

**DATA MINING MENGGUNAKAN METODE CLUSTERING DAN  
ASOSIASI PADA DATA TRANSAKSI PENJUALAN**

**SKRIPSI**

Diajukan untuk Memenuhi Salah Satu Syarat Memperoleh  
Gelar Sarjana Komputer Program Studi Teknik Informatika



STATE ISLAMIC UNIVERSITY  
SUNAN KALIJAGA  
YOGYAKARTA

Disusun oleh:

**Arawinda Khairunnisa**

**19106050033**

**PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS SAINS DAN TEKNOLOGI  
UNIVERSITAS ISLAM NEGERI SUNAN KALIJAGA  
YOGYAKARTA**

**2023**

## HALAMAN PENGESAHAN



KEMENTERIAN AGAMA  
UNIVERSITAS ISLAM NEGERI SUNAN KALIJAGA  
FAKULTAS SAINS DAN TEKNOLOGI  
Jl. Marsda Adisucipto Telp. (0274) 540971 Fax. (0274) 519739 Yogyakarta 55281

### PENGESAHAN TUGAS AKHIR

Nomor : B-2309/Un.02/DST/PP.00.9/08/2023

Tugas Akhir dengan judul : Data Mining Menggunakan Metode Clustering dan Asosiasi Pada Data Transaksi Penjualan

yang dipersiapkan dan disusun oleh:

Nama : ARAWINDA KHAIRUNNISA  
Nomor Induk Mahasiswa : 19106050033  
Telah diujikan pada : Jumat, 18 Agustus 2023  
Nilai ujian Tugas Akhir : A

dinyatakan telah diterima oleh Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta

### TIM UJIAN TUGAS AKHIR



Ketua Sidang

Dr. Ir. Shofwatul 'Uyun, S.T., M.Kom., IPM., ASEAN Eng.  
SIGNED

Valid ID: 64e6ed6896f8a



Penguji I

Muhammad Didik Rohmad Wahyudi, S.T.,  
MT.  
SIGNED

Valid ID: 64e5df092f9dd



Penguji II

Eulis Yulianti Faridah, M.Ds  
SIGNED

Valid ID: 64e5f7eada5b4



Yogyakarta, 18 Agustus 2023

UIN Sunan Kalijaga  
Dekan Fakultas Sains dan Teknologi

Prof. Dr. Dra. Hj. Khurul Wardati, M.Si.  
SIGNED

Valid ID: 64e70367d9343

## SURAT PERNYATAAN KEASLIAN

### SURAT PERNYATAAN KEASLIAN SKRIPSI

Yang bertanda tangan di bawah ini:

Nama : Arawinda Khairunnisa  
NIM : 19106050033  
Program Studi : Teknik Informatika  
Fakultas : Sains dan Teknologi

Menyatakan dengan sesungguhnya, bahwa skripsi saya yang berjudul: *Data Mining Menggunakan Metode Clustering dan Asosiasi pada Data Transaksi Penjualan* adalah hasil karya pribadi dan sepanjang pengetahuan penyusun tidak berisi materi yang dipublikasikan atau ditulis orang lain, kecuali bagian-bagian tertentu yang penyusun ambil sebagai acuan.

Apabila terbukti pernyataan ini tidak benar, maka sepenuhnya menjadi tanggungjawab penyusun.

Yogyakarta, 14 Agustus 2023

Yang menyatakan,

STATE ISLAMIC UNIVERSITY  
SUNAN KALIJAGA  
YOGYAKARTA



Arawinda Khairunnisa  
NIM 191060500

## SURAT PERSETUJUAN SKRIPSI



KEMENTERIAN AGAMA  
UNIVERSITAS ISLAM NEGERI SUNAN KALIJAGA  
FAKULTAS SAINS DAN TEKNOLOGI  
Jl. Marsda Adisucipto Telp. (0274) 515856 Yogyakarta 55281

### SURAT PERSETUJUAN SKRIPSI/TUGAS AKHIR

Hal : Persetujuan Skripsi  
Lamp : -

Kepada  
Yth. Dekan Fakultas Sains dan Teknologi  
UIN Sunan Kalijaga Yogyakarta  
Di Yogyakarta

*Assalamualaikum wr.wb.*

Setelah membaca, meneliti, memberikan petunjuk dan mengoreksi serta mengadakan perbaikan seperlunya, maka saya selaku pembimbing berpendapat bahwa skripsi Saudara:

Nama : Arawinda Khairunnisa  
NIM : 19106050033  
Judul Skripsi : *Data Mining* Menggunakan Metode *Clustering* dan Asosiasi pada Data Transaksi Penjualan

Sudah dapat diajukan kembali kepada Program Studi Teknik Informatika Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta sebagai salah satu syarat untuk memperoleh gelar Sarjana Strata Satu dalam Program Studi Teknik Informatika.

Dengan ini saya berharap agar skripsi tersebut di atas dapat segera dimunaqosyahkan. Atas perhatiannya saya ucapkan terimakasih.

*Wassalamu'alaikum wr.wb.*

Yogyakarta, 14 Agustus 2023  
Pembimbing,

Dr. Ir. Shofwatul'Uyun, S.T., M.Kom  
NIP. 19820511 200604 2 002

## KATA PENGANTAR

*Alhamdulillah* rabbi'l'alam, puji syukur senantiasa penulis panjatkan kehadirat Allah SWT, yang telah melimpahkan taufiq dan hidayah-Nya sehingga penulis dapat menyelesaikan Tugas Akhir yang berjudul **“Data mining Menggunakan Metode Clustering dan Asosiasi pada Data Transaksi Penjualan”**. Skripsi ini disusun untuk memenuhi salah satu syarat menyelesaikan pendidikan Strata 1 pada Fakultas Sains dan Teknologi Universitas Islam Negeri Sunan Kalijaga Yogyakarta. Penyusunan Skripsi ini tentu tidak lepas dari bimbingan dan bantuan dari beberapa pihak. Oleh karena itu, pada kesempatan ini penulis mengucapkan terima kasih kepada semua pihak yang telah membantu hingga selesainya penyusunan Skripsi, terutama kepada:

1. Allah SWT yang telah memberikan rahmat dan hidayah-Nya, sehingga penulis dapat menyelesaikan skripsi ini.
2. Bapak Prof. Phil Al Makin, MA., selaku Rektor UIN Sunan Kalijaga.
3. Ibu Prof. Dr. Dra. Hj. Khurul Wardati, M.Si., selaku Dekan Fakultas Sains dan Teknologi UIN Sunan Kalijaga.
4. Ibu Maria Ulfa Siregar, S.Kom., MIT., Ph.D. selaku Ketua Program Studi Teknik Informatika Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta.
5. Bapak Ir. Aulia Faqih Rifa'i, M.Kom. selaku Dosen Pembimbing Akademik.
6. Ibu Dr. Ir. Shofwatul 'Uyun, S.T., M.Kom. selaku Dosen Pembimbing Skripsi yang telah memberikan arahan, bimbingan, masukan, serta motivasi selama penyusunan tugas akhir.
7. Dosen Penguji yang telah bersedia memberikan saran dan masukan terkait dengan tugas akhir supaya lebih baik.
8. Bapak/Ibu Dosen Teknik Informatika UIN Sunan Kalijaga yang telah membimbing dan memberikan banyak ilmu selama masa perkuliahan.



9. Kedua orang tua penulis dan semua keluarga yang senantiasa memberikan semangat, motivasi, serta doa.
10. Untuk kakak saya Alm Azalia Khairunnisa yang telah memberi warna dalam hidup penulis hingga akhir masa hidupnya.
11. Pemilik Toserba Delima yang telah banyak membantu dan bekerja sama untuk mengumpulkan data yang penulis perlukan.
12. Keluarga besar yang telah memberikan dukungan, semangat, serta doa.
13. Sahabat-sahabat FGDM Ridha, Jul, Ipeh, Alya dan Dea yang menemani suka dan duka selama kuliah.
14. Semua kucing yang telah rela dirinya menjadi objek hiburan sehingga membantu penulis dalam menangkal stress.
15. 8 Anak Nyasar yang selalu memberikan hiburan, semangat dan motivasi melalui lagu dan kontennya.
16. Teman-teman KKN, terutama mba Ami, Rafika dan Putri yang menjadikan KKN berkesan.
17. Teman-teman TIF angkatan 2019 UIN Sunan Kalijaga Yogyakarta.
18. Semua pihak yang telah membantu dalam penyusunan skripsi.

Penulis menyadari dalam penyusunan skripsi ini masih terdapat banyak kekurangan dari segi apapun. Oleh karena itu berbagai bentuk kritik dan saran dari berbagai pihak yang bertujuan untuk menyempurnakan penulisan skripsi ini, penulis akan menerima dengan senang hati supaya dapat dijadikan evaluasi bagi penelitian selanjutnya. Semoga skripsi ini dapat bermanfaat bagi semua pihak yang membutuhkan terutama dalam bidang Teknik Informatika.

Yogyakarta, 1 Agustus 2023

Arawinda Khairunnisa

19106050033

## HALAMAN PERSEMBAHAN

Dengan mengucapkan syukur Alhamdulillah, penulis mempersembahkan tugas akhir ini untuk semua orang yang telah berkontribusi dalam menyelesaikan tugas akhir ini, diantaranya adalah sebagai berikut :

1. Bapak saya, Joko Sri Purwoko yang selalu mendoakan, memberi motivasi, kasih sayang, semangat, dukungan material dan moral kepada penulis, serta senantiasa menjadi teman sehoobi dan teman bercanda.
2. Ibu saya, Artri Adelina yang selalu mendoakan, memberi kasih sayang, motivasi, dan semangat kepada penulis, serta selalu menjawab pertanyaan *random* penulis, teman penyayang bumi dan menjadi guru agama nomor 1 bagi penulis.
3. Kakak saya, Alm Azalia Khairunnisa yang telah memberi warna dalam hidup penulis, sebagai panutan, mengajari berbagai pelajaran hidup yang berharga selama masa hidupnya.
4. Diri saya sendiri, karena telah mampu berjuang hingga titik ini dengan berbagai masalah yang harus dihadapi.
5. Seluruh keluarga besar yang senantiasa memberi dukungan, semangat serta doa.

STATE ISLAMIC UNIVERSITY  
SUNAN KALIJAGA  
YOGYAKARTA

## HALAMAN MOTTO

“Allah tidak membebani seseorang, melainkan sesuai dengan kesanggupannya”

**Q.S. Al-Baqarah : 286**

“Apapun yang jadi takdirmu pasti akan mencari jalannya sendiri untuk menemukanmu”

**Ali bin Abi Thalib**

*“Even though we can’t see the end, we must go to the end”*

**Stray Kids - Glow**

*“There is no need to rush. Do not compare yourself with others. It is okay to run slower. Take it easy, just look ahead and run”*

**Stray Kids – My Pace**

*“We’ve all made mistakes. It’s okay, we just have to experience it and grow up. If you fall behind for a while, you can take a break. You’re doing well”*

**Stray Kids – Grow Up**

*“You’re amazing as you are, don’t you ever tell yourself that you’re not enough”*

**Stray Kids - Youtiful**

*“Life has ups and downs, just realized nothing lasts forever”*

**G-idle – Change**



## DAFTAR ISI

<b>HALAMAN JUDUL .....</b>	<b>i</b>
<b>HALAMAN PENGESAHAN.....</b>	<b>ii</b>
<b>SURAT PERNYATAAN KEASLIAN .....</b>	<b>iii</b>
<b>SURAT PERSETUJUAN SKRIPSI.....</b>	<b>iv</b>
<b>KATA PENGANTAR.....</b>	<b>v</b>
<b>HALAMAN PERSEMBAHAN.....</b>	<b>vii</b>
<b>HALAMAN MOTTO .....</b>	<b>viii</b>
<b>DAFTAR ISI.....</b>	<b>ix</b>
<b>DAFTAR TABEL.....</b>	<b>xi</b>
<b>DAFTAR GAMBAR.....</b>	<b>xii</b>
<b>INTISARI .....</b>	<b>xiv</b>
<b>ABSTRACT .....</b>	<b>xv</b>
<b>BAB I PENDAHULUAN.....</b>	<b>16</b>
1.1. Latar Belakang .....	16
1.2. Rumusan Masalah.....	18
1.3. Batasan Masalah .....	18
1.4. Tujuan Penelitian .....	19
1.5. Manfaat Penelitian .....	19
1.6. Keaslian Penelitian.....	19
<b>BAB II TINJAUAN PUSTAKA DAN LANDASAN TEORI.....</b>	<b>6</b>
2.1. Tinjauan Pustaka .....	6
2.2. Landasan Teori .....	9
2.2.1. Data mining.....	10
2.2.2. Clustering .....	15
2.2.3. K-Means Clustering .....	16
2.2.4. Association Rule .....	21
2.2.5. Frequent Pattern Growth (FP-Growth) .....	23

2.1.6.	Python .....	24
<b>BAB III</b>	<b>METODE PENELITIAN .....</b>	<b>26</b>
3.1.	Alat dan Bahan Penelitian .....	26
3.2.	Objek Penelitian .....	27
3.3.	Metode Penelitian .....	27
3.4.	Tahapan Penelitian.....	27
<b>BAB IV</b>	<b>HASIL DAN PEMBAHASAN.....</b>	<b>30</b>
4.1.	Hasil.....	30
4.1.1.	Pengumpulan Data .....	30
4.1.2.	Data Preparation / Pre-processing .....	31
4.1.3.	Data Mining .....	38
4.1.4.	Interpretasi.....	46
4.2.	Pembahasan .....	47
4.2.1.	Penerapan Algoritma K-Means Clustering Pada Python .....	47
4.2.2.	Penerapan Algoritma FP-Growth Pada Python.....	51
4.2.3.	Analisis Perbandingan.....	52
<b>BAB V</b>	<b>PENUTUP.....</b>	<b>55</b>
5.1.	Kesimpulan.....	55
5.2.	Saran .....	56
<b>DAFTAR PUSTAKA</b>	<b>.....</b>	<b>58</b>
<b>CURRICULUM VITAE</b>	<b>.....</b>	<b>62</b>

STATE ISLAMIC UNIVERSITY  
SUNAN KALIJAGA  
YOGYAKARTA

## DAFTAR TABEL

Tabel 2.1 Tinjauan Pustaka .....	8
Tabel 3.1 Alat Penelitian .....	26
Tabel 3.2 Bahan Penelitian.....	26
Tabel 4.1 Sampel Data Standarisasi <i>Scaling</i> .....	39
Tabel 4.2 Centroid Awal .....	40
Tabel 4.3 Nilai Jarak ke Centroid.....	40
Tabel 4.4 Centroid Baru .....	41
Tabel 4.5 Niali Jarak ke Centroid Baru .....	41
Tabel 4.6 Sampel Data Transaksi.....	42
Tabel 4.7 <i>Descanding Support Count</i> .....	43
Tabel 4.8 Tabel Transaksi Sesuai <i>Descanding Support Count</i> .....	43
Tabel 4.9 <i>Frequent Pattern</i> .....	45
Tabel 4.10 <i>Association Rules</i> .....	46
Tabel 4.11 <i>Frequent Itemset</i> Tanpa dan Dengan Klaster .....	53
Tabel 4.12 Analisis Perbandingan.....	54

STATE ISLAMIC UNIVERSITY  
SUNAN KALIJAGA  
YOGYAKARTA

## DAFTAR GAMBAR

Gambar 2.1 <i>Data Science</i> .....	10
Gambar 2.2 Alur KDD .....	11
Gambar 2.3 K-Means <i>Clustering</i> .....	16
Gambar 2.4 <i>Elbow Curve</i> Dataset A .....	19
Gambar 2.5 <i>Elbow Curve</i> Dataset B.....	19
Gambar 3.1 Tahap Penelitian .....	27
Gambar 4.1 Data Mentah .....	30
Gambar 4.2 Mengubah Tipe Data .....	32
Gambar 4.3 Atribut <i>Monetary</i> .....	33
Gambar 4.4 Atribut <i>Frequency</i> .....	33
Gambar 4.5 Atribut <i>Recency</i> .....	34
Gambar 4.6 Atribut Baru .....	34
Gambar 4.7 <i>Outliers</i> .....	35
Gambar 4.8 Boxplot <i>Outliers</i> .....	36
Gambar 4.9 Standarisasi <i>Scaling</i> .....	37
Gambar 4.10 <i>Reshape</i> Data .....	38
Gambar 4.11 Mengubah Nilai Data.....	38
Gambar 4.12 FP- <i>Tree</i> Transaksi No 1.....	44
Gambar 4.13 FP- <i>Tree</i> Transaksi No 2.....	44
Gambar 4.14 FP- <i>Tree</i> Pembacaan Semua Transaksi .....	44
Gambar 4.15 <i>Elbow Curve</i> .....	47
Gambar 4.16 Nilai <i>Silhouette</i> .....	48
Gambar 4.17 Nilai Klaster Optimal.....	49
Gambar 4.18 Boxplot <i>Monetary</i> .....	49
Gambar 4.19 Boxplot <i>Frequency</i> .....	50
Gambar 4.20 Boxplot <i>Recency</i> .....	50
Gambar 4.21 <i>Frequent Itemset</i> .....	51

Gambar 4.22 Aturan Asosiatif.....	52
Gambar 4.23 Proses <i>Filtering</i> .....	52



# DATA MINING MENGGUNAKAN METODE *CLUSTERING* DAN ASOSIASI PADA DATA TRANSAKSI PENJUALAN

Arawinda Khairunnisa

19106050033

## INTISARI

Di jaman sekarang ini, sangat mudah bagi seseorang untuk membuat data atau mendatakan sesuatu. Namun tidak semua orang dapat mengolah data tersebut menjadi sebuah informasi atau pengetahuan baru yang dapat sangat berguna. Hal ini dikarenakan keterbatasan ilmu, alat atau karena merasa tidak membutuhkannya. Oleh sebab itu skala volume data saat ini berjumlah sangat besar dan jika data tidak diolah menjadi suatu informasi baru maka data tersebut hanya akan menjadi sampah dalam memori. Salah satu cara untuk mengubah data menjadi informasi baru adalah dengan memanfaatkan *data mining*.

Untuk mengatasi masalah tersebut, maka pada penelitian ini akan dilakukannya *data mining* menggunakan metode *clustering* dan metode asosiasi dengan algoritma K-Means dan FP-Growth dalam mengolah data transaksi toserba Delima pada bulan Februari 2023 menjadi data yang dapat dimanfaatkan lebih lanjut oleh pihak toko. Metode asosiasi digunakan untuk menemukan pola asosiasi dari data transaksi dan metode *clustering* digunakan untuk meningkatkan akurasi dari hasil metode asosiasi. Kedua algoritma ini peneliti pilih karena dari beberapa studi terbukti bahwa kedua algoritma ini merupakan algoritma terbaik untuk melakukan *clustering* dan asosiasi / analisis keranjang pasar.

Pada penelitian ini, kedua algoritma tersebut berhasil diimplementasikan. Dengan menggunakan nilai klaster = 2, nilai *minimum support* 0.008, nilai *minimum confidence* 20% dan *lift* > 1, menghasilkan pola asosiasi serta jumlah *rules* yang berbeda pada data tanpa *clustering* dan data hasil *clustering*. Pola asosiasi terkuat pada tiap data adalah: data tanpa klaster (Vit 600ml → Minuman Dingin) dengan *support* 0.013 dan *confidence* 55%, klaster<sub>0</sub> (Vit 600ml → Minuman Dingin) dengan *support* 0.010 dan *confidence* 50% dan klaster<sub>1</sub> (Gula Pasir 1kg, Tepung Gandum → Minyak Sawit) dengan *support* 0.020 dan *confidence* 100%. Dengan melakukan *data clustering* sebelum menemukan pola asosiasi, terbukti dapat meningkatkan akurasi dari 0,06887 menjadi 28,6392.

**Kata kunci** : *Data mining*, *clustering*, Asosiasi, Analisis keranjang pasar, Algoritma K-Means *Clustering*, Algoritma FP-Growth.



## DATA MINING USES CLUSTERING AND ASSOCIATION METHOD ON SALES TRANSACTION DATA

Arawinda Khairunnisa

19106050033

### ABSTRACT

In this day and age, it is very easy for someone to create data or record something. However, not everyone can process the data into new information or knowledge that can be very useful. This is due to limited knowledge, tools or because they feel they do not need it. Therefore, the scale of the current data volume is very large and if the data is not processed into new information, then the data will only become garbage in memory. One way to turn data into new information is to use data mining.

To overcome this problem, this research will carry out data mining using the clustering method and the association method with the K-Means and FP-Growth algorithms in processing Delima department store transaction data in February 2023 into data that can be utilized further by the store. The association method is used to find association patterns from transaction data and the clustering method is used to increase the accuracy of the association method results. The researchers chose these two algorithms because from several studies it has been proven that these two algorithms are the best algorithms for clustering and association / market basket analysis.

In this study, both algorithms were successfully implemented. By using a cluster value = 2, minimum support 0.008, minimum confidence 20% and a lift > 1, it produces an association pattern and a different number of rules on data without clustering and data from clustering results. The strongest association patterns for each data are: data without clusters (Vit 600ml → Cold Drinks) with support value 0.013 and confidence value 55%, cluster<sub>0</sub> (600ml Vit → Cold Drinks) with support value 0.010 and confidence value 50% and cluster<sub>1</sub> (1kg Granulated Sugar, Wheat Flour → Palm Oil) with support value 0.020 and confidence value 100%. By doing data clustering before finding association patterns, it is proven to increase accuracy from 0.06887 to 28.6392.

**Keywords :** Data mining, clustering, association, Market basket analysis, K-Means Clustering algorithm, FP-Growth algorithm.

# BAB I

## PENDAHULUAN

### 1.1. Latar Belakang

Sekarang ini data sudah menjadi bagian penting dalam kehidupan manusia. Dengan adanya data, orang-orang mampu mengetahui berbagai macam informasi berdasarkan apa yang mereka butuhkan. Data juga memiliki berbagai macam fungsi yang dapat membantu manusia untuk mendapatkan dan merangkum penelitian. Saat ini, data dijadikan sebagai sebuah keterangan dan bahkan menjadi dasar untuk mempermudah orang-orang dalam mencari ataupun mengamati suatu hal (Amira, n.d.). Data dapat didefinisikan sebagai sekumpulan informasi atau nilai yang diperoleh dari pengamatan (observasi) suatu objek atau peristiwa yang dapat berupa kata-kata, kalimat, angka, simbol dan lainnya, yang kemudian diolah menjadi bentuk lain yang lebih kompleks, dapat berupa informasi, *database*, dan lainnya. Namun tidak semua orang mampu dalam pengolahan data tersebut dikarenakan keterbatasan ilmu, alat atau bahkan karena merasa tidak membutuhkan. Menurut John Naisbitt, kita terbenam di dalam data namun kita kekurangan informasi dan pengetahuan (Larose, 2005). Dengan meningkatnya kebutuhan data dan segala sesuatu yang didatangkan, maka skala volume data saat ini berjumlah sangat besar namun data-data tersebut hanya menjadi sampah dalam memori penyimpanan saja, apabila tidak diolah menjadi informasi. Oleh karena itu, dari permasalahan tersebut muncul cabang ilmu yang dapat mengatasi masalah dalam pengambilan informasi atau pola yang penting dan menarik dari sebuah data dengan jumlah besar yang disebut dengan *data mining*.

*Data mining* atau penambangan data dapat didefinisikan sebagai proses seleksi, eksplorasi dan pemodelan dari data berjumlah besar untuk menemukan pola atau kecenderungan yang biasanya tidak disadari keberadaannya. Sehingga *data mining* dapat digunakan oleh perusahaan, kelompok maupun individu untuk mendapatkan informasi yang dapat meningkatkan proses bisnis. *Data mining* dapat dikatakan sebagai proses mengekstrak pengetahuan dari sejumlah data yang

tersedia. Pengetahuan yang dihasilkan dari proses *data mining* harus baru, mudah dimengerti dan bermanfaat (Gambetta, 2012).

*Data mining* memiliki beberapa metode, yaitu estimasi, *forecasting*, klasifikasi, *clustering* dan asosiasi (Larose, 2005). Metode yang umum digunakan untuk pencarian suatu pola dalam data guna untuk membantu dalam membuat strategi perusahaan adalah metode asosiasi atau bisa juga disebut *market basket analysis*. Dengan menggunakan metode ini dapat diketahui pola atau kebiasaan dari konsumen seperti barang apa yang sering dibeli secara bersamaan. Namun jika data yang diolah terlalu besar, pendekatan asosiasi menjadi kurang akurat, sehingga perlu dilakukannya segmentasi atau *clustering* agar dataset menjadi lebih kecil dan hasil akurasi yang didapat menjadi lebih tinggi (Nurelasari, 2017). Maka pada penelitian ini akan dilakukan komparasi atau perbandingan antara penerapan *data mining* pada transaksi penjualan menggunakan metode asosiasi dengan menggunakan metode *clustering* dan tanpa *clustering*. Perbandingan dilihat dari segi cara metode ini bekerja (akankah menggabungkan metode *clustering* dan asosiasi dapat berjalan dengan baik), pola asosiasi yang dihasilkan, waktu, serta akurasi hasil.

Pada metode asosiasi terdapat beberapa algoritma yang dapat digunakan seperti algoritma apriori, FP-Growth, eclat, fuzzy c *covering* dan lainnya, namun penulis memilih untuk menggunakan algoritma FP-Growth untuk perhitungan pada metode asosiasi dikarenakan algoritma FP-Growth terbukti lebih efisien dan lebih cepat proses penghitungannya (Kahar, 2021). Sedangkan untuk metode *clustering* penulis memilih untuk menggunakan algoritma K-Means *clustering* karena merupakan salah satu algoritma yang populer digunakan untuk mengklaster data serta terbukti dapat menghasilkan data yang lebih akurat (Muningsih, 2018), membutuhkan waktu lebih singkat serta lebih mampu untuk mengolah data yang berjumlah besar dibandingkan dengan algoritma lainnya (Kamila et al., 2019).

Studi kasus dalam penelitian ini adalah Toserba Delima yang merupakan suatu toko yang menyediakan kebutuhan sehari-hari, seperti alat tulis, alat kebutuhan rumah tangga, makanan, minuman dan lain-lain. Toserba Delima yang terletak di Jl. Bimo Kurdo No.64, Demangan, Kec. Gondokusuman, Kota

Yogyakarta, Daerah Istimewa Yogyakarta 55221, dimana daerah tersebut merupakan daerah yang cukup strategis untuk tempat berbisnis, dikarenakan dekat dengan Universitas Islam Negeri Sunan Kalijaga dan SD Muhammadiyah Sopen. Faktor ini membuat Toserba Delima selalu ramai oleh pembeli, mulai dari siswa SD, mahasiswa UIN hingga pembisnis disekitarnya. Dalam sebulan Toserba Delima mampu mencatat lebih dari 20.000 *record* data transaksi, maka akan sangat disayangkan jika data transaksi ini hanya dipakai sebatas untuk menghitung laba dan stok barang, karena dengan data tersebut dapat dimanfaatkan untuk kepentingan usaha lainnya, seperti strategi peletakan barang, membuat paket *bundle* dan lainnya. Data yang digunakan pada penelitian ini adalah data transaksi dari Toserba Delima di bulan Februari 2023.

### **1.2. Rumusan Masalah**

Berdasarkan latar belakang diatas, maka dapat dirumuskan permasalahan yang akan diselesaikan dalam penelitian ini yaitu:

1. Bagaimana menerapkan algoritma K-Means dan FP-Growth untuk menganalisis pola pembelian konsumen?
2. Bagaimana perbandingan pola asosiasi yang dihasilkan dari data tanpa *clustering* dengan data hasil *clustering*?

### **1.3. Batasan Masalah**

Dalam penelitian ini terdapat beberapa batasan masalah yang akan dibahas supaya penelitian dapat dilakukan dapat tercapai dan sesuai dengan yang diharapkan, antara lain sebagai berikut :

1. Data yang diolah adalah data transaksi Toserba Delima selama sebulan pada bulan Februari 2023 dengan total data sebanyak 20.749 data.
2. Metode *data mining* yang digunakan adalah metode *clustering* dan asosiasi.
3. Penelitian dengan mengolah data dengan menggunakan algoritma FP-Growth dan algoritma K Means *Clustering*.
4. Menggunakan bahasa pemrograman Phyton untuk mengolah data.

#### 1.4. Tujuan Penelitian

Tujuan penelitian ditentukan sebagai langkah awal untuk memulai suatu penelitian karena dengan adanya tujuan, langkah-langkah penelitian menjadi lebih jelas dan terarah.

1. Mampu menerapkan algoritma K-Means dan FP-Growth untuk menganalisis pola pembelian konsumen.
2. Dapat menganalisis perbandingan pola asosiasi yang dihasilkan dari data tanpa *clustering* dengan data hasil *clustering*.

#### 1.5. Manfaat Penelitian

Dengan adanya penelitian ini diharapkan dapat memberikan manfaat untuk berbagai pihak yang terkait didalamnya, diantaranya:

1. Bagi Peneliti
  - a. Sebagai salah satu sarana untuk menerapkan keilmuan teknik informatika yang telah dipelajari sebelumnya selama masa perkuliahan agar lebih memahami cara pengaplikasiannya
  - b. Menambah pengetahuan peneliti dengan mengetahui perbandingan dari hasil algoritma K-Means dan algoritma FP-Growth.
2. Bagi Toserba Delima

Pihak toko akan mendapatkan beberapa informasi berikut yang dapat digunakan sebagai pertimbangan pengambilan keputusan terkait strategi pemasaran produk:

- a. Produk apa saja yang sering dibeli bersamaan oleh konsumen.
  - b. Produk apa saja yang sering dan jarang dibeli oleh konsumen.
3. Bagi Pihak Lain

Hasil dari penelitian ini diharapkan menjadi informasi ilmiah, menambah pengetahuan sekaligus sebagai referensi pada penelitian selanjutnya.

#### 1.6. Keaslian Penelitian

Penelitian mengenai *data mining* sudah banyak dilakukan, namun penelitian mengenai implementasi *data mining* dengan metode *clustering* dan asosiasi menggunakan algoritma K-Means *Clustering* dan algoritma FP-Growth

untuk mengetahui pola pembelian konsumen dari data transaksi pada Toserba Delima menggunakan bahasa pemrograman Python belum pernah ditemukan oleh peneliti.





## BAB V

### PENUTUP

#### 5.1. Kesimpulan

Berdasarkan hasil dari penelitian yang telah dilakukan, maka dapat diambil beberapa kesimpulan sebagai berikut:

1. Implementasi *data mining* dengan menggunakan algoritma K-Means dan FP-Growth berhasil peneliti lakukan, dengan menggunakan atribut faktur, tanggal, barang, jumlah, dan harga pada data transaksi toserba Delima pada bulan Februari 2023 yang kemudian data dibagi menjadi 2 kluster berdasarkan hasil penghitungan k optimal menggunakan *elbow* dan *silhouette method*. Kemudian data transaksi keseluruhan dan data hasil *clustering*, dilakukan pencarian pola asosiasi dengan nilai *minimum support* 0.008, nilai *minimum confidence* 20%, dan *lift* > 1 yang kemudian menghasilkan beberapa *association rule*. Dari hasil pengolahan tersebut diketahui pula pola pembelian konsumen pada toserba Delima pada bulan Februari 2023. Berikut *rule* terkuat dari setiap data yang dilakukan uji coba:
  - Hasil dari FP-Growth dengan menggunakan keseluruhan data transaksi (tidak dilakukannya segmentasi) memperoleh *rule* terkuat Vit 600ml → Minuman Dingin dengan nilai *support* 0.013309 dan nilai *confidence* 55% yang berarti dalam keseluruhan data transaksi jumlah transaksi yang mengandung item tersebut ada sebanyak 1% dan konsumen yang membeli Vit 600ml mempunyai kemungkinan 55% untuk memilih Vit 600ml yang berada didalam kulkas.
  - Hasil dari FP-Growth dengan menggunakan data klaster0 memperoleh *rule* terkuat Vit 600ml → Minuman Dingin dengan nilai *support* 0.010967 dan nilai *confidence* 50% yang berarti dalam keseluruhan data transaksi jumlah transaksi yang mengandung item

tersebut ada sebanyak 1% dan konsumen yang membeli Vit 600ml mempunyai kemungkinan 50% untuk membeli Vit 600ml yang dingin.

- Hasil dari FP-Growth dengan menggunakan data klaster<sub>1</sub> memperoleh *rule* terkuat Gula Pasir 1kg, Tepung Gandum → Minyak Sawit dengan nilai *support* 0.020365 dan nilai *confidence* 100% yang berarti dalam keseluruhan data transaksi jumlah transaksi yang mengandung item tersebut ada sebanyak 2% dan konsumen yang membeli Gula Pasir 1kg, Tepung Gandum pasti juga membeli Minyak Sawit.
2. Dari hasil pencarian pola asosiasi pada data tanpa *clustering* dan hasil *clustering* dengan menggunakan nilai *minimum support* 0.008, *minimum confidence* 20% dan *lift* > 1 dapat dianalisis perbandiannya sebagai berikut:
- Dari segi pola asosiasi, semua data menghasilkan pola asosiasi yang berbeda.
  - Dari segi jumlah *rule* yang dihasilkan, pola asosiasi menggunakan data hasil *clustering* memiliki jumlah *rule* terbanyak yang terdapat pada klaster<sub>1</sub>, dengan menghasilkan 88 *rule*
  - Dari segi akurasi, pola asosiasi dengan data hasil *clustering* memiliki nilai akurasi tertinggi yang terdapat pada klaster<sub>1</sub> dengan nilai akurasi sebesar 28,6392.
  - Dari segi waktu, pencarian pola asosiasi dengan data tanpa *clustering* lebih seingkat diakrenakan hanya menggunakan 1 algoritma.

## 5.2. Saran

Pada penelitian ini tentunya tidak terlepas dari kekurangan dan kesalahan. Oleh karena itu peneliti menyarankan adanya perbaikan yang dapat dijadikan masukan untuk peneliti selanjutnya, diantaranya sebagai berikut:

1. Menggunakan data dengan rentang waktu yang lebih lama, agar pada

atribut *recency* tiap data memiliki nilai yang jauh berbeda.

2. Dilakukannya pengkategorian barang seperti makanan, minuman, alat tulis dan lain-lain agar setiap kategori mendapatkan pola asosiasinya.
3. Menggunakan algoritma, bahasa pemrograman atau tools yang berbeda, misalnya menggunakan algoritma K-Medoids dan Apriori menggunakan php atau *tools* RapidMiner untuk menghasilkan penelitian yang lebih beragam dan variatif.
4. Dari *coding* yang telah dibuat pada penelitian ini, dapat diimplementasikan menjadi sebuah aplikasi, agar pihak toko dapat dengan mudah untuk mengetahui pola asosiasi dari data transaksi penjualannya setiap saat jika membutuhkan.



## DAFTAR PUSTAKA

- Ambarwati, T. (2022). *Implementasi Data Mining Dengan Metode Algoritma Apriori Untuk Mengetahui Pola Pembelian Konsumen (Studi Kasus : Super Dazzle)* [Universitas Islam Negeri Sunan Kalijaga Yogyakarta].
- Amira. (n.d.). *Pengertian Data: Fungsi, Manfaat, Jenis, dan Contohnya*. Gramedia Blog. Retrieved June 15, 2023, from <https://www.gramedia.com/literasi/pengertian-data/>
- Aulia, T. E. (2022). *Cara Mendeteksi dan Menangani Outlier*. Pacmann. <https://pacmann.io/blog/cara-mendeteksi-dan-menangani-outlier-saat-melakukan-data-analysis>
- Davies, D. L., & Bouldin, D. W. (1979). *A Cluster Separation Measure*. PAMI, 1(2), 224–227.
- Debatty, T., Mees, W., Michiardi, P., & Thonnard, O. (2014). *Determining the k in k-means with mapreduce*. Proceedings of the EDBT/ICDT, (ISSN 1613, 19–28.
- Dharshinni, N. P., Bangun, E., Karunia, S., Damayanti, R., Rophe, G., & Pandapotan, R. (2020). *Menu Package Recommendation using Combination of K-Means and FP-Growth Algorithms at Bakery Stores*. Jurnal Mantik, 4(2), 1272–1277.
- Dino, L. (2022). *Association mining — Lift in Python*. Medium. Retrieved June 5, 2023, from <https://medium.com/@24littledino/association-mining-lift-in-python-28c62a36678>
- Fahlevi, A. (2021). *Proses Data Mining KDD*. Binus University. <https://sis.binus.ac.id/2021/09/30/proses-data-mining-kdd/>
- Gambetta, W. (2012). *Pohon keputusan (decision tree)*.
- Gogia, N. (2019). *Why Scaling is Important in Machine Learning?* Analytics Vidhya. Retrieved June 15, 2023, from <https://medium.com/analytics-vidhya/why-scaling-is-important-in-machine-learning-ae5781d161a#:~:text=MinMaxScaler preserves the shape of,MinMaxScaler is 0 to 1.>
- Gunadi, G., & Sensuse, D. I. (2012). *Penerapan Metode Data Mining Market Basket Analysis Terhadap Data Penjualan Produk Buku dengan Menggunakan Algoritma Apriori dan Frequent Pattern Growth (FP-Growth) Studi Kasus Percetakan PT. Gramedia*. Telematika MKOM, 1(1), 118–132.

- Han, J., Kamber, M., & Pei, J. (2012). *Data Mining Concepts and Techniques* (Third).
- Hermawati, F. A. (2013). *Data Mining*. In CV Andi Offset (Vol. 2).
- Hornngren, Charles, T. (2012). *Akuntansi Biaya* (12th ed.). Erlangga.
- Kadir, A. (2018). *Dasar Pemrograman Python 3*. Andi Offset.
- Kahar, R. N. (2021). *Perbandingan Analisa Market Basket Terhadap Data Penjualan Minimarket Dengan Algoritma Apriori, Fp-growth, dan Fuzzy C-covering*. Prosiding Seminar Nasional Informatika Dan Sistem Informatika, 6(1), 50–59.
- Kamila, I., Khairunnisa, U., & Mustakim. (2019). *Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokan Data Transaksi Bongkar Muat di Provinsi Riau*. Jurnal Ilmiah Rekayasa Manajemen Sistem Informasi, 5(1), 119–125.
- Kaura, M., & Kanga, S. (2016). *Market Basket Analysis: Identify the Changing Trends of Market Data Using Association Rule Mining*. Procedia Computer Science, 85, 78–85.
- Khairu, N. N. (2020). *Unsupervised Learning: K-means Clustering using Python (Case: Online Retail Dataset)*. Medium. Retrieved Mei 1, 2023, from <https://nzlul.medium.com/unsupervised-learning-k-means-clustering-using-python-case-online-retail-dataset-df7d18599a52>
- Kukuh, R. (2018). *Melakukan Feature Scaling pada Dataset*. Machine Learning ID. Retrieved June 10, 2023, from <https://medium.com/machine-learning-id/melakukan-feature-scaling-pada-dataset-229531bb08de>
- Kurniawan, A., Susanto, E. H., & Alfian, F. (2022). *Implementasi Algoritma Fp Growth Untuk Prediksi Penjualan di Sistem Transaksi Penjualan Produk Koperasi Sekolah Berbasis Web SMA Negeri 1 Giri*. Informatika Dan Komputer, 12(1), 41–48.
- Kusrini, & Luthfi, E. T. (2009). *Algoritma Data Mining* (T. A. Prabawati (ed.)). CV. ANDI
- Larose, D. T. (2005). *Discovering knowledge in data: an introduction to data mining*. Jhon Wiley & Sons Inc.
- Madhulatha, T. S. (2011). *Comparison between K-Means and K-Medoids Clustering Algorithms*. International Journal of Advanced Computing (IJAC), 3(2), 65–69.
- Maimon, O., & Rokach, L. (2006). *Data Mining and Knowledge Discovery Handbook*.
- Mamahit, N., & Qoiriah, A. (2020). *Penerapan Algoritma Fp-Growth dan K-Means pada Data Transaksi Minimarket*. Journal of Informatics and Computer Science (JINACS), 1(02), 78–83.
- Muningsih, E. (2018). *Komparasi Metode Clustering K-Means dan K-Medoids Dengan*



- Model Fuzzy Rfm Untuk Pengelompokan Pelanggan*. *Evolusi*, 6(2), 106–113.
- Nobert, P. (2021). *MACHINE LEARNING – SUPERVISED AND UNSUPERVISED LEARNING*. Binus University. Retrieved Mei 1, 2023, from <https://student-activity.binus.ac.id/himmat/2021/03/machine-learning-supervised-and-unsupervised-learning/>
- Nosrati, M. (2011). *Python: An appropriate language for real world programming*. *World Applied Programming*, 1(2), 110–117.
- Nurelasari, E. (2017). *Algoritma Asosiasi K-Means dan FP-Growth untuk Analisis Keranjang Pasar pada Penjualan Produk Aluminium*. *Information Management for Educators and Professionals*, 1(2), 179–186.
- Prasetya, R. E. (2021). *[Belajar DM] Metode dan Peran Data Mining*. Rudy Eko Prasetya Blog. Retrieved June 15, 2023, from <https://rudyekoprasetya.wordpress.com/2021/02/06/belajar-dm-metode-dan-peran-data-mining/>
- Purnia, D. S., & Warnilah, A. I. (2017). *Implementasi Data Mining Pada Penjualan Kacamata Menggunakan Algoritma Apriori*. *IJCIT (Indonesian Journal on Computer and Information Technology)*, 2(2), 31–39.
- Putri, Z. K., Iskandar, I., & Nazir, A. (2021). *Implementasi Algoritma FP-Growth untuk Menemukan Pola Keterkaitan Antara Matakuliah Pemrograman dan Matakuliah Matematika*. *Hasil Penelitian Ilmu Komputer Dan Teknologi Informasi*, 7(2), 51–57.
- PyShark. (n.d.). *Dunn Index for K-Means Clustering Evaluation*. PyShark. Retrieved July 15, 2032, from [https://pyshark.com/dunn-index-for-k-means-clustering-evaluation-using-python/#:~:text=The Dunn Index \(DI\) is,a given number of clusters.](https://pyshark.com/dunn-index-for-k-means-clustering-evaluation-using-python/#:~:text=The Dunn Index (DI) is,a given number of clusters.)
- Rusdaman, D., & Setiyono, A. (2018). *Algoritma fp-growth dalam Penempatan Lokasi Barang di Gudang Pt. XYZ*. *Ilmu Pengetahuan Dan Teknologi Komputer*, 4(1), 63–70.
- Setyorini, S. G., Sari, E. K., Elita, L. R., & Putri, S. A. (2021). *Market Basket Analysis with K-Means and FP-Growth Algorithm as Citra Mustika Pandawa Company Analisis Keranjang Pasar Menggunakan Algoritma K-Means dan*. *Institute of Research and Publication Indonesia*, 1(April), 41–46.
- Singh, K., Malik, D., & Sharma, N. (2011). *Evolving limitations in K-means algorithm in data mining and their removal*. *IJCEM International Journal of Computational Engineering & Management*, 12(ISSN (Online): 2230-7893), 105–109.



- Wandi, N., Hendrawan, R. A., & Mukhlason, A. (2012). *Pengembangan Sistem Rekomendasi Penelusuran Buku Dengan Penggalian Association Rule Menggunakan Algoritma Apriori (Studi Kasus Badan Perpustakaan Dan Kearsipan Provinsi Jawa Timur)*. Institut Teknologi Sepuluh Nopember, 1(1).
- Witten, I. H., Frank, E., & Hall, M. A. (2011). *Data Mining: Practical Machine Learning Tools and Techniques*.
- Xu, R., & Wunsch, D. (2008). *Clustering*. John Wiley & Sons, Inc.

