dari kehidupan sehari-hari. Salah satunya adalah *MPStore*, sebuah aplikasi yang memiliki lebih dari 250.000 pengguna aktif di seluruh Indonesia. Mengingat jumlah pengguna yang besar, pemahaman tentang perilaku pengguna menjadi penting untuk meningkatkan layanan dan memaksimalkan potensi aplikasi. Oleh karena itu, penelitian ini bertujuan untuk menganalisis segmentasi pengguna dan pola hubungan antar pengguna dalam aplikasi *MPStore*. Penelitian ini berpotensi memberikan wawasan berharga bagi PT. Mitra Pedagang Indonesia, perusahaan pengembang aplikasi *MPStore*, dalam mengoptimalkan strategi pemasaran dan meningkatkan kualitas layanan. Untuk mencapai tujuan ini, penelitian ini menggunakan metode *K-Means* untuk mengelompokkan pengguna berdasarkan karakteristik dan preferensi pengguna. Selain itu, metode *Apriori* digunakan untuk mengidentifikasi pola pembelian bersama antara produk yang ditawarkan oleh aplikasi. Hasil klusterisasi menunjukkan bahwa 3 kluster adalah yang terbaik dengan nilai *Silhouette Coefficient* 0,91. Analisis *Apriori* pada data transaksi setiap kluster mengidentifikasi terdapat 2 *rules* pada kluster 1, 1069 *rules* pada kluster 2, dan 8 *rules* pada 3 klaster.

Kata kunci : *Clustering, Asosiasi, K-Means, Apriori*

1. PENDAHULUAN

Dalam era digitalisasi yang berkembang pesat, aplikasi perangkat lunak hadir dalam berbagai bentuk dan fungsi, mengakomodasi kebutuhan beragam pengguna. Menurut *Bankmycell* 6.92 miliar atau 85.82% orang memiliki *smartphone* pada tahun 2023 di dunia[1].

Perusahaan-perusahaan berlomba-lomba memanfaatkan aplikasi perangkat lunak sebagai alat penting dalam berkomunikasi dengan pelanggan dan bersaing dalam pasar yang semakin ketat**.**

Salah satu cara meningkatkan daya saing adalah dengan menganalisis data pengguna untuk memahami preferensi, perilaku, dan pola penggunaan aplikasi. Pengguna aplikasi sering kali memiliki karakteristik yang berbeda dan memanfaatkan aplikasi dengan cara yang beragam. Oleh karena itu, segmen pengguna dan pola hubungan di antaranya dalam aplikasi menjadi sangat penting. Dengan pemahaman yang lebih dalam tentang segmen pengguna, perusahaan dapat merancang strategi yang lebih efektif dan mengoptimalkan pengalaman pengguna.

Salah satu perusahaan yang bergerak di ranah perangkat lunak yaitu PT. Mitra Pedagang Indonesia dengan aplikasi *MPStore* yang memiliki potensi besar. Dengan lebih dari 250.000 pengguna aktif di seluruh Indonesia, *MPStore* menjadi sumber data yang kaya. Data ini mencakup preferensi pengguna, pola pembelian, dan interaksi pengguna dengan berbagai layanan yang disediakan oleh aplikasi. Analisis data dari *MPStore* memberikan pemahaman yang lebih mendalam tentang perilaku pengguna, membantu PT. Mitra Pedagang Indonesia untuk mengoptimalkan strategi pemasaran dan meningkatkan kualitas layanan.

Penelitian terdahulu oleh Siregar menunjukkan pentingnya analisis data dalam bisnis dengan menggunakan metode *K-Means* untuk mengidentifikasi barang laris, membantu manajemen toko mengelola stok dan strategi pemasaran[2].

Pada penelitian Pandey juga menyebutkan bahwa *K-Means* sangat efisien untuk *dataset* besar[3]. Penelitian lain oleh Putra dkk menggunakan algoritma *Apriori* pada data penjualan retail untuk menemukan pola pembelian konsumen, yang bermanfaat dalam mengoptimalkan strategi pemasaran dan manajemen stok, serta meningkatkan efisiensi operasional dan omset perusahaan[4].

Berdasarkan uraian di atas, terlihat bahwa perusahaan memerlukan pemahaman yang lebih dalam tentang karakteristik pengguna dan pola pembelian untuk meningkatkan daya saing. Oleh karena itu, penelitian ini berfokus pada klusterisasi pengguna aplikasi *MPStore* dan analisis pola pembelian menggunakan metode *K-Means* dan *Apriori*. Dengan pendekatan ini, diharapkan penelitian dapat memberikan wawasan yang signifikan untuk mendukung pengembangan dan pengelolaan aplikasi *MPStore*, meningkatkan kepuasan pengguna, serta membantu PT. Mitra Pedagang Indonesia dalam merancang strategi bisnis yang lebih efektif dan efisien.

2. TINJAUAN PUSTAKA

2.1. *K-Means*

Algoritma *K-Means* adalah algoritma iteratif yang mencoba untuk membagi *dataset* menjadi k kelompok *subdata* yang telah ditentukan sebelumnya, yang berbeda satu sama lain tanpa tumpang tindih, di mana setiap titik data hanya termasuk dalam satu kelompok. Algoritma ini berupaya membuat titik-titik data dalam *intra-cluster* menjadi serupa sebanyak mungkin, sambil tetap menjaga kelompok-kelompok itu sendiri sejauh mungkin berbeda[5].

Tahapan algoritma *K-Means* meliputi:

a. Menentukan banyaknya/jumlah *Cluster* k[6]**.**
b. Menentukan nilai pusat (*Centroid*). Pada iterasi awal, nilai awal *Centroid* dipilih secara acak. Sementara itu, untuk iterasi berikutnya, nilai Centroid diperbarui menggunakan persamaan berikut[6]:

$$v_{ij} = \frac{1}{N_i}\sum_{k=0}^{Ni} X_{kj} \quad (1)$$

Di mana:

- V_{ij} adalah *centroid* atau rata-rata *Cluster* ke-i untuk variabel ke-j
- N_i adalah jumlah data anggota *Cluster* ke-i
- i,k adalah indeks dari *Cluster*
- X_{kj} adalah Nilai data ke-k yang ada di dalam *Cluster* tersebut untuk variabel ke-j

c. Mengukur jarak antara setiap objek dengan titik *Centroid* menggunakan metode *Euclidean Distance*[6].

$$d_{ij} = \sqrt{\sum_{k=1}^{P}\left(x_{ik} - y_{jk}\right)^2} \quad (2)$$

Di mana:

- d_{ij} adalah jarak objek antara objek i dan j
- P adalah dimensi data
- X_{ik} adalah koordinat dari objek i pada dimensi k
- X_{jk} adalah koordinat dari objek j pada dimensi k

d. Mengelompokkan objek berdasarkan jarak minimum. Hasil dari keanggotaan data pada *distance* matriks berupa nilai 0 atau 1, di mana nilai 1 menunjukkan data dialokasikan ke suatu klaster, sedangkan nilai 0 menunjukkan data dialokasikan ke klaster lainnya[6].
e. Kembali ke langkah kedua dan ulangi proses hingga nilai *Centroid* stabil dan tidak ada perubahan dalam anggota klaster[6].

2.2. *Silhoutte Coefficient*

Silhouette Coefficient digunakan untuk mengevaluasi kualitas dan kekuatan klaster, dengan menilai seberapa baik suatu objek ditempatkan dalam klaster tertentu. Metode ini mengombinasikan pendekatan separasi dan kohesi. Kohesi mengukur sejauh mana data berdekatan dengan pusat klaster yang diikutinya, sedangkan separasi mengukur jarak antara pusat klaster tersebut dengan pusat klaster lainnya[7].

Langkah perhitungan *Silhouette Coefficient* dimulai dengan menghitung jarak rata-rata dari data ke-i ke seluruh data lain dalam klaster yang sama. Dalam hal ini, diasumsikan data ke-i berada pada klaster A. Perhitungan ini dilakukan menggunakan persamaan (3).

$$a(i) = \frac{1}{|A|-1}\sum_{j\in A, j\neq 1} d(i,j) \quad (3)$$

Di mana:

- A adalah banyaknya data di *cluster* A

Langkah berikutnya adalah menghitung nilai b(i), yang merupakan jarak rata-rata minimum antara data ke-i dan semua data di klaster yang berbeda. Misalkan klaster yang berbeda dari klaster A adalah klaster C. Perhitungan dilakukan menggunakan persamaan berikut[8].

$$d(i,C) = \frac{1}{|C|}\sum_{j\in C} d(i,j) \quad (4)$$

Di mana:

- C adalah banyaknya data di kluster C

Setelah menghitung d(i,C) untuk semua klaster C ≠ A, selanjutnya memilih nilai jarak paling minimum sebagai nilai b(i).

$$b(i) = \min_{c\neq A} d(i,j) \quad (5)$$

Jika klaster B memiliki jarak terkecil, maka d(i,B)=b(i). Klaster B disebut tetangga data ke-i dan menjadi klaster terbaik kedua untuk data tersebut setelah klaster A. Setelah nilai a(i) dan b(i) diketahui, langkah terakhir adalah menghitung *Silhouette Coefficient*[8].

$$s(i) = \frac{b(i)-a(i)}{\max\{a(i)-b(i)\}} \quad (6)$$

Nilai s(i) berada antara -1 dan 1, di mana setiap nilai diinterpretasi sebagai berikut:

- s(i) = 1 => data ke-i digolongkan dengan baik (dalam A)
- s(i) = 0 => data ke-i berada di tengah antara dua klaster (A dan B)
- s(i) = -1 => data ke-i digolongkan dengan lemah (dekat ke klaster B daripada A)

Berikut adalah penafsiran nilai Silhoutte Coefficient yang ditunjukkan pada Tabel 1.

Tabel 1. Interpretasi nilai *Silhoutte Coefficient*

Silhouette Coefficient	**Interpretasi**
0.71 – 1.00	Struktur yang dihasilkan kuat
0.51 – 0..70	Struktur yang dihasilkan baik
0.26 – 0.50	Struktur yang dihasilkan lemah
< 26	Tidak terstruktur

2.3. *Apriori*

Algoritma *Apriori* adalah metode untuk menemukan pola dengan frekuensi tinggi. Pola-pola ini merupakan kombinasi item dalam suatu *database* yang memiliki nilai frekuensi atau *support* di atas ambang batas tertentu, yang dikenal sebagai *minimum support*[20].

Algoritma *Apriori* merupakan bagian dari teknik *Association Rules Mining* (ARM) dan termasuk dalam metode *data mining*. Aturan asosiasi yang dihasilkan algoritma ini berbentuk logika *if-then*[9].

Terdapat dua tolak ukur dalam membentuk aturan dalam penerapan algoritma apriori yaitu *support* dan *confidence*.

Support atau bisa juga disebut nilai penunjang adalah persentase dari laporan atau *record* yang di dalamnya mengandung kombinasi item[9].

Pada persamaan (7) adalah rumus untuk mencari nilai *support* dan untuk persamaan (8) adalah rumus untuk mendapatkan nilai *support* dari suatu kombinasi item.

$$Support(A) = \frac{Jumlah\ transaksi\ mengandung\ A}{Total\ transaksi} \quad (7)$$

$$Support(A,B) = \frac{Jumlah\ transaksi\ mengandung\ A\ dan\ B}{Total\ transaksi} \quad (8)$$

Confidence atau biasa disebut nilai kepastian adalah Kuatnya hubungan antar item dalam aturan asosiasi[9].

Persamaan (9) dan persamaan (10) adalah rumus untuk mendapatkan nilai *confidence.*

$$Confidence(A,B) = \frac{Jumlah\ transaksi\ mengandung\ A\ dan\ B}{Total\ transaksi\ mengandung\ A} \quad (9)$$

$$Confidence(A => B) = \frac{Support(A,B)}{Support(A)} \quad (10)$$

Dan untuk mendapatkan nilai persentase *confidence* dapat menggunakan rumus berikut:

$$Confidence(A => B) = \frac{Support(A,B)}{Support(A)} * 100\% \quad (11)$$

Algoritma *apriori* memiliki beberapa langkah-langkah dalam penerapannya yaitu sebagai berikut

a. Melakukan pemindaian database untuk menemukan kandidat *1-itemset* (C1) dan menghitung nilai *support*-nya. Kemudian, nilai *support* dibandingkan dengan *minimum support* yang telah ditentukan sebelumnya. Jika nilai *support* memenuhi atau melebihi *minimum support*, *itemset* tersebut akan dimasukkan ke dalam *large-itemset* 1 (L1)[9].
b. Itemset yang tidak memenuhi kriteria untuk *large-itemset* akan dihapus dan tidak digunakan dalam iterasi berikutnya (*pruning process*)[9].
c. Proses pembentukan kandidat baru (*joining*) dan *pruning* dilakukan berulang kali hingga tidak ada lagi kandidat yang dapat dibentuk[9].
d. Langkah terakhir adalah membentuk *association rule* dari semua *large-itemset* yang memenuhi nilai *minimum support*. Nilai *confidence* juga dihitung untuk setiap aturan. Jika nilai *confidence* kurang dari *minimum confidence* yang telah ditentukan, aturan tersebut tidak digunakan atau tidak termasuk dalam aturan asosiasi yang dipakai[9].

2.4. Penelitian Terkait

Penelitian ini juga mengambil informasi dari berbagai penelitian terdahulu yang relevan. Melalui tinjauan ini, kita dapat memahami konteks dan latar belakang topik yang sedang diteliti, serta mengidentifikasi celah pengetahuan yang masih ada.

Penelitian yang dilakukan oleh Ibnu Haidar berjudul "Implementasi Algoritma *Apriori* untuk Mencari Pola Transaksi Penjualan (Studi Kasus: *Carroll Kitchen*)" menghasilkan sebuah sistem yang menganalisis data transaksi penjualan kafe *Carroll Kitchen* dengan algoritma *apriori*. Dari total data transaksi penjualan selama satu tahun (Januari-Desember 2020) dengan jumlah 5.355 data, ditemukan 24 aturan asosiasi. Salah satu contohnya adalah jika konsumen membeli Teh manis (dingin), maka 71,05% membeli Nasi goreng jambal. Hasil ini dapat membantu kafe dalam pengambilan keputusan bisnis[9].

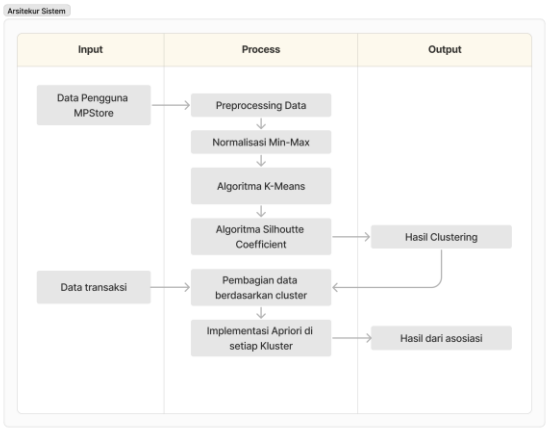
Penelitian yang dilakukan oleh Afolabi dkk berjudul "Analisis Kepuasan Pelanggan Menggunakan Kombinasi *K-Means Clustering* dan *Association Rule Mining Technique*". Penelitian ini menggunakan data kuesioner dari 200 responden tentang pengalaman responded dengan penyedia layanan jaringan seluler di Nigeria. Penelitian ini menemukan hubungan kuat antara sektor pelanggan tertentu, yaitu mahasiswi, dan beberapa atribut kepuasan pelanggan di sektor ini. Penelitian ini juga menemukan bahwa pelanggan di sektor ini cenderung menggunakan MTN sebagai penyedia layanan jaringan seluler, terutama dengan perangkat Blackberry. Penelitian ini menyimpulkan bahwa menggunakan *association rules* dalam kombinasi dengan *k-means* dapat membantu mengungkap hubungan-hubungan menarik antara item-item dalam data kuesioner. Penelitian ini juga memberikan saran bagi penyedia layanan jaringan seluler lainnya untuk menargetkan sektor pelanggan ini dan menyediakan cakupan jaringan yang luas untuk meningkatkan pangsa pasar dan keuntungan[10].

Penelitian yang dilakukan oleh Mohammad Ali Farajian dan Shahriar Mohammadi ini mengusulkan kerangka kerja baru untuk menganalisis perilaku pelanggan perbankan dengan menggunakan metode klasterisasi *K-means* dan induksi aturan asosiasi. Data yang digunakan dalam penelitian ini berasal dari sebuah perusahaan kartu debit Iran yang mencakup informasi akun dan transaksi pelanggan. Penelitian ini menggunakan algoritma *K-means* untuk membagi pelanggan menjadi tiga kelompok berdasarkan skor perilaku pelanggan (CP) dan nilai RFM. Selanjutnya, penelitian ini menggunakan induksi aturan asosiasi Apriori untuk menggambarkan karakteristik masing-masing kelompok pelanggan. Hasil penelitian ini menunjukkan bahwa pelanggan dapat diklasifikasikan menjadi tiga kelompok yang menguntungkan: pengguna manja, pengguna *transaktor*, dan pengguna jarang.

Penelitian ini juga menemukan beberapa aturan asosiasi yang menggambarkan profil pelanggan dalam hal jenis terminal, segmen hari, segmen waktu, jenis transaksi, jenis kelamin, rentang usia, dan status perkawinan. Penelitian ini memberikan wawasan tentang perilaku pelanggan perbankan dan memberikan saran untuk meningkatkan hubungan pelanggan dan strategi pemasaran. Jurnal ini juga berisi banyak referensi yang mendukung penelitian ini[11].

3. METODE PENELITIAN

Pada bab ini menjelaskan tentang arsitektur sistem yang dibangun sebelum dilakukannya tahapan uji coba. Desain rancangan sistem pada penelitian ini dengan mencantumkan semua tahapan secara garis besar yang ditunjukkan pada Gambar 1.



Gambar 1. Arsitektur sistem

Alur yang ditunjukkan pada gambar di atas merupakan diagram IPO yaitu diagram yang menunjukkan jalan alur keseluruhan. Terdapat 3 tahapan yaitu *input*, proses, output. Bagian *input*, bagian ini merupakan bagian awal yaitu terdapat proses penginputan data. Untuk datanya sendiri merupakan data dari pengguna *MPStore* dan data transaksi. Untuk data penggunanya sendiri adalah data pengguna sampai oktober 2023 dan data transaksi yang digunakan adalah data transaksi bulan agustus 2023.

Proses inti penelitian dimulai dengan tahapan *preprocessing* data guna memastikan data siap untuk digunakan secara optimal sesuai dengan skenario penelitian. Langkah berikutnya adalah normalisasi data menggunakan metode *min-max*, yang membantu dalam mengatur skala data untuk keperluan analisis selanjutnya. Setelah tahap normalisasi, langkah selanjutnya yaitu melakukan perhitungan menggunakan *K-Means* untuk mendapatkan hasil *clustering* dengan menggunakan beberapa jumlah kluster. Perhitungan *K-Means* dilakukan beberapa kali dengan jumlah kluster yang berbeda. Setelah itu hasil dari perhitungan *K-Means* dievaluasi menggunakan algoritma *Silhoutte Coefficient* untuk menentukan hasil *clustering* dengan jumlah kluster terbaik. Hasil dari proses ini menjadi landasan untuk mengurai data transaksi menjadi kelompok-kelompok atau kluster yang nantinya diaplikasikan metode *Apriori* guna mendalami produk-produk yang sering dibeli dalam setiap kluster atau kelompok tertentu.

Pada bagian *output* menghasilkan hasil dari *clustering* menggunakan *K-Means* dan hasil dari algoritma *Apriori* yakni produk-produk yang sering dibeli dari setiap kluster.

4. HASIL DAN PEMBAHASAN

4.1. *Input* Data

Dataset yang menjadi dasar penelitian ini berasal dari *MPStore*, terkait dengan perilaku pengguna pada bulan Oktober 2023. Terdapat dua jenis data yang digunakan dalam penelitian ini. Pertama, data yang digunakan untuk perhitungan dengan metode *K-Means* terdiri dari atribut pengguna aplikasi *MPStore* dan yang kedua adalah data transaksi pengguna *MPStore* digunakan untuk perhitungan algoritma *Apriori*.

Data pengguna *MPStore* yang digunakan dalam perhitungan *K-Means* merupakan data sebanyak 72.941 yang memuat sejumlah atribut terkait aktivitas pengguna.

Tabel 2. Data pengguna *MPStore* (5 contoh data dari total 72.941)

IDRESELLER	**f1**	**f2**	**f3**	**f4**	**f5**	**f6**
AA0024	0,00	509.333,33	2,67	0,00	1,00	15,67
AA0064	2,00	0,00	0,00	0,00	0,00	0,00
AA0065	2,00	0,00	0,00	0,00	0,00	0,00
AA0191	0,00	0,00	0,00	0,00	0,00	53,33
AA0229	2,00	0,00	0,00	0,00	0,00	1,00

Di mana:

- f1: Jumlah *downline*
- f2: Rata-rata pengisian saldo
- f3: Rata-rata transaksi QRIS
- f4: Rata-rata pencatatan kasir
- f5:Rata-rata transaksi tagihan
- f6: Rata-rata transaksi digital

Sementara data transaksi pengguna *MPStore* yang digunakan untuk perhitungan algoritma *Apriori* merujuk pada transaksi yang dilakukan oleh pengguna aplikasi. Data ini diurai kembali berdasarkan kluster hasil dari *K-Means*, memungkinkan analisis *Apriori* dilakukan di setiap kluster untuk memahami pola transaksi yang spesifik pada masing-masing kelompok pengguna.

Tabel 3. Data transaksi (contoh 8 dari total 200 data)

Transaksi	**Kode Produk**
1	DN20, DN200, DN50
2	DN50
3	DN200, DN25, DN50, DN500, DN70, DN75
4	DN20, DN60
5	DN50, DN60
6	DN20, DN200, DN50, DN70
7	DN50
8	DN25, DN30, DN50

4.2. *Preprocessing* Data

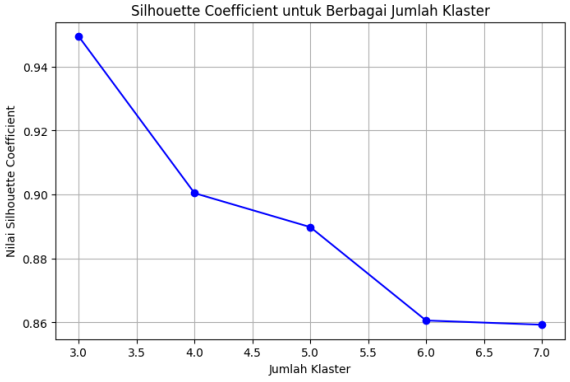
Preprocessing data merupakan tahap awal yang krusial sebelum memulai perhitungan dalam penelitian ini. Dalam penelitian ini, sebelum dilakukan perhitungan dengan metode *K-Means*, data pengguna melalui proses ini. Penghapusan data kosong dari data pengguna yang tidak ada interaksi apapun terhadap aplikasi. Hal ini dilakukan dengan pertimbangan bahwa data yang kosong pada semua fitur tidak memberikan informasi yang bermanfaat dalam analisis selanjutnya. Penghapusan data kosong ini diharapkan dapat mempersiapkan *dataset* yang lebih bersih dan lebih informatif untuk proses selanjutnya dalam penelitian ini.

4.3. Normalisasi Data

Setelah tahap *preprocessing* data, langkah selanjutnya yang dilakukan adalah normalisasi menggunakan metode *Min-Max*. Pada penelitian ini, normalisasi *Min-Max* diterapkan pada data pengguna aplikasi *MPStore* yang telah melewati tahap *preprocessing* sebelumnya. Tujuan dari normalisasi ini adalah untuk memastikan setiap atribut memiliki skala yang seragam, memudahkan proses perhitungan *K-Means*.

4.4. Implementasi Sistem

Pada penelitian ini, program *Python* digunakan untuk mendukung proses analisis data dengan data yang lebih lengkap dari perhitungan sebelumnya. *Python* dipilih karena memiliki pustaka-pustaka yang mumpuni untuk kebutuhan analisis data. Implementasi ini mencakup penerapan algoritma *K-Means* untuk klasterisasi pengguna dan algoritma *Apriori* untuk analisis pola pembelian. Hasil dari program divisualisasikan untuk mempermudah interpretasi data.



Gambar 2. Hasil *silhouette coefficient* dari setiap jumlah kluster yang digunakan

Klasterisasi dilakukan untuk mengelompokkan pengguna aplikasi *MPStore* berdasarkan karakteristik dan aktivitas pengguna. Setelah dilakukan *preprocessing* dan normalisasi data, algoritma *K-Means* diterapkan pada data pengguna sejumlah 72.941 data dengan jumlah klaster bervariasi, yaitu 3 hingga 7 klaster. Berdasarkan nilai *Silhouette Coefficient*, jumlah klaster terbaik adalah 3 klaster dengan nilai 0,91, yang menunjukkan struktur klaster yang kuat.

Tabel 4. Hasil dari *clustering* 3 kluster

Kluster	Banyak anggota
C1	22193
C2	6
C3	318

Setelah klasterisasi, algoritma *Apriori* diterapkan pada data transaksi pengguna di setiap klaster (C1, C2, dan C3). Analisis ini bertujuan untuk menemukan pola pembelian yang signifikan dalam setiap klaster, menggunakan parameter minimum *support* sebesar 5% dan minimum *confidence* sebesar 50%. Data transaksi dipisah berdasarkan kluster dan ditemukan untuk jumlah data dapat dilihat pada tabel berikut.

Tabel 5. Data transaksi setiap kluster

Kluster	Banyak Data Transaksi
C1	3489
C2	2073
C3	462

Setelah data transaksi dipisahkan berdasarkan klaster, algoritma *Apriori* diterapkan pada masing-masing klaster untuk menganalisis pola pembelian. Analisis ini dilakukan untuk menemukan aturan asosiasi yang signifikan, dengan parameter minimum *support* sebesar 5% dan minimum *confidence* sebesar 50%. Berikut jumlah aturan asosiasi yang ditemukan dari setiap kluster.

Tabel 6. Jumlah aturan asosiasi setiap kluster

Kluster	Jumlah Aturan Asosiasi
C1	2
C2	1069
C3	8

Tabel 7. Aturan asosiasi kluster 1

Item	*Support Antecedent %*	*Support Item %*	*Confidence %*
PLN50 => PLN20	18	18,6	50
SP5 => SP10	11,84	14,47	53,27

Tabel 8. Aturan asoasiasi kluster 2 (8 Data Contoh dari Total 1069)

Item	*Support Antecedent %*	*Support Item %*	*Confidence %*
AX10 => DN100	11,82	26,58	50,61
AX10 => DN50	11,82	33,04	56,73
AX10 => PLN20	11,82	30,78	54,29
BANKBCA014 => DN50	7,86	33,04	66,26

Item	*Support Antecedent %*	*Support Item %*	*Confidence %*
BANKBCA014 => I10	7,86	25,08	64,42
BANKBRI002 => DN100	16,59	26,58	54,36
BANKBRI002 => DN50	16,59	33,04	59,3
BANKBRI002 => I10	16,59	25,08	54,36

Tabel 9. Aturan asosiasi kluster 3

Item	*Support Antecedent %*	*Support Item %*	*Confidence %*
AX10 => I10	11,9	24,68	52,73
DN100 => X10	11,04	20,35	52,94
I25 => I10	11,04	24,68	52,94
I5 => I10	10,82	24,68	62
SP5 => SP10	14,29	21,21	56,06
X5 => X10	8,23	20,35	68,42
X50 => X10	10,61	20,35	59,18
X50 => X5	10,61	8,23	57,14
X5 => X50	8,23	10,61	73,68

Pada klaster C1, ditemukan 2 aturan asosiasi yang signifikan. Salah satu contohnya adalah aturan bahwa jika seorang pengguna membeli produk PLN50, maka terdapat peluang sebesar 50% (*confidence*) juga membeli produk PLN20.

Pada klaster C2, ditemukan total 1069 aturan asosiasi. Sebagai contoh, produk AX10 memiliki asosiasi yang kuat dengan produk DN50, dengan *confidence* sebesar 56.73%, dan dengan produk PLN20, dengan *confidence* sebesar 54.29%.

Pada klaster C3, ditemukan 8 aturan asosiasi. Salah satu pola pembelian menunjukkan bahwa jika seorang pengguna membeli produk X5, maka terdapat peluang sebesar 73.68% (*confidence*) juga membeli produk X50.

Hasil analisis *Apriori* ini memberikan wawasan yang berharga dalam memahami pola pembelian di setiap klaster. Pola-pola ini dapat dimanfaatkan oleh PT. Mitra Pedagang Indonesia untuk menyusun strategi pemasaran yang lebih personal, meningkatkan efisiensi pengelolaan stok barang, dan mengidentifikasi peluang promosi yang lebih spesifik pada setiap kelompok pelanggan.

5. KESIMPULAN DAN SARAN

Penelitian ini berhasil melakukan klasterisasi pengguna aplikasi *MPStore* menggunakan metode *K-Means* dan analisis pola pembelian menggunakan algoritma *Apriori*. Algoritma *K-Means* efektif dalam mengelompokkan pengguna dengan nilai *Silhouette Coefficient* terbaik sebesar 0,91 untuk tiga klaster. Analisis pola pembelian menggunakan algoritma *apriori* juga berhasil mengidentifikasi aturan asosiasi yang signifikan dalam setiap klaster, yang dapat digunakan untuk menyusun strategi pemasaran yang lebih personal dan efisien.

Berdasarkan hasil penelitian ini, beberapa saran yang dapat diberikan yaitu PT. Mitra Pedagang Indonesia dapat menggunakan hasil klasterisasi dan analisis pola pembelian untuk menyusun strategi pemasaran yang lebih personal dan tepat sasaran. Dengan memahami karakteristik dan preferensi pengguna, perusahaan dapat meningkatkan kualitas layanan yang ditawarkan oleh aplikasi *MPStore*. Hal ini dapat dilakukan dengan menyesuaikan fitur-fitur aplikasi sesuai dengan kebutuhan dan keinginan pengguna dalam setiap klaster.

DAFTAR PUSTAKA

[1] A. Turner, "How Many People Have Smartphones Worldwide (Oct 2023)." Diakses: 13 Oktober 2023. [Daring]. Tersedia pada: https://www.bankmycell.com/blog/how-many-phones-are-in-the-world

[2] M. H. Siregar, "Klasterisasi Penjualan Alat-Alat Bangunan Menggunakan Metode K-Means (Studi Kasus Di Toko Adi Bangunan)," *Jurnal Teknologi Dan Open Source*, vol. 1, no. 2, hlm. 83–91, Des 2018.

[3] A. K. PANDEY, "A Simple Explanation of K-Means Clustering." Diakses: 7 Desember 2023. [Daring]. Tersedia pada: https://www.analyticsvidhya.com/blog/2020/10/a-simple-explanation-of-k-means-clustering/

[4] J. Lasmana Putra, M. Raharjo, T. Alfian Armawan Sandi, dan R. Prasetyo, "Implementasi Algoritma Apriori Terhadap Data Penjualan Pada Perusahaan Retail," *Jurnal PILAR Nusa Mandiri*, vol. 15, no. 1, hlm. 85, Mar 2019, [Daring]. Tersedia pada: https://www.kaggle.com.

[5] I. Dabbura, "K-means Clustering: Algorithm, Applications, Evaluation Methods, and Drawbacks." Diakses: 8 Oktober 2023. [Daring]. Tersedia pada: https://towardsdatascience.com/k-means-clustering-algorithm-applications-evaluation-methods-and-drawbacks-aa03e644b48a

[6] M. A. K. Lutfi dan A. Nilogiri, "Implementasi Algoritma K-Means Clustering Untuk Pengelompokan Minat Konsumen Pada Produk Online Shop," *Universitas Muhammadiyah Jember*, 2019.

[7] D. Ayu, I. C. Dewi, dan K. Pramita, "Analisis Perbandingan Metode Elbow dan Sillhouette pada Algoritma Clustering K-Medoids dalam Pengelompokan Produksi Kerajinan Bali," 2019.

[8] R. Hidayati, A. Zubair, A. H. Pratama, dan L. Indana, "Analisis Silhouette Coefficient pada 6 Perhitungan Jarak K-Means Clustering," *Techno.Com*, vol. 20, no. 2, hlm. 186–197, Mei 2021, doi: 10.33633/tc.v20i2.4556.

[9] I. Haidar, "Implementasi Algoritma Apriori Untuk Mencari Pola Transaksi Penjualan (Studi Kasus: Carroll Kitchen)," 2021.

[10] I. Afolabi dan F. Adegoke, "Analysis Of Customer Satisfaction For Competitive Advantage Using Clustering And Association Rules," 2014. [Daring]. Tersedia pada: www.iaset.us

[11] M. A. Farajian dan S. Mohammadi, "Mining the banking customer behavior using clustering and association rules methods," *International Journal of Industrial Engineering & Production Research*, vol. 21, no. 4, hlm. 239–245, 2010