

**PENENTUAN LOYALITAS PELANGGAN DI BATUBARA MART
MENGUNAKAN *K-MEANS CLUSTERING***

SKRIPSI

**Oleh:
REYNALDY LANGGENG PHAMBUDY
NIM. 15650075**



**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2022**

**PENENTUAN LOYALITAS PELANGGAN DI BATUBARA MART
MENGUNAKAN *K-MEANS CLUSTERING***

SKRIPSI

**Oleh:
REYNALDY LANGGENG PHAMBUDY
NIM. 15650075**

Diajukan Kepada:

**Fakultas Sains dan Teknologi
Universitas Islam Negeri (UIN) Maulana Malik Ibrahim Malang
Untuk Memenuhi Salah Satu Persyaratan dalam
Memperoleh Gelar Sarjana Komputer (S.KOM)**

**PROGRAM STUDI INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2022**

HALAMAN PERSETUJUAN

PENENTUAN LOYALITAS PELANGGAN DI BATUBARA MART MENGUNAKAN *K-MEANS CLUSTERING*

SKRIPSI

Oleh :
REYNALDY LANGGENG PHAMBUDY
NIM. 15650075

Telah Diperiksa dan Disetujui untuk Diuji
Tanggal : Desember 2022

Pembimbing I,


Fajar Rohman Hariri, M.Kom
NIP. 19890515 201801 1 001

Pembimbing II,


Dr. Muhammad Faisal, M.T
NIP. 19740510 200501 1 007

Mengetahui,
Ketua Program Studi Teknik Informatika
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang




Dr. Fachrul Kurniawan ST., M.MT., IPM
NIP. 19771020 200912 1 001

HALAMAN PENGESAHAN

PENENTUAN LOYALITAS PELANGGAN DI BATUBARA MART MENGUNAKAN *K-MEANS CLUSTERING*

SKRIPSI

Oleh :

Reynaldy Langgeng Phambudy
NIM. 15650075

Telah Dipertahankan di Depan Dewan Penguji
dan Dinyatakan Diterima Sebagai Salah Satu Persyaratan
untuk Memperoleh Gelar Sarjana Komputer (S.Kom)
Pada Tanggal Desember 2022

Susunan Dewan Penguji

- | | | |
|------------------------|--|---------|
| 1. Ketua Penguji | : <u>Prof. Dr. Suhartono, M.Kom</u>
NIP. 19680519 200312 1 001 | (.....) |
| 2. Anggota Penguji I | : <u>Ajib Hanani, M.T</u>
NIDT. 19840731 201608011 076 | (.....) |
| 3. Anggota Penguji II | : <u>Fajar Rohman Hariri, M. Kom</u>
NIP. 19890515 201801 1 001 | (.....) |
| 4. Anggota Penguji III | : <u>Dr. Muhammad Faisal, M.T</u>
NIP. 19740510 200501 1 007 | (.....) |

Mengetahui,
Ketua Program Studi Teknik Informatika
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang



Dr. Fachrul Kurniawan ST., M.MT., IPM
NIP. 19771020 200912 1 001

HALAMAN PERSEMBAHAN

اَلْحَمْدُ لِلّٰهِ رَبِّ الْعَالَمِيْنَ

Puji syukur kehadiran Allah SWT, shalawat serta salam kepada Rasulullah SAW. Penulis persembahkan skripsi ini kepada:

Kedua orang tua penulis tercinta, Bapak Purwo Santoso, dan Ibu Hamimah. Keluarga di rumah yang selalu mendukung penulis, memberikan do'a, dorongan, serta motivasi yang tiada henti. Dan terimakasih juga kepada saudara-saudara saya yang senantiasa mengingatkan saya ketika melakukan suatu kesalahan, membantu saya dalam segala hal, dan mendukung saya dalam menyelesaikan tugas akhir ini.

Dosen pembimbing penulis, Bapak Fajar Rohman Hariri, M.Kom dan Bapak Dr. Muhammad Faisal, M.T yang telah dengan sabar membimbing penelitian skripsi ini dan selalu memberikan masukan, dorongan, serta semangat untuk menjalani setiap tahapan skripsi.

Seluruh Dosen Teknik Informatika Universitas Islam Negeri (UIN) Maulana Malik Ibrahim Malang telah membimbing dan memberikan ilmunya yang bermanfaat. Seluruh Keluarga Teknik Informatika, terutama teman-teman Teknik Informatika angkatan 2015 yang telah memberikan semangat dan juga motivasi. Serta teman-teman penulis yang terus mendorong penulis untuk melewati tahapan skripsi.

HALAMAN PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan dibawah ini:

Nama : Reynaldy Langgeng Phambudy
NIM : 15650075
Fakultas/Jurusan : Sains dan Teknologi / Teknik Informatika
Judul Skripsi : Penentuan Loyalitas Pelanggan Di Batubara Mart
Menggunakan *K-Means Clustering*.

Menyatakan dengan sebenarnya bahwa Skripsi yang saya tulis ini benar-benar merupakan hasil karya sendiri, bukan merupakan pengambilalihan data, tulisan atau pikiran orang lain yang saya akui sebagai hasil tulisan atau pikiran saya sendiri, kecuali dengan mencantumkan sumber cuplikan pada daftar pustaka.
Apabila dikemudian hari terbukti atau dapat dibuktikan Skripsi ini hasil jiplakan, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Malang, Desember 2022

Yang membuat Pernyataan,



Reynaldy Langgeng Phambudy
NIM. 15650075

KATA PENGANTAR

Puji syukur bagi Allah SWT yang telah mencurahkan nikmat serta kekuatan sehingga penulis dapat menyelesaikan skripsi dengan baik. Tidak lupa juga shalawat serta salam kepada nabi Muhammad SAW yang telah membimbing umatnya ke jalan yang benar.

Dalam mengerjakan skripsi ini, tidak lepas dari bantuan, bimbingan, serta dukungan berbagai pihak baik secara moril maupun materil. Oleh karenanya, izinkan penulis mengucapkan terima kasih yang sebesar-besarnya kepada:

1. Prof. Dr. H. M. Zainuddin, M.A, selaku Rektor Universitas Islam Negeri Maulana Malik Ibrahim Malang.
2. Dr. Sri Harini, M.Si, selaku Dekan Fakultas Sains dan Teknologi.
3. Dr. Fachrul Kurniawan ST., M.MT., IPM, selaku Ketua Jurusan Teknik Informatika.
4. Fajar Rohman Hariri, M.Kom selaku Dosen Pembimbing I yang telah meluangkan waktu untuk membimbing serta memberi masukan kepada penulis dalam mengerjakan skripsi ini.
5. Dr. Muhammad Faisal, M.T selaku Dosen Pembimbing II yang juga senantiasa memberikan masukan dan petunjuk tentang penulisan skripsi ini.
6. Seluruh Dosen Teknik Informatika yang telah mencurahkan ilmunya kepada penulis selama masa kuliah.
7. Bapak Purwo Santoso, dan Ibu Hamimah, serta keluarga besar penulis yang selalu meberikan doa, nasihat, dan dukungan kepada penulis selama masa kuliah.
8. Teman-teman Teknik Informatika 2015, yang telah membantu penulis dalam menyelesaikan skripsi.
9. Sahabat dan teman – teman yang telah memberikan semangat, dorongan, dan dukungan dalam menyelesaikan skripsi, serta diri sendiri yang sudah berusaha dan tidak menyerah untuk menyelesaikan skripsi.

Dalam Penulisan skripsi ini, masih terdapat kekurangan. Namun penulis berharap, skripsi ini dapat memberikan manfaat kepada pembaca dan kepada penulis. Aamiin.

Malang, Desember 2022

Penulis

DAFTAR ISI

COVER	
HALAMAN JUDUL	ii
HALAMAN PERSETUJUAN	iii
HALAMAN PENGESAHAN	iv
HALAMAN PERSEMBAHAN	v
HALAMAN PERNYATAAN KEASLIAN TULISAN	vi
KATA PENGANTAR	viii
DAFTAR ISI	viii
DAFTAR TABEL	x
DAFTAR GAMBAR	xi
ABSTRAK	xiii
ABSTRACT	xiv
المخلص	xv
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Pernyataan Masalah	4
1.3 Tujuan Penelitian	4
1.4 Manfaat Penelitian	4
1.5 Batasan Masalah	5
1.6 Sistematika Penulisan	5
BAB II KAJIAN PUSTAKA	7
2.1 Loyalitas Pelanggan	8
2.2 <i>Clustering</i>	8
2.3 <i>Preprocessing</i>	10
2.4 <i>K-Means Clustering</i>	10
2.5 Metode Elbow	14
2.6 Evaluasi <i>Clustering Davies Bouldin Index (DBI)</i>	15
BAB III METODE PENELITIAN	17
3.1 Objek Penelitian	17
3.2 Alat Penelitian	17
3.2.1 <i>Hardware</i>	17

3.2.2	<i>Software</i>	17
3.3	Prosedur Penelitian	18
3.3.1	Desain Sistem	18
BAB IV HASIL DAN PEMBAHASAN		36
4.1	Hasil Penelitian	36
4.1.1	<i>Import Library</i>	36
4.1.2	<i>Preprocessing</i>	38
4.1.3	<i>Model Building</i>	47
4.1.4	<i>Evaluation</i>	56
4.2	Pembahasan	59
4.3	Integrasi Islam	60
BAB V PENUTUP		63
5.1	Kesimpulan	63
5.2	Saran	64
DAFTAR PUSTAKA		

DAFTAR TABEL

Tabel 3. 1 Dataset.....	20
Tabel 3. 2 Dataset Hasil Data <i>Selection</i>	21
Tabel 3. 3 Hasil Data <i>Preprocessing</i>	21
Tabel 3. 4 Data Transformation RFM.....	22
Tabel 3. 5 Dataset Hasil Normalisasi Transformasi RFM	25
Tabel 3. 6 Iterasi 1.....	25
Tabel 3. 7 Iterasi 2.....	25
Tabel 3. 8 Iterasi 3.....	25
Tabel 3. 9 Alokasi Kluster dengan Euclidean distance (Iterasi 1)	27
Tabel 3. 10 Alokasi Kluster dengan Euclidean distance (Iterasi 2)	28
Tabel 3. 11 Alokasi Kluster dengan Euclidean distance (Iterasi 3)	28
Tabel 3. 12 Hasil Clustering dari data RFM	29

DAFTAR GAMBAR

Gambar 2. 1 Flowchart K-Means.....	14
Gambar 3. 1 <i>Flowchart</i> Sistem	18
Gambar 3. 2 Proses RFM	19
Gambar 3. 3 Diagram Alir K-Means.....	24
Gambar 3. 4 <i>Flowchart</i> Uji DBI	31
Gambar 4. 1 <i>Source Code Import Library</i>	36
Gambar 4. 2 <i>Source Code Load Dataset</i>	37
Gambar 4. 3 <i>Load Dataset</i>	37
Gambar 4. 4 <i>Source Code Data Cleansing</i>	38
Gambar 4. 5 Hasil <i>Data Cleansing</i>	39
Gambar 4. 6 <i>Source Code Recency</i>	40
Gambar 4. 7 Hasil <i>Recency</i>	40
Gambar 4. 8 <i>Source Code Menghitung Last Transaction Date</i>	41
Gambar 4. 9 Hasil <i>Menghitung Last Transaction Date</i>	41
Gambar 4. 10 <i>Source Code Extract Jumlah Hari</i>	41
Gambar 4. 11 Hasil <i>Extract Jumlah Hari</i>	42
Gambar 4. 12 <i>Source Code Frequency</i>	42
Gambar 4. 13 Hasil <i>Frequency</i>	43
Gambar 4. 14 <i>Source Code Merging Dataframe</i>	43
Gambar 4. 15 Hasil <i>Merging Dataframe</i>	43
Gambar 4. 16 <i>Source Code Monetary</i>	44
Gambar 4. 17 Hasil <i>Monetary</i>	44
Gambar 4. 18 <i>Source Code Merging Pada Dataframe</i>	45
Gambar 4. 19 Hasil <i>Merging Atribut Recency, Frequency dan Monetary</i>	45
Gambar 4. 20 <i>Source Code Rescaling</i>	46
Gambar 4. 21 <i>Source Code Rescaling Pada Dataframe</i>	46
Gambar 4. 22 Hasil <i>Dataframe Rescaling</i>	46
Gambar 4. 23 <i>Source Code K-Means Clustering Dengan Nilai k=2</i>	48
Gambar 4. 24 Hasil <i>Penentuan Nilai k=2</i>	48
Gambar 4. 25 <i>Source Code Visualisasi Nilai k=2</i>	48
Gambar 4. 26 Hasil <i>K-Means Clustering</i>	49
Gambar 4. 27 <i>Source Code Iterasi</i>	50
Gambar 4. 28 <i>Source Code Iterasi-1</i>	50
Gambar 4. 29 <i>Source Code Hasil Cluster Pada Iterasi-1</i>	51
Gambar 4. 30 Hasil <i>Cluster Pada Iterasi-1</i>	51
Gambar 4. 31 <i>Source Code Hasil Visualisasi Cluster Pada Iterasi-1</i>	51
Gambar 4. 32 Hasil <i>Visualisasi Cluster Pada Iterasi-1</i>	52
Gambar 4. 33 <i>Source Code Iterasi-2</i>	53
Gambar 4. 34 <i>Source Code Hasil Cluster Pada Iterasi-2</i>	53
Gambar 4. 35 Hasil <i>Cluster Pada Iterasi-2</i>	54
Gambar 4. 36 <i>Source Code Hasil Visualisasi Cluster Pada Iterasi-2</i>	54
Gambar 4. 37 Hasil <i>Visualisasi Cluster Pada Iterasi-2</i>	55

Gambar 4. 38 <i>Source Code Evaluation</i>	56
Gambar 4. 39 Hasil <i>Evaluation</i>	57
Gambar 4. 40 <i>Source Code Evaluation</i> Metode Elbow	57
Gambar 4. 41 Hasil Visualisasi <i>Cluster</i> Pada Iterasi-2	58
Gambar 4. 42 <i>Source Code Evaluation</i> Metode DBI.....	58
Gambar 4. 43 Hasil <i>Evaluation</i> Metode DBI.....	59

ABSTRAK

Phambudy, Reynaldy Langgeng. 2022. **Penentuan Loyalitas Pelanggan Di Batubara Mart Menggunakan K-Means Clustering**. Skripsi. Program Studi Teknik Informatika, Fakultas Sains dan Teknologi. Universitas Islam Negeri Maulana Malik Ibrahim Malang. (I) Fajar Rohman Hariri, M.Kom., (II) Dr. Muhammad Faisal, M.T.

Kata Kunci: *Clustering*, Davies Boulding Index, Elbow, *K-means Clustering*, Loyalitas Pelanggan, RFM model

Batubara Mart merupakan salah satu toko yang menjual berbagai macam kebutuhan kelengkapan sehari-hari seperti makanan, minuman, peralatan rumah tangga, dan lainnya yang ditempatkan pada rak-rak atau etalase. Mengingat banyaknya pesaing bisnis pada usaha yang sama, pemilik toko harus dapat melakukan upaya untuk mencegah terjadinya perpindahan pelanggan, karena pelanggan yang puas dapat memberikan banyak keuntungan bagi toko tersebut seperti dapat memungkinkan tercapainya loyalitas pelanggan. Untuk mengetahui loyalitas pelanggan dapat memanfaatkan data transaksi yang telah diolah dengan menggunakan teknik *clustering*. Dalam penelitian ini, penulis menggunakan metode *k-means clustering* dengan menggunakan data transaksi dari bulan Januari sampai Agustus 2022. Metode *k-means clustering* diimplementasikan menggunakan bahasa pemrograman Python. Hasil dari penelitian setelah dilakukan pengujian Elbow dan *Davies Bouldin Index* (DBI) dengan nilai $k=2$ dan iterasi-3 mendapatkan hasil bahwa pada metode Elbow terdapat *cluster* terbaik yang diperoleh pada *cluster* 4 dengan nilai Elbow 7.75. Karena *cluster* terbaik merupakan *cluster* yang membentuk kurva. Sedangkan pada metode DBI didapatkan hasil *cluster* yang optimal terdapat pada *cluster* 8 dengan nilai DBI 0.05, karena semakin kecil nilai DBI maka *cluster* akan semakin optimal.

ABSTRACT

Phambudy, Reynaldy Langgeng. 2022. **Determination Of Customer Loyalty at Batubara Mart Using K-Means Clustering**. Thesis. Department of Informatics Engineering, Faculty of Science and Technology. Maulana Malik Ibrahim State Islamic University Malang. (I) Fajar Rohman Hariri, M.Kom., (II) Dr. Muhammad Faisal, M.T.

Kata Kunci: *Clustering*, Customer Loyalty, Davies Boulding Index, Elbow, *K-means Clustering*, RFM model

Batubara Mart is a shop that sells various kinds of daily necessities such as food, drinks, household appliances, and others which are placed on shelves or window displays. Given the many business competitors in the same business, shop owners must be able to make efforts to prevent customer movement, because satisfied customers can provide many benefits for the store such as being able to achieve customer loyalty. To find out customer loyalty, you can use transaction data that has been processed using clustering techniques. In this study, the authors used the k-means clustering method using transaction data from January to August 2022. The k-means clustering method is implemented using the Python programming language. The results of the research after testing the Elbow and Davies Bouldin Index (DBI) with a value of $k=2$ and iteration-3 get the result that the Elbow method has the best cluster obtained in cluster 4 with an Elbow value of 7.75. Because the best cluster is a cluster that forms a curve. While in the DBI method, the optimal cluster results are found in cluster 8 with a DBI value of 0.05, because the smaller the DBI value, the cluster will be more optimal.

الملخص

فامبودي, راينالدي لانغجنغ. 2022, تحديد ولاء العملاء في Batubara Mart يستخدم K-means Clustering . البحث الجامعي. قسم التقنية المعلوماتية كلية العلوم الكلمات الرئيسية. موالنا مالك إبراهيم الإسلامية الحكومية مالنج. (I) فجر الرحمن الحريري الماجستير. (II) الدكتور. محمد فيصل الماجستير.

الكلمات الرئيسية: التجميع, Elbow, DBI, K-Means Clustering, RFM, ولاء الزبون

Batubara Mart هو متجر يبيع أنواعًا مختلفة من الضروريات اليومية مثل الطعام والمشروبات والأجهزة المنزلية وغيرها التي توضع على الأرفف أو النوافذ. بالنظر إلى العديد من المنافسين في الأعمال التجارية في نفس العمل، يجب أن يكون أصحاب المتاجر قادرين على بذل الجهود لمنع حركة العملاء، لأن العملاء الراضين يمكن أن يوفر العديد من الفوائد للمخزن مثل القدرة على تحقيق ولاء العملاء. لمعرفة ولاء العملاء، يمكنك استخدام بيانات المعاملات التي تمت معالجتها باستخدام تقنيات التجميع. في هذه الدراسة، استخدم المؤلفون طريقة K-means Clustering باستخدام بيانات المعاملات من يناير إلى أغسطس 2022. يتم تنفيذ طريقة K-means Clustering باستخدام لغة برمجة Python. نتائج البحث بعد اختبار DBI و Elbow بقيمة $k=2$ و $k=3$ Iteration تحصل على نتيجة أن طريقة Elbow لديها أفضل مجموعة تم الحصول عليها في Cluster 4 بقيمة Elbow تبلغ 7,75. لأن أفضل كتلة هي الكتلة التي تشكل منحني. بينما في طريقة DBI، يتم العثور على نتائج الكتلة المثلى في الكتلة 8 بقيمة DBI تبلغ 0,05، لأنه كلما كانت قيمة DBI أصغر، زادت الكتلة المثلى.

BAB I

PENDAHULUAN

1.1 Latar Belakang

Kajian pendidikan merupakan kerangka pengetahuan yang diperoleh melalui pendidikan. Dengan memiliki ilmu pendidikan dapat berkembang sesuai dengan agama dan sains (Daulay & Salminawati, 2022). Keterkaitan antara ilmu Sains dan Islam merupakan perwujudan antara agama dengan sains (Minarno, 2017). Selain itu dengan adanya hubungan antara ilmu agama dengan ilmu umum itu mempengaruhi kemajuan terutam bidang sains yang begitu pesat (Harahap, 2022). Semakin canggih teknologi, akan memberikan peluang misalnya pemasaran produk yang melibatkan antara pemilik dan pembeli (Irfanita, 2020). Maka dari itu terdapat dalil yang membatasi model ikatan sependapat dengan insan yaitu aktifitas jual beli (Afandi, 2009). Begitu juga terdapat dalam perintah Allah yang tertulis dalam surat Al-baqarah ayat 275:

الَّذِينَ يَأْكُلُونَ الرِّبَا لَا يَقُومُونَ إِلَّا كَمَا يَقُومُ الَّذِي يَتَخَبَّطُهُ الشَّيْطَانُ مِنَ الْمَسِّ ۚ ذَٰلِكَ بِأَنَّهُمْ قَالُوا إِنَّمَا الْبَيْعُ مِثْلُ
الرِّبَا ۗ وَأَحَلَّ اللَّهُ الْبَيْعَ وَحَرَّمَ الرِّبَا ۚ فَمَنْ جَاءَهُ مَوْعِظَةٌ مِنْ رَبِّهِ فَانْتَهَىٰ فَلَهُ مَا سَلَفَ وَأَمْرُهُ إِلَى اللَّهِ ۗ وَمَنْ عَادَ فَأُولَٰئِكَ
أَصْحَابُ النَّارِ ۖ هُمْ فِيهَا خَالِدُونَ

// Hai manusia janganlah mencuri hak orang lain, akan terkena penyakit gangguan kejiwaan. Dimana kondisi manusia yang terkena penyakit gila, diakibatkan karena kegiatan jual beli dilakukan dengan tidak jujur atau biasa disebut riba. Sementara itu di dalam al-qur'an telah menjelaskan bahwa kegiatan jual beli itu kegiatan yang halal sedangkan yang haram itu adalah riba. Barang siapa yang kembali mengambil riba, maka orang itu termasuk penghuni neraka. Sedangkan orang yang berusaha menjauhi larangannya, maka segala urusan nya kembali kepada Allah SWT (Q.S. Al-Baqarah ayat 275).

Kemudian aktifitas perdangan merupakan pekerjaan yang mulia, apalagi Baginda Rasulullah SAW merupakan salah satu saudagar dan Rasulullah SAW selalu mendoakan para saudagar agar memiliki sifat jujur dan baik. dimana dalam hadist Rasulullah SAW menazarkan kepada orang-orang yang berdagang dengan kejujuran, sebagaimana sabda antara lain:

عن عبد الله بن عمر رضي الله عنه قال: قال رسول الله صلى الله عليه وسلم:
التَّاجِرُ الْأَمِينُ الصَّدُوقُ الْمُسْلِمُ مَعَ الشُّهَدَاءِ الشُّهَدَاءُ - وفي رواية: مع النبيين
و الصديقين و الشهداء - يَوْمَ الْقِيَامَةِ : رواه ابن ماجه والحاكم والدارقطني وغيرهم

“Dari ‘Abdullah bin ‘Umar radhiallahu ‘anhu bahwa Rasulullah bersabda, “Seorang pedagang Muslim yang jujur dan amanah (terpercaya) akan (dikumpulkan) bersama para Nabi, orang-orang shiddiq dan orang-orang yang syahid pada hari kiamat (di Surga).”(HR Ibnu Majah, al-Hakim dan ad-Daraquthni).

Aktifitas jual beli merupakan aktifitas yang sering dilakukan oleh manusia untuk membeli kebutuhan sehari-sehari. Berikutnya melakukan aktifitas jual beli dengan melibatkan kegiatan transaksi yang digunakan untuk bertukar barang sesuai dengan yang dibeli. Selain itu juga jual beli ini dapat dilakukan di dunia bisnis yang memberikan peluang untuk membantu pedagang dalam memasarkan produk yang nantinya akan dibeli oleh pelanggan.

Pelanggan merupakan seseorang baik itu individu maupun kelompok yang dapat melakukan pembelian produk sesuai dengan kebutuhan maupun dilihat dari harga. Selanjutnya pelanggan juga dapat memberikan keuntungan bagi pihak perusahaan atau pemilik usaha untuk mencapai suatu loyalitas pelanggan **(Hamdani, 2020)**. Kemudian loyalitas pelanggan merupakan tahapan awal untuk mencoba sebuah produk dengan melihat kepuasan pelanggan dengan mengarahkan pelanggan untuk pembelian ulang dari produk tersebut. Loyalitas

tersebut digunakan untuk membantu perusahaan dalam mencapai tujuan dengan produk yang sejenis. serta memperbaiki kondisi perekonomian yang sedang mengalami penurunan saat ini (**Putri, 2017**).

Batu Bara Mart adalah toko yang menjual berbagai kebutuhan sehari-hari misalnya makanan, minuman, peralatan rumah tangga , yang ditempatkan pada etalase. Meninjau adanya pesaing yang ketat di dunia bisnis, pemilik toko harus mempertahankan agar pelanggannya tidak berpindah ke toko lain, dengan memberikan kepuasan pelanggan untuk tercapainya loyalitas pelanggan. Maka dari itu pemilik toko akan memberikan hadiah dengan pelayanan yang prima serta promo spesial untuk pelanggan yang loyal, tetapi terdapat masalah lainnya saat pihak toko belum mengetahui nama pelanggan yang loyal maupun kurang loyal, sehingga pemilik toko kesulitan dalam memberikan hadiah untuk pelanggan yang loyal.

Berdasarkan permasalahan yang ada, penulis serta pihak toko berencana memecahkan data pelanggan yang memiliki loyalitas tinggi terhadap kesejahteraan toko. Pada penelitian ini dilakukan proses pengelompokan data dengan menggunakan teknik data mining. Data mining sendiri merupakan teknik yang sering digunakan untuk membantu dalam pengambilan terutama di dunia bisnis, metode yang digunakan yaitu clustering. Dimana metode clustering ini memiliki kelebihan salah satunya mengidentifikasi objek yang memiliki kesamaan karakteristik tertentu, serta menggunakan karakteristik sebagai vector atau centroid (**Hamdani, 2020**). Sebab itu, penulis akan melakukan pengamatan dengan judul **“Penentuan Loyalitas Pelanggan di Batubara Mart Menggunakan K-Means**

Clustering". Penelitian ini memiliki tujuan yaitu membangun sistem yang dapat mengelompokkan dan mengetahui tingkat loyalitas pelanggan berdasarkan transaksi pembelian agar pihak agen dapat mempertahankan pelanggan yang potensial tersebut dengan cara memberikan pelayanan prima dan memberikan hadiah berupa promo spesial.

1.2 Pernyataan Masalah

Dari penjabaran latar belakang, selanjutnya dilakukan perumusan masalahnya adalah bagaimana cara mengimplementasikan metode *K-Means Clustering* pada Penentuan Loyalitas Pelanggan di Batubara Mart?

1.3 Tujuan Penelitian

Menurut penjabaran yang diuraikan, terdapat keutungan yang diperoleh yaitu Mengimplementasikan metode *K-Means Clustering* pada Penentuan Loyalitas Pelanggan di Batubara Mart.

1.4 Manfaat Penelitian

Dari penelitian yang dilakukan diperoleh manfaat antara lain:

1. Bagi Penulis

Penelitian yang dilakukan memberikan wawasan dan pengetahuan berkaitan dengan evaluasi memakai teknik *K-means clustering*.

2. Bagi Perancang

Pengkajian dijadikan sebagai acuan untuk mengembangkan aplikasi sesuai dengan kebutuhan pengguna.

3. Bagi Pengguna

Penelitian yang dilakukan ini memberikan kemudahan dan membantu pihak toko batu bara mart untuk mengetahui pelanggan yang loyal dengan memberikan pelayanan yang baik.

1.5 Batasan Masalah

Batasan masalah dalam penelitian ini adalah:

1. Data yang digunakan adalah data transaksi periode Januari sampai Agustus 2022 pada Batubara Mart.
2. Dalam penelitian ini menggunakan Algoritma *K-Means Clustering* yang bertujuan untuk mengelompokkan loyalitas *customer* pada Batubara Mart.
3. Dalam penelitian ini menggunakan metode Elbow untuk mendapatkan jumlah kluster terbaik dan evaluasi *Davies Bouldin Index*.

1.6 Sistematika Penulisan

Urutan pencatatan tugas akhir yang biasanya terdiri dari beberapa bagian diantaranya Bab Pendahuluan, Pengkajian pustaka, metode penelitian, lalu hasil dan pembahasan, bahkan penutup. Berikut penjabaran secara detail berkaitan dengan pencatatan tugas akhir, antara lain:

BAB I: PENDAHULUAN

Pendahuluan berisi latar belakang masalah yang menjadi landasan penelitian, rumusan masalah, tujuan dan manfaat dari penelitian, kemudian batasan-batasan masalah dalam penelitian, dan sistematika penulisan.

BAB II: KAJIAN PUSTAKA

Kajian pustaka akan membahas mengenai analisis permasalahan-permasalahan pada penelitian dan penelitian-penelitian terkait Implementasi *K-Means Clustering* pada Penentuan Loyalitas Pelanggan di Batubara Mart.

BAB III: METODE PENELITIAN

Metode penelitian akan membahas tentang penjelasan rancangan struktur program dan antarmuka dari Sistem Penentuan Loyalitas Pelanggan di Batubara Mart yang dibuat.

BAB IV: HASIL DAN PEMBAHASAN

Hasil dan pembahasan menjelaskan tentang implementasi sistem berdasarkan rancangan yang telah dibuat dan pengujian aplikasi terhadap data uji untuk membuktikan bahwa aplikasi sesuai dengan tujuan penelitian yang disebutkan.

BAB V: PENUTUP

Bab terakhir atau penutup bab akan membahas tentang kesimpulan dari penelitian yang telah dilakukan dan saran yang dapat dilakukan untuk mengembangkan penelitian ini di masa mendatang.

BAB II

KAJIAN PUSTAKA

Penelitian pertama dilakukan oleh (Putra & Cendra Wadisman, 2018) yang membahas mengenai penggunaan teknik data mining untuk pemilihan pelanggan di tempat laundry. Kemudian dilakukan pengumpulan data dengan melalui observasi maupun wawancara, kemudian dilakukan analisa sesuai dengan penentuan pelanggan. Hasil dari penelitian tersebut menghasilkan pelanggan potensial dengan hasil yang akurat.

Selanjutnya penelitian kedua dilakukan oleh (Satriawan, 2021), yang membahas mengenai adanya peningkatan penggunaan operator seluler, lalu di tahun berikutnya terjadi penurunan jumlah pelanggan. untuk mengatasi kendala tersebut perlu dilakukan penerapan data mining dengan clustering. Dimana algoritma yang digunakan yaitu Kmeans yang digunakan sebagai pembentukan klaster sesuai dengan model RFM. Hasil yang diperoleh menunjukkan bahwa segmentasi pasar dapat dilihat dari profile pelanggan loyal.

Dan penelitian yang terakhir dilakukan oleh (Ardimansyah, 2021), dengan membahas mengenai penggunaan PLN pascabayar, terdapat kendala salah satunya pelanggan tidak tertib dalam membayar tagihan listrik terutama pelanggan pascabayar. kemudian dilakuakn pengelompokkan pelanggan dengan metode K-means clustering. Dengan menggunakan clustering akan menghasilkan cluster pelanggan loyal dan pelanggan yang menunggak tagihan listrik. Hasil yang

diperoleh diharapkan dapat membantu pihak PLN dalam mengetahui pelanggan yang loyal.

Dari penelitian-penelitian terdahulu diatas, adapun beberapa kesimpulan perbedaan dengan penelitian yang sedang dilakukan adalah sebagai berikut:

1. Pada penelitian pertama dan ketiga, penelitian terdahulu tidak ada penentuan jumlah cluster terbaik, sedangkan pada penelitian ini dilakukan dengan menggunakan metode tambahan yaitu elbow untuk mendapatkan informasi jumlah cluster terbaik dengan melihat persentase hasil perbandingan antar cluster.
2. Pada penelitian kedua, penelitian terdahulu terdapat metode yang digunakan untuk menguji kualitas cluster yang dihasilkan, sedangkan dalam penelitian ini hasil clustering tidak dinilai kualitasnya, hal ini dikarenakan pada penelitian ini sudah menggunakan metode elbow untuk menentukan jumlah cluster terbaik. Pada penelitian kedua ini juga pembentukan cluster didasarkan pada model RFM.

2.1 Loyalitas Pelanggan

Loyalitas pelanggan merupakan kewajiban dalam membeli sebuah produk yang diminati walaupun di masa yang akan datang memiliki potensi pasar yang menyebabkan perubahan. Pelanggan yang loyal akan memberikan keuntungan untuk perusahaan dikarenakan pelanggan yang loyal dapat memperkenalkan produk yang dirasakan rekannya atau keluarga. Selain itu pelanggan loyal akan menggunakan produk dari perusahaan jika produk tersebut sesuai dengan

pelanggan, dan pelanggan akan memakai produk itu tanpa menggunakan produk lain (Ardimansyah, 2021).

Dan loyalitas konsumen sangat penting dikarenakan dengan konsumen menggunakan produk secara rutin dapat meningkatkan kepuasan pelanggan pada perusahaan tersebut, dari kepuasan pelanggan tersebut dapat timbul rasa ingin memberikan rekomendasi produk ke orang lain (Lintas Media Danawa, 2022).

Ada lima cara untuk memelihara loyalitas konsumen yaitu antara lain (Sasongko, 2021):

1. Cara yang pertama, pelanggan memiliki hak dalam melakukan pembelian produk, untuk menjaga hak pelanggan perusahaan lebih memperhatikan apa yang diinginkan pelanggan saat pembelian suatu produk.
2. Kemudian cara selanjutnya, dilakukan pendekatan dengan melakukan inovasi pada produk sesuai yang diharapkan oleh pelanggan.
3. Berikutnya melakukan pengukuran pelanggan untuk melihat apakah ada kenaikan dalam kepuasan atau tidak, dan dilaksanakan secara terus terutama mengenai loyalitas pelanggan.
4. mewujudkan biaya dengan memberikan harga yang terjangkau khususnya untuk pelanggan yang loyal.
5. Dan terakhir dengan cara memberikan pemberian hadiah untuk pelanggan yang loyal misalnya dengan memberikan promo dan pelayanan yang prima.

2.2 *Clustering*

Metode clustering merupakan metode yang termasuk dalam teknik data mining yang digunakan untuk mengelompokkan data yang memiliki kemiripan dengan kelompok lain. Selain itu cluster biasanya mengumpulkan data berupa record dengan melihat kemiripan dengan lain, serta memiliki ketidakmiripan record dalam cluster lain (Sulistiyowati, 2018).

Kemudian menurut penelitian lain mengatakan, clustering merupakan teknik dalam data mining dengan mencari pola pada data serta memiliki karakteristik dengan satu data lain yang diperoleh (Dinata, Safwandi, Hasdyna, & Azizah, 2020).

Lalu clustering tersebut biasanya digunakan dalam membagi rangkaian data menjadi beberapa group sesuai kesamaan yang ditentukan. Setelah itu dilakukan pengumpulan data dengan objek data yang sama dengan lainnya (Rahmah, 2020).

2.3 *Preprocessing*

Preprocessing data merupakan tahapan proses data yang akan menghasilkan kumpulan data akhir. Dimana tahapan preprocessing pada penelitian ini terdiri dari proses pembersihan data hingga pemecahan masalah. Ada tahapan preprocessing yang akan digunakan antara lain:

1. *Data Cleansing*

Biasanya data cleansing diterapkan untuk mengurangi data yang kurang relevan atau data yang tidak sesuai (Dharshinni & Fandi, 2022).

Tahapan proses cleansing sendiri terdiri dari analisis kualitas dengan

melakukan pengubahan, koreksi, bahkan menghapus data. (Suhandi, Kurniati, & Norma, 2020).

2. *Data Preparation*

Selanjutnya melakukan perbaikan dari permasalahan yang terjadi dengan tahapan modeling. (Suhandi, Kurniati, & Norma, 2020).

2.4 **K-Means Clustering**

K-means clustering merupakan metode dengan menerapkan data clustering sebagai pengelompokan data baik satu cluster maupun berkelompok. kemudian dalam pengelompokan data yang mempunyai karakter yang sama akan di kelompokkan dalam satu kelompok, sedangkan pengelompokan yang berbeda akan di kelompokkan dengan cluster yang lainnya, sehingga data dalam satu kelompok memiliki variasi. Informasi yang berupa data akan disesuaikan dengan cluster yang sama, sedangkan dengan informasi yang berkarakter berbeda akan disesuaikan dengan cluster yang berbeda (Sulistiyowati, 2018).

Selanjutnya setelah melakukan pengumpulan data, dilakukan partisi data ke bentuk satu cluster, sehingga satu cluster yang sama memiliki karakteristik yang sama juga, melainkan dengan karakteristik berbeda akan dikelompokkan dengan kelompok yang berbeda (Muningsih, Maryani, & Handayani, 2021).

Tahapan clustering dengan K-means sebagai berikut (Sulistiyowati, 2018):

1. Pertama, melakukan jumlah cluster sesuai dataset yang dipakai.

2. Kemudian yang kedua menentukan nilai centroid dilakukan secara acak sesuai target awal k-means, dengan mencari terlebih dahulu jarak antara kelompok sesuai dengan iterasi, seperti pada persamaan 2.1 berikut ini:

$$\frac{(\text{Jumlah Data})}{(\text{Jumlah cluster}+1)} \quad (2.1)$$

Keterangan:

Jumlah data: Jumlah data yang akan digunakan.

Jumlah cluster: Jumlah kelompok yang telah ditentukan sebelumnya

Seperti sangat tinggi, tinggi, normal, rendah, sangat rendah.

Setelah itu di tahap iterasi terlebih dahulu melakukan perhitungan dengan mencari nilai rata-rata, dapat dilihat pada persamaan 2.2:

$$V_{ij} = \frac{1}{N_i} = \sum_{k=0}^{N_i} x_{kj} \quad (2.2)$$

Keterangan:

V_{ij} : centroid rata-rata cluster ke-i untuk variabel ke-j

N_i : jumlah anggota cluster ke-I

i,k: indeks dari cluster

j: indeks dari variabel

X_{kj} : nilai data ke-k variabel ke-j dalam cluster tersebut.

3. Selanjutnya melakukan record dengan menghitung jarak terdekat dengan centroid atau titik pusat menggunakan euclidean distance, dapat dilihat pada persamaan 2.3:

$$De = \sqrt{(x_i - s_i)^2 + (y_i - t_i)^2} \quad (2.3)$$

Keterangan:

De: Euclidean Distance I: banyaknya objek

(x, y): koordinat objek

(s, t): koordinat centroid

4. Lalu mengelompokkan objek sesuai jarak dengan titik pusat terdekat.
5. Dan tahapan akhir mengulangi langkah ke-2 dengan iterasi hingga centroid yang bernilai.

Proses cara kerja K-means dengan menggunakan euclidean distance untuk menghitung jarak dengan menetapkan terlebih dahulu jumlah cluster yang akan dibentuk, seperti 3, 5, atau 7 cluster. kemudian ditentukan titik pusat secara awal. lalu jika nilai jarak sudah diketahui, maka mengelompokkan setiap data berdasarkan jarak yang dekat dari titik pusat. jika terdapat data yang belum masuk ke cluster, mengulangi proses penentuan centroid, sehingga tidak ada data yang belum masuk cluster. Metode k-means memiliki ciri khas yang pertama mudah dalam mengelompokkan data menjadi sederhana, yang kedua k-means tidak melakukan segmentasi data sesuai pola kelompok, dan terakhir k-means mengalami kendala saat mengelompokkan data yang mengandung outlier.



Gambar 2.1 Flowchart K-Means

2.5 Metode Elbow

Metode elbow biasanya metode yang sering digunakan untuk menentukan jumlah banyaknya cluster yang akan terbentuk di suatu titik. Serta memberikan pengetahuan dalam memilih nilai klaster, dan menambahkan nilai cluster yang dijadikan sebagai model data untuk menentukan cluster terbaik. selain itu melakukan perhitungan persentase menerapkan visualisasi berupa grafik sebagai informasi. Apabila hasil perhitungan cluster pertama dengan hasil cluster kedua akan mengalami penurunan nilai, sehingga cluster tersebut termasuk cluster yang baik (Satriawan, 2021).

Setelah itu metode elbow perlu melakukan perhitungan dengan nilai SSE setiap cluster sesuai dengan titik pusatnya (Nurdiana, Nilogiri, & Rahman, 2022).

Pada metode elbow juga memiliki tujuan sebagai penentuan jumlah cluster terbaik. Adapun tahapan perhitungan metode elbow sebagai berikut (Nurdiana, Nilogiri, & Rahman, 2022):

1. Menetapkan nilai kluster terlebih dahulu.
2. Kemudian menaikkan nilai kluster sesuai yang sudah ditentukan.
3. Selanjutnya menghitung nilai SSE untuk tiap klaster.

$$SSE = \sum_{k=1}^k \sum_{x_i \in S_k} \|X_i - C_k\|_2^2 \quad (2.4)$$

X_i = Nilai attribute informasi ke - i

C_k = Angka attribute titik pusat kluster ke - i

4. Lalu menghitung nilai SSE hingga kluster ditetapkan.
5. Menganalisa hasil nilai SSE kluster yang turun secara teliti.
6. Dan terakhir kluster akan membentuk kurva.

2.6 Evaluasi *Clustering Davies Bouldin Index (DBI)*

Davies Bouldin Index (DBI) merupakan metode yang termasuk evaluasi internal dengan melakukan pengukuran klaster berdasarkan nilai separasi dan kohesi. pengukuran menggunakan kohesi yaitu adanya pendekatan data dengan pusat klaster sesuai klaster yang dipakai. sedangkan separasi dilakukan pengukuran jarak dengan titik pusat dari tiap klasternya. Davies Bouldin Index (DBI) sangat membantu dalam menyempurnakan kinerja pada algoritma itu sendiri untuk menghasilkan cluster yang baik (Wen Gie & Deny Jollyta, 2020). Berikutnya dilakukan evaluasi dengan skema evaluasi cluster, dimana dilihat dari

baik atau tidaknya kuantita data klaster tersebut. Semakin kecil nilai DBI yang diperoleh, hasil akan menunjukkan skema cluster yang optimal (Jollyta, Efendi, & Zarlis, 2019). Rumus DBI, (R_{ij}) untuk mengetahui nilai perbandingan antara cluster ke-i dan cluster ke-j adalah:

$$R_{ij} = \frac{SSW_I + SSW_J}{SSB_{I,J}} \quad (2.5)$$

Persamaan untuk menghitung nilai Davies Bouldin Index (DBI) berikut ini:

$$R_{ij} = \frac{1}{k} \sum_{i=1}^k \max_{j \neq i} (R_{ij}) \quad (2.6)$$

Keterangan :

k = jumlah cluster yang digunakan

Semakin rendah nilai Davies Bouldin Index (DBI) yang diperoleh, maka semakin baik kualitas cluster yang diperoleh dari suatu clustering data.

Dalam menghitung DBI diperlukan alat ukur untuk klasterisasi, dilihat dari akurasi klaster dengan jarak yang paling kecil dan menunjukkan nilai optimal (Situmorang, Arifin, Rusilpan, & Juliane, 2022).

BAB III

METODE PENELITIAN

3.1 Objek Penelitian

Objek penelitian ini digunakan untuk, pembatasan pembahasan topik penelitian yang kemudian dianalisa terkaita Penerapan K-means clustering pada penentuan loyalitas pelanggan. Objek penelitian yang dipakai yaitu perusahaan Batu Bara Mart yang berlokasi di Lingkungan 01, ciriung, Cibinong, Kabupaten Bogor, Jawa Barat 16918

3.2 Alat Penelitian

Penelitian ini membutuhkan dua jenis alat bantu untuk mempermudah penelitian yaitu menggunakan perangkat keras dan perangkat lunak:

3.2.1 *Hardware*

Alat perangkat keras yang mendukung aktifitas penelitian, antara lain:

1. Intel Core i3-3320
2. Memori RAM 4 GB
3. Harddisk 20 GB

3.2.2 *Software*

Dan perangkat lunak yang digunakan untuk penelitian ini antara lain:

1. Sistem Operasi Microsoft Windows 10
2. Google Chrome
3. Sublime Text 3
4. Visual Studio Code

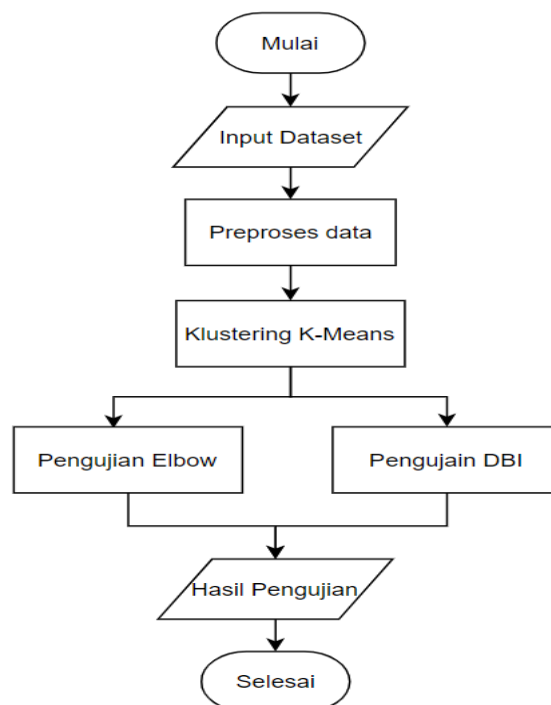
5. XAMPP
6. Draw.io
7. Balsamiq

3.3 Prosedur Penelitian

Dalam prosedur penelitian ini terdiri dari desain sistem dan pengujian sistem.

3.3.1 Desain Sistem

Pada prosedur penelitian terdapat desain sistem yang mana terdiri dari 2 gambaran flowchart seperti flowchat proses K-means dan flowchart sistem. Gambaran flowchart proses K-means, dapat dilihat pada Gambar 3.1 dan Gambar 3.2 itu flowchart proses K-means.

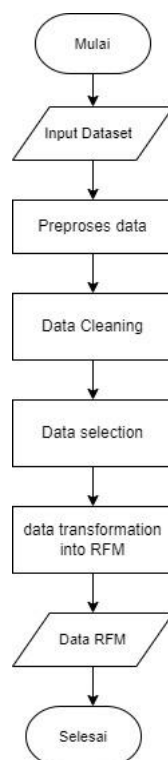


Gambar 3. 1 *Flowchart* Sistem

Pada penelitian ini, tahapan utama untuk mengetahui loyalitas customer adalah melakukan input dataset untuk dilakukan *preprocessing* data yang menghasilkan data RFM dari beberapa *customer* yang kemudian akan diproses ke dalam algoritma *k-means clustering*. Pada tahap *k-means clustering* nantinya akan diperoleh nilai jarak terdekat dari masing-masing centroid, setelah itu dilakukan pengujian elbow dan DBI untuk mengetahui hasil clustering terbaik.

1. RFM

Gambar 3.2 merupakan diagram alir proses RFM, ada tahap awal penelitian ini, dataset yang digunakan akan diubah terlebih dulu ke dalam bentuk RFM (*Recency, Frequency, Monetary*). Hal ini dilakukan untuk mempermudah proses klustering dalam bentuk kunjungan, transaksi dan biaya yang telah dilakukan oleh customer.



Gambar 3. 2 Proses RFM

2. Input Dataset

Dataset yang digunakan memiliki kolom product, SKU, supplier, reference no, date, quantity, total unit, unit purchase price, subtotal. Dataset ini nantinya akan diproses dan ditransformasikan ke dalam bentuk RFM dan kolom yang tidak digunakan dieliminasi.

Tabel 3. 1 Dataset

Product	SKU	Supplier	Reference No	Date	Quantity	Total Unit	Unit Purchase Price	Subtotal
OREO MINI ORIGIN AL 23GR	7622 2105 1528 5	ADIT ORE O	PO202 1/4484	29- 12- 2021	48 Pc(s)	0 Pc(s)	1366, 75	65604
MILKU STROB ERI 200ML	8998 8662 0273 2	ARIM BI WIN GS	PO202 1/4492	30- 12- 2021	36 Pc(s)	0 Pc(s)	2425	87300
INDOMI E KALDU UDANG 65GR	0896 8601 0138	DAY U PT INDO MAR CO MIE	PO202 1/4308	15- 12- 2021	80 Pc(s)	2 Pc(s)	1800, 63	14044 9,14
DETTO L LIQUID ANTISE PTIK 95ML	8993 5600 2202 0	BAM BAN G DETT OL	PO202 1/3133	10- 09- 2021	3 Pc(s)	0 Pc(s)	16221 ,67	48665 ,01

3. Praproses Data

Data yang sudah diperoleh dari data transaksi pelanggan akan melalui praproses data yang meliputi (Sudriyanto, 2017):

- Langkah pertama melakukan pembersihan data untuk mengurangi nilai yang kosong maupun data yang tidak valid.
- Kemudian setelah dilakukan pembersihan data, selanjutnya menentukan atribut dengan model RFM, misalnya atribut transaksi, jumlah frekuensi transaksi hingga jumlah nominal setiap pelanggan.

Tabel 3. 2 Dataset Hasil Data Selection

Bulan	Tanggal	Supplier	Unit	Pu	Total Price	Bulan	Tanggal	Supplier	Unit	Pu	Total Price	Bulan	Tanggal	Supplier	Unit	Pu	Total Price	Bulan	Tanggal	Supplier	Unit	Pu	Total Price
April	08-04-2	ADLUCKBUN	3	4700		April	01-03-2	ADLUCKBUN	8	6210		Februari	24-02-2	ADIT OREO	10	16401		Februari	12-01-2	AANKANTONG	10	52697	
	08-04-2	ADLUCKBUN	3	78900			24-03-2	ADIT OREO	20	16390			24-02-2	ADIT OREO	20	15106			12-01-2	AANKANTONG	10	52697	
	08-04-2	ADLUCKBUN	3	42000			24-03-2	ADIT OREO	24	33362			24-02-2	ADIT OREO	24	161964			12-01-2	AANKANTONG	5	25610	
	08-04-2	ADLUCKBUN	3	78900			24-03-2	ADIT OREO	6	36973			24-02-2	ADIT OREO	6	36973			12-01-2	AANKANTONG	10	7387	
	08-04-2	ADLUCKBUN	3	66600			24-03-2	ADIT OREO	20	48664			24-02-2	ADIT OREO	16	66394			12-01-2	AANKANTONG	5	25610	
	08-04-2	ADLUCKBUN	3	45900			24-03-2	ADIT OREO	6	19891			24-02-2	ADIT OREO	6	12198			12-01-2	AANKANTONG	10	1232	
	08-04-2	ADLUCKBUN	3	62500			24-03-2	ADIT OREO	0	0			24-02-2	ADIT OREO	10	16401			12-01-2	AANKANTONG	10	52697	
	08-04-2	ADLUCKBUN	3	66357			24-03-2	ADIT OREO	6	19891			24-02-2	ADIT OREO	4	3732,8			23-01-2	ADIT OREO	12	11778	
	08-04-2	ADLUCKBUN	3	21600			24-03-2	ADIT OREO	12	74461			24-02-2	ADIT OREO	12	73946			23-01-2	ADIT OREO	144	238174	
	08-04-2	ADLUCKBUN	3	32100			24-03-2	ADIT OREO	12	73946			24-02-2	ADIT OREO	12	0			23-01-2	ADIT OREO	36	53042	
	08-04-2	ADLUCKBUN	1	27500			24-03-2	ADIT OREO	12	214322			24-02-2	ADIT OREO	0	0			23-01-2	ADIT OREO	24	423844	
	21-04-2	ADIT OREO	24	57814			24-03-2	ADIT OREO	2	18900			24-02-2	ADIT OREO	20	16390			23-01-2	ADIT OREO	6	36972	
	21-04-2	ADIT OREO	36	58454			24-03-2	ADIT OREO	144	236174			23-02-2	AGUS KEPALA D.	3	37500			23-01-2	ADIT OREO	4	9732,2	
	21-04-2	ADIT OREO	24	144559			24-03-2	ADIT OREO	30	43203			23-02-2	AGUS KEPALA D.	6	36000			23-01-2	ADIT OREO	6	36973	
	21-04-2	ADIT OREO	6	65859			24-03-2	ADIT OREO	30	43203			23-02-2	AGUS KEPALA D.	6	41804			23-01-2	ADIT OREO	0	0	
	21-04-2	ADIT OREO	24	168630			24-03-2	ADIT OREO	6	76302			23-02-2	AGUS KEPALA D.	3	15000			23-01-2	ADIT OREO	6	124197	
	21-04-2	ADIT OREO	36	58454			24-03-2	ADIT OREO	24	33362			17-02-2	AGUS SGM	12	343780			23-01-2	ADIT OREO	6	76309	
	21-04-2	ADIT OREO	10	15313			24-03-2	ADIT OREO	0	0			17-02-2	AGUS SGM	50	46406			23-01-2	ADIT OREO	6	36972	
	21-04-2	ADIT OREO	36	58454			24-03-2	ADIT OREO	0	0			17-02-2	AGUS SGM	0	0			23-01-2	ADIT OREO	6	34197	
	21-04-2	ADIT OREO	6	54001			24-03-2	ADIT OREO	0	0			17-02-2	AGUS SGM	0	0			23-01-2	ADIT OREO	4	9732,8	
	21-04-2	ADIT OREO	20	31825			23-03-2	AGUS KEPALA	12	72000			17-02-2	AGUS SGM	3	117399			23-01-2	ADIT OREO	24	33361	
	21-04-2	ADIT OREO	36	58454			23-03-2	AGUS KEPALA	3	21600			17-02-2	AGUS SGM	24	835725			22-01-2	AGUNG SENSI	12	114252	
	21-04-2	ADIT OREO	10	15313			23-03-2	AGUS KEPALA	3	37500			17-02-2	AGUS SGM	0	0			22-01-2	AGUNG SENSI	12	114252	
	21-04-2	ADIT OREO	24	57814			23-03-2	AGUS KEPALA	3	15000			17-02-2	AGUS SGM	60	36406			09-01-2	AGUNG SENSI	24	39468	
	21-04-2	ADIT OREO	6	36604			31-03-2	AGUS SGM	10	732712			17-02-2	AGUS SGM	48	633902			09-01-2	AGUNG SENSI	0	0	
	21-04-2	ADIT OREO	36	58454			31-03-2	AGUS SGM	36	16406			03-02-2	AGUS SGM	0	0			08-01-2	AGUNG SENSI	6	69550	
	21-04-2	ADIT OREO	12	73207			31-03-2	AGUS SGM	12	343780			03-02-2	AGUS SGM	0	0			25-01-2	AGUS SGM	480	16406	
	21-04-2	ADIT OREO	24	153113			31-03-2	AGUS SGM	12	164666			03-02-2	AGUS SGM	12	617175			20-01-2	AGUS SGM	10	77252	
	21-04-2	ADIT OREO	6	36604			31-03-2	AGUS SGM	0	0			03-02-2	AGUS SGM	0	0			20-01-2	AGUS SGM	0	0	
	21-04-2	ADIT OREO	20	16226			31-03-2	AGUS SGM	6	115528			03-02-2	AGUS SGM	0	0			20-01-2	AGUS SGM	0	0	
	21-04-2	ADIT OREO	16	107894			31-03-2	AGUS SGM	21	16406			03-02-2	AGUS SGM	48	26406			20-01-2	AGUS SGM	0	0	
	07-04-2	ADIT OREO	12	107908			31-03-2	AGUS SGM	0	0			03-02-2	AGUS SGM	21	732197			20-01-2	AGUS SGM	30	26406	
	07-04-2	ADIT OREO	20	32042			31-03-2	AGUS SGM	24	353881			03-02-2	AGUS SGM	0	0			20-01-2	AGUS SGM	0	0	
	07-04-2	ADIT OREO	6	54001			31-03-2	AGUS SGM	6	154037			03-02-2	AGUS SGM	0	0			20-01-2	AGUS SGM	0	0	
	07-04-2	ADIT OREO	24	146414			31-03-2	AGUS SGM	42	26406			03-02-2	AGUS SGM	3	107399			20-01-2	AGUS SGM	10	727124	
	07-04-2	ADIT OREO	0	0			31-03-2	AGUS SGM	0	0			03-02-2	AGUS SGM	0	0			20-01-2	AGUS SGM	0	0	
	07-04-2	ADIT OREO	144	22307			31-03-2	AGUS SGM	10	77252			03-02-2	AGUS SGM	0	0			20-01-2	AGUS SGM	21	158465	

- Tahapan berikutnya dengan mempersiapkan data agar tidak ada data yang double atau redudansi dalam melakukan proses klaterisasi. tahapan preprocessing ini menggunakan 10 pelanggan berdasarkan transaksi pembelian, kunjungan serta total pembayaran selama januari-desember.

Tabel 3. 3 Hasil Data Preprocessing

Bulan	Tanggal	Supplier	Unit Purchase	Total Price
Januari	12-01-2021	AAN KANTONG	10	52697
	12-01-2021	AAN KANTONG	10	93577
	12-01-2021	AAN KANTONG	10	52697
	12-01-2021	AAN KANTONG	5	25610
	12-01-2021	AAN KANTONG	10	7387
	12-01-2021	AAN KANTONG	5	25610

d. *Transformation*, yaitu transformasi data ke dalam bentuk terukur sehingga dapat digunakan sebagai atribut untuk proses klasterisasi. Pada tahap ini, dataset yang sudah direduksi kemudian ditransformasikan ke dalam bentuk model RFM. Nilai RFM ini diperoleh dari hasil transformasi berdasarkan total kunjungan pelanggan (*Recency*), total transaksi (*Frequency*), dan jumlah pengeluaran biaya (*monetary*).

- 1) *Recency* adalah rentang waktu (hari, bulan, tahun) dari transaksi akhir sampai saat ini oleh pelanggan membeli.
- 2) *Frequency* atau frekuensi adalah total transaksi atau transaksi rata-rata dalam satu kali periode.
- 3) *Monetary* atau moneter adalah biaya total pelanggan di sekali waktu.

Tabel berikut merupakan hasil Transformation menggunakan RFM selama bulan Januari - Desember 2021.

Tabel 3. 4 Data Transformation RFM

Nama	R	F	M
ADIT OREO	361	9103	28783208,19
ADI LUCKY BUNNY	155	772	9832396,71
AGUS SGM	305	10728	268498303,9
ARIMBI WINGS	3445	158075	271260922,7
BAMBANG DETTOL	168	949	8926048,61
BAYU	457	3806	29054573,25
CHANDRA NIVEA	257	1787	19571413,48
DAYU PT INDOMARCO MIE	1019	43548	103903449,2
DESKA ARTABOGA	1433	16731	60276131,21
DESTI	110	4212	10937476,03

- e. Selanjutnya melakukan normalisasi yang bertujuan untuk mengurangi kesalahan pada data serta tidak konsistensinya data, dengan cara menghitung memakai min-max scaling.

$$X' = \frac{(x - X_{min})}{(X_{max} - X_{min})} \quad (3.1)$$

Keterangan :

X' = Hasil normalisasi data

X_{max} = Data maksimum dari data X

X_{min} = Data minimum dari data X

Berikut ini adalah contoh perhitungan dari min-maxscaling untuk data RFM pada pelanggan Adit Oreo

- Recency

$$X = (1 - (14 - 0)) / (188 - 0) = 0,92$$

- Frequency

$$X' = \frac{(962 - 0)}{(2543200 - 0)} = 0,33$$

- Monetary

$$X' = \frac{(286 - 0)}{(5276 - 0)} = 286$$

4. Klusterisasi

K-means cluster ialah teknik algoritma untuk analisis non hirarki yang digunakan untuk mengelompokkan ke dalam satu hingga banyak klaster sesuai dengan karakteristiknya.



Gambar 3. 3 Diagram Alir K-Means

Kemudian tahapan penggunaan clustering memanfaatkan algoritma *K-means*, diantaranya:

a) Input Dataset

Proses pertama dilakukan penggunaan dataset yang sudah ditransformasikan ke dalam RFM. kemudian nilai RFM dihasilkan dari rekapitulasi banyaknya kunjungan pelanggan, kemudian dilakukan transaksi, dan pembiayaan belanja konsumen. setelah itu perlu dilakukan normalisasi.

Tabel 3. 5 Dataset Hasil Normalisasi Transformasi RFM

Nama	R	F	M
ADIT OREO	10,897692	0,893699179	9102,973
ADI LUCKY BUNNY	11,631572	0,239479712	771,9727
AGUS SGM	10,979795	9,786671966	10727,97
ARIMBI WINGS	0	8,897352035	158075
BAMBANG DETTOL	11,545658	0,226921893	948,9727
BAYU	10,542455	0,892257034	3805,973
CHANDRA NIVEA	11,250161	0,584975509	1786,973
DAYU PT INDOMARCO MIE	8,471646	3,392143131	43547,97
DESKA ARTABOGA	7,0876404	2,023580976	27601,67
DESTI	11,792258	0,265070581	4211,973

b) Partisi Data

Proses selanjutnya dilakukan transformasi RFM dengan cara penentuan jumlah centroid dan titik pusat.

c) Inisiasi Jumlah Cluster

Kemudian proses inisiasi banyak nya cluster dengan menggunakan $K=2$ dengan iterasi = 3

Tabel 3. 6 Iterasi 1

	R	F	M
C1	10,97979	9,786671966	10727,97268
C2	11,25016	0,584975509	1786,972682

Tabel 3. 7 Iterasi 2

C1	0	8,897352035	158074,9727
C2	8,471646	3,392143131	43547,97268

Tabel 3. 8 Iterasi 3

	R	F	M
C1	11,54566	0,226921893	948,9726822
C2	7,08764	2,023580976	27601,67176

d) Menentukan Titik Centroid

Titik centroid ditentukan dengan pemilihan data yang dijadikan sebagai C1 ataupun C2, ataupun pemilihan titik pusat kluster yang nanti akan diproses ke dalam algoritma.

e) Menghitung Data berdasarkan Centroid

Titik pusat atau centroid yang telah ditentukan kemudian dihitung jaraknya terhadap distribusi data disekitarnya. Perhitungan ini bisa menggunakan Euclidean distance, ataupun jaccard distance. Pada penelitian ini perhitungan yang dilakukan menggunakan Euclidean distance.

f) Alokasi seluruh data kedalam Cluster

Hitung *centroid*/rata-rata dari data yang ada di masing-masing cluster menggunakan perhitungan *Euclidean distance*

$$d(x,y) = \sqrt{((R - Rc)^2) + ((F - Fc)^2) + ((M - Mc)^2)} \quad (3.2)$$

Berikut ini adalah contoh perhitungan euclidean distance untuk customer adit oreo untuk C1 dan C2 pada iterasi 1.

$$\begin{aligned} dC1(Adit oreo) &= \\ \sqrt{((10,8977 - 10,9798)^2) + ((0,8936 - 9,786)^2) + ((9102,97 - 10727,97)^2)} &= 1625,02 \\ dC2(Adit oreo) &= \\ \sqrt{((10,8977 - 11,2502)^2) + ((0,8936 - 0,5849)^2) + ((9102,97 - 1786,97)^2)} &= 7316 \end{aligned}$$

g) Menentukan Centroid Baru

Setelah dilakukan perhitungan terhadap titik centroid C1 dan C2, pada iterasi selanjutnya dilakukan kembali pemilihan centroid seperti yang tercantum pada tabel 4, 5, dan 6.

h) Lakukan sebanyak iterasi

Lakukan perulangan langkah pada bagian a-g hingga iterasi yang diinginkan.

i) Uji Elbow dan *Davies Bouldin Index* (DBI)

Pengujian elbow dan DBI akan dilakukan setelah mendapat hasil clustering pada masing-masing centroid dan setelah dilakukan beberapa kali iterasi.

j) Hasil Cluster

Berdasarkan hasil clustering yang telah dilakukan terhadap loyalitas customer, tabel 7 hingga 9 menyatakan hasil klustering untuk 3 kali iterasi, iterasi 1 dan iterasi 2 memiliki kluster yang cocok dan sesuai.

Tabel 3. 9 Alokasi Kluster dengan Euclidean distance (Iterasi 1)

Nama	C1	C2	Jarak terdekat	Kluster
ADIT OREO	1625,0243	7316	1625,0243	1
ADI LUCKY BUNNY	9956,0046	1015	1015	2
AGUS SGM	0	8941,005	0	1
ARIMBI WINGS	147347	156288	147347	1
BAMBANG DETTOL	9779,0047	838,0001	838,0001	2
BAYU	6922,0057	2019	2019	2
CHANDRA NIVEA	8941,0047	0	0	2
DAYU PT INDOMARCO	32820,001	41761	32820,001	1
DESKA ARTABOGA	16873,701	25814,7	16873,701	1
DESTI	6516,007	2425	2425	2

Tabel 3. 10 Alokasi Kluster dengan Euclidean distance (Iterasi 2)

Nama	C1	C2	Jarak terdekat	Kluster
ADIT OREO	148972	34445	34445	2
ADI LUCKY BUNNY	157303	42776	42776	2
AGUS SGM	147347	32820	32820	2
ARIMBI WINGS	0	114527	0	1
BAMBANG DETTOL	157126	42599	42599	2
BAYU	154269	39742	39742	2
CHANDRA NIVEA	156288	41761	41761	2
DAYU PT INDOMARCO	114527	0	0	2
DESKA ARTABOGA	130473,3	15946,3	15946,3	2
DESTI	153863	39336	39336	2

Tabel 3. 11 Alokasi Kluster dengan Euclidean distance (Iterasi 3)

Nama	C1	C2	Jarak terdekat	Kluster
ADIT OREO	8154,0001	18498,7	8154,0001	1
ADI LUCKY BUNNY	177,00002	26829,7	177,00002	1
AGUS SGM	9779,0047	16873,7	9779,0047	1
ARIMBI WINGS	157126	130473,3	130473,3	2
BAMBANG DETTOL	0	26652,7	0	1
BAYU	2857,0003	23795,7	2857,0003	1
CHANDRA NIVEA	838,00013	25814,7	838,00013	1
DAYU PT INDOMARCO	42599	15946,3	15946,3	2
DESKA ARTABOGA	26652,7	0	0	2
DESTI	3263	23389,7	3263	1

Tabel 3. 12 Hasil Clustering dari data RFM

Nama	R	F	M	Iterasi 1	Iterasi 2	Iterasi 3
ADIT OREO	10,89 7692	0,8936 99179	9102, 973	1	2	1
ADI LUCKY BUNNY	11,63 1572	0,2394 79712	771, 9727	2	2	1
AGUS SGM	10,97 9795	9,7866 71966	1072 7,97	1	2	1
ARIMBI WINGS	0	8,8973 52035	158 075	1	1	2
BAMBANG DETTOL	11,54 5658	0,2269 21893	948, 9727	2	2	1
BAYU	10,54 2455	0,8922 57034	3805, 973	2	2	1
CHANDRA NIVEA	11,25 0161	0,5849 75509	1786, 973	2	2	1
DAYU PT INDOMARCO MIE	8,471 646	3,3921 43131	43547, 97	1	2	2
DESKA ARTABOGA	7,087 6404	2,0235 80976	27601, 67	1	2	2
DESTI	11,792 258	0,2650 70581	4211, 973	2	2	1

5. Pengujian kluster dengan menguji nilai rata-rata menggunakan pengujian elbow dan metode *Davies-Bouldin-Index*.

a. Pengujian Elbow

Pengujian teknik elbow ialah biasa dipakai penentuan kalster dengan melakukan tolak ukur sesuai dengan jumlah cluster yang akan membentuk satu titik pusat. Selanjutnya melakukan perbandingan dengan menghitung SSE setiap nilai klaster, jika jumlah nilai klaster k, akan semakin kecil hasilnya. untk rumus SSE dapat dilihat di persamaan berikut ini.

$$SSE = K = \sum_{K=1}^K \sum_{xi} |xi - ck|^2 \quad (3.3)$$

Keterangan:

K = cluster ke-c

X_i = jarak data obyek ke-i

C_k = pusat cluster ke-i

Contoh perhitungan SSE untuk iterasi 1

$$SSE = 10 \times (7316 - 86900,9)^2$$

$$SSE = 6,334 \times 10^{10}$$

Contoh perhitungan SSE untuk iterasi 2

$$SSE = 10 \times (15946,3 - 86900,9)^2$$

$$SSE = 5,035 \times 10^{10}$$

Contoh perhitungan SSE untuk iterasi 3

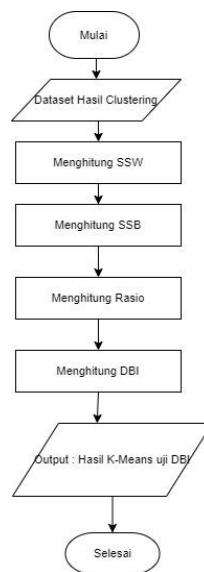
$$SSE = 10 \times (8154 - 86900,9)^2$$

$$SSE = 4,679 \times 10^{10}$$

Berikutnya setelah menentukan klaster, lalu dilakukan 3 kali untuk di uji coba. kemudian masukan nilai K berawal dari K nya yaitu 2. Hasil ujicoba yang dilakukan menghasilkan nilai SSE yang menunjukkan penurunan pada iterasi yaitu 2, sehingga jumlah cluster saat iterasi 2. etode Elbow dalam penelitian kali ini dilakukan sebanyak 3 kali ujicoba. Pengujian nilai K berawal dari K= 2.

b. Evaluasi *Davies-Bouldin-Index*

Tahapan selanjutnya dilakukan evaluasi DBI dengan menganalisis setiap clustering. kemudian dilakukan pendekatan dengan kohensi dan separasi. cluster dapat dikemukakan dengan hasil optimal, apabila hasil klaster untuk kohesi itu rendah tetapi sebaliknya apabila untuk separasi itu memperoleh nilai tinggi.



Gambar 3. 4 *Flowchart Uji DBI*

Berikut tahapan perhitungan dengan evalasui DBI diantaranya:

1. Menghitung terlebih dahulu dengan melakukan perhitungan SSW (Sum Of Square Within Cluster), dimana satu cluster atau lebih di lihat kemiripan setiap clusternya. Jika nilai kecil, maka kemiripannya juga semakin besar. menghitung SSW bertujuan untuk mengetahui nilai kohesi dalam cluster.

$$SSW_i = \frac{1}{m_i} \sum_{j=1}^{m_i} (x_j, c_i) \quad (3.4)$$

Keterangan:

m_i = jumlah data dalam *cluster ke-i*

x = data dalam *cluster*

(x,c) = jarak data ke centroid

x_j = data pada *cluster* tersebut

c_i = centroid *cluster ke-i*

Contoh perhitungan SSW untuk iterasi-1:

$$\text{SSW cluster 1} = \frac{1}{2} \times (1625+0+147347+32820+16873) = 39733,145$$

$$\begin{aligned} \text{SSW cluster 2} &= \frac{1}{8} \times (1015+838+2019+0+2425) \\ &= 1320,33 \end{aligned}$$

Contoh perhitungan SSW untuk iterasi-2:

$$\text{SSW cluster 1} = \frac{1}{2} \times (147347+0) = 73673,5$$

$$\begin{aligned} \text{SSW cluster 2} &= \frac{1}{8} \times \\ &(34445+42776+42599+39742+41761+0+15946+39336) \\ &= 32075,663 \end{aligned}$$

Contoh perhitungan SSW untuk iterasi-3:

$$\text{SSW cluster 1} = \frac{1}{6} \times (8154+177+9779+0+2857+838+3263) = 3581,14$$

$$\text{SSW cluster 2} = \frac{1}{4} \times (130473+15946,301+0) = 48806,534$$

2. Kemudian menghitung SSB (Sum Of Square between cluster) dengan melihat jarak setiap cluster , serta bertujuan untuk menghasilkan nilai separasi. Nilai separasi diperoleh dari perbedaan satu cluster dengan lainnya.

$$SSB_{i,j} = d (ci, cj) \quad (3.5)$$

Keterangan :

C_i = Cluster 1

C_j = Cluster lainnya

$d (ci,j)$ = Jarak antara centroid satu dengan lainnya

Contoh perhitungan SSB pada iterasi-1:

$d(x,y)=$

$$\sqrt{((10,97 - 11,25)^2) + ((9,78 - 0,58)^2) + ((10727,97 - 1786,97)^2)}$$

$$= 8941$$

Contoh perhitungan SSB pada iterasi-2:

$$d(x,y)= \sqrt{((0 - 8471)^2) + ((8,897 - 3,392)^2) + ((158074 - 43547)^2)}$$

$$= 114527$$

Contoh perhitungan SSB pada iterasi-3:

$d(x,y)=$

$$\sqrt{((1433 - 168)^2) + ((16731 - 949)^2) + ((60276131,2 - 8926048,61)^2)}$$

$$= 26552,7$$

3. Setelah itu dilakukan perhitungan rasio yang diperoleh dari perumpamaan

kluster yang satu dengan lain. sehingga banyaknya kohesi lebih kecil, sedangkan hasil separasi memiliki nilai yang besar.

$$R_{i,j} = \frac{SSW_i + SSW_j}{SSB_{i,j}} \quad (3.6)$$

Keterangan:

$R_{i,j}$ = Rasio antar *cluster*

SSW_i = *cluster* 1

SSW_j = *cluster* 2

SSB_{ij} = separasi dari *cluster* 1 dan 2

Contoh perhitungan rasio pada iterasi-1:

$$R = \frac{39733,145 + 1259}{8941} = 4,58$$

Contoh perhitungan rasio pada iterasi-2:

$$R = \frac{73673,5 + 32075,663}{114527} = 0,923$$

Contoh perhitungan rasio pada iterasi-3:

$$R = \frac{3581,14 + 48806,53}{26652,7} = 1,96$$

4. Menghitung DBI (*Davis Bouldin Index*)

Dan tahapan akhir melakukan DBI, dimana jika hasil perhitungan memperoleh nilai mendekati nol, maka yang menghasilkan clustering semakin baik.

$$DBI = \frac{1}{k} \sum \max_{i \neq j} (R_{i,j}) \quad (3.7)$$

Keterangan :

k = kluster yang ada

$R_{i,j}$, =rasio antara kluster i dan j

Max = dicari rasio antar kluster yang terbesar

Contoh perhitungan rasio pada iterasi-2

$$DBI = \frac{1}{2} \times 0,923 = 0,461$$

Contoh perhitungan rasio pada iterasi-3

$$DBI = \frac{1}{2} \times 1,96 = 0,98$$

Dengan menggunakan evaluasi DBI menunjukkan bahwa pada klaster 1 dan 2 menghasilkan efek pengelompokkan dikarenakan menghasilkan nilai kurang dari 1. Sehingga ujicoba yang sudah dilakukan sebelumnya sebanding dengan hasil menggunakan metode elbow.

BAB IV

HASIL DAN PEMBAHASAN

4.1 Hasil Penelitian

Hasil penelitian dalam implementasi *k-means clustering* pada penentuan loyalitas pelanggan di Batubara Mart dilakukan menggunakan data transaksi periode Januari sampai Agustus 2022 sebagai dataset penelitian. Penelitian ini untuk melakukan *clustering* diimplementasikan menggunakan bahasa pemrograman Python, sehingga dihasilkan penentuan loyalitas pelanggan di Batubara Mart berdasarkan hasil implementasi metode *k-means clustering*.

4.1.1 Import Library

Library merupakan kumpulan modul yang berisi kumpulan kode yang digunakan berulang kali dalam program yang berbeda. Dalam penelitian ini untuk mengimplementasikan metode *k-means clustering* dengan memanggil *library* numpy, pandas, matplotlib, seaborn, datetime yang digunakan untuk dataframe dan visualisasi. Kemudian, untuk melakukan *clustering* menggunakan *library* sklearn, MinMaxScaler dan KMeans. Berikut merupakan *source code library* yang digunakan dapat dilihat pada Gambar 4.1.

```
# Library untuk dataframe dan visualisasi
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import datetime as dt

# Import library untuk Clustering
import sklearn
from sklearn.preprocessing import MinMaxScaler
from sklearn.cluster import KMeans
```

Gambar 4. 1 Source Code Import Library

Setelah melakukan *import library*, *data set* yang telah disiapkan dimasukan dengan cara memilih dataset tersebut. Berikut merupakan *source code* untuk melakukan *load* dataset dapat dilihat pada Gambar 4.2.

```
# Load dataset
df1 = pd.read_excel('Documents/Tokomiring - Bulan 01.xlsx', skiprows=1, skipfooter=1)
df2 = pd.read_excel('Documents/Tokomiring - Bulan 02.xlsx', skiprows=1, skipfooter=1)
df3 = pd.read_excel('Documents/Tokomiring - Bulan 03.xlsx', skiprows=1, skipfooter=1)
df4 = pd.read_excel('Documents/Tokomiring - Bulan 04.xlsx', skiprows=1, skipfooter=1)
df5 = pd.read_excel('Documents/Tokomiring - Bulan 05.xlsx', skiprows=1, skipfooter=1)
df6 = pd.read_excel('Documents/Tokomiring - Bulan 06.xlsx', skiprows=1, skipfooter=1)
df7 = pd.read_excel('Documents/Tokomiring - Bulan 07.xlsx', skiprows=1, skipfooter=1)
df8 = pd.read_excel('Documents/Tokomiring - Bulan 08.xlsx', skiprows=1, skipfooter=1)
df = pd.concat([df1, df2, df3, df4, df5, df6, df7, df8]).reset_index(drop=True)
df
```

Gambar 4. 2 Source Code Load Dataset

Maka dataset yang telah dimasukan akan ditampilkan yang terdiri dari dataset transaksi bulan Januari sampai Agustus dengan atribut yang ada dalam data tersebut yaitu *Date*, *Invoice No*, *Customer name*, *Location*, *Payment status*, *Payment method*, *Total amount*, *Total paid*, *Net Amount before shipping*, *Total item*, dan *Added By*. Dataset berjumlah 8000 baris. Hasil *load* dataset dapat dilihat pada Gambar 4.3.

	Date	Invoice No.	Customer name	Location	Payment status	Payment Method	Total amount	Total paid	Net Amount Before Shipping	Total Items	Added By
0	31-01-2022 20:58	119561	leni marlina	Toko Miring	Paid	Cash	138133	138133.0	138133	9	Miss SRI RAHAYU
1	31-01-2022 20:58	119560	leni marlina	Toko Miring	Paid	Cash	107439	107439.0	107439	10	Miss SRI RAHAYU
2	31-01-2022 20:58	119559	leni marlina	Toko Miring	Paid	Cash	103537	103537.0	103537	1	Miss SRI RAHAYU
3	31-01-2022 20:58	119558	UTAMI GUSTI WIJAYA	Toko Miring	Paid	Card	100876	100876.0	100876	13	MISS SOFIA
4	31-01-2022 20:58	119557	Walk in Customer	Toko Miring	Paid	Cash	108433	108433.0	108433	2	MISS SUCI ANAIA
...	...	...	...	...	...	...	...	...	...	...	...
7995	26-08-2022 20:40	192713	NADIA	Toko Miring	Paid	Cash	19142	19142.0	19142	7	MISS META PUTRI
7996	26-08-2022 20:40	192712	Walk in Customer	Toko Miring	Paid	Cash	7235	7235.0	7235	1	MISS SUCI ANAIA
7997	26-08-2022 20:40	192711	PUTRI ANGGRAINI	Toko Miring	Paid	Card	115960	115960.0	115960	14	Miss SRI RAHAYU
7998	26-08-2022 20:37	192710	Walk in Customer	Toko Miring	Paid	Cash	1885	1885.0	1885	1	MISS META PUTRI
7999	26-08-2022 20:37	192709	Walk in Customer	Toko Miring	Paid	Cash	19621	19621.0	19621	3	MISS META PUTRI

Gambar 4. 3 Load Dataset

4.1.2 *Preprocessing*

Preprocessing ialah proses pengumpulan data mentah yang akan menghasilkan data sesungguhnya. Pada pengamatan ini dataset yang telah dikumpulkan berupa data transaksi dari bulan Januari sampai Agustus 2022 kemudian dilakukan pemrosesan pembersihan dari kata yang tidak sesuai, sehingga memudahkan proses penerapan sistem.

4.1.2.1 *Data Cleansing*

Data *cleansing* merupakan tahap untuk menyiapkan data dengan menghapus atau memodifikais data yang salah, tidak relevan atau adanya duplikasi data. Dalam penelitian ini data *cleansing* dilakukan dengan melihat tingkatan kelengkapan data yang hilang sesuai dataset yang digunakan. Selanjutnya terdapat *source code* yang digunakan untuk data *cleansing*, seperti Gambar 4.4.

```
# Melihat persentase missing value yang terdapat pada dataset
df_null = round(100*(df.isnull().sum())/len(df), 2)
df_null
```

Gambar 4. 4 *Source Code Data Cleansing*

Dari hasil *cleansing* yang ada pada Gambar 4.5 dijelaskan bahwa tidak terdapat atribut yang memiliki *missing value*, maka tidak perlu dilakukan data *cleansing*. Namun, jika terdapat atribut yang memiliki *missing value* menggunakan fungsi `df=df.dropna ()` yang digunakan untuk menghapus data tersebut.

Date	0.00
Invoice No.	0.00
Customer name	0.00
Location	0.00
Payment status	0.00
Payment Method	0.35
Total amount	0.00
Total paid	0.35
Net Amount Before Shipping	0.00
Total Items	0.00
Added By	0.00
dtype: float64	

Gambar 4. 5 Hasil Data *Cleansing*

4.1.2.2 Data *Preparation*

Data *preparation* merupakan tahapan untuk penggabungan pada data mentah menjadi data yang siap untuk digunakan dalam visualisasi. Dalam data *preparation* terdiri dari *recency*, *frequency* dan *monetary*.

1. Membuat Atribut *Recency*

Proses data *preparation* yang pertama yaitu membuat atribut baru *recency* untuk menghitung jumlah hari sejak pembelian terakhir. Untuk menghitung *recency*, perhitungannya adalah jarak antara tanggal data pelanggan melakukan transaksi dengan tanggal terakhir data yang terdata. Contohnya Nadia melakukan pembelian pada tanggal 26 agustus 2022 sedangkan data tersebut berakhir pada tanggal 30 agustus 2022, maka *recency* yang didapatkan adalah 3 hari atau 3. Pada proses data *preparation* dilakukan dengan membuat fungsi untuk membuat atribut baru yaitu *recency* berdasarkan atribut “Date”. Kemudian menampilkan pembayaran terakhir yang dilakukan oleh pelanggan Berikutnya

menghitung selisih antara max_date dengan date. Berikut merupakan *source code* data preparation yang ada di gambar 4.6.

```
# Membuat atribut baru : Recency
df['Date'] = pd.to_datetime(df['Date'],format='%d-%m-%Y')

# Menampilkan tanggal maksimal untuk mengetahui tanggal terakhir kali melakukan transaksi
max_date = max(df['Date'])
max_date

Timestamp('2022-07-30 00:00:00')

# Menghitung selisih antara max_date dengan Date
df['Recency'] = max_date - df['Date']
df
```

Gambar 4. 6 Source Code Recency

	Date	Invoice No.	Customer name	Location	Payment status	Payment Method	Total amount	Total paid	Net Amount Before Shipping	Total Items	Added By	Recency
0	2022-01-31 20:58:00	119561	Ieni marlina	Toko Miring	Paid	Cash	138133	138133.0	138133	9	Miss SRI RAHAYU	210 days 12:43:00
1	2022-01-31 20:58:00	119560	Ieni marlina	Toko Miring	Paid	Cash	107439	107439.0	107439	10	Miss SRI RAHAYU	210 days 12:43:00
2	2022-01-31 20:58:00	119559	Ieni marlina	Toko Miring	Paid	Cash	103537	103537.0	103537	1	Miss SRI RAHAYU	210 days 12:43:00
3	2022-01-31 20:58:00	119558	UTAMI GUSTI WIJAYA	Toko Miring	Paid	Card	100876	100876.0	100876	13	MISS SOFIA	210 days 12:43:00
4	2022-01-31 20:58:00	119557	Walk in Customer	Toko Miring	Paid	Cash	108433	108433.0	108433	2	MISS SUCI ANAIA	210 days 12:43:00
...	...	...	...	...	...	...	...	...	...	...	...	...
7995	2022-08-26 20:40:00	192713	NADIA	Toko Miring	Paid	Cash	19142	19142.0	19142	7	MISS META PUTRI	3 days 13:01:00
7996	2022-08-26 20:40:00	192712	Walk in Customer	Toko Miring	Paid	Cash	7235	7235.0	7235	1	MISS SUCI ANAIA	3 days 13:01:00
7997	2022-08-26 20:40:00	192711	PUTRI ANGGRAINI	Toko Miring	Paid	Card	115960	115960.0	115960	14	Miss SRI RAHAYU	3 days 13:01:00
7998	2022-08-26 20:37:00	192710	Walk in Customer	Toko Miring	Paid	Cash	1885	1885.0	1885	1	MISS META PUTRI	3 days 13:04:00
7999	2022-08-26 20:37:00	192709	Walk in Customer	Toko Miring	Paid	Cash	19621	19621.0	19621	3	MISS META PUTRI	3 days 13:04:00

Gambar 4. 7 Hasil Recency

Selanjutnya, akan dihitung waktu terakhir transaksi yang dilakukan pelanggan dengan mendeskripsikan berkaitan dengan pembelian terakhir. Berikut *source code* transaksi terakhir yang dilakukan pelanggan dapat dilihat pada Gambar 4.8. Hasil dari perhitungan waktu terakhir transaksi pelanggan seperti di gambar 4.9.

```
rfm_p = df.groupby('Supplier')['Recency'].min()
rfm_p = rfm_p.reset_index()
rfm_p
```

Gambar 4. 8 Source Code Menghitung Last Transaction Date

	Customer name	Recency
0	A.RIVAI	30 days 01:02:00
1	AAK	121 days 13:26:00
2	AAN	1 days 13:56:00
3	ABANG	212 days 13:56:00
4	ABAS	91 days 15:45:00
...	...	...
1582	yoce	183 days 15:47:00
1583	yuli	60 days 14:08:00
1584	yulia alpina	29 days 18:42:00
1585	yullya	92 days 19:33:00
1586	yusrizal	1 days 21:51:00

Gambar 4. 9 Hasil Menghitung Last Transaction Date

Sehingga pada atribut *Recency* hanya diperlukan data jumlah hari, maka akan dilakukan ekstraksi jumlah hari. Berikut *source code* ekstraksi jumlah hari dapat dilihat pada Gambar 4.10. Dan hasilnya dapat dilihat pada Gambar 4.11.

```
rfm_p['Recency'] = rfm_p['Recency'].dt.days
rfm_p
```

Gambar 4. 10 Source Code Extract Jumlah Hari

	Customer name	Recency
0	A.RIVAI	30
1	AAK	121
2	AAN	1
3	ABANG	212
4	ABAS	91
...	...	...
1582	yoce	183
1583	yuli	60
1584	yulia alpina	29
1585	yullya	92
1586	yusrizal	1

Gambar 4. 11 Hasil *Extract* Jumlah Hari

2. Membuat Atribut *Frequency*

Atribut *Frequency* digunakan untuk mengetahui jumlah proses transaksi yang dilakukan oleh pelanggan. Untuk membuat atribut *Frequency* dengan detail dan perhitungan menggunakan *source code* yang dapat dilihat pada Gambar 4.12. Maka didapatkan hasil *Frequency* yang dapat dilihat pada Gambar 4.13.

```
: # Membuat atribut baru : Frequency
rfm_f = df.groupby('Customer name')['Invoice No.'].count()
rfm_f = rfm_f.reset_index()
rfm_f.columns = ['Customer name', 'Frequency']
rfm_f
```

Gambar 4. 12 *Source Code Frequency*

	Customer name	Frequency
0	A.RIVAI	1
1	AAK	4
2	AAN	7
3	ABANG	1
4	ABAS	1
...	...	...
1582	yoce	1
1583	yuli	1
1584	yulia alpina	1
1585	yullya	1
1586	yusrizal	3

Gambar 4. 13 Hasil *Frequency*

Kemudian dilakukan *merging* yaitu menggabungkan dua *dataframe* menggunakan *source code* yang dapat dilihat pada Gambar 4.14.

```
# Menggabungkan (merging) dua dataframe
rfm = pd.merge(rfm_p, rfm_f, on='Customer name', how='inner')
rfm
```

Gambar 4. 14 *Source Code Merging Dataframe*

	Customer name	Recency	Frequency
0	A.RIVAI	30	1
1	AAK	121	4
2	AAN	1	7
3	ABANG	212	1
4	ABAS	91	1
...	...	...	...
1582	yoce	183	1
1583	yuli	60	1
1584	yulia alpina	29	1
1585	yullya	92	1
1586	yusrizal	1	3

Gambar 4. 15 Hasil *Merging Dataframe*

3. Membuat Atribut *Monetary*

Atribut *Monetary* merupakan atribut yang digunakan untuk menghitung

jumlah total transaksi (kontribusi pendapatan). *Source code* untuk membuat atribut *Monetary* seperti di gambar 4.16 serta hasilnya tercantum pada gambar 4.17.

```
# Membuat atribut baru : Monetary
df['Monetary'] = df['Total amount']
rfm_m = df.groupby('Customer name')['Monetary'].sum()
rfm_m = rfm_m.reset_index()
rfm_m
```

Gambar 4. 16 *Source Code Monetary*

	Customer name	Monetary
0	A.RIVAI	26190
1	AAK	114264
2	AAN	760698
3	ABANG	27966
4	ABAS	12352
...	...	...
1582	yoce	20329
1583	yuli	13580
1584	yulia alpina	45193
1585	yullya	23639
1586	yusrizal	335217

Gambar 4. 17 Hasil *Monetary*

Kemudian, melakukan *merging* atribut *Recency*, *Frequency* dan *Monetary* terhadap dataframe dengan *source code* yang digunakan pada Gambar 4.18. Hasil *merging* atribut *Recency*, *Frequency* dan *Monetary* dapat dilihat pada Gambar 4.19.

```
# Menggabungkan dataframe
rfm = pd.merge(rfm, rfm_m, on='Customer name', how='inner')
rfm
```

Gambar 4. 18 Source Code Merging Pada Dataframe

	Customer name	Recency	Frequency	Monetary
0	A.RIVAI	30	1	26190
1	AAK	121	4	114264
2	AAN	1	7	760698
3	ABANG	212	1	27966
4	ABAS	91	1	12352
...	...	...	...	...
1582	yoce	183	1	20329
1583	yuli	60	1	13580
1584	yulia alpina	29	1	45193
1585	yullya	92	1	23639
1586	yusrizal	1	3	335217

Gambar 4. 19 Hasil Merging Atribut Recency, Frequency dan Monetary

Selanjutnya, dilakukan *rescaling* atribut yang bertujuan untuk menormalisasi atribut agar memiliki skala yang sebanding. Dalam penelitian ini menggunakan MinMaxScaler. MinMaxScaler digunakan untuk mengubah data berada pada rentang 0 sampai 1. Hasilnya terdapat 1587 baris dan 3 kolom. Berikut merupakan *source code rescaling* yang dapat dilihat pada Gambar 4.20.

```
# Rescaling Atribute
# Normalisasi atribut agar memiliki skala yg sebanding
rfm_df = rfm[['Recency', 'Frequency', 'Monetary']]
# Instantiate
scaler = MinMaxScaler()
# fit_transform
rfm_df_scaled = scaler.fit_transform(rfm_df)
rfm_df_scaled.shape
```

Gambar 4. 20 Source Code Rescaling

Hasil dari *rescaling* ditampilkan dalam dataframe dengan *columns* *Recency*, *Frequency* dan *Monetary* menggunakan *source code* pada Gambar 4.21.

```
rfm_df_scaled = pd.DataFrame(rfm_df_scaled)
rfm_df_scaled.columns = ['Recency', 'Frequency', 'Monetary']
rfm_df_scaled
```

Gambar 4. 21 Source Code Rescaling Pada Dataframe

Sehingga didapatkan hasil *recaling* yang dapat dilihat pada Gambar 4.22 berupa tabel yang berisi nilai atribut *Recency*, *Frequency* dan *Monetary* yang berjumlah 1587 baris dan 3 kolom.

	Recency	Frequency	Monetary
0	0.141509	0.00000	0.000219
1	0.570755	0.00114	0.001045
2	0.004717	0.00228	0.007109
3	1.000000	0.00000	0.000236
4	0.429245	0.00000	0.000090
...	...	...	...
1582	0.863208	0.00000	0.000164
1583	0.283019	0.00000	0.000101
1584	0.136792	0.00000	0.000398
1585	0.433962	0.00000	0.000195
1586	0.004717	0.00076	0.003118

Gambar 4. 22 Hasil Dataframe Rescaling

4.1.3 Model Building

1. K-Means Clustering dengan nilai $k=2$

Pada tahap *k-means clustering* dilakukan penentuan nilai k . Dalam penelitian ini menggunakan nilai $k=2$. Kemudian hasilnya ditampilkan menggunakan fungsi `print`. Berikut merupakan *source code* penentuan nilai $k=2$ yang dapat dilihat pada Gambar 4.23. Dengan penentuan nilai $k=2$ dihasilkan bahwa iterasi yang akan dilakukan sebanyak 3 kali yang dapat dilihat pada Gambar 4.24.

```
from math import sqrt, floor
import numpy as np

# fungsi naive_sharding dengan parameter
def naive_sharding(ds, k):
    """
    Create cluster centroids using deterministic naive sharding algorithm.

    Parameters
    -----
    ds : numpy array
        The dataset to be used for centroid initialization.
    k : int
        The desired number of clusters for which centroids are required.
    Returns
    -----
    centroids : numpy array
        Collection of k centroids as a numpy array.
    """

    def _get_mean(sums, step):
        """Vectorizable ufunc for getting means of summed shard columns."""
        return sums/step

    n = np.shape(ds)[1]
    m = np.shape(ds)[0]
    centroids = np.zeros((k, n))

    composite = np.mat(np.sum(ds, axis=1))
    ds = np.append(composite.T, ds, axis=1)
    ds.sort(axis=0)

    step = floor(m/k)
    vfunc = np.vectorize(_get_mean)

    for j in range(k):
        if j == k-1:
            centroids[j:] = vfunc(np.sum(ds[j*step:,1:], axis=0), step)
        else:
            centroids[j:] = vfunc(np.sum(ds[j*step:(j+1)*step,1:], axis=0), step)

    return centroids #return data
```

```
naive_centroids = naive_sharding(rfm_df_scaled.to_numpy(), 2) #menjalankan fungsi
kmeans = KMeans(n_clusters=2, init=naive_centroids, n_init=1).fit(rfm_df_scaled)
centroid = kmeans.cluster_centers_
print(' First centroid using naive sharding:')
print(naive_centroids)
print('-----')
print('Current center (Centroid):')
print(kmeans.cluster_centers_)
print('Diperoleh pada iterasi ke ', kmeans.n_iter_)
```

Gambar 4. 23 Source Code K-Means Clustering Dengan Nilai k=2

```
First centroid using naive sharding:
[[7.34730781e-02 8.38453489e-05 4.93485919e-04]
 [6.04570654e-01 2.97531210e-03 5.10671970e-03]]
-----
Current center (Centroid):
[[1.12540647e-01 2.22513417e-03 3.68538698e-03]
 [6.93587043e-01 4.36499739e-04 1.40748070e-03]]
Diperoleh pada iterasi ke 2
```

Gambar 4. 24 Hasil Penentuan Nilai k=2

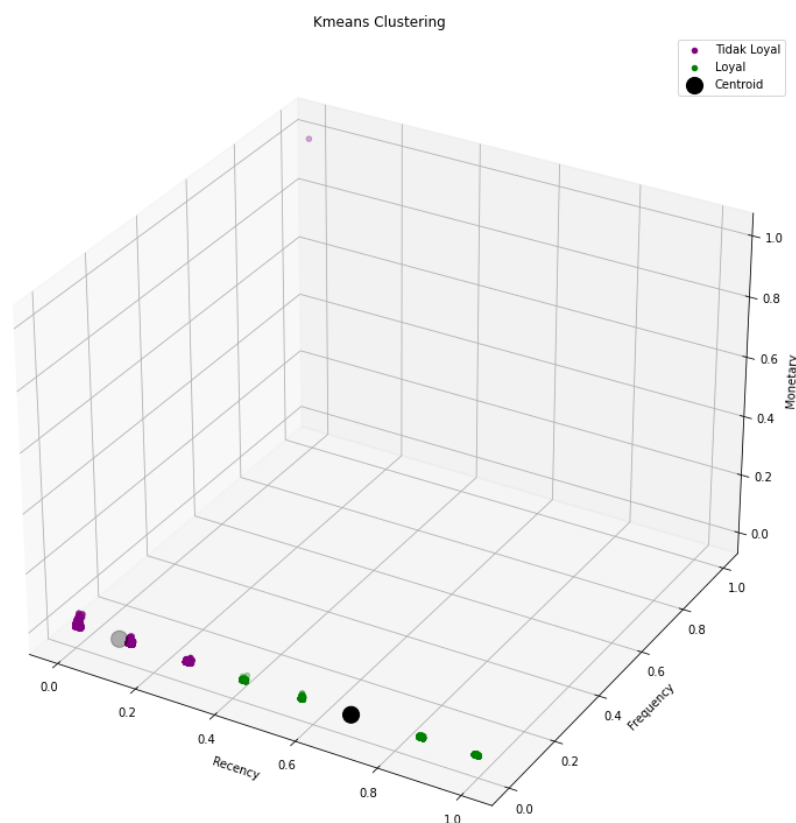
Selanjutnya, ditampilkan hasil dari k-means clustering dengan nilai k=2 untuk mendapatkan *cluster* dengan x-label yaitu atribut *Recency*, y-label yaitu *Frequency* dan z-label yaitu *Monetary*. Kemudian ditampilkan menggunakan perintah `plt.show()`. *Source code* hasil k-means clustering dapat dilihat pada Gambar 4.25.

```
result = rfm_df_scaled.copy()
result['Cluster'] = kmeans.labels_
result
```

```
data1 = result[result.Cluster==0]
data2 = result[result.Cluster==1]
fig = plt.figure(figsize=(20,13))
kplot = plt.axes(projection='3d')
# Data for three-dimensional scattered points
kplot.scatter3D(data1.Recency, data1.Frequency, data1.Monetary, c='purple', label = 'Tidak Loyal')
kplot.scatter3D(data2.Recency, data2.Frequency, data2.Monetary, c='green', label = 'Loyal')
kplot.scatter(centroid[:,0], centroid[:,1], centroid[:,2], c='black', label='Centroid', s=200)
kplot.set_xlabel('Recency')
kplot.set_ylabel('Frequency')
kplot.set_zlabel('Monetary')
plt.legend()
plt.title("Kmeans Clustering")
plt.show()
```

Gambar 4. 25 Source Code Visualisasi Nilai k=2

Dari Gambar 4.26 merupakan hasil *clustering* kegiatan transaksi dieksekusi memakai algoritma *k-means clustering*. Label Cluster 1 disimbolkan dengan tanda titik purple, cluster 2 disimbolkan dengan tanda titik green dan *Centroid* disimbolkan dengan titik *black*. Pada *centroid* menghasilkan nilai 0.69. Titik centroid terletak pada nilai *Recency* 0.69, nilai *Frequency* 0.14 dan nilai *Monetary* 0.0. Cluster terbanyak pada *k-means clustering* yaitu cluster 1 yang datanya sedikit menyebar. Sedangkan cluster 2 tidak terlalu banyak dan menyebar.



Gambar 4. 26 Hasil K-Means Clustering

2. Manualisasi perhitungan setiap iterasi

Pada tahap ini dilakukan perhitungan pada setiap iterasi dengan

menghitung jarak antara dua pusat cluster ke titik data dan menemukan *cluster* mana yang paling dekat dengan titik data. *Source code* manualisasi perhitungan setiap iterasi dapat dilihat pada Gambar 4.27.

```
from scipy.spatial.distance import euclidean
def assign_cluster(data, centers):
    cluster_labels = []
    min_distance = []
    for point in data:
        # Menghitung jarak antara dua pusat cluster ke titik data
        distances = [euclidean(point, center) for center in centers]
        # Menemukan cluster mana yang paling dekat dengan titik data
        cluster_labels.append(distances.index(min(distances)))
        min_distance.append(min(distances))
    return cluster_labels, min_distance
```

Gambar 4. 27 Source Code Iterasi

Iterasi-1

Kemudian, ditampilkan hasil centroid pada iterasi 1 menggunakan *source code* pada Gambar 4.28. Serta dihasilkan nilai *Recency*, *Frequency* dan *Monetary* pada *centroid* iterasi-1.

```
print('Centroid iterasi-1:')
centroid1 = pd.DataFrame(naive_centroids)
centroid1
```

Centroid iterasi-1:

	0	1	2
0	0.073473	0.000084	0.000493
1	0.604571	0.002975	0.005107

Gambar 4. 28 Source Code Iterasi-1

Selanjutnya, dilakukan penentuan *cluster* dan *distance* pada dataframe dengan menghasilkan nilai *Recency*, *Frequency*, *Monetary*, Jarak terdekat dan *Cluster*. *Source code* yang digunakan untuk penentuan *cluster* dan *distance* pada iterasi-1 dapat dilihat pada Gambar 4.29. Hasil

dari *cluster* pada iterasi-1 dapat dilihat pada Gambar 4.30.

```
cluster1 = assign_cluster(rfm_df_scaled.values, centroid1.values)[0]
distance1 = assign_cluster(rfm_df_scaled.values, centroid1.values)[1]
result1 = pd.DataFrame({'Recency':rfm_df_scaled['Recency'], 'Frequency':rfm_df_scaled['Frequency'], 'Monetary':
result1
```

Gambar 4. 29 Source Code Hasil Cluster Pada Iterasi-1

	Recency	Frequency	Monetary	Jarak Terdekat	Cluster
0	0.141509	0.00000	0.000219	0.068037	0
1	0.570755	0.00114	0.001045	0.034108	1
2	0.004717	0.00228	0.007109	0.069109	0
3	1.000000	0.00000	0.000236	0.395471	1
4	0.429245	0.00000	0.000090	0.175422	1
...	...	...	...	...	...
1582	0.863208	0.00000	0.000164	0.258701	1
1583	0.283019	0.00000	0.000101	0.209546	0
1584	0.136792	0.00000	0.000398	0.063320	0
1585	0.433962	0.00000	0.000195	0.170705	1
1586	0.004717	0.00076	0.003118	0.068809	0

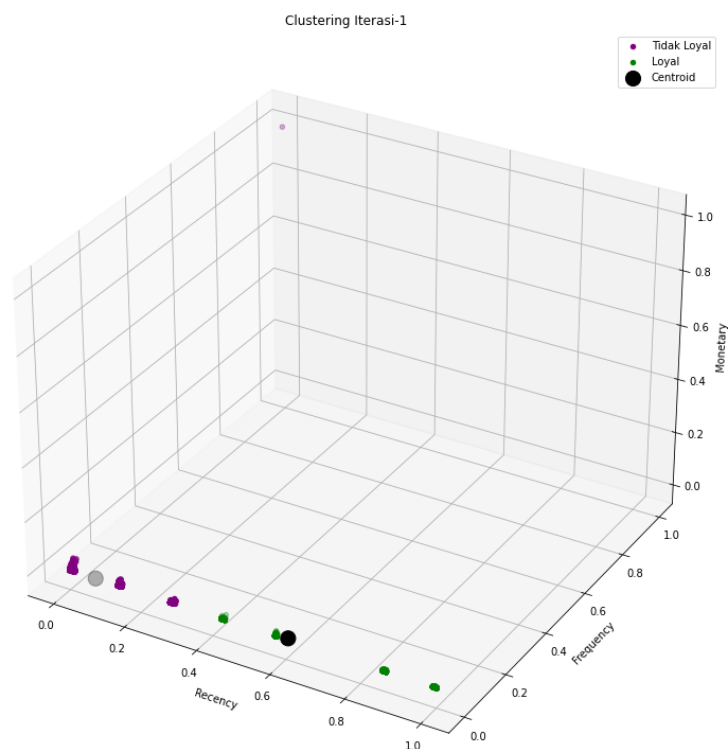
Gambar 4. 30 Hasil Cluster Pada Iterasi-1

Dari hasil *cluster* pada iterasi-1 dilakukan visualisasi dengan menghasilkan data menggunakan *source code* pada Gambar 4.31.

```
data1 = result1[result1.Cluster==0]
data2 = result1[result1.Cluster==1]
fig = plt.figure(figsize=(20,13))
kplot = plt.axes(projection='3d')
# Data for three-dimensional scattered points
kplot.scatter3D(data1.Recency, data1.Frequency, data1.Monetary, c='purple', label = 'Tidak Loyal')
kplot.scatter3D(data2.Recency, data2.Frequency, data2.Monetary, c = 'green', label = 'Loyal')
kplot.scatter(centroid1[0], centroid1[1], centroid1[2], c='black', label='Centroid', s=200)
kplot.set_xlabel('Recency')
kplot.set_ylabel('Frequency')
kplot.set_zlabel('Monetary')
plt.legend()
plt.title("Clustering Iterasi-1")
plt.show()
```

Gambar 4. 31 Source Code Hasil Visualisasi Cluster Pada Iterasi-1

Hasil dari titik tersebar tiga dimensi pada cluster iterasi-1 dapat dilihat pada Gambar 4.32. Titik centroid terletak pada nilai *Recency* 0,60, nilai *Frequency* 0.01 dan nilai *Monetary* 0.01. Cluster terbanyak pada iterasi-1 yaitu cluster 2 yang datanya menyebar. Sedangkan cluster 1 tidak terlalu banyak dan dekat dengan titik *centroid*.



Gambar 4. 32 Hasil Visualisasi *Cluster* Pada Iterasi-1

Iterasi-2

Pada iterasi-2 untuk menentukan titik centroid dilakukan menggunakan *source code* pada Gambar 4.33. hasilnya terdapat 2 data pada centroid iterasi-2 yaitu cluster 0 dan cluster 1.

```
print('Centroid iterasi-2:')
centroid2 = rfm_df_scaled.groupby(cluster1).mean()
centroid2
```

Centroid iterasi-2:

	Recency	Frequency	Monetary
0	0.112541	0.002225	0.003685
1	0.693587	0.000436	0.001407

Gambar 4. 33 Source Code Iterasi-2

Selanjutnya, dilakukan penentuan *cluster* dan *distance* pada dataframe dengan menghasilkan nilai *Recency*, *Frequency*, *Monetary*, Jarak terdekat dan *Cluster*. Source code yang digunakan untuk penentuan *cluster* dan *distance* pada iterasi-2 dapat dilihat pada Gambar 4.34. Hasil dari *cluster* pada iterasi-2 dapat dilihat pada Gambar 4.35.

```
cluster2 = assign_cluster(rfm_df_scaled.values, centroid2.values)[0]
distance2 = assign_cluster(rfm_df_scaled.values, centroid2.values)[1]
result2 = pd.DataFrame({'Recency':rfm_df_scaled['Recency'], 'Frequency':rfm_df_scaled['Frequency'], 'Monetary': rfm_df_scaled['Monetary'], 'Jarak Terdekat': distance2, 'Cluster': cluster2})
result2
```

Gambar 4. 34 Source Code Hasil Cluster Pada Iterasi-2

	Recency	Frequency	Monetary	Jarak Terdekat	Cluster
0	0.141509	0.00000	0.000219	0.029260	0
1	0.570755	0.00114	0.001045	0.122835	1
2	0.004717	0.00228	0.007109	0.107878	0
3	1.000000	0.00000	0.000236	0.306416	1
4	0.429245	0.00000	0.000090	0.264345	1
...	...	...	...	...	...
1582	0.863208	0.00000	0.000164	0.169626	1
1583	0.283019	0.00000	0.000101	0.170530	0
1584	0.136792	0.00000	0.000398	0.024575	0
1585	0.433962	0.00000	0.000195	0.259628	1
1586	0.004717	0.00076	0.003118	0.107835	0

Gambar 4. 35 Hasil *Cluster* Pada Iterasi-2

Dari hasil *cluster* pada iterasi-2 dilakukan visualisasi dengan menghasilkan data untuk titik tersebar tiga dimensi menggunakan *source code* pada Gambar 4.36.

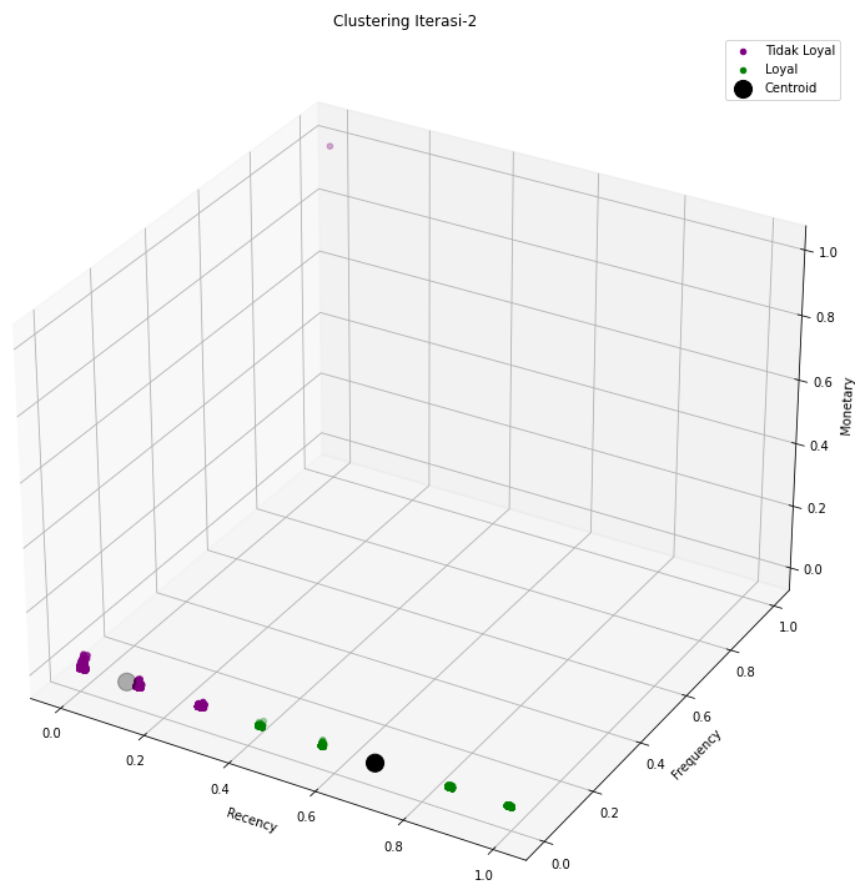
```

data1 = result2[result2.Cluster==0]
data2 = result2[result2.Cluster==1]
fig = plt.figure(figsize=(20,13))
kplot = plt.axes(projection='3d')
# Data for three-dimensional scattered points
kplot.scatter3D(data1.Recency, data1.Frequency, data1.Monetary, c='purple', label = 'Cluster 1')
kplot.scatter3D(data2.Recency, data2.Frequency, data2.Monetary, c = 'green', label = 'Cluster 2')
kplot.scatter(centroid2.Recency, centroid2.Frequency, centroid2.Monetary, c='black', label='Centroid', s=200)
kplot.set_xlabel('Recency')
kplot.set_ylabel('Frequency')
kplot.set_zlabel('Monetary')
plt.legend()
plt.title("Clustering Iterasi-2")
plt.show()

```

Gambar 4. 36 *Source Code* Hasil Visualisasi *Cluster* Pada Iterasi-2

Hasil dari titik tersebar tiga dimensi pada cluster iterasi-2 dapat dilihat pada Gambar 4.37. Terdapat titik centroid yang disimbolkan dengan titik hitam, cluster 1 dengan simbol titik ungu dan cluster 2 dengan simbol titik hijau. Titik centroid terletak pada nilai *Recency* 0.69, nilai *Frequency* 0.01 dan nilai *Monetary* 0.01. Cluster terbanyak pada iterasi-2 yaitu cluster 2 yang datanya menyebar. Sedangkan cluster 1 tidak terlalu banyak dan dekat dengan titik *centroid*.



Gambar 4. 37 Hasil Visualisasi *Cluster* Pada Iterasi-2

4.1.4 Evaluation

Pada tahap ini dilakukan evaluasi metode *k-means clustering* untuk menguji cluster yang dihasilkan menggunakan Elbow dan *Davies Bouldin Index* (DBI). Berikut merupakan *source code* untuk melakukan evaluasi menggunakan metode Elbow dan DBI seperti yang ada di Gambar 4.38.

```
from sklearn.metrics import davies_bouldin_score
sse,db = {}, {}
cluster, k_value = [], []
for k in range(2, 26):
    # seed of 10 for reproducibility.
    Kmeans = KMeans(n_clusters=k, n_init=1, random_state=0).fit(rfm_df_scaled)
    clusters = Kmeans.labels_
    cluster.append(clusters)
    k_value.append(k)
    sse[k] = Kmeans.inertia_ # Inertia: Sum of distances of samples to their closest cluster center
    db[k] = davies_bouldin_score(rfm_df_scaled, clusters)
    print('-----')
    print('n_cluster          : ', k)
    print('Elbow score            : ', sse[k])
    print('Davies Bouldin Index score: ', db[k])
```

Gambar 4. 38 Source Code Evaluation

Maka dihasilkan *cluster* dari metode *k-means clustering*, terdapat jumlah cluster 2 sampai 13 yang memiliki nilai pada metode Elbow dan *Davies Bouldin Index* (DBI) yang dapat dilihat pada Gambar 4.39.

```
-----
n_cluster          : 2
Elbow score        : 44.08415561910874
Davies Bouldin Index score: 0.5275732024790732
-----
n_cluster          : 3
Elbow score        : 16.680583405543466
Davies Bouldin Index score: 0.3973613754470821
-----
n_cluster          : 4
Elbow score        : 7.746001066833155
Davies Bouldin Index score: 0.4176201989724663
-----
n_cluster          : 5
Elbow score        : 5.1101813566756
Davies Bouldin Index score: 0.3003326543803647
-----
```

```

n_cluster          : 6
Elbow score         : 3.2515485592983016
Davies Bouldin Index score: 0.14180633557011973
-----
n_cluster          : 7
Elbow score         : 2.0264162392173453
Davies Bouldin Index score: 0.06696680196854574
-----
n_cluster          : 8
Elbow score         : 0.041232606135466185
Davies Bouldin Index score: 0.05055512396210433
-----
n_cluster          : 9
Elbow score         : 0.03511606172264818
Davies Bouldin Index score: 0.3011589090866469
-----
n_cluster          : 10
Elbow score         : 0.033550966775578825
Davies Bouldin Index score: 0.36727132301054527
-----
n_cluster          : 11
Elbow score         : 0.03028610580491179
Davies Bouldin Index score: 0.48401017188989964
-----
n_cluster          : 12
Elbow score         : 0.027631769995889836
Davies Bouldin Index score: 0.5453469208435404
-----
n_cluster          : 13
Elbow score         : 0.021587835399057342
Davies Bouldin Index score: 0.5021942224513333
-----

```

Gambar 4. 39 Hasil *Evaluation*

1. Pengujian ELBOW

Pengujian Elbow merupakan sebagai penentu jumlah kluster terbaik/paling optimum. *Source code* untuk pengujian Elbow dapat dilihat pada Gambar 4.40.

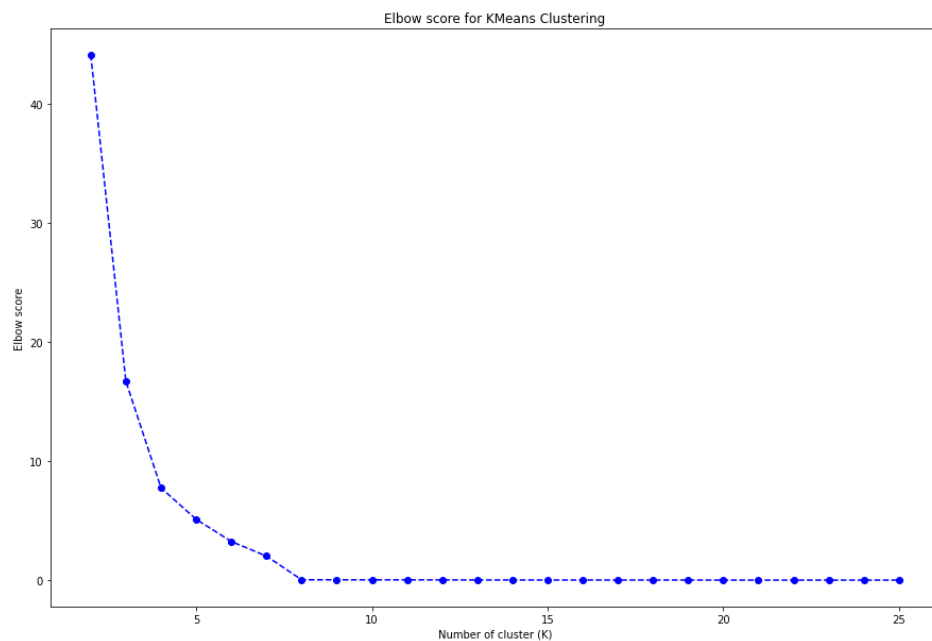
```

plt.figure(figsize=(15,10))
plt.plot(list(sse.keys()), list(sse.values()), linestyle='--', marker='o', color='b');
plt.xlabel('Number of cluster (K)');
plt.ylabel('Elbow score');
plt.title('Elbow score for KMeans Clustering');

```

Gambar 4. 40 *Source Code Evaluation Metode Elbow*

Dari Gambar 4.41 merupakan hasil *evaluation* metode Elbow didapatkan hasil cluster terbaik yaitu pada cluster 4 dengan nilai Elbow 7,75. Karena *cluster* terbaik merupakan *cluster* yang membentuk kurva.



Gambar 4. 41 Hasil *Evaluation* Metode Elbow

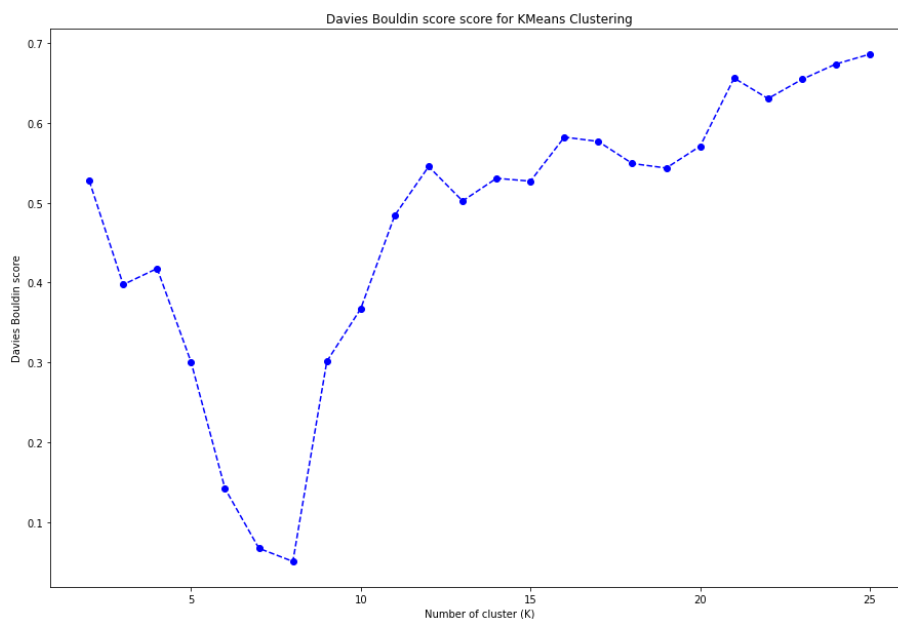
2. Pengujian *Davies Bouldin Index* (DBI)

Pengujian DBI dilakukan untuk mengukur evaluasi cluster yang didasarkan pada nilai separasi dan kohesi. *Source code* yang digunakan dapat dilihat pada Gambar 4.42.

```
plt.figure(figsize=(15,10))
plt.plot(list(db.keys()), list(db.values()), linestyle='--', marker='o', color='b');
plt.xlabel('Number of cluster (K)');
plt.ylabel('Davies Bouldin score');
plt.title('Davies Bouldin score score for KMeans Clustering');
```

Gambar 4. 42 *Source Code Evaluation* Metode DBI

Hasil dari pengujian metode DBI dapat dilihat pada Gambar 4.43. dari gambar tersebut dapat dijelaskan bahwa terdapat *cluster* yang optimal terdapat pada cluster 8 dengan nilai DBI 0.05, karena semakin kecil nilai DBI maka *cluster* semakin optimal.



Gambar 4. 43 Hasil *Evaluation* Metode DBI

4.2 Pembahasan

Dari hasil penelitian yang telah dilakukan pada implementasi *K-Means Clustering* untuk menentukan loyalitas pelanggan di Batubara Mart didapatkan hasil *cluster*. Dalam penelitian ini dataset yang digunakan merupakan data transaksi bulan Januari sampai Agustus 2022. Data tersebut akan digunakan untuk menentukan *cluster* yang diimplementasikan menggunakan bahasa pemrograman Python untuk melakukan *clustering* dengan metode *k-means clustering*. Hasil metode *k-means clustering* dengan nilai $k=2$ didapatkan hasil bahwa titik centroid terletak pada nilai *Recency* 0.69, nilai *Frequency* 0.01 dan nilai *Monetary* 0.01

. Cluster terbanyak pada *k-means clustering* yaitu cluster 1 yang datanya berdekatan dan tidak terlalu menyebar. Sedangkan cluster 2 tidak terlalu banyak dan menyebar dengan titik *centroid*. Dalam penelitian ini dilakukan iterasi sebanyak 3 kali dan menghasilkan *cluster* pada masing-masing iterasi. Kemudian dilakukan evaluasi dengan melakukan pengujian menggunakan metode Elbow dan metode DBI.

Pada metode Elbow didapatkan hasil evaluasi yaitu hasil cluster terbaik terdapat pada cluster 4 dengan nilai Elbow 7,75. Karena *cluster* terbaik merupakan *cluster* yang membentuk kurva. Sedangkan pada metode DBI didapatkan hasil *cluster* yang optimal terdapat pada cluster 8 dengan nilai DBI 0.05, karena semakin kecil nilai DBI maka *cluster* semakin optimal.

4.3 Integrasi Islam

Pengkajian yang sudah dilakukan akan membantu memaksimalkan kepuasan pelanggan. Serta mengoptimalkan pelayanan dengan memberikan promo sesuai dengan produk yang dijual. Seperti di dalam al-qur'an yang dijadikan sebagai pedoman manusia tertuli surat al-Baqarah ayat 275 :

الَّذِينَ يَأْكُلُونَ الرِّبَا لَا يَتَغَيَّرُونَ إِلَّا كَمَا يَقُومُ الَّذِي يَتَخَبَّطُهُ الشَّيْطَانُ مِنَ الْمَسِّ ۚ ذَٰلِكَ بِأَنَّهُمْ قَالُوا إِنَّمَا الْبَيْعُ مِثْلُ
الرِّبَا ۚ وَأَحَلَّ اللَّهُ الْبَيْعَ وَحَرَّمَ الرِّبَا ۚ فَمَنْ جَاءَهُ مَوْعِظَةٌ مِنْ رَبِّهِ فَانْتَهَىٰ فَلَهُ مَا سَلَفَ وَأَمْرُهُ إِلَى اللَّهِ ۗ وَمَنْ عَادَ فَأُولَٰئِكَ
أَصْحَابُ النَّارِ ۖ هُمْ فِيهَا خَالِدُونَ

// Hai orang-orang yang mengambil hak orang lain, tidak akan dapat berdiri melainkan kemasukan syaitan lantaran terkena penyakit gila. Keadaan orang yang terkena penyakit gila, diakibatkan karena kegiatan jual beli dilakukan dengan tidak jujur atau biasa disebut riba. Sementara itu di dalam al-qur'an telah menjelaskan bahwa kegiatan jual beli itu kegiatan yang halal sedangkan yang haram itu adalah riba. Barang siapa yang kembali mengambil riba, maka orang itu termasuk penghuni neraka. Sedangkan orang yang berusaha

menjauhi larangannya, maka segala urusan nya kembali kepada Allah SWT (Q.S. Al-Baqarah ayat 275).

Tafsir dari surat diatas yaitu orang-orang yang memakan hasil dari transaksi riba dengan mengambil kelebihan dari modalnya dan memanfaatkan barang tersebut untuk mengambil keuntungan setinggi mungkin dari orang yang membutuhkan dengan mengeksploitasi aspek kebutuhannya. Allah sebenarnya sudah menghalalkan berbagai macam bentuk jual beli. Namun, Allah sepenuhnya mengharamkan segala macam jual beli yang berhubungan dengan riba. Umat yang melanggar larangan dari tuhan dalam bentuk riba ini, maka orang tersebut adalah calon penghuni neraka dan mereka akan kekal di dalam sana.

Makna dari ayat tersebut mengatakan bahwa kegiatan jual beli adalah kegiatan yang halal sesuai dengan ketentuan islam. Dengan membeli barang atau produk dari proses serah terima antar penjual dengan pembeli. Pengamatan dilakukan diharapkan dapat membantu pihak toko untuk memunculkan kepuasan konsumen dari tingkal loyalitas pelanggan. Loyalitas adalah perilaku yang murni muncul dari benak pelanggan sehingga tidak bisa dipaksakan.

Sebagaimana bagi umat Islam yang beriman, mengaku bahwasannya keterkaitan iman dan agama itu tanpa adanya dusta. Yang dimaksud keimanan itu yaitu loyalitas yang dimiliki oleh hamba kepada yang Maha Pencipta-Nya. Seperti ayat Allah surat al-hujurat ayat 15:

إِنَّمَا الْمُؤْمِنُونَ الَّذِينَ آمَنُوا بِاللَّهِ وَرَسُولِهِ ثُمَّ لَمْ يَرْتَابُوا وَجَاهَدُوا بِأَمْوَالِهِمْ وَأَنْفُسِهِمْ فِي سَبِيلِ اللَّهِ ۚ أُولَٰئِكَ هُمُ الصَّادِقُونَ

// Sebaik-baiknya orang yang beriman, hanya orang beriman lah percaya kepada yang Maha Kuasa dan Rasul-Nya. Dan orang yang beriman tidak ragu-ragu serta berjuang di jalan Allah (Q.S. Al-Hujurat ayat 15).

Tafsir dari surat diatas yaitu setiap umat beragama yang beriman merupakan

orang-orang yang percaya kepada Tuhan dan para Rasul-Nya. Dan mereka tidak akan memiliki keraguan dalam hati dan pikirannya sehingga mereka akan berjuang untuk mendapatkan cinta dan kasih dari Tuhan dan Rasul-Nya dengan bersedekah berupa harta dan jiwa kepada jalan yang telah di perlihatkan oleh para Rasul-Nya. Maka dari itu umat yang melakukan hal tersebut merupakan orang yang baik dan benar.

Pada surat tersebut dapat jika dihubungkan dengan loyalitas pelanggan, yang mana, apabila pelanggan loyal akan percaya dengan satu produk dan memakai secara terus menerus sesuai dengan harga yang tersedia. Faktor utama yang menghasilkan keuntungan itu salah satunya loyalitas pelanggan yang setia dengan produknya. Dan loyalitas pelanggan terbentuk dari kepuasan pelanggan bahkan dilihat dari kualitas layanan perusahaan yang memberikan pelayanan yang baik untuk pelanggan.

BAB V

PENUTUP

5.1 Kesimpulan

Setelah dilakukan penelitian terhadap implementasi metode *K-Means Clustering* pada penentuan loyalitas pelanggan di Batubara Mart dengan menggunakan dataset pada bulan Januari sampai Agustus 2022 didapatkan hasil penelitian.

Dari hasil penelitian tersebut dapat dibuat kesimpulan bahwa Dalam mengimplementasikan metode *K-Means Clustering* pada Penentuan Loyalitas Pelanggan di Batubara Mart dilakukan dengan beberapa tahapan yaitu mengumpulkan dataset transaksi bulan Januari sampai Agustus 2022, *preprocessing* untuk membersihkan data agar terhindar dari duplikasi data atau *missing value*. Dalam *preprocessing* terdiri dari tahap data *cleansing* dan data *preparation*. Hasil data *preparation* digunakan untuk melakukan *clustering* menggunakan bahasa pemrograman Python dengan mengimplementasikan metode *K-Means Clustering*. Selanjutnya dilakukan evaluasi model menggunakan metode Elbow dan DBI. Hasil yang didapatkan dengan menerapkan metode *K-Means Clustering* pada Penentuan Loyalitas Pelanggan di Batubara Mart yaitu pada metode Elbow terdapat *cluster* terbaik yang diperoleh pada cluster 4 dengan nilai Elbow 7,75. Karena *cluster* terbaik merupakan *cluster* yang membentuk kurva. Sedangkan pada metode DBI didapatkan hasil *cluster* yang optimal terdapat pada cluster 8 dengan nilai DBI 0.05, karena semakin kecil nilai DBI maka *cluster* semakin optimal.

5.2 Saran

Dari kesimpulan yang telah dibuat berdasarkan hasil penelitian, maka dibuat saran yang berguna dalam pengembangan penelitian selanjutnya. Saran yang diberikan yaitu:

1. Dalam penelitian selanjutnya dapat mengembangkan *clustering* dengan data serupa namun dengan metode yang lebih baik, seperti *K-Nearest Neighbor*. Karena metode KNN cukup sederhana untuk dijalankan, cukup dapat dikenali dan kuat untuk menangani pembatasan informasi (Arinal & Sentosa, 2022).
2. Dalam penelitian selanjutnya dapat dikembangkan menjadi aplikasi berbasis *website* untuk sistem pendukung keputusan, sehingga pelanggan yang loyal akan diberikan *reward*.

DAFTAR PUSTAKA

- Ardimansyah, F. A. (2021). Implementasi Metode K-Means Clustering Pada Aplikasi Analisis Loyalitas Pelanggan PT. PLN (Persero) ULP Maros Berbasis Progressive Web Apps. *SISITI: Seminar Ilmiah Sistem Informasi dan Teknologi Informasi*. Vol. 10. No. 1.
- Arinal, V., & Sentosa, E. (2022). Klasifikasi Tingkat Kesejahteraan RW 006 Kelurahan Kalideres Jakarta Barat dengan Metode K-Nearest Neighbor. *Jurnal Pendidikan dan Konseling*.
- Daulay, A. R., & Salminawati. (2022). INTEGRASI ILMU AGAMA DAN SAINS TERHADAP PENDIDIKAN ISLAM DI ERA MODERN. *JOSR: Journal of Social Research*, 717-724.
- Dharshinni, N. P., & Fandi, C. (2022). Penerapan Metode K-Medoids Clustering Untuk Mengelompokkan Ketahanan Pangan. *JURNAL MEDIA INFORMATIKA BUDIDARMA*.
- Dinata, R. K., Safwandi, Hasdyna, N., & Azizah, N. (2020). Analisis K-Means Clustering pada Data Sepeda Motor. *Informatics Journal*.
- Hamdani, R. (2020). Pengelompokan Loyalitas Pelanggan Dengan Menggunakan Kombinasi Rfm Dan Algoritma K-Means. *Journal Of Informatic Pelita Nusantara*, 5(1, ISSN 2541-3724), 7-13.
- Harahap, S. R. (2022). Integrasi Keilmuan antara Sains dan Teknologi dengan Agama. *Book Chapter of Proceedings Journey-Liaison Academia and Society*.
- Irfanita, M. Y. (2020, Agustus 7). *OPINI - PERKEMBANGAN TEKNOLOGI PADA BIDANG PEMASARAN*. Retrieved from KUMPARAN.COM: <https://kumparan.com/fanongg/opini-perkembangan-teknologi-pada-bidang-pemasaran-1twri96vyYl/4>
- Jollyta, Efendi, & Zarlis. (2019). Optimasi Cluster Pada Data Stunting: Teknik Evaluasi Cluster Sum of Square Error dan Davies Bouldin Index. *Prosiding Seminar Nasional Riset Information* , 918-926.
- Lintas Media Danawa. (2022, September 14). *Customer Loyalty*. Retrieved from LMD Lintas Media Danawa: <https://lintasmediadanawa.com/berita/customer-loyalty/>

- Minarno, E. B. (2017). Integrasi Sains-Islam dan Implementasinya Integrasi Sains-Islam dan Implementasinya. *Seminar Nasional Teknologi Informasi, Komunikasi dan Industri (SNTIKI)* 9.
- Muningsih, E., Maryani, I., & Handayani, V. R. (2021). Penerapan Metode K-Means dan Optimasi Jumlah Cluster dengan Index Davies Bouldin untuk Clustering Propinsi Berdasarkan Potensi Desa. *Evolusi: Jurnal Sains dan Manajemen* .
- Nosra, A., Arifianto, D., & Rahman, M. (2022). Penerapan Metode Cosine Similarity Untuk Meningkatkan Kinerja K-Means Pada Pengelompokan Wilayah Penanganan Covid Di Dki Jakarta. *Jurnal Smart Teknologi*.
- Nurdiana, N., Nilogiri, A., & Rahman, M. (2022). Penerapan Algoritma Fuzzy C-Means dan Metode Elbow untuk Mengelompokkan Provinsi di Indonesia Berdasarkan Indeks Demokrasi Indonesia. *Jurnal Smart Teknologi*.
- Putra, R., & Cendra Wadisman. (2018). Implementasi Data Mining Pemilihan Pelanggan Potensial Menggunakan Algoritma K Means. *INTECOMS: Journal of Information Technology and Computer Science* 1.1, 72-77.
- Putri, Y. L. (2017). Pengaruh Kualitas Pelayanan Terhadap Loyalitas Pelanggan Dengan Kepuasan Sebagai Variabel Intervening. *Among Makarti*, 10(19), 70-90.
- Rahmah, S. A. (2020). KLASERISASI POLA PENJUALAN PESTISIDA MENGGUNAKAN METODE K-MEANS CLUSTERING (STUDI KASUS DI TOKO JUANDA TANI KECAMATAN HUTABAYU RAJA). *Djtechno : Journal of Information Technology Research*.
- Sasongko, S. R. (2021). FAKTOR-FAKTOR KEPUASAN PELANGGAN DAN LOYALITAS PELANGGAN (LITERATURE REVIEW MANAJEMEN PEMASARAN). *JIMT Jurnal Ilmu Manajemen Terapan*.
- Satriawan, M. A. (2021). Segmentasi Pelanggan Telkomsel Menggunakan Metode Clustering Dengan Rfm Model Dan Algoritma K-means. *eProceedings of Engineering* 8.2 .
- Situmorang, A., Arifin, Rusilpan, I., & Juliane, C. (2022). Analisa dan Penerapan Metode Algoritma K-Means Clustering Untuk Mengidentifikasi Rekomendasi Kategori Baru Pada List Movie IMDb. *JURNAL MEDIA INFORMATIKA BUDIDARMA*, 2171-2179.
- Suhandha, Y., Kurniati, I., & Norma, S. (2020). Penerapan Metode Crisp-DM Dengan Algoritma K-Means Clustering Untuk Segmentasi Mahasiswa

Berdasarkan Kualitas Akademik. *Jurnal Teknologi Informatika dan Komputer MH Thamrin*.

Sulistyowati. (2018). ANALISA SEGMENTASI KONSUMEN MENGGUNAKAN ALGORITMA K-MEANS CLUSTERING. *Prosiding Seminar Nasional Sains dan Teknologi Terapan*.

Wen Gie, & Deny Jollyta. (2020). Perbandingan Euclidean dan Manhattan Untuk Optimasi Cluster Menggunakan Davies Bouldin Index: Status Covid-19 Wilayah Riau. *Prosiding Seminar Nasional Riset Dan Information Science*, 187-191.