

## ANALISIS PEMILIHAN CLUSTER OPTIMAL DALAM SEGMENTASI PELANGGAN TOKO RETAIL

Santi Ika Murpratiwi<sup>1)</sup>, I Gusti Agung Indrawan<sup>2)</sup>, Arik Aranta<sup>3)</sup>

<sup>1,2</sup> Program Studi Teknik Informatika, STMIK STIKOM INDONESIA

<sup>3</sup> Fakultas Teknik, Universitas Mataram

Email: santiika@stiki-indonesia.ac.id, agung.indrawan@stiki-indonesia.ac.id, arikaranta@unram.ac.id

### ABSTRAK

Saat ini pemanfaatan data menjadi fokus dalam bidang pemasaran khususnya untuk menyusun strategi. Agar strategi pemasaran bisa tepat sasaran dibutuhkan segmentasi pelanggan. Data mining khususnya clustering mampu membantu proses segmentasi pelanggan. Dalam penelitian ini, data mining diimplementasikan untuk segmentasi pelanggan UD. XYZ dengan metode K-Means, K-medoids, dan Means. Tujuan penelitian ini adalah mencari metode dan nilai k terbaik yang dihasilkan dari ketiga metode clustering yang digabungkan dengan model RFM. Data transaksi berjumlah 153.492 data diimplementasikan ke dalam model RFM menjadi 10.145 data untuk dilakukan identifikasi pelanggan potensial yang selanjutnya digunakan sebagai bahan pengolahan data. Inisialisasi cluster awal pada metode K-Medoids, X-Means, dan K-Means dilakukan secara random. Nilai k dalam penelitian ini diinisialisasi dari 1 sampai 10. Nilai k diimplementasikan secara berulang dan dihitung validasi cluster menggunakan metode *David Bouldin Index (DBI)* dan jarak rata-rata cluster dengan centroid. Hasil penelitian menunjukkan K-medoids memiliki nilai validitas yang lebih baik dibandingkan dengan X-Means dan K-Means. Rata-rata nilai DBI yang dihasilkan metode K-Medoids adalah 0,540778. Jumlah cluster terbaik yang dihasilkan adalah 5 cluster. Hal ini ditentukan dengan mempertimbangkan jumlah persebaran data pada k = 5 yang menghasilkan nilai sama pada metode K-Medoids, X-Means, dan K-Means. Tingkatan pelanggan yang terbentuk adalah *About To Sleep, Customer Needing Attention, Recent Customer, Potential Loyalist, dan Loyal Customers*.

**Kata kunci:** Segmentasi Pelanggan, K-Medoids, X-Means, dan K-Means, DBI

### ABSTRACT

*Data usage has become a focus in the field of marketing, especially for formulating strategies. In order for the marketing strategy to be right on target, customer segmentation is needed. Data mining, especially clustering, can help the customer segmentation process. In this research, data mining is implemented for UD customer segmentation. XYZ using the K-Means, K-medoids, and Means methods. The direction of this study was to find the best method and k value generated from the three clustering methods. This research presents the data mining process by combining the RFM model with the K-Medoids, X-Means, and K-Means algorithms. Transaction data that has been implemented into the RFM model is used as data processing material. Transaction data totaling 153,492 is implemented into the RFM model to 10,145 data for identification of potential customers. Initial cluster initialization in the K-Means, K-Medoids, and X-Means methods was carried out randomly. The k value in this study is initialized from 1 to 10. The k value is implemented repeatedly and the cluster validation is calculated using the David Bouldin Index (DBI) method and the average distance of the cluster to the centroid. The results showed that K-medoids have a good validity value compared to K-Means and X-Means. The average DBI value produced by the K-Medoids method is 0.540778. The best number of clusters produced is 5 clusters, this is determined by considering the amount of data distribution at k = 5 which produces the same value in the K-Means, K-Medoids, and X-Means methods. The levels of customers formed are About To Sleep, Customer Needing Attention, Recent Customers, Potential Loyalists, and Loyal Customers.*

**Keywords :** Customer Segmentation, K-Medoids, X-Means, dan K-Means, DBI

## 1. PENDAHULUAN

Data menjadi hal penting dan berharga di era teknologi informasi karena data diperlukan dalam melakukan penyusunan strategi hingga pengambilan keputusan [1]. Data yang terorganisir mampu membantu pihak manajerial dalam mengambil kesimpulan proses bisnis [2]. Saat ini data menjadi asset yang penting karena data dapat menyebabkan kerugian jika tidak digunakan dengan bijak dan data dapat memberikan keuntungan jika dilakukan pengolahan dan analisis dengan baik [3]. Teknologi pengolahan data saat ini terus berkembang dengan sangat pesat. Pengolahan data yang sedang pesat dikembangkan dan dimanfaatkan saat ini adalah data mining.

Data mining merupakan upaya memanfaatkan data dengan menambang informasi penting yang tersimpan dalam data dengan jumlah besar [4]. Terdapat 5 (Lima) fungsi data mining diantaranya klasifikasi, clustering, asosiasi, *sequencing*, dan *forecasting*. Data mining dimanfaatkan oleh berbagai bidang seperti pemasaran, bisnis, pemerintahan, kedokteran, hingga pendidikan. Contoh implementasinya adalah untuk perkiraan cuaca [5], memprediksi tingkat penularan covid [6], klasifikasi perilaku siswa [7], klasifikasi kanker payudara [8], hingga segmentasi pelanggan [9].

Dalam penelitian ini dilakukan segmentasi pelanggan dengan objek penelitian yaitu UD. XYZ yang merupakan sebuah toko retail bahan dan perlengkapan kue yang ada di Kota Denpasar. Saat ini menjamurnya usaha retail membuat UD. XYZ harus menyiapkan strategi promosi dan pemasaran untuk menjaga stabilitas usahanya. Dengan data penjualan yang dimiliki, segmentasi pelanggan dilakukan untuk membantu membaca karakteristik dan perilaku pelanggan dengan mengimplementasikan fungsi clustering pada data mining.

Dataset yang digunakan adalah data transaksi dan data pelanggan dimana sebelum dilakukan clustering, data akan diinisialisasikan terlebih dahulu untuk mengubah data menjadi numerik dan mudah dilakukan clustering [10]. Proses inialisasi data dilakukan dengan menggunakan model RFM RFM (*Recency, Frequency, Monetary*). Pada proses implementasi RFM, data transaksi pelanggan dibagi menjadi 3 (tiga) atribut yaitu *Recency* untuk menghitung jarak belanja terakhir dengan hari ini, *Frequency* untuk menghitung jumlah berapa kali seorang pelanggan melakukan transaksi, dan *Monetary* untuk mengetahui berapa nominal uang yang sudah dibelanjakan seorang pelanggan dalam waktu tertentu [11]. Hasil dari implementasi model RFM digunakan untuk proses clustering data.

Clustering merupakan upaya penambangan data dengan mengelompokkan data ke dalam cluster yang memiliki karakteristik yang mirip [12]. Dengan clustering data dikelompokkan berdasarkan jarak paling dekat dengan objek lain pada dataset tersebut dan data akan diambil secara acak [13]. Ada beberapa metode clustering yang dapat digunakan untuk mengelompokkan dataset dalam jumlah besar. Pada penelitian ini metode clustering yang digunakan untuk proses segmentasi customer adalah K-Medoids, X-Means, dan K-Means dengan alasan ketiga metode tersebut memiliki efektifitas yang tinggi dalam membagi data sesuai dengan jumlah cluster yang diinginkan oleh user dan ketiga metode tersebut mampu diimplementasikan dalam data dengan jumlah kecil maupun besar.

K-means adalah metode clustering yang sangat populer karena efektivitasnya dalam membuat cluster dengan membagi data ke dalam beberapa partisi untuk data dalam jumlah kecil maupun besar [14]. Sedangkan metode K-Medoids merupakan pengembangan dari metode K-Means, dimana K-Medoids dibuat untuk mengatasi kelemahan K-Means yaitu sangat sensitive terhadap data outlier [15]. Dan metode X-Means merupakan pengembangan dari algoritma K-means. X-means membentuk cluster awal menggunakan algoritma K-means. Setiap cluster awal yang terbentuk dibagi menjadi dua cluster berdasarkan kriteria BIC. Proses ini berulang hingga tidak ada lagi cluster yang dapat dibagi. X-means membutuhkan komputasi yang lebih sedikit daripada k-means dan mampu mengoptimalkan jumlah cluster yang terbentuk [16]. Ketiga metode ini mengharuskan user menentukan jumlah cluster terlebih dahulu. Penetapan jumlah cluster awal sangat penting karena akan mempengaruhi hasil akhir clustering [17]. Penentuan jumlah cluster yang kurang tepat dapat menghasilkan cluster yang tidak optimal dan dapat mempengaruhi informasi cluster yang terbentuk.

Solusi untuk mengatasi kesalahan dalam menentukan jumlah cluster awal, maka dalam penelitian ini akan dilakukan analisis nilai  $k$  yang paling optimal terhadap tiga metode clustering dalam melakukan segmentasi pelanggan UD. XYZ. Dalam penelitian ini, dilakukan analisis terhadap ketiga metode tersebut dengan mengganti nilai  $k$  dari nilai 1 sampai 10 dan membandingkan berdasarkan nilai rata-rata antar centroid dan nilai *Davies bouldin index (DBI)*. Nilai  $k$  optimal yang didapat menjadi acuan untuk melakukan clustering terhadap data transaksi UD. XYZ.

Banyak penelitian yang sudah memanfaatkan data mining untuk segmentasi pelanggan. Salah satunya yaitu penelitian oleh Debora Agustina, dkk pada tahun 2007 yang memanfaatkan metode *K-Means Hierarchical* berdasarkan *recency*, *frequency*, dan *profit* untuk segmentasi pelanggan PT. XYZ. Dalam penelitian tersebut didapat hasil bahwa customer PT. XYZ dapat dikelompokkan menjadi dua dengan detail 102 pelanggan dalam cluster *profitable* dan 61 pelanggan dalam cluster *less profitable/unprofitable customer* [18]. Dalam data mining, clustering menjadi salah satu fungsi yang cocok untuk segmentasi pelanggan dengan metode populernya yaitu K-Means. Penelitian tahun 2018 oleh Yanuar Wicaksono [19] menyatakan bahwa segmentasi pelanggan menggunakan multikriteria dengan metode K-Means mampu menghasilkan cluster yang lebih baik. Implementasi K-means dilakukan pula oleh Nisa Hanum Hanani dkk untuk melakukan segmentasi pelanggan produk digital service Indihome dan didapatkan hasil pelanggan dibagi ke dalam 3(tiga) kelompok berdasarkan tingkat keuntungan yang didapat perusahaan [20]. K-means menjadi metode favorit dalam melakukan segmentasi pelanggan, namun dengan seiring bertambahnya waktu metode K-means terus dikembangkan. Salah satu metode yang dikembangkan untuk mengatasi kelemahan K-Means adalah K-Medoids dan X-Means. Metode K-Medoids juga digunakan untuk melakukan segmentasi pelanggan seperti yang dilakukan oleh Siti Monalisa dan Imelda Erza pada BC 4 HPAI Pekanbaru. K-medoids mampu membagi menjadi 2 Cluster Terbaik agen biasa dan 3 Cluster Terbaik agen stok berdasarkan *DBI (Davies Bouldin Index)* [21]. K-medoids juga diimplementasikan oleh Putri Eka Prakasawati dkk untuk melakukan segmentasi pelanggan berdasarkan produk dengan jumlah pembelian dan area. Didapatkan hasil pelanggan terbagi menjadi 5 cluster [22]. Sedangkan metode X-Means diimplementasikan untuk segmentasi potensi sekolah menengah unggul di Kabupaten Banyumas, oleh Aditya Wijayanto dan Rifki Adytama. Dalam penelitian tersebut sekolah menengah dikelompokkan berdasarkan prestasi, fasilitas, jumlah siswa, dan jumlah guru dan didapatkan hasil empat cluster dengan nilai DBI sebesar 0.846 [23]. Selain itu, Purwa Hasan Putra dan M. Syahputra Novelan mengelompokkan siswa berdasarkan minat bakat dan didapatkan hasil bahwa X-Means mampu mengelompokkan data siswa menjadi 3 dan 4 kelompok [24].

Dalam mengukur efektivitas sebuah metode clustering dilakukan dengan teknik komparasi metode. Dalam penelitiannya, Mediana Aryuni dkk mengimplementasikan metode K-means dan K-medoids untuk segmentasi customer pada Bank XYZ dan didapatkan hasil berdasarkan nilai DBI yang didapat K-Means lebih unggul dibanding K-Medoids [25]. Dalam topik lain, Stephanie Monica dkk membandingkan metode K-Means dan X-Means untuk mengelompokkan objek wisata di Provinsi Bali dan didapatkan hasil metode X-Means menghasilkan distribusi cluster yang lebih merata [26]. Topik penelitian tentang teknologi clustering selain mengimplementasikan dan membandingkan metode juga dapat dilakukan dengan melakukan penentuan jumlah cluster awal. Ni putu Eka Merlina dkk melakukan penelitian untuk mencari nilai k terbaik pada metode K-Means dengan menerapkan metode Elbow dan didapatkan hasil  $k=3$  sebagai jumlah cluster ideal yang selanjutnya akan diimplementasikan ke dalam studi kasus [10]. Penelitian lainnya oleh Rifki Adhitama dkk, dilakukan pencarian cluster ideal untuk menentukan jumlah SMK di Jawa Tengah dengan bantuan metode X-Means dan K-means dan didapatkan hasil jumlah cluster idealnya adalah 4 dilihat dari persebaran nilai DBI yang dihasilkan dengan implementasi X-Means dan K-Means [27].

Dari semua penelitian sebelumnya dapat disimpulkan bahwa hal terpenting dalam melakukan clustering data adalah menentukan metode dan nilai  $k$  awal. Sehingga dalam penelitian ini dilakukan analisis jumlah cluster optimal untuk segmentasi pelanggan toko retail dengan menggunakan 3(tiga) metode clustering yaitu K-Medoids, X-Means, dan K-Means.

## A. Model RFM

Model RFM (*Recency*, *Frequency*, dan *Monetary*) merupakan model untuk mengidentifikasi kelompok pelanggan dengan perlakuan khusus. Model RFM memungkinkan diimplementasikan pada bidang pemasaran karena membutuhkan data history seperti riwayat pembelian dan riwayat penelusuran [28]. Model RFM banyak digunakan karena dinilai sederhana, tidak membutuhkan perangkat lunak statistik khusus, skala yang digunakan numerik yang objektif, dan hasil segmentasi RFM mudah dipahami dan ditafsirkan [11]. Pada Tabel 1 merupakan detail atribut dari model RFM.

Tabel 1. Atribut Model RFM

No.	Attribute	Keterangan
1.	Recency (R)	Jarak terakhir belanja dengan tanggal yang ditentukan

- |                  |   |
|------------------|---|
| 2. Frequency (F) | Jumlah Transaksi yang dilakukan pelanggan |
| 3. Monetary (M)  | Jumlah uang yang dibelanjakan             |

## B. Clustering

Konsep data mining yang mampu mengelompokkan data ke dalam kelompok yang serupa disebut dengan clustering. Clustering mampu membagi data ke dalam sub set dengan tujuan setiap sub set dapat diambil dan dimanfaatkan informasinya. Algoritma clustering terdiri dari dua bagian yaitu secara hirarkis dan secara partitional. Algoritma hirarkis menemukan cluster secara berurutan dimana cluster ditetapkan sebelumnya, sedangkan algoritma partitional menentukan semua kelompok pada waktu tertentu [12].

### a) K-means

Metode populer untuk clustering saat ini adalah K-Means. K-Means merupakan metode untuk menganalisa data tanpa supervise (*unsupervised*) dan dengan sistem partisi. Algoritma ini memiliki inputan data dan nilai  $k$  dengan arti data atau objek akan didistribusikan ke dalam  $k$  cluster dan setiap cluster memiliki nilai titik pusat (*centroid*) yang merepresentasikan cluster tersebut [14].

Metode K-Means dapat diimplementasikan dalam berbagai bidang termasuk citra maupun suara. K-Means memiliki keunggulan diantaranya mudah diimplementasikan, waktu perhitungan dan komputasi yang relative lebih cepat, dan umum digunakan. Dibalik kelebihanannya K-Means memiliki beberapa kekurangan diantaranya sulitnya menentukan nilai  $k$  yang tepat agar hasil clustering optimal, K-means akan susah divisualisasikan jika atribut yang digunakan terlalu banyak karena dalam metode K-means atribut yang digunakan mencerminkan dimensi data pada K-Means, dan proses clustering dilakukan dengan cara random yang mengakibatkan hasil clustering yang berbeda-beda [29].

### b) K-Medoids

Algoritma K-Medoids atau *Partitioning Around Medoids* (PAM) merupakan metode yang dikembangkan untuk mengatasi kekurangan metode K-Means yaitu untuk mengoptimalkan kriteria penyekatan objektif yang dilakukan berdasarkan kedekatan jarak antar objek. Menurut Han dan Kamber (2006) dalam algoritma K-Medoids yang terdiri dari  $n$  objek dipartisi menjadi  $k$  cluster dimana jumlah  $k \leq n$ .

Algoritma K-medoids dipilih karena memiliki kelebihan diantaranya pusat cluster pada K-Medoids adalah suatu data yang secara riil berada di tengah cluster (*median*), posisi medoid juga tidak terlalu terpengaruh terhadap adanya *noise* ataupun distribusi jarak yang berbeda-beda antara data dengan pusat cluster, dan tidak sensitive terhadap outlier. Di samping kelebihan tersebut, K-Medoids juga memiliki kekurangan diantaranya relative kurang efisien, proses yang rumit, sulit menentukan nilai  $k$  yang tepat, optimal untuk cluster yang terpisah, dan ukuran dataset kecil.

### c) X-Means

Menurut Pelleg pada tahun 2000, dilakukan ekstensi terhadap metode K-means untuk menghasilkan proses komputasi yang lebih cepat dan melakukan estimasi jumlah klaster yang paling efisien dalam pemrosesan data. Penelitian tersebut menghasilkan metode X-Means (*Extended K-Means*). X-Means diklaim mampu digunakan pada data set kecil dan proses komputasi yang lebih cepat daripada K-Means [30]. Algoritma X-Means terdiri dari 2 operasi yang terus diulang hingga eksekusi berakhir.

### d) Davies Bouldin Index (DBI)

*Davies bouldin index (DBI)* merupakan ukuran untuk mengevaluasi kinerja sebuah metode clustering yang diperkenalkan oleh David L. Davies dan Donald W. Bouldin. Davies Bouldin Index digunakan untuk mengevaluasi cluster secara umum berdasarkan kuantitas dan kedekatan antar anggota cluster. Semakin kecil nilai Davies Bouldin Index, maka semakin baik cluster yang dihasilkan [31]. Metode DBI dipilih karena dapat diimplementasikan dalam semua data baik ukuran besar dan kecil sehingga sangat cocok dan dapat diterima untuk perhitungan akurasi clustering [32].

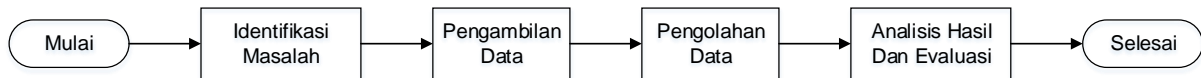
## 2. METODE

### A. Dataset Penelitian

Dalam penelitian ini data yang dipergunakan meliputi data transaksi UD. XYZ dalam kurun waktu satu tahun dari Januari 2009 hingga Desember 2009 dan data pelanggan. Data yang digunakan merupakan arsip data transaksi UD. XYZ sebanyak 153.492 data.

### B. Alur Penelitian

Alur penelitian dibuat untuk membantu penulis dalam menjalankan penelitiannya. Pada Gambar 1 merupakan skema alur penelitian.



Gambar 1. Alur Penelitian

#### a) Identifikasi Masalah

Studi kasus dalam penelitian ini adalah usaha retail perlengkapan dan bahan kue yang ada di Kota Denpasar yaitu UD. XYZ. UD. XYZ menjual produknya secara offline melalui toko dan online melalui whatsapp, media sosial, dan website yang disediakan untuk memberikan pelayanan terbaik kepada customer. Dengan bertambahnya customer tiap tahunnya, UD. XYZ mengalami kesusahan dalam mengelompokkan customer yang berpotensi untuk memberikan keuntungan.

#### b) Pengambilan Data

Proses pengambilan data dilakukan secara langsung ke UD. XYZ. Adapun data yang dipergunakan dalam penelitian ini meliputi data transaksi UD. XYZ dalam kurun waktu satu tahun dari Januari 2009 hingga Desember 2009, data pelanggan, dan informasi proses bisnis UD. XYZ. Dalam hal penggunaan data, digunakan arsip data transaksi UD. XYZ sebagai bahan pengolahan data sebanyak 153.492 data. Pada data transaksi ini terdiri dari 10(sepuluh) atribut yang ditunjukkan pada Tabel 2. Sedangkan Tabel 3 merupakan cuplikan data transaksi UD. XYZ.

Tabel 2. Atribut Data Transaksi

Atribut	Type Data	Keterangan
Tanggal, bulan, tahun, jam, menit	Date Time	Tanggal dilakukan transaksi
ID Kasir	Text	Petugas yang melakukan transaksi
ID Toko	Text	Tempat dilakukan transaksi
No Order	Text	Nomor transaksi Penjualan
Kode Barang	Text	Kode Barang yang ditransaksikan
Jumlah Barang	Number	Jumlah Barang per kode barang
Total Biaya	Number	Total Harga per kode barang
Kode customer	Text	Kode untuk masing-masing pelanggan
Discount	Number	Potongan Harga
Status	Text	Status pembayaran

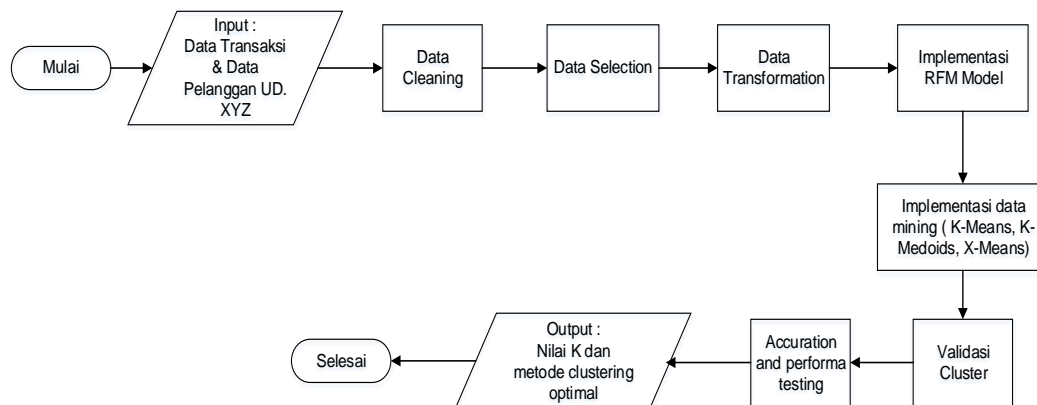
Tabel 3. Data Transaksi Toko Retail

Tgl	Toko	KD Kasir	Kode Barang	Jumlah Barang	Total Biaya	No Order	Kode Customer	Status	Disc
02/01/...	Toko Gatsu	BAMBANG	10137	25	1.618.750	52887	CO312	0	6.250
02/01/...	Gudang Gatsu	BAMBANG	50053	15	1.937.250	52888	CO311	0	20.250
02/01/...	Gudang Gatsu	BAMBANG	01954	5	617.500	52889	CO321	0	0

#### c) Pengolahan Data

Proses pengolahan data dimulai dengan *data cleaning*, *data selection*, *data transformation*, implementasi RFM Model, Implementasi clustering, validasi cluster, dan perhitungan akurasi. Berikut merupakan detail masing-masing proses pengolahan data pada penelitian ini.

1. *Input*, merupakan proses penginputan data yang digunakan dalam proses pengolahan data yaitu data transaksi dan data pelanggan UD. XYZ dalam kurun waktu satu tahun. Data masih dalam bentuk file excel dengan format .xls.
2. *Data cleaning*, merupakan proses dalam menghilangkan data yang redundan, atau tidak memiliki nilai. Data ganda dan data yang memiliki atribut kosong dihilangkan dengan tujuan mempermudah pada proses berikutnya.
3. *Data selection* (pemilihan data), pada proses ini dilakukan pemilihan atribut yang digunakan untuk mengolah data.
4. *Data Transformasi*, proses pengolahan data ini dilakukan untuk mempermudah proses clustering. Semua data ditransformasikan ke dalam bobot yang seragam dan seimbang.
5. Implementasi model RFM, pada tahap ini semua data ditransformasikan dengan dibobotkan menggunakan metode RFM. Data yang bersifat variatif diubah menjadi satu tipe data yaitu numerik. Proses implementasi model RFM dilakukan dengan bantuan Microsoft Excel.
6. Implementasi clustering data, pada proses ini diimplementasikan metode clustering dengan menggunakan 3(tiga) metode yaitu K-Medoids, X-Means, dan K-Means. Proses implementasinya dibantu dengan menggunakan *software RapidMiner*.
7. *Validasi cluster*, dilakukan dengan melihat dan mengevaluasi persebaran cluster pada masing-masing metode.
8. *Accuration and performance testing*, perhitungan akurasi dilakukan dengan menghitung jarak rata-rata antar cluster dan menghitung nilai *David Bouldin Index (DBI)*.
9. *Output*, dari penelitian yang dilakukan diharapkan menghasilkan nilai *k* dan metode clustering mana yang paling optimal untuk diimplementasikan ke data transaksi UD. XYZ.  
Pada Gambar 2 merupakan alur pengolahan data dalam penelitian ini.



Gambar 2. Alur Pengolahan Data

### C. Alur Pengujian

Alur pengujian dalam penelitian ini dibagi menjadi 3(tiga) dengan detail sebagai berikut.

1. Alur pertama, data transaksi dilakukan implementasi terhadap metode clustering K-Medoids, X-Means, dan K-Means dengan menggunakan nilai *k* atau cluster awal. Nilai cluster awal dilakukan percobaan dari nilai 0 hingga 10.
2. Alur kedua, setiap proses implementasi metode dilakukan perhitungan rata-rata jarak antar cluster dan perhitungan nilai *Davies bouldin index (DBI)*.
3. Alur Ketiga, setelah dildapat nilai akurasi masing-masing cluster yang terbentuk dilakukan analisis terhadap nilai akurasi yang dihasilkan untuk menentukan nilai *k* atau cluster awal optimal untuk data transaksi UD. XYZ.

### 3. HASIL DAN PEMBAHASAN

#### A. Preprocessing Data

Tabel 4. Data Cleaning

Tgl	Toko	KD Kasir	Kode Barang	Jumlah Barang	Total Biaya	No Order	Kode Customer	Status	Disc
02/01/...	Toko 17 Ritel	BAMBANG	-	1	655.000	52852	C0745	0	0
02/01/...	Gudang Gatsu	BAMBANG	10134	50	3.700.000	52821	C0784	0	0
02/01/...	-	BAMBANG	10137	3	195.000	52811	C1221	0	0
02/01/...	Toko 17 Ritel	BAMBANG	20003	3	832.500	52812	C1111	0	0

Proses pembersihan data dilakukan untuk menghilangkan baris yang bernilai kosong dan data yang tidak valid. Data transaksi perusahaan retail dilakukan cleaning untuk menghilangkan baris yang memiliki nilai null. Dalam proses ini data yang sebelumnya 153.492 data menjadi 93.685 data. Data pada Tabel 4 yang kosong dihapus untuk mempermudah proses *data selection*. Pemilihan data (*data selection*) mengacu pada atribut model RFM yaitu jumlah frekuensi transaksi, jumlah nominal transaksi untuk setiap pelanggan, dan jarak waktu belanja customer terakhir. Proses ini mereduksi dan menghapus kolom data yang tidak terpakai dan tidak sesuai dengan kebutuhan untuk clustering data. Pada data transaksi UD. XYZ atribut yang dihilangkan adalah *Toko*, *KD Kasir*, *Kode Barang*, *Jumlah Barang*, *Status*, dan *Disc*. Sample proses data cleaning seperti pada Tabel 4.

Dalam tahapan pemilihan data dilakukan pengelompokan data transaksi berdasarkan *kode customer* untuk mengambil *frequency* belanja customer, jumlah nominal transaksi, dan jarak belanja terakhir. Pada proses pemilihan ini data transaksi UD. XYZ tersisa hanya 10.145 data transaksi. Pada Tabel 5 merupakan hasil dari pembersihan data hingga pemilihan data.

Tabel 5. Hasil Data Selection

Tgl	Total Biaya	No Order	Kode Customer
02/01/2009	1.618.750	52887	CO312
02/01/2009	1.937.250	52888	CO311
02/01/2009	617.500	52889	CO321

#### B. Hasil Implementasi RFM

Setelah melalui *preprocessing data*, dilanjutkan dengan mengimplementasikan model RFM dengan menerapkan atributnya pada data transaksi. Penentuan bobot model RFM dilakukan dengan menggali informasi kepada pihak UD. XYZ melalui wawancara. Penentuan bobot model RFM menggunakan metode pembobotan konvensional dimana atribut yang memiliki prioritas paling tinggi akan diberikan nilai bobot paling tinggi, sedangkan atribut yang memiliki prioritas paling rendah diberikan nilai bobot paling rendah.

Hasil implementasi RFM selanjutnya digunakan sebagai bahan untuk implementasi metode clustering untuk melakukan segmentasi pelanggan. Proses implementasi metode clustering menggunakan visualisasi 2 dimensi sehingga atribut yang digunakan adalah nilai *Recency (R)* dan rata-rata nilai *Frequency (F)* ditambah *Monetary (M)*. Sehingga cuplikan data yang digunakan untuk proses clustering ditampilkan pada Tabel 6.

Tabel 6. Bahan Clustering Data

Kode Customer	Nilai Recency (R)	Nilai Frequency (F) + Nilai Monetary (M)
CO312	4	2.5
CO311	1	1.5
CO321	2	1

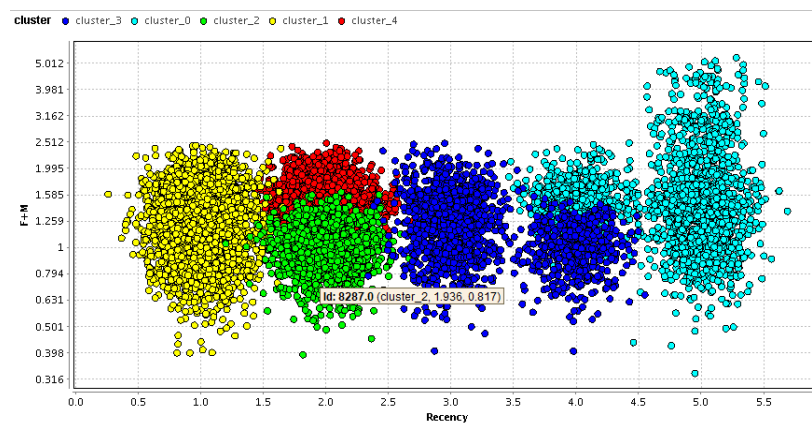
#### C. Hasil Implementasi Clustering

Proses clustering dilakukan dengan menggunakan metode K-Medoids, X-Means, dan K-Means dimana dilakukan proses iterasi nilai  $k$  dari 1 hingga 10. Sebagai sample proses clustering, nilai  $k=5$  membentuk segmentasi pelanggan menjadi 5 tingkatan seperti pada Tabel 7.

Tabel 7 . Hasil Segmentasi Pelanggan Dengan K-Means

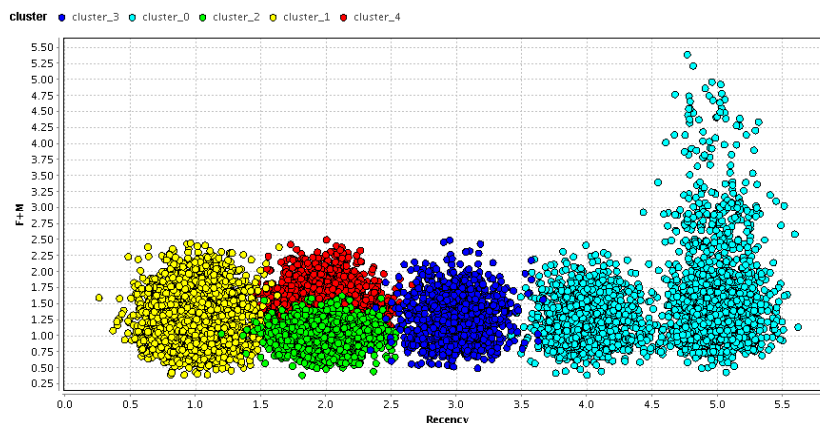
Cluster Ke-	Jumlah Anggota	Tingkatan Customer
1	1977 (19.49%)	About To Sleep
2	3319 (32.71%)	Customer Needing Attention
3	1505 (14.83%)	Recent Customer
4	1575(15.53%)	Potential Loyalist
5	1769(17.44%)	Loyal Customers

Pada Gambar 3 merupakan hasil clustering dengan K-Means menggunakan parameter  $k=5$ . Visualisasi persebaran data hasil K-Means ditampilkan pada Gambar 3.



Gambar 3. Hasil Implementasi K-Means

Pada Gambar 4 merupakan hasil clustering dengan K-Medoids menggunakan parameter  $k=5$ . Visualisasi persebaran data hasil K-Medoids ditampilkan pada Gambar 4 berikut ini.



Gambar 4. Hasil Implementasi K-Medoids

Dalam clustering K-Medoids dengan parameter  $k=5$  ini membentuk segmentasi pelanggan menjadi 5 tingkatan seperti pada Tabel 8.

Tabel 8 . Hasil Segmentasi Pelanggan Dengan K-Medoids

Cluster Ke-	Jumlah Anggota	Tingkatan Customer
1	2495 (24.60%)	About To Sleep
2	3319 (32.71%)	Customer Needing Attention
3	1505 (14.83%)	Recent Customer



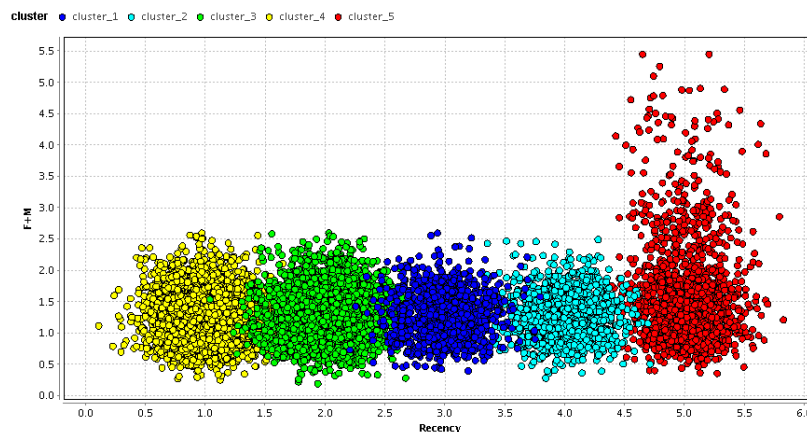
4	1057 (10.42%)	Potential Loyalist
5	1769 (17.44%)	Loyal Customers

Dalam clustering X-Means dengan parameter k=5 ini membentuk segmentasi pelanggan menjadi 5 tingkatan seperti pada Tabel 9.

Tabel 9 . Hasil Segmentasi Pelanggan Dengan X-Means

Cluster Ke-	Jumlah Anggota	Tingkatan Customer
1	1977 (19.49%)	About To Sleep
2	3319 (32.71%)	Customer Needing Attention
3	1505 (14.83%)	Recent Customer
4	1575 (15.53%)	Potential Loyalist
5	1769 (17.44%)	Loyal Customers

Pada Gambar 4 merupakan hasil clustering dengan X-Means menggunakan parameter k=5. Visualisasi persebaran data hasil X-Means ditampilkan pada Gambar 5 berikut ini.



Gambar 5. Hasil Implementasi X-Means

Dengan k=5 maka terbentuk 5(lima) tingkatan loyalitas pelanggan dengan detail keterangan seperti pada Tabel 10 berikut ini.

Tabel 10. Tingkat Loyalitas Pelanggan

Cluster	Label	Keterangan
0	About To Sleep	Pelanggan pernah melakukan transaksi namun lama tidak melakukan transaksi kembali. Butuh untuk diajak kembali agar tidak kehilangan pelanggan
1	Customer Needing Attention	Pelanggan di atas rata-rata umum, namun belum melakukan transaksi baru-baru ini
2	Recent Customer	Pelanggan yang melakukan transaksi baru-baru ini namun tidak begitu sering
3	Potential Loyalist	Pelanggan melakukan pembelian baru-baru ini, sering melakukan transaksi, dan nominal belanja yang cukup
4	Loyal Customers	Pelanggan yang belanja secara rutin dan responsive terhadap promosi yang diberikan

#### D. Analisis Akurasi Clustering

Tabel 11. Hasil Perhitungan Nilai Akurasi Clustering

Nilai k	K-Means		K-Medoids		X-Means	
	Rata-rata Jarak Objek ke Centroid	DBI	Rata-rata Jarak Objek ke Centroid	DBI	Rata-rata Jarak Objek ke Centroid	DBI
1	2,000	?	0,56	?	0,133	?

2	1,478	0,507	0,05	0,514	0,038	0,49
3	0,759	0,765	0,031	0,593	0,03	0,708
4	0,644	0,735	0,025	0,72	0,015	0,606
5	0,406	0,582	0,018	0,697	0,014	0,629
6	0,317	0,701	0,019	0,362	0,009	0,536
7	0,232	0,632	0,011	0,481	0,006	0,534
8	0,221	0,605	0,014	0,535	0,007	0,429
9	0,168	0,554	0,007	0,425	0,005	0,512
10	0,091	0,478	0,01	0,54	0,004	0,443

Proses perhitungan nilai akurasi clustering dilihat berdasarkan rata-rata jarak objek ke *centroid* dan nilai DBI. Setelah dilakukan perhitungan dan visualisasi segmentasi pelanggan dengan menggunakan clustering data metode K-Medoids, X-Means, dan K-Means didapatkan hasil seperti pada Tabel 11. Dilihat dari hasil pada Tabel 15, dapat disimpulkan bahwa berdasarkan nilai rata-rata perolehan nilai DBI K-Medoids dapat diperhitungkan untuk menjadi metode yang dapat digunakan untuk segmentasi pelanggan UD. XYZ dengan rata-rata nilai DBI 0,540778. Jika dilihat dari Tabel 7, Tabel 8, dan Tabel 9, cluster yang terbentuk pada cluster 2, 3, dan 5 menghasilkan jumlah anggota yang sama pada ketiga metode yang diimplementasikan. Hal ini dapat dijadikan sebagai bahan pertimbangan penggunaan metode dengan nilai  $k=5$ , karena mampu memberikan persebaran nilai yang merata dan sama.

#### 4. SIMPULAN DAN SARAN

Penelitian ini bertujuan untuk melakukan segmentasi pelanggan dan mencari metode dan nilai  $k$  optimal untuk mengukur loyalitas pelanggan terhadap UD XYZ yang merupakan toko retail bahan dan perlengkapan kue dengan mengkombinasikan model RFM dan teknik *clustering*. Atribut *Recency*, *Frequency*, dan *Monetary* dipilih sebagai atribut untuk clustering data. Proses validasi cluster dilakukan dengan menghitung nilai *Davies Bouldin Index (DBI)* untuk mencari jumlah cluster yang paling optimal. Dimana perbandingan hasil DBI menunjukkan bahwa K-Medoids memiliki nilai DBI yang rendah sehingga mampu menghasilkan cluster yang optimal jika dibandingkan dengan hasil perhitungan DBI metode K-Means dan X-Means. Untuk nilai  $k$  dapat disimpulkan bahwa nilai  $k$  bisa dipilih dengan nilai 5 ke atas. Namun berdasarkan analisa penulis nilai  $k=5$  sudah cukup mewakili persebaran data pada segmentasi pelanggan UD. XYZ. Sehingga dapat diambil kesimpulan bahwa penggunaan model RFM dan metode *clustering* khususnya algoritma K-Medoids, dapat membantu UD. XYZ dalam melakukan segmentasi pelanggan potensial dan dapat digunakan oleh bagian pemasaran untuk menyusun strategi pemasaran yang tepat untuk kelima kelompok yang terbentuk.

#### DAFTAR PUSTAKA

- [1] S. S. Utami, "Peranan Sistem Informasi Manajemen untuk Pengambilan Keputusan Pengusaha Kecil," *J. Ekon. dan Kewirausahaan*, vol. 11, no. 2, pp. 142–150, 2011.
- [2] N. W. Suryathi, D. P. Darmawan, and W. Suartana, "Kinerja Keuangan sebagai Dasar Pengambilan Keputusan Investasi di Dhyana Pura Beach Resort Seminyak Kuta Badung Pendahuluan," *J. Manaj. Agribisnis*, vol. 1, no. 2, pp. 24–42, 2013.
- [3] V. S. Moertini, "Data Mining Sebagai Solusi Bisnis," *Integral*, vol. 7, no. 1, pp. 44–56, 2002.
- [4] M. S. Chen, J. Han, and P. S. Yu, "Data mining: An overview from a database perspective," *IEEE Trans. Knowl. Data Eng.*, vol. 8, no. 6, pp. 866–883, 1996, doi: 10.1109/69.553155.
- [5] S. Mujasah, "Pemanfaatan Data Mining Untuk Prakiraan Cuaca," *J. Meteorol. dan Geofis.*, vol. 12, no. 2, pp. 189–195, 2011, doi: 10.31172/jmg.v12i2.100.
- [6] A. F. Watratan, A. Puspita. B, and D. Moeis, "Implementasi Algoritma Naive Bayes Untuk Memprediksi Tingkat Penyebaran Covid-19 Di Indonesia," *J. Appl. Comput. Sci. Technol.*, vol. 1, no. 1, pp. 7–14, 2020, doi: 10.52158/jacost.v1i1.9.
- [7] H. Noor, "Optimasi Model Klasifikasi C4.5 Dan Particle Swarm Optimization Untuk Prediksi Siswa Bermasalah," *J. Ilm. Technol.*, vol. 9, no. 4, p. 228, 2018, doi: 10.31602/tji.v9i4.1537.
- [8] H. Harafani and H. A. Al-kautsar, "MENINGKATKAN KINERJA K-NN UNTUK KLASIFIKASI KANKER PAYUDARA DENGAN SELEKSI FITUR," *J. Pendidik. Teknol. dan Kejur.*, vol. 18, no. 1, pp. 99–110, 2021.
- [9] B. E. Adiana, I. Soesanti, and A. E. Permasari, "Analisis Segmentasi Pelanggan Menggunakan Kombinasi Rfm Model Dan Teknik Clustering," *J. Terap. Teknol. Inf.*, vol. 2, no.

- 1, pp. 23–32, 2018, doi: 10.21460/jutei.2018.21.76.
- [10] N. P. E. Merliana, Ernawati, and A. J. Santoso, “Analisa Penentuan Jumlah Cluster Terbaik pada Metode K-Means,” in *Prosiding Seminar Nasional Multi Disiplin Ilmu & Call For Papers UNISBANK (SENDI\_U)*, pp. 978–979.
- [11] J. Wei, S. Lin, and H. Wu, “A review of the application of RFM model,” *African J. Bus. Manag.*, vol. 4, no. 19, pp. 4199–4206, 2010.
- [12] A. K. Jain, M. N. Murty, and P. J. Flynn, “Data Clustering : A Review,” *ACM Comput. Surv.*, vol. 31, no. 3, pp. 264–323, 2000, doi: <https://doi.org/10.1145/331499.331504>.
- [13] R. A. Indraputra and R. Fitriana, “K-Means Clustering Data COVID-19,” *J. Tek. Ind.*, vol. 10, no. 3, pp. 275–282, 2020.
- [14] A. Yudi, “K-Means – Penerapan , Permasalahan dan Metode Terkait,” *J. Sist. dan Inform.*, vol. 3, no. Pebruari 2007, pp. 47–60, 2007.
- [15] H. Park and C. Jun, “A simple and fast algorithm for K-medoids clustering,” *Expert Syst. Appl.*, vol. 36, no. 2, pp. 3336–3341, 2009, doi: 10.1016/j.eswa.2008.01.039.
- [16] T. Ishioka, “AN EXPANSION OF X-MEANS FOR AUTOMATICALLY DETERMINING THE OPTIMAL NUMBER OF CLUSTERS — PROGRESSIVE ITERATIONS OF Å -MEANS AND MERGING OF THE CLUSTERS —,” in *Proceedings of the Fourth IASTED International Conference*, 2005, vol. July 4-6, pp. 91–96.
- [17] S. S. P. Luka, I. M. Candiasa, and K. Y. E. Aryanto, “ANALISIS PEMBENTUKAN KELOMPOK DISKUSI PANEL SISWA MENGGUNAKAN ALGORITMA FUZZY C-MEANS DAN K-MEANS,” *J. Pendidik. Teknol. dan Kejur.*, vol. 16, no. 2, pp. 267–277, 2019.
- [18] D. Agustina, D. Pujotomo, and D. Puspitasari, “Pengembangan Strategi Hubungan Pelanggan Berdasarkan Segmentasi Pelanggan Menggunakan Data Mining,” *Ind. Eng. Online Journal*, vol. 4, no. 2, pp. 1–8, 2016.
- [19] Y. Wicaksono, “Segmentasi Pelanggan Bisnis Dengan Multi Kriteria Menggunakan K-Means,” *Indones. J. od Bus. Intell.*, vol. 1, no. 2, pp. 45–53, 2018, doi: <http://dx.doi.org/10.21927/ijubi.v1i2.872>.
- [20] N. H. Harani, C. Prianto, and F. A. Nugraha, “Segmentasi Pelanggan Produk Digital Service Indihome Menggunakan Algoritma K-Means Berbasis Python,” *Manaj. Inform.*, vol. 10, no. 2, pp. 133–146, 2020, doi: 10.34010/jamika.v10i2.
- [21] S. Monalisa and I. Erza, “Analisis Loyalitas Agen Biasa dan Agenstok Menggunakan Model RFM ( Recency , Frequency , Monetary ) dan Algoritma K-Medoids pada BC 4 HPAI Pekanbaru,” *Techno.com*, vol. 20, no. 1, pp. 109–121, 2021, doi: <https://doi.org/10.33633/tc.v20i1.4219>.
- [22] P. E. Prakasawati, Y. H. Chrisnanto, and A. I. Hadiana, “SEGMENTASI PELANGGAN BERDASARKAN PRODUK MENGGUNAKAN METODE K- MEDOIDS,” in *KOMIK (Konferensi Nasional Teknologi Informasi dan Komputer)*, 2019, vol. 3, pp. 335–339, doi: 10.30865/komik.v3i1.1610.
- [23] A. Wijayanto and R. Adhitama, “Penggunaan X-Means Clustering Method untuk Mengelompokkan Potensi Sekolah Menengah Unggul di Kabupaten Banyumas,” *J. Informatics, Inf. Syst. Softw. Eng. Appl.*, vol. 2, no. 1, pp. 80–88, 2019.
- [24] P. H. Putra and M. S. Novelan, “Analysis Of The Use Of X-Means Method In Grouping Interest And Talent Data Students,” *J. Ipteks Terap.*, vol. 14, no. 2, pp. 152–159, 2020, doi: <http://doi.org/10.22216/jit.2020.v14i2.5414>.
- [25] M. Aryuni, E. D. Madyatmadja, and E. Miranda, “Customer Segmentation in XYZ Bank using K-Means and K-Medoids Clustering,” in *2018 International Conference on Information Management and Technology (ICIMTech)*, 2018, no. September, pp. 412–416, doi: 10.1109/ICIMTech.2018.8528086.
- [26] S. Monica and F. Natalia, “Clustering Tourism Object in Bali Province Using K- Means and X-Means Clustering Algorithm,” in *2018 IEEE 20th International Conference on High Performance Computing and Communications; IEEE 16th International Conference on Smart City; IEEE 4th International Conference on Data Science and Systems (HPCC/SmartCity/DSS)*, 2018, pp. 1462–1467, doi: 10.1109/HPCC/SmartCity/DSS.2018.00241.
- [27] R. Adhitama, A. Burhanuddin, and R. Ananda, “PENENTUAN JUMLAH CLUSTER IDEAL SMK DI JAWA TENGAH DENGAN METODE X-MEANS CLUSTERING DAN K-MEANS CLUSTERING DETERMINING VOCATIONAL IDEAL CLUSTER NUMBER IN CENTRAL JAVA WITH X-MEANS CLUSTERING AND K-MEANS CLUSTERING METHODS,” *JIKO (Jurnal Inform. dan Komput.*, vol. 3, no. 1, pp. 1–5, 2020, doi: 10.33387/jiko.

- [28] Sudriyanto, "CLUSTERING LOYALITAS PELANGGAN DENGAN METODE RFM (RECENCY, FREQUENCY, MONETARY ) DAN FUZZY C-MEANS," *Pros. SNATIF Ke-4*, pp. 815–822, 2017.
- [29] S. Shukla and S. Naganna, "A Review ON K-means DATA Clustering APPROACH," *Int. J. Inf. Comput. Technol.*, vol. 4, no. 17, pp. 1847–1860, 2014.
- [30] D. Pelleg and A. Moore, "X-means: Extending K-means with Efficient Estimation of the Number of Clusters," *Icml*, vol. 1, no. June, pp. 727–734, 2000.
- [31] A. Bates and J. Kalita, "Counting Clusters in Twitter Posts," in *ICTCS '16: Proceedings of the Second International Conference on Information and Communication Technology for Competitive Strategies*, 2016, pp. 1–9, doi: <https://doi.org/10.1145/2905055.2905295>.
- [32] J. Xiao, J. Lu, and X. Li, "Davies Bouldin Index based hierarchical initialization K-means," *Intell. Data Anal.*, vol. 21, no. 6, pp. 1327–1338, 2017, doi: 10.3233/IDA-163129.