

IMPLEMENTASI ALGORITMA K-MEANS DALAM PENGELOMPOKKAN TINGKAT TRANSAKSI PRODUK PPOB NUSANTARA

Muhammad Rosyad Dzikrillah¹, Deden Hardan Gutama², Dhina Puspasari Wijaya³, Dita Danianti⁴

^{1,2,3,4} Program Studi Informatika, Fakultas Komputer dan Teknik, Universitas Alma Ata

¹193200060@almaata.ac.id, ²hardan@almaata.ac.id, ³dhina.puspa@almaata.ac.id, ⁴dita@almaata.ac.id

Abstrak

PPOB Nusantara merupakan penyedia aplikasi berbasis layanan pembayaran online dengan lebih dari 800 produk. Dengan produk dan transaksi yang sudah relatif banyak sulit memperoleh informasi berupa jumlah transaksi setiap kategori produk yang akan digunakan untuk menjaga stabilitas ataupun promosi produk. Di PPOB Nusantara proses penggalan data menjadi informasi sebelumnya masih dilakukan manual menggunakan *software* pengolah angka, sehingga rentan kesalahan dan membutuhkan waktu dan proses yang cukup lama. Dalam penelitian ini dilakukan perancangan sistem klusterisasi berbasis *website* untuk mengelompokkan produk berdasarkan tingkat transaksi menggunakan algoritma K-Means. Klaster yang ditentukan adalah 5 (*Very Low, Low, Medium, High, Very High*) dengan data yang digunakan adalah data transaksi periode bulan September 2023 dengan total 91.923 *record*. Atribut yang digunakan adalah kategori dan jumlah transaksi. Hasil pengujian sistem klusterisasi didapatkan 26 kategori dengan 15 kategori dan 157 produk masuk ke dalam klaster *Very Low*, 5 kategori dan 23 produk di klaster *Low*, 3 kategori dan 5 produk di klaster *Medium*, 2 kategori dan 2 produk di klaster *High*, dan 1 kategori dan 1 produk di klaster *Very High*. Selanjutnya, hasil klusterisasi akan digunakan tim marketing untuk merancang strategi pemasaran yang tepat seperti penyesuaian harga, segmentasi pasar, pengembangan produk, dan pengiriman pesan promosi berdasarkan klaster yang ada.

Kata kunci: clustering, data mining, k-means, data transaksi

1. Pendahuluan

Payment Point Online Bank atau disingkat PPOB adalah satu kesatuan sistem yang berfungsi sebagai media interaksi sistem pembayaran tagihan apapun secara online dengan pihak bank sebagai penyelenggara sekaligus penampung dana pelanggan (Purwatiningsih et al., 2021). PPOB Nusantara merupakan salah satu produk dari PT Bakoel Nusantara yang bergerak di bidang penyedia layanan pembayaran untuk berbagai macam jenis transaksi secara online berbasis aplikasi (*web* dan *mobile*). Saat ini PPOB Nusantara memiliki 831 produk dengan 37 kategori yang berupa tagihan pasca bayar dan pra-bayar seperti pembayaran PLN, PDAM, TELKOM, Pulsa, dan lain-lain.

Masalah utama yang dimiliki PPOB Nusantara adalah produk dan transaksi sudah sangat banyak, sehingga tim *frontliner* sulit mengelompokkan produk berdasarkan volume transaksinya yang dimana informasi tersebut akan digunakan oleh tim marketing untuk menjadwalkan strategi promosi. Untuk mengelompokkan data dalam jumlah banyak dengan karakteristik yang sama dapat menggunakan data mining *clustering*. Data mining atau sering disebut Knowledge Discovery in Database (KDD) adalah salah satu bidang *computer science* dimana merupakan kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan

keteraturan, pola hubungan dalam himpunan data yang berukuran besar (Hossain et al., 2019).

Dalam data mining ada beberapa pendekatan seperti klusterisasi, klasifikasi, dan asosiasi (Pradana et al., 2022). Dalam pengolahan data penjualan ditinjau dari beberapa penelitian sebelumnya seperti yang dilakukan oleh (Indarwati et al., 2019) mengimplementasikan metode regression pada data penjualan *smartphone* untuk memprediksi penjualan *smartphone* dengan hasil yang masuk dalam kategori sangat baik. Penelitian lain yang dilakukan oleh (Indriyani & Irfiani, 2019) menerapkan *clustering* pada data penjualan toko perlengkapan *outdoor*, yang menghasilkan tiga *cluster* dimana terdapat 2 jenis barang paling laris, 8 jenis barang cukup laris, dan 18 jenis barang kurang laris.

Dari beberapa penelitian terdahulu, metode yang paling relevan dengan penelitian ini yaitu *clustering*, karena dapat membentuk beberapa kelompok yang merepresentasikan tingkat transaksi. *Clustering* merupakan salah satu teknik data mining yang membagi data ke dalam beberapa kelompok (*cluster*) dimana setiap *cluster* dapat ditempati beberapa anggota secara bersamaan (Yulyantari & ADH, 2018).

Ditinjau dari penelitian sebelumnya, beberapa algoritma *clustering* diantaranya ada K-Means, K-Medoids, dan Fuzzy C-Means. K-Means dan K-Medoids termasuk ke dalam *hard clustering*, dimana

setiap titik data hanya dapat menempati satu *cluster* berbeda dengan Fuzzy C-Means yang termasuk ke dalam *soft clustering*. Perbandingan antara algoritma K-Means dan K-Medoids sebelumnya telah dilakukan. Pada penelitian Aryuni et. al (2018) algoritma K-Means dibandingkan dengan algoritma K-Medoids dengan hasil nilai *average within centeroid distance* (W) sebesar 35.241 untuk K-Means dan sebesar 88.849 untuk K-Medoids. Dengan total waktu eksekusi dari proses *elbow method* berturut-turut untuk K-Means adalah 3 detik sedangkan K-Medoids 13 menit 44 detik sehingga dapat ditarik kesimpulan bahwa algoritma K-Means memiliki performa yang lebih baik dari sisi nilai *average within centeroid distance* dan kompleksitas waktu (Aryuni et al., 2018). Dalam kasus *clustering* data transaksi pada penelitian Kamila, Insanul et. al (2019) algoritma K-Means dan K-Medoids tidak menunjukkan perbedaan signifikan dalam proses pengelompokan data. Sedangkan dari segi kompleksitas waktu K-Means lebih cepat dengan waktu rata-rata 1 detik sedangkan K-Medoids 1 menit 38 detik dengan data yang diproses sebanyak 3.089 record (Kamila et al., 2019).

Pada penelitian ini algoritma yang akan digunakan adalah K-Means karena termasuk dalam *hard clustering* dan juga relatif lebih cepat daripada K-Medoids. K-Means adalah salah satu metode pengelompokan non-hierarki yang dapat mempartisi data ke dalam dua kelompok atau lebih (Premasundari & Yamini, 2019). Adapun langkah-langkah pada algoritma K-Means adalah sebagai berikut (Astuti et al., 2019):

1. Menentukan nilai k sebagai jumlah klaster yang ingin dibentuk
2. Inisialisasi k pusat *cluster*, dapat dilakukan dengan berbagai cara. Pada umumnya dengan cara mengambil secara acak dari data yang ada.
3. Menghitung jarak setiap data input terhadap masing-masing *centroid* menggunakan rumus jarak Euclidean (*Euclidean Distance*) hingga ditemukan jarak yang paling dekat dari setiap data dengan *centroid*. Persamaan (1) merupakan formula perhitungan *Euclidian Distance* (Astuti et al., 2019).

$$De = \sqrt{(X_i - S_i)^2 + (Y_i - t_i)^2} \quad (1)$$

Dengan:

- De adalah *euclidian distance*
 - i adalah banyaknya objek
 - (x, y) merupakan koordinat objek
 - (s, t) merupakan *centeroid*
4. Mengklasifikasi setiap data berdasarkan kedekatannya dengan *centroid* (jarak paling kecil)
 5. Memperbarui nilai *centroid*. Nilai *centroid*

baru diperoleh dari rata-rata *cluster* yang bersangkutan, dengan rumus yang ditunjukkan pada persamaan (2).

$$v_{ij} = \frac{1}{N_i} \sum_{k=0}^{N_i} X_{kj} \quad (2)$$

Dengan:

- v_{ij} adalah *centroid* atau rata-rata *cluster* ke-i untuk variabel ke-j
 - N_i adalah jumlah data yang menjadi anggota klaster ke-i
 - i,k adalah indeks dari *cluster*
 - j adalah indeks dari variabel
 - x_{kj} adalah nilai data ke-k yang ada di dalam klaster tersebut untuk variabel ke-j
6. Melakukan perulangan dari langkah 2 hingga 5, sampai anggota tiap klaster tidak ada yang berubah.

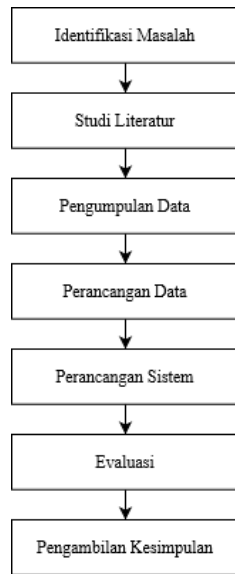
Pada penelitian sebelumnya algoritma K-Means dibandingkan dengan algoritma K-Medoids dengan hasil nilai *average within centeroid distance* (W) sebesar 35.241 untuk K-Means dan sebesar 88.849 untuk K-Medoids. Dengan total waktu eksekusi dari proses *elbow method* berturut-turut untuk K-Means adalah 3 detik sedangkan K-Medoids 13 menit 44 detik sehingga dapat ditarik kesimpulan bahwa algoritma K-Means memiliki performa yang lebih baik dari sisi nilai *average within centeroid distance* dan kompleksitas waktu (Kamila et al., 2019).

Algoritma K-Means sebelumnya telah digunakan pada penelitian Handoko Suhandino et. al (2020) untuk menentukan tingkat penjualan paket data Telkomsel. Algoritma K-Means diimplementasikan pada sistem berbasis *website* dengan hasil pengujian memiliki presentase kesesuaian sebesar 100% dengan perhitungan manual. Kelompok yang ditentukan yaitu penjualan rendah, penjualan sedang, dan penjualan tinggi. Selanjutnya untuk hasil klasterisasi dengan kelompok penjualan rendah dan sedang akan diusulkan untuk diadakan promosi (Handoko et al., 2020).

Berdasarkan permasalahan dan penelitian terdahulu sebelumnya, maka pada penelitian ini akan dibangun sistem berbasis web dengan algoritma K-Means untuk mengelompokkan data transaksi PPOB Nusantara menjadi 5 kelompok yaitu *very low*, *low*, *medium*, *hgh*, dan *very high*.

2. Metode

Jenis penelitian yang digunakan pada penelitian ini adalah penelitian kuantitatif. Penelitian kuantitatif merupakan penelitian yang bersifat objektif, dapat diperkirakan, dapat diukur instrumennya berupa data atau statistik yang valid dan tepat (Danuri & Maisaroh, 2019). Adapun desain penelitian pada penelitian ini seperti yang ada di gambar 1.



Gambar 1. Desain Penelitian

Langkah-langkah pada gambar 1 dijabarkan sebagai berikut:

1. Identifikasi masalah: Pada tahap ini dilakukan proses perumusan masalah. Kesimpulan dari masalah yang ditemui yaitu proses pengelompokan data transaksi produk dari PPOB Nusantara yang akan digunakan untuk pengambilan keputusan dalam strategi marketing masih dilakukan manual sehingga menghambat kinerja dari tim marketing maupun *frontliner* dan memiliki resiko *human error* yang menyebabkan tingkat akurasi informasi berkurang. Maka dari itu dibutuhkan sistem pengelompokan data transaksi berdasarkan tingkat ataupun volume transaksi yang dapat mengelompokkan data secara cepat.
2. Studi literatur: Pada studi literatur, penggunaan metode K-Means diperoleh dari jurnal ilmiah dan penelitian terdahulu. Target dari studi literatur adalah memperoleh metode yang tepat dengan masalah yang akan diselesaikan. Dengan banyaknya data yang dimiliki PPOB Nusantara maka metode yang dapat menyelesaikan pengelompokan secara cepat yaitu K-Means.
3. Pengumpulan data: Pengumpulan data dilakukan dengan observasi ke kantor dan wawancara dengan manajemen PT Bakoel Nusantara. Sedangkan untuk data yang akan diolah menggunakan data sekunder dari tim PPOB Nusantara
4. Perancangan data: Pada tahap ini dilakukan penyesuaian data mentah menjadi data yang siap diolah menggunakan algoritma K-Means.
5. Perancangan sistem: Tahap ini berupa perencanaan serta perancangan sistem yang akan dibangun pada penelitian ini. Proses ini menggunakan metode *waterfall*.

6. Evaluasi: Pada tahap ini akan dilakukan evaluasi pada sistem yang telah dirancang dengan pengujian menggunakan *black-box testing*.
7. Pengambilan kesimpulan: Pada tahap ini akan dilakukan penarikan kesimpulan dari penelitian yang telah dilakukan.

2.1 Sumber Data

Subjek pada penelitian ini adalah PPOB Nusantara dengan populasi yaitu data transaksi pada bulan september 2023 dengan total 91.923 *record*. Teknik pengumpulan data dengan wawancara tidak terstruktur kepada HR Manager PT Bakoel Nusantara, observasi, dan data primer. Data primer merupakan data yang langsung dikumpulkan peneliti juga disebut sebagai data asli atau baru yang belum pernah dipublikasi sebelumnya (Ulfa, 2021). Adapun contoh data primer yang bisa dilihat pada tabel 1.

Tabel 1. Data Transaksi PPOB Nusantara

Kode Produk	Nama Produk	Kategori	Status
PULSA	Pulsa	PULSA	refund
PLNPREPAID	PLN Prabayar	PLN	success
...
DANA50	DANA 50RB	DANA	success

Data transaksi yang diperoleh kemudian dilakukan penyesuaian pada tahap perancangan data. Gambaran data yang nantinya diproses dengan algoritma K-Means adalah seperti di tabel 2.

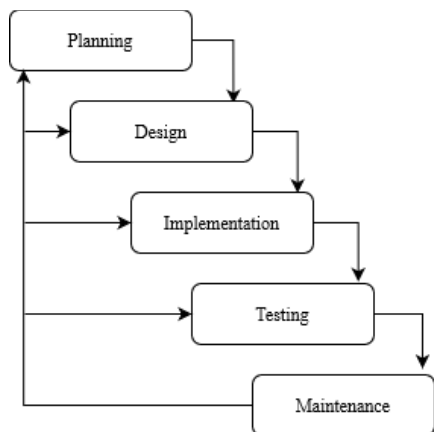
Tabel 2. Proses Transformasi Data

Kategori	Jumlah	Log
VPLN	1	0,0
R_GAME	1	0,0
KVISION_DENOM	4	0,6
GODRIVE	8	0,9
Telkomvision	12	1,1
ETOLL	13	1,1
Top TV	34	1,5
TIKET_KERETA_API	67	1,8
TV_BERLANGGANAN	70	1,8
...
TELCO_POSTPAID	175	2,2

2.2 Pengembangan Sistem

Metode pengembangan sistem pada penelitian ini menggunakan metode *waterfall*. Metode *waterfall* adalah pendekatan yang menggambarkan pengembangan perangkat lunak secara sistematis dan berurutan, langkah demi langkah. Dimulai dari tahap spesifikasi kebutuhan pengguna, metode ini melanjutkan ke tahapan perencanaan seperti perencanaan, permodelan, konstruksi, pengujian sistem, dan akhirnya penyerahan sistem kepada pengguna (Kurniawan et al., 2020).

Adapun tahapan dari metode waterfall seperti pada gambar 2 (Ihsani et al., 2022).



Gambar 2. Tahapan Metode Waterfall

Penjelasan dari tahapan waterfall dirincikan sebagai berikut:

1. Planning

Tahap ini melibatkan pengumpulan dan pemahaman terhadap kebutuhan pengguna atau pemangku kepentingan terkait sistem yang akan dikembangkan.

Hasil dari wawancara didapatkan tujuan dan persyaratan sistem. Tujuan dari sistem untuk mengklasterisasi data transaksi berdasarkan kategori produk. *Output* yang di dapat adalah informasi berupa kelompok produk berdasarkan tingkat/volume transaksi.

Adapun beberapa persyaratan sistem yang diharapkan:

- Memiliki sistem autentifikasi untuk *user* (target user: marketing, *frontliner*).
- Dapat mengunggah data transaksi (dalam bentuk file CSV) untuk klasterisasi.
- Dapat menampilkan, mengedit, dan menambah produk.
- Dapat melakukan klasterisasi data transaksi.
- Dapat menampilkan hasil klasterisasi

2. Design

Pada tahap ini, dilakukan perancangan sistem secara menyeluruh berdasarkan kebutuhan yang telah diidentifikasi. Perancangan mencakup perancangan arsitektur sistem, perancangan antarmuka pengguna, perancangan basis data, dan perancangan modul atau komponen sistem.

3. Implementation

Tahap implementasi melibatkan pengkodean atau penulisan program berdasarkan rancangan yang telah dibuat. Kode program yang dihasilkan harus mengikuti standar yang telah ditetapkan dan memenuhi kebutuhan yang telah didefinisikan. Proses pengkodean menggunakan bahasa pemrograman python dan *framework* flask serta MySQL sebagai *Database Management System*. Pada

tahap ini juga dilakukan implementasi algoritma K-Means pada sistem yang dirancang.

4. Testing

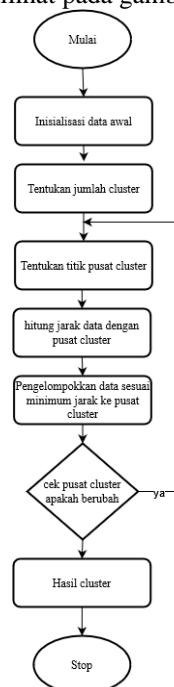
Tahap pengujian dilakukan untuk memastikan bahwa perangkat lunak yang dikembangkan berfungsi dengan baik dan memenuhi persyaratan yang telah ditetapkan sebelumnya. Pengujian dapat meliputi pengujian unit, pengujian integrasi, pengujian sistem, dan pengujian penerimaan. Pengujian yang digunakan pada penelitian ini adalah *black-box testing*.

5. Maintenance

Setelah perangkat lunak digunakan, tahap pemeliharaan dilakukan untuk memperbaiki kesalahan (*bug*) yang mungkin muncul, mengoptimalkan kinerja, dan menghadapi perubahan atau peningkatan yang dibutuhkan.

2.3 Flowchart

Flowchart adalah representasi grafis berupa diagram yang digunakan untuk menggambarkan urutan langkah-langkah atau proses dalam suatu sistem, algoritma, atau tugas. *Flowchart* digunakan untuk menyajikan informasi secara visual dan sistematis, sehingga memudahkan pemahaman dan komunikasi mengenai proses yang terjadi (Gutama, 2019). Adapun *flowchart* dari metode yang digunakan bisa dilihat pada gambar 3.



Gambar 3. Flowchart K-Means

3. Hasil dan Pembahasan

3.1 Perhitungan Manual

Perhitungan manual merupakan gambaran mengenai proses algoritma K-Means yang diimplementasikan pada sistem yang dirancang. Proses perhitungan manual menggunakan data

transaksi yang berasal dari sumber data yaitu transaksi bulan september 2023.

Tabel 3. Data Transaksi September

Tgl	Kode Produk	Nama Produk	Kategori	Status
01/09/2023	PULSA	Pulsa Telepon Prabayar	PULSA	REFUND
01/09/2023	PLNPRE PAIDB	PLN Prepaid	PLN_BUK OPIN	SUCCESS
01/09/2023	PULSA	Pulsa Telepon Prabayar	PULSA	SUCCESS
01/09/2023	PULSA	Pulsa Telepon Prabayar	PULSA	SUCCESS
01/09/2023	PLNPOS TPAIDB	PLN Pasca Bayar	PLN_BUK OPIN	SUCCESS
01/09/2023	PULSA	Pulsa Telepon Prabayar	PULSA	SUCCESS
01/09/2023	PULSA	Pulsa Telepon Prabayar	PULSA	SUCCESS
01/09/2023	PULSA	Pulsa Telepon Prabayar	PULSA	SUCCESS
01/09/2023	PLNPRE PAIDB	PLN Prepaid	PLN_BUK OPIN	SUCCESS
...
30/09/2023	PULSA	Pulsa Telepon Prabayar	PULSA	SUCCESS

Data transaksi kemudian dilakukan proses transformasi dengan atribut kategori, jumlah, dan nilai logaritma natural dari jumlah, menjadi seperti tabel 4.

Tabel 4. Transformasi Data

No	Kategori	Jumlah	Log_value
1	VPLN	1	0,0
2	R_GAME	1	0,0
3	KVISION_DENOM	4	0,6
4	GODRIVE	8	0,9
5	Telkomvision	12	1,1
6	ETOLL	13	1,1
7	Top TV	34	1,5
8	TIKET_KERETA_API	67	1,8
9	TV_BERLANGGANAN	70	1,8
10	TELCO_POSTPAID	175	2,2
11	GOPAY	190	2,3
12	GRAB	212	2,3
13	BPJSTK	214	2,3
14	PSW_II	267	2,4
15	TELKOM_GROUPB	270	2,4
16	MULTI_FINANCE	414	2,6
17	PGN	418	2,6
18	PBB	594	2,8
19	SHOPEE	609	2,8
20	TRFBANK	1387	3,1
21	DANA	1784	3,3
22	BPJS_KES	3059	3,5
23	TELKOM_GROUP	3292	3,5
24	PDAM_II	8155	3,9
25	PULSA	12830	4,1
26	PLN_BUKOPIN	57843	4,8

Selanjutnya yaitu menentukan nilai K dan kondisi berhenti. Nilai K yang ditentukan adalah 5. K=5 (Very Low, Low, Medium, High, Very High). Sedangkan untuk kondisi berhenti adalah sebagai berikut:

- Kondisi 1: Ketika pusat cluster tidak berubah.
- Kondisi 2: maksimal iterasi 100 kali.

Tahap berikutnya adalah inialisasi pusat kluster (centroids) dengan memilih secara acak (dinamis) dari data. Pusat kluster dapat dilihat di tabel 5.

Tabel 5. Insialiasasi Centroids

Kategori	Centroids	Jumlah	Log_value
GODRIVE	1	8	0,9
TELCO_POSTPAID	2	175	2,2
MULTI_FINANCE	3	414	2,6
TRFBANK	4	1387	3,1
PULSA	5	12830	4,1

Selanjutnya menghitung jarak antar data dengan pusat kluster (centroids) dengan rumus *eculidance distance*. Misalkan untuk kategori PULSA

$$C1 = \sqrt{(12830 - 8)^2 + (4,1 - 0,9)^2} = 12822$$

$$C2 = \sqrt{(12830 - 175)^2 + (4,1 - 2,2)^2} = 12655$$

$$C3 = \sqrt{(12830 - 414)^2 + (4,1 - 2,6)^2} = 12416$$

$$C4 = \sqrt{(12830 - 1387)^2 + (4,1 - 3,1)^2} = 11443$$

$$C5 = \sqrt{(12830 - 12830)^2 + (4,1 - 4,1)^2} = 0$$

Begitu juga dengan data lain.

Selanjutnya yaitu mengklasifikasi data berdasarkan jarak terdekat dengan pusat kluster seperti pada tabel 6.

Tabel 6. Klasifikasi Data

Produk Kategori	C1	C2	C3	C4	C5
VPLN	1				
R_GAME	1				
KVISION_DENOM	1				
GODRIVE	1				
Telkomvision	1				
ETOLL	1				
Top TV	1				
TIKET_KERETA_API	1				
TV_BERLANGGANAN	1				
TELCO_POSTPAID		1			
GOPAY		1			
GRAB		1			
BPJSTK		1			
PSW_II		1			
TELKOM_GROUPB		1			
MULTI_FINANCE			1		
PGN			1		
PBB			1		
SHOPEE			1		
TRFBANK				1	
DANA				1	
BPJS_KES				1	
TELKOM_GROUP				1	
PDAM_II					1
PULSA					1
PLN_BUKOPIN					1

Tahap berikutnya memperbarui nilai centroids (pusat kluster) berdasarkan rata-rata yang diperoleh dari hasil klusterisasi.

Tabel 7. Update Pusat Kluster

Centroids	Jumlah	Log_value
1	29,71428571	1,271561
2	221,33333333	2,339403
3	508,75	2,699145
4	2380,5	3,349128
5	26276	4,260634

Langkah terakhir mengulangi tahap perhitungan jarak antar data dengan pusat kluster dan *update* pusat kluster sampai kondisi berhenti. Dari proses perhitungan didapatkan hasil seperti pada tabel 8.

Tabel 8. Hasil Klasterisasi.

No	Kategori	Jumlah	Log_value	Kluster
1	VPLN	1	0,0	1
2	R_GAME	1	0,0	1
3	KVISION_DENOM	4	0,6	1
4	GODRIVE	8	0,9	1
5	Telkomvision	12	1,1	1
6	ETOLL	13	1,1	1
7	Top TV	34	1,5	1
8	TIKET_KERETA_A PI	67	1,8	1
9	TV_BERLANGGAN AN	70	1,8	1
10	TELCO_POSTPAID	175	2,2	1
11	GOPAY	190	2,3	1
12	GRAB	212	2,3	1
13	BPJSTK	214	2,3	1
14	PSW_II	267	2,4	1
15	TELKOM_GROUPB	270	2,4	1
16	MULTI_FINANCE	414	2,6	2
17	PGN	418	2,6	2
18	PBB	594	2,8	2
19	SHOPEE	609	2,8	2
20	TRFBANK	1387	3,1	2
21	DANA	1784	3,3	3
22	BPJS_KES	3059	3,5	3
23	TELKOM_GROUP	3292	3,5	3
24	PDAM_II	8155	3,9	4
25	PULSA	12830	4,1	4
26	PLN_BUKOPIN	57843	4,8	5

Setelah klasterisasi selesai, setiap kluster dihitung rata-ratanya untuk kemudian dilabeli berdasarkan tingkat transaksinya.

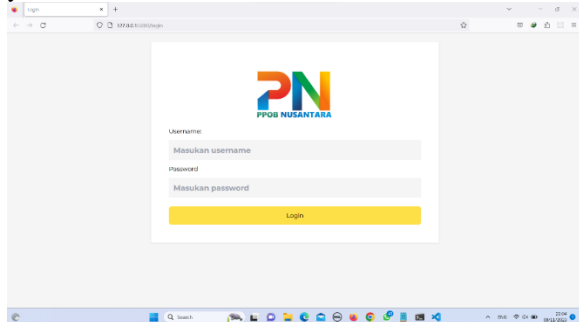
Tabel 9. Melabeli Setiap Kluster

Cluster	Average (jumlah)	Average (log_value)	Label
1	157,8666667	1,878368	Very Low
2	1093,5	2,987969	Low
3	3175,5	3,50152	Medium
4	10492,5	4,009825	High
5	57843	4,8	Very High

3.2 Hasil Antarmuka Sistem

1. Halaman Login

Saat user mengakses sistem klasterisasi maka akan diarahkan ke halaman *login*. Pada tampilan ini user harus memasukkan *username* dan *password*. Jika *username* dan *password* sesuai dan lolos verifikasi maka akan masuk ke halaman selanjutnya, yaitu halaman *dashboard*.

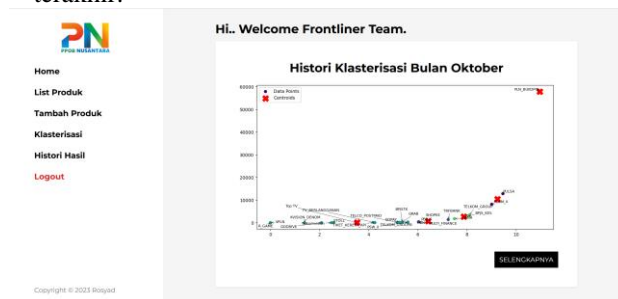


Gambar 4. Halaman Login

2. Halaman Dashboard

Halaman pertama yang tampil setelah user *login* adalah halaman *dashboard*. Pada halaman ini

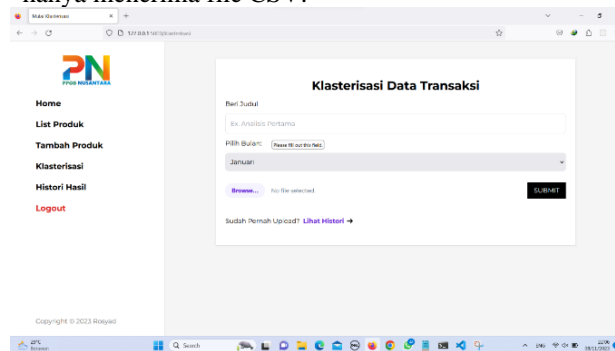
ditampilkan informasi berupa grafik hasil klasterisasi terakhir.



Gambar 5. Halaman Dashboard

3. Halaman Upload Data Transaksi

Adalah halaman untuk memulai tahap klasterisasi yaitu form upload data transaksi. Dengan inputan judul, bulan, dan file data transaksi yang hanya menerima file CSV.



Gambar 6. Halaman Upload Data Transaksi

4. Halaman Hasil

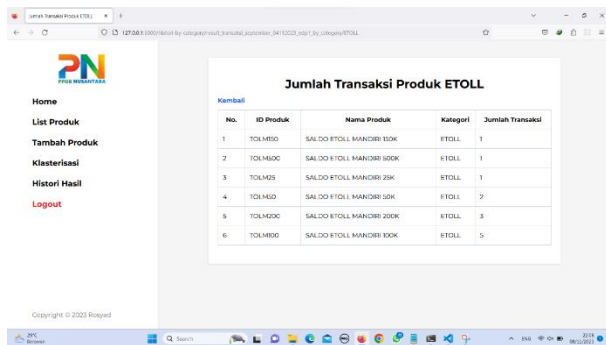
Tampilan hasil dari klasterisasi seperti yang ada di gambar 8. Hasil yang ditampilkan berupa klasifikasi kategori berdasarkan tingkat transaksinya. Dengan setiap kategori dapat di klik untuk melihat anggota produk beserta jumlah transaksinya. Tabel hasil klasterisasi diurutkan dari yang terendah yaitu *very low, low, medium, high, very high* dikarenakan transaksi dengan tingkat penjualan rendah akan lebih diperhatikan untuk segera dilakukan evaluasi dan promosi.

No	Kategori	Jumlah Transaksi
1	VPLN	1
2	R_GAME	1
3	KVISION_DENOM	4
4	GODRIVE	8
5	Telkomvision	12
6	ETOLL	13
7	Top TV	34
8	TIKET_KERETA_API	67
9	TV_BERLANGGANAN	70

Gambar 7. Halaman Hasil

5. Halaman Sub Kategori

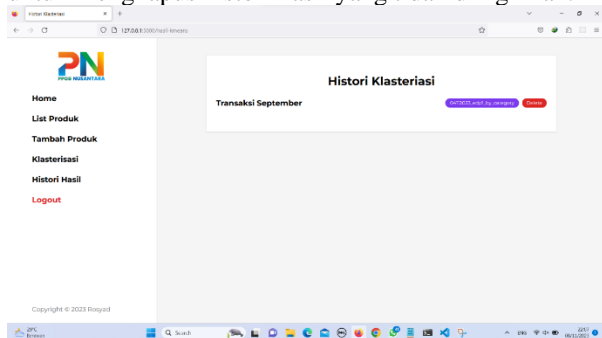
Pada tampilan ini adalah halaman sub kategori yang berisi anggota produk beserta jumlah transaksinya, diurutkan secara *ascending*.



Gambar 8. Halaman Sub Kategori

6. Halaman Histori

Seluruh hasil klasterisasi akan tersimpan di dalam halaman histori klasterisasi. Dengan judul diambil dari nama bulan yang telah diinputkan dan juga terdapat judul yang dimasukkan oleh pengguna untuk membedakan jika dilakukan klasterisasi lebih dari satu untuk bulan yang sama. Terdapat fitur *delete* untuk menghapus histori hasil yang tidak diinginkan.



Gambar 9. Halaman Histori Hasil

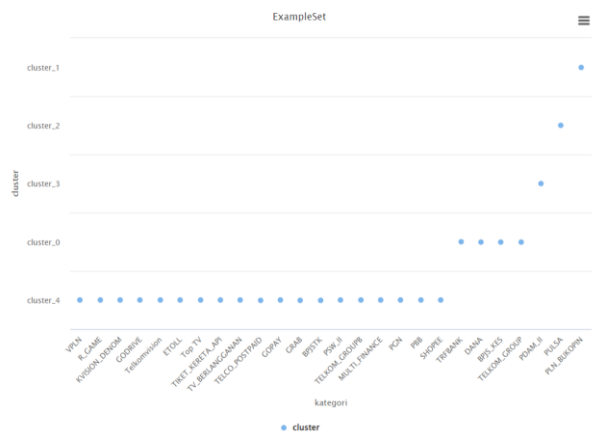
3.3 Pengujian

Pengujian metode K-Means pada penelitian ini dilakukan dengan membandingkan perhitungan manual pada tabel 8 dengan hasil proses menggunakan RapidMiner. Hasil dari proses K-Means di RapidMiner bisa dilihat pada tabel 10.

Tabel 10. Hasil Perhitungan RapidMiner

No	Kategori	Jumlah	Log_value	Klaster
1	TRFBANK	1387	3.1	0
2	DANA	1784	3.3	0
3	BPJS_KES	3059	3.5	0
4	TELKOM_GROUP	3292	3.5	0
5	PLN_BUKOPIN	57843	4.8	1
6	PULSA	12830	4.1	2
7	PDAM_II	8155	3.9	3
8	VPLN	1	0	4
9	R_GAME	1	0	4
10	KVISION_DENOM	4	0.6	4
11	GODRIVE	8	0.9	4
12	Telkomvision	12	1.1	4
13	ETOLL	13	1.1	4
14	Top TV	34	1.5	4
15	TIKET_KERETA_A PI	67	1.8	4
16	TV_BERLANGGAN AN	70	1.8	4
17	TELCO_POSTPAID	175	2.2	4
18	GOPAY	190	2.3	4
19	GRAB	212	2.3	4
20	BPJSTK	214	2.3	4
21	PSW_II	267	2.4	4
22	TELKOM_GROUPB	270	2.4	4
23	MULTI_FINANCE	414	2.6	4
24	PGN	418	2.6	4

25	PBB	594	2.8	4
26	SHOPEE	609	2.8	4



Gambar 10. Grafik Hasil Perhitungan

Dari hasil perbandingan terdapat sedikit perbedaan. Perbedaan disebabkan inisialisasi *centroids* awal yang dilakukan secara acak.

Selanjutnya dilakukan pengujian sistem menggunakan *blackbox testing*. *Blackbox Testing* merupakan sebuah pendekatan yang digunakan untuk menguji perangkat lunak tanpa memperhatikan detail internal dari perangkat lunak tersebut. Pengujian dilakukan dengan memeriksa hasil keluaran berdasarkan masukan yang diberikan, tanpa memperhatikan bagaimana kode program bekerja di dalamnya. Proses *Black Box Testing* melibatkan pengujian program yang telah dibuat dengan memberikan data ke setiap formulir yang ada. Tujuan dari pengujian ini adalah untuk memastikan bahwa program berjalan sesuai dengan kebutuhan yang ditetapkan oleh perusahaan (Ningrum et al., 2019). Adapun hasil *blackbox testing* pada sistem yang dirancang dapat dilihat pada tabel 11.

Tabel 11. Pengujian Blackbox

No	Menu	Aksi	Tampilan	Hasil
1	Halaman Login	Login dan logout	- Form input username, password, button login	SUKSES
2	Dashboard	Menekan menu mulai klasterisasi, histori transaksi, dan menu di sidebar	- Menampilkan menu, ringkasan produk, dan histori klasterisasi	SUKSES
3	Klasterisasi	Memasukkan inputan dan menekan tombol submit	- Form input upload data - Memproses data dan redirect ke halaman hasil	SUKSES
4	Hasil	Menekan kategori, menekan button klaster by produk	- Tabel berdasarkan klaster anggota dari kategori tertentu dengan	SUKSES

			jumlah kluster	
			- Menampilkan hasil klasterisasi berdasarkan ID Produk	
5	Histori	Menekan judul	- Menampilkan daftar histori hasil	SUKSES
			- Redirect ke hasil klasterisasi	
6	Produk	Tambah data, edit data, hapus data	- Menampilkan daftar produk	SUKSES
			- Redirect ke form edit	
			- Redirect ke halaman tambah produk	
			- Menghapus produk	

4. Kesimpulan dan Saran

Berdasarkan hasil perancangan diperoleh hasil berupa sistem klasterisasi tingkat transaksi produk PPOB Nusantara berbasis website. Sistem ini dirancang untuk membantu tim *frontliner* dan marketing dalam mengelompokkan produk berdasarkan volume transaksi. Berdasarkan uji coba yang dilakukan dapat disimpulkan bahwa sistem klasterisasi tingkat transaksi produk PPOB Nusantara berbasis website menggunakan algoritma K-Means dapat menyelesaikan permasalahan pengelompokkan produk berdasarkan tingkat transaksi. Terdapat fitur tambah produk dan upload data transaksi berdasarkan bulan, sehingga sistem yang dirancang dapat digunakan secara berkelanjutan. Berdasarkan hasil pengujian yang dilakukan menggunakan data transaksi bulan september sebanyak 91.923 record dengan 26 kategori dan 5 cluster di dapatkan hasil 15 kategori dan 157 produk masuk ke dalam klaster *Very Low*, 5 kategori dan 23 produk di klaster *Low*, 3 kategori dan 5 produk di klaster *Medium*, 2 kategori dan 2 produk di klaster *High*, dan 1 kategori dan 1 produk di klaster *Very High*, hasil klasterisasi selanjutnya akan digunakan tim marketing untuk merencanakan strategi promosi secara berkala mulai dari kelompok transaksi yang paling rendah.

Pada penelitian ini inialisasi nilai *centroids* awal masih dilakukan secara acak sehingga ketika dilakukan proses yang sama, terdapat kemungkinan hasil klaster yang sedikit berbeda. Dan untuk jumlah k yang ditentukan masih statis, yaitu k=5 sesuai permintaan dari tim marketing. Pada penelitian selanjutnya mungkin dapat ditambahkan metode inialisasi *centroids* awal dan metode penentuan k yang optimal seperti *elbow* atau metode yang lain.

Daftar Pustaka:

Aryuni, M., Didik Madyatmadja, E., & Miranda, E. (2018). *PENERAPAN K-MEANS DAN K-MEDOIDS CLUSTERING PADA DATA*

INTERNET BANKING DI BANK XYZ APPLICATION OF K-MEANS AND K-MEDOIDS CLUSTERING ON INTERNET BANKING DATA IN XYZ BANK.

Astuti, D., Rahmat Iskandar, A., & Febrianti, A. (2019). *Penentuan Strategi Promosi Usaha Mikro Kecil Dan Menengah (UMKM) Menggunakan Metode CRISP-DM dengan Algoritma K-Means Clustering*. 1(2), 60–072. <https://doi.org/10.20895/INISTA.V1I2>

Danuri, & Maisaroh, S. (2019). *Metode Penelitian Pendidikan*.

Gutama, D. H. (2019). PERANCANGAN SISTEM PELELANGAN BERITA BERBASIS WEBSITE. *Indonesian Journal of Business Intelligence (IJUBI)*, 2(1), 32. <https://doi.org/10.21927/ijubi.v2i1.1023>

Handoko, S., Fauziah, F., & Handayani, E. T. E. (2020). IMPLEMENTASI DATA MINING UNTUK MENENTUKAN TINGKAT PENJUALAN PAKET DATA TELKOMSEL MENGGUNAKAN METODE K-MEANS CLUSTERING. *Jurnal Ilmiah Teknologi Dan Rekayasa*, 25(1), 76–88. <https://doi.org/10.35760/tr.2020.v25i1.2677>

Hossain, M. Z., Akhtar, M. N., Ahmad, R. B., & Rahman, M. (2019). A dynamic K-means clustering for data mining. *Indonesian Journal of Electrical Engineering and Computer Science*, 13(2), 521–526. <https://doi.org/10.11591/ijeecs.v13.i2.pp521-526>

Ihsani, I., Pramuntadi, A., Gutama, D. H., & Wijaya, D. P. (2022). IMPLEMENTASI ALGORITMA GENETIKA DALAM PENENTUAN RUTE OPTIMAL UNTUK KURIR KANTOR POS BERBASIS WEB (STUDI KASUS: KANTOR POS WATES). *Indonesian Journal of Business Intelligence (IJUBI)*, 5(2), 76. <https://doi.org/10.21927/ijubi.v5i2.2662>

Indarwati, T., Irawati, T., & Rimawati, E. (2019). PENGGUNAAN METODE LINEAR REGRESSION UNTUK PREDIKSI PENJUALAN SMARTPHONE. *Jurnal Teknologi Informasi Dan Komunikasi (TIKOMSiN)*, 6(2). <https://doi.org/10.30646/tikomsin.v6i2.369>

Indriyani, F., & Irfiani, E. (2019). *Clustering Data Penjualan pada Toko Perlengkapan Outdoor Menggunakan Metode K-Means* (Vol. 7, Issue 2).

Kamila, I., Khairunnisa, U., & Mustakim. (2019). Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokan Data Transaksi Bongkar Muat di Provinsi Riau. *Jurnal Ilmiah Rekayasa Dan Manajemen Sistem Informasi*, 5(1), 119–125.

Kurniawan, H., Apriliah, W., Kurniawan, I., & Firmansyah, D. (2020). Penerapan Metode

- Waterfall Dalam Perancangan Sistem Informasi Penggajian Pada SMK Bina Karya Karawang. *Jurnal Interkom: Jurnal Publikasi Ilmiah Bidang Teknologi Informasi Dan Komunikasi*, 14(4), 13–23.
<https://doi.org/10.35969/interkom.v14i4.58>
- Ningrum, F. C., Suherman, D., Aryanti, S., Prasetya, H. A., & Saifudin, A. (2019). *Pengujian Black Box pada Aplikasi Sistem Seleksi Sales Terbaik Menggunakan Teknik Equivalence Partitions*. 4(4).
<http://openjournal.unpam.ac.id/index.php/informatika>
- Pradana, M. G., Saputro, P. H., & Wijaya, D. P. (2022). KOMPARASI METODE SUPPORT VECTOR MACHINE DAN NAÏVE BAYES DALAM KLASIFIKASI PELUANG PENYAKIT SERANGAN JANTUNG. *Indonesian Journal of Business Intelligence (IJUBI)*, 5(2), 87.
<https://doi.org/10.21927/ijubi.v5i2.2659>
- Premasundari, M., & Yamini, C. (2019). A VIOLENT CRIME ANALYSIS USING FUZZY C-MEANS CLUSTERING APPROACH. *ICTACT JOURNAL ON SOFT COMPUTING*, 3.
<https://doi.org/10.21917/ijsc.2019.0270>
- Purwatiningsih, A., Lestari, A. W., & Kamaluddin. (2021). *ANALISIS PENGARUH KUALITAS PELAYANAN SISTEM PAYMENT POINT ONLINE BANK (PPOB) TERHADAP KEPUASAN PELANGGAN (STUDI PADA PT. PLN PERSERO AREA PELAYANAN PELANGGAN DAN JARINGAN DI PROBOLINGGO)*. 11, 140–150.
- Ulfa, R. (2021). VARIABEL PENELITIAN DALAM PENELITIAN PENDIDIKAN. *Jurnal Pendidikan Dan Keislaman*, 342–351.
- Yulyantari, L. M., & ADH, I. P. W. (2018). *Manajemen Model pada Sitem Pendukung Keputusan*. Penerbit Andi.

