

BAB II

LANDASAN TEORI

2.1. Tinjauan Pustaka

Dr. Han Jiawei(2021) Seorang ahli terkemuka dalam bidang data mining, Dr. Han Jiawei telah memberikan kontribusi besar dalam pengembangan algoritma data mining termasuk *FP-Growth*. Menurut Dr. Han, implementasi *FP-Growth* dalam analisis pola transaksional telah membuka peluang baru dalam pemahaman perilaku konsumen, yang sangat relevan dalam konteks strategi pemasaran.

Dr. Rajkumar Kannan(2021) Seorang pakar dalam bidang analisis data dan kecerdasan buatan, Dr. Kannan telah menyelidiki aplikasi metode asosiasi dan *Clustering* dalam berbagai konteks bisnis, termasuk pemasaran. Menurutnya, kombinasi antara *FP-Growth* dan *K-Means* dapat memberikan nilai tambah yang signifikan dalam pengembangan strategi pemasaran yang berbasis data.

Berikut dua teknik ini secara terpisah sebelum mengintegrasikannya dalam konteks strategi pemasaran.

1. *FP-Growth (Frequent Pattern Growth)*:

1. *FP-Growth* adalah metode yang *efisien* untuk menemukan pola yang sering muncul dalam kumpulan data transaksional.
2. Dalam konteks KKGJ Mart, *FP-Growth* dapat digunakan untuk mengidentifikasi pola pembelian yang sering terjadi di antara

produk-produk yang ditawarkan oleh toko tersebut. Misalnya, mungkin ada kecenderungan bahwa pelanggan yang membeli susu juga cenderung membeli roti.

3. Dengan mengetahui pola-pola ini, KKGJ Mart dapat mengoptimalkan penempatan produk, strategi harga, dan promosi untuk meningkatkan penjualan dan kepuasan pelanggan.

2. *K-Means Clustering*:

1. *K-Means* adalah algoritma *Clustering* yang populer untuk mengelompokkan data ke dalam kelompok-kelompok yang homogen berdasarkan atribut-atribut tertentu.

2. Dalam konteks KKGJ Mart, *K-Means* dapat digunakan untuk mengelompokkan pelanggan berdasarkan pola pembelian mereka. Misalnya, pelanggan dapat dikelompokkan berdasarkan jenis produk yang sering mereka beli atau berdasarkan total belanjaan bulanan mereka.

3. Dengan informasi ini, KKGJ Mart dapat menyesuaikan strategi pemasaran mereka untuk setiap kelompok pelanggan, misalnya dengan menawarkan diskon atau promosi khusus untuk setiap kelompok.

3 Integrasi *FP-Growth* dengan *K-Means*:

Setelah mengidentifikasi pola-pola pembelian dengan *FP-Growth*, kita dapat menggunakan informasi ini sebagai fitur untuk melaksanakan *Clustering* dengan *K-Means*. Ini akan membantu dalam mengidentifikasi kelompok pelanggan yang memiliki preferensi produk yang serupa, sehingga memungkinkan KKGJ Mart untuk menyesuaikan strategi pemasaran mereka dengan lebih baik.

2.1.1 Tinjauan Tentang Data Penjualan

Data penjualan merupakan bagian penting dalam penelitian yang menggunakan metode asosiasi dengan *FP-Growth* dan *Clustering* dengan *K-Means* untuk mendukung strategi *marketing*. Data penjualan merupakan informasi yang penting untuk mengetahui tindakan pelanggan dalam membeli produk atau layanan. Tinjauan tentang data penjualan bertujuan untuk mengumpulkan informasi mengenai data penjualan, teori kapabilitas, teori analisis, teori big data, teori data dan informasi, serta teori pertahanan negara.

Data penjualan merupakan informasi yang menggambarkan jumlah barang atau jasa yang dibeli oleh pelanggan dalam suatu waktu tertentu. Data penjualan dapat digunakan untuk mengetahui trend penjualan, sebuah data yang penting dalam pengembangan strategi *marketing*. Trend penjualan dapat menjadi informasi yang penting bagi perusahaan untuk mengetahui apa yang harus dilakukan untuk memperbaiki produk atau layanan yang dijual.

Data penjualan juga dapat digunakan untuk menentukan target pasar, yang merupakan bagian penting dalam pengembangan strategi marketing. Target pasar adalah grup pelanggan yang dibuat oleh perusahaan untuk menjadi pilihan pembeli produk atau layanan yang dijual.

2.1.2 Data Mining

Data *mining* adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam database. Data *Mining* adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, machine learning untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar (Setyawati, Nur kholis and Anjumi, 2021).

Sementara itu, Dunham (2021) berpendapat bahwa data mining adalah proses menemukan pola dan hubungan yang menarik dalam data yang besar, dengan menggunakan teknik seperti asosiasi, klasifikasi, *Clustering*, dan prediksi. Salah satu kesulitan untuk mendefinisikan Data Mining adalah kenyataan bahwa Data Mining mewarisi banyak aspek dan teknik dari bidang-bidang ilmu yang sudah mapan terlebih dahulu. Berawal dari beberapa disiplin ilmu, Data Mining bertujuan untuk memperbaiki teknik tradisional sehingga bisa menangani:

1. Jumlah data yang sangat besar
2. Dimensi data yang tinggi
3. Data yang heterogen dan berbeda sifat

Pengelompokan Data Mining dibagi menjadi beberapa kelompok yaitu :

1.Deskripsi

Deskripsi merupakan cara untuk menggambarkan pola dan kecenderungan yang terdapat dalam data yang dimiliki.

2.Estimasi

Estimasi hampir sama dengan klasifikasi, kecuali variabel target estimasi lebih ke arah numerik daripada ke arah kategori. Model yang dibangun menggunakan *record* lengkap yang menyediakan nilai variabel target sebagai nilai prediksi.

3.Prediksi

Prediksi menerka sebuah nilai yang belum diketahui dan juga memperkirakan nilai untuk masa mendatang.

4.Klasifikasi

Dalam klasifikasi terdapat target variabel kategori, misal penggolongan pendapatan dapat dipisahkan dalam tiga kategori, yaitu tinggi, sedang, dan rendah.

5.Pengklasteran

Pengklasteran merupakan pengelompokan *record*, pengamatan, atau memperhatikan dari membentuk kelas objek-objek yang memiliki kemiripan.

6. Asosiasi

Asosiasi bertugas menemukan atribut yang muncul dalam satu waktu. Dalam dunia bisnis lebih umum disebut analisis keranjang belanja.

2.1.3 Metode CLUSTERING DENGAN K-MEANS

Metode *Clustering* dengan *K-Means* dapat digunakan dalam pemasaran produk untuk meningkatkan pemahaman terhadap pelanggan dan strategi pemasaran.

Dengan mengelompokkan pelanggan berdasarkan perilaku pembelian, harga, atau karakteristik lainnya, pemasar dapat menyusun kampanye yang lebih personal dan relevan. Selain itu, analisis produk bersama, pengukuran kinerja kampanye, dan prediksi tren pasar juga dapat ditingkatkan menggunakan *K-Means*.

K-Means Cluster adalah teknik analisis *cluster* yang digunakan untuk menempatkan objek dalam satu atau lebih *cluster* berdasarkan karakteristiknya, sehingga objek dengan karakteristik tertentu ditempatkan dalam cluster yang sama dengan objek lain yang memiliki karakteristik tersebut (Sari & Sukestiyarno, 2021).

Dalam metode ini, pada proses pemisahan data ke dalam kelompok-kelompok sehingga data yang memiliki karakter sama dikelompokkan ke dalam satu kelompok yang sama dan data yang memiliki karakter berbeda dikelompokkan ke dalam kelompok yang lain. Adapun manfaat pengelompokan data ini adalah untuk meminimalkan fungsi objektif yang diatur pada proses pengelompokan, yang biasanya berupaya meminimalisir variasi data di dalam suatu kelompok dan

mengoptimalkan variasi data yang berbeda kelompok. Langkah yang dilakukan untuk membentuk *Clustering* pada metode *K-Means* adalah:

1. Menentukan banyaknya cluster (k) untuk jumlah cluster dari dataset yang ada.
2. Menentukan k sebagai Centroid, biasanya dilakukan secara acak (random).
3. Hitung jarak data dengan centroid menggunakan rumus jarak menggunakan rumus Euclidean (persamaan 1)

$$d_{euclidean}(x,y) = \sqrt{\sum_i (x_i - y_i)^2} \dots (1)$$

dimana d merupakan titik dokumen, x_i merupakan data kriteria dan y_i merupakan centroid pada cluster ke-i.

1. Kelompokkan data berdasarkan jarak terdekat dengan centroid.
2. Kemudian perbaharui atau menentukan nilai centroid baru dengan lokasi dari pusat cluster menggunakan persamaan 2:

$$\mu_k = \frac{1}{N_k} \sum_{q=1}^{N_k} x_q \dots (2)$$

Dimana μ_k = titik centroid dari cluster ke-K

N_k = banyaknya data pada cluster ke-K

x_q = data ke-q pada cluster ke-K

3. Lakukan langkah 3 sampai 5 sampai anggota tiap *cluster* tidak ada yang berubah.

2.1.4 Metode FP-Growth

Seorang ahli terkemuka dalam bidang data *mining*, Dr. Han Jiawei(2021) telah memberikan kontribusi besar dalam pengembangan algoritma data mining termasuk *FP-Growth*. Menurut Dr. Han, implementasi *FP-Growth* dalam analisis pola transaksional telah membuka peluang baru dalam pemahaman perilaku konsumen, yang sangat relevan dalam konteks strategi pemasaran.

Menurut Han et al. (2021), *FP-Growth* adalah algoritma *mining frequent pattern tree* yang efisien untuk menemukan pola frequent itemset tanpa perlu menghasilkan kandidat yang berlebihan. Algoritma ini bekerja dengan membangun struktur data pohon *FP-Tree* dan melakukan penambangan pola *frequent itemset* secara langsung dari *FP-Tree* tersebut. Karakteristik algoritma *FP-Growth* adalah struktur dari data digunakan adalah *tree* atau disebut dengan *FP-Tree*. Dengan menggunakan *FP-Tree*, algoritma ini dapat langsung mengekstrak *frequent itemset* dari *FP-Tree*.

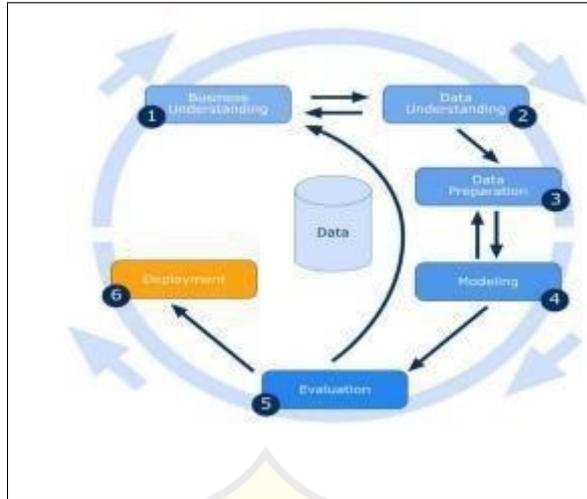
FP-Tree merupakan struktur penyimpanan data yang dimampatkan. *FP-tree* dibangun dengan memetakan setiap data transaksi ke dalam setiap lintasan tertentu dalam *FP-tree*. Karena dalam setiap transaksi yang dipetakan, mungkin ada transaksi yang memiliki item yang sama, maka lintasannya memungkinkan untuk saling menimpa. Semakin banyak data transaksi yang memiliki item yang sama,

maka proses pemampatan dengan struktur data *FP-tree* semakin efektif. Kelebihan dari *FP-tree* adalah hanya memerlukan dua kali pemindaian data transaksi yang terbukti sangat efisien. Metode *FP-Growth* dapat dibagi menjadi 3 tahapan yaitu:

1. Tahap pembangkitan conditional pattern base, merupakan sub database yang berisi *prefix path* dan *suffix pattern* yang di dapat dari dari *FP-tree* yang telah dibentuk.
2. Tahap pembangkitan conditional *FP-Tree*, *supportcount* dari setiap item pada setiap *conditional pattern base* dijumlahkan, kemudian setiap item yang memiliki jumlah *support* lebih besar atau sama dengan nilai minimum *support* yang akan di bangkitkan.
3. Tahap pencarian *frequent itemset*, jika conditional *FP-tree* adalah lintasan tunggal (*single path*), maka bisa didapat *frequent pattern* dari kombinasi item untuk setiap conditional *FP-Tree*, dan jikan bukan lintasan tunggal, maka lakukan pembangkitan secara rekursif.

2.2 CRISP-DM

Menurut Hasanah, dkk (2021:104) *CRISP-DM* merupakan metode yang menggunakan model proses pengembangan data yang banyak digunakan para ahli untuk memecahkan masalah. Metodologi ini terdiri dari enam tahapan yaitu *Business Understanding*, *Data Understanding*, *Data Preparation*, *Modelling*, *Evaluation*, dan *Deployment*. Proses metodologi ini terdiri dari 6 tahapan yang dapat dijelaskan sebagai berikut.



Gambar 2.1 Metodologi *CRISP-DM* (Hasanah,2021)

1. *Business Understanding* (Pemahaman Bisnis)

Beberapa hal yang dilakukan pada tahap ini seperti memahami kebutuhan serta tujuan dari sudut pandang bisnis selanjutnya mengartikan pengetahuan ke dalam bentuk pendefinisian masalah pada data mining dan kemudian menentukan rencana serta strategi untuk mencapai tujuan data mining.

2. *Data Understanding* (Pemahaman Data)

Tahapan ini diawali dengan Memahami dataset dan atribut-atribut yang tersedia.

3. *Data Preparation* (Persiapan Data)

Dalam tahapan ini yaitu membangun dataset akhir dari berupa data mentah. Ada beberapa hal yang akan dilakukan mencakup melakukan

pembersihan data (*Data Cleaning*), melakukan pemilihan data (*Data Selection*), *record* dan atribut-atribut, dan juga melakukan transformasi terhadap data (*Data Transformation*) untuk dijadikan masukan dalam tahap pemodelan.

4. *Modelling* (Pemodelan)

Pada tahap ini dilakukan metode statistika dan *Machine Learning* untuk penentuan terhadap teknik data *mining*, alat bantu data *mining*, dan algoritma data mining yang akan diterapkan. Lalu selanjutnya adalah melakukan penerapan teknik dan algoritma data mining tersebut kepada data dengan bantuan alat bantu. Jika diperlukan penyesuaian data terhadap teknik data mining tertentu, dapat kembali ke tahap data preparation.

5. *Evaluation* (Pengujian)

Melakukan interpretasi terhadap hasil dari data mining yang dihasilkan dalam proses pemodelan pada tahap sebelumnya. Evaluasi dilakukan terhadap model yang diterapkan pada tahap sebelumnya dengan tujuan agar model yang ditentukan dapat sesuai dengan tujuan yang ingin dicapai dalam tahap pertama.

6. *Deployment* (Penyebaran)

Tahap deployment atau rencana penggunaan model adalah tahap yang paling dihargai dari proses *CRISP-DM*. Perencanaan untuk Deployment dimulai selama Business Understanding dan harus 23

menggabungkan tidak hanya bagaimana untuk menghasilkan nilai model, tetapi juga bagaimana mengkonversi skor keputusan, dan bagaimana untuk menggabungkan keputusan dalam sistem operasional. Pada akhirnya, rencana sistem Deployment mengakui bahwa tidak ada model yang statis. Model tersebut dibangun dari data yang diwakili data pada waktu tertentu, sehingga perubahan waktu dapat menyebabkan berubahnya karakteristik data. Model Pun harus dipantau dan mungkin diganti dengan model yang sudah diperbaiki.

2.3 UML(Unified Modelling Language)

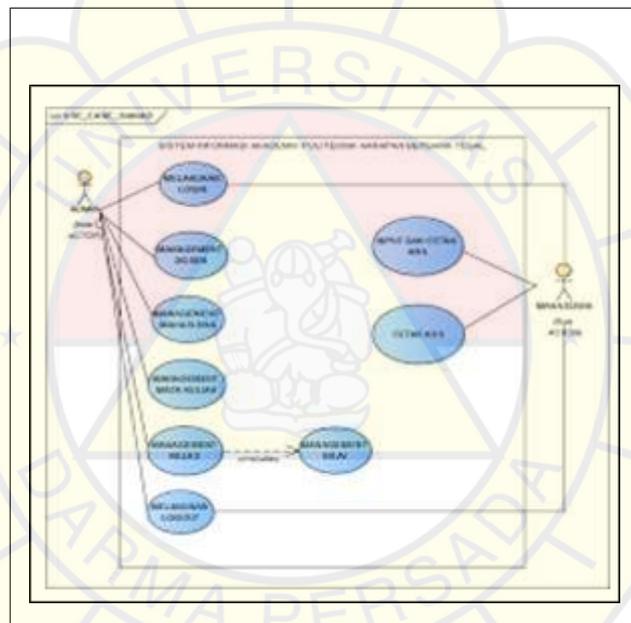
Menurut Muslihudin, M., & Oktafianto Dalam Buku Berjudul “Analisis dan Perancangan Sistem Informasi Menggunakan Model Terstruktur dan UML”. UML singkatan dari *Unified Modeling Language* yang berarti bahasa pemodelan standar. UML memiliki sintaks dan semantik. Ketika kita membuat model menggunakan konsep UML ada aturan-aturan yang harus diikuti. UML bukan hanya sekedar diagram, tetapi juga menceritakan konteks dalam membuat suatu sistem yang akan dibuat.

Dalam pandangan Jacobson (2021), UML adalah bahasa visual untuk menggambarkan, membangun, dan mendokumentasikan artefak dari sistem perangkat lunak. UML membantu untuk memahami, merancang, mengkonfigurasi, dan memelihara aplikasi perangkat lunak yang kompleks. Terdapat beberapa jenis-jenis diagram di dalam UML sebagai berikut:

2.3.1. Use Case Diagram

Menurut Sommerville (2021), use case diagram adalah diagram yang menggambarkan hubungan antara aktor-aktor dengan aktivitas-aktivitas sistem yang mereka lakukan. Diagram ini merupakan bagian dari pemodelan Unified Modeling Language (UML) dan digunakan untuk menangkap persyaratan fungsional dari sebuah sistem.

Contoh Diagram Use Case :



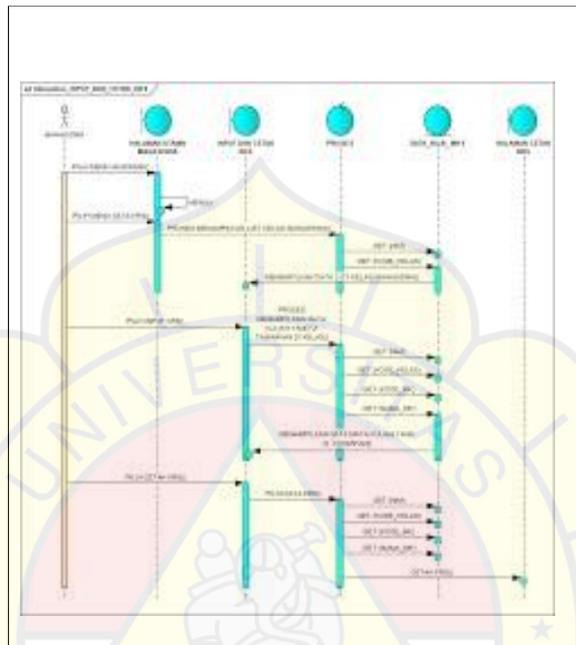
Gambar 2.2 Diagram Use Case (M Teguh Prihandoyo, 2021)

2.3.2. Activity Diagram

Menurut Rumbaugh dkk. (2021), activity diagram adalah diagram alur kerja yang menggambarkan aliran aktivitas dalam sebuah proses, termasuk urutan aktivitas, pengambilan keputusan, dan percabangan paralel. Diagram ini membantu memvisualisasikan urutan langkah-langkah dalam suatu proses bisnis

mendeskripsikan waktu hidup objek serta pesan yang dikirimkan serta diterima antar objek.

Contoh Diagram *Sequence*:



Gambar 2.4 Diagram *Sequence* (M Teguh Prihandoyo, 2021)

2.4 Website

Menurut Sitinjak, dkk (2020) *Website* adalah sering juga disebut web, dapat diartikan suatu kumpulan-kumpulan halaman yang menampilkan berbagai macam informasi teks, data, gambar diam maupun bergerak, data animasi, suara, video maupun gabungan dari semuanya, baik itu yang bersifat statis maupun yang dinamis, yang dimana membentuk satu rangkaian bangunan yang saling berkaitan dimana masing-masing dihubungkan dengan jaringan halaman atau hyperlink.

2.5 PHP

Menurut Rahmasari, T.(2019:414) PHP merupakan singkatan dari “Hypertext Preprocessor”. PHP adalah sebuah bahasa scripting yang terpasang pada HTML. Sebagian besar sintaknya mirip dengan bahasa pemrograman C, Java, ASP dan Perl ditambah beberapa fungsi PHP yang Spesifik dan mudah dimengerti. PHP digunakan untuk membuat tampilan web menjadi lebih dinamis, dengan PHP anda bisa menampilkan atau menjalankan beberapa file dalam 1 file dengan cara di include dan require. PHP itu sendiri sudah dapat berinteraksi dengan beberapa database walaupun dengan kelengkapan yang berbeda yaitu seperti DBMS, MySQL, Oracle.

2.6 Database

Menurut Ultariani, N. dkk (2020:221) Database adalah suatu aplikasi yang menyimpan sekumpulan data. Setiap database mempunyai API tertentu untuk membuat, mengakses, mengatur, mencari, menyalin data yang ada di dalamnya. Database yaitu kumpulan file-file yang berhubungan satu dengan yang lainnya, diatur sedemikian rupa sehingga dapat digunakan oleh beberapa program aplikasi database.

2.7 MySQL

Menurut Betha Sidik dalam buku yang “Pemrograman web dengan PHP (2012 : 333)” menyebutkan bahwa : “MySQL merupakan software database yang termasuk paling populer di lingkungan Linux, kepopuleran ini karena ditunjang karena performansi query dari database nya yang saat itu bisa dikatakan paling 28

cepat dan jarang bermasalah”. MySQL merupakan aplikasi database server. SQL kepanjangan dari Structured Query Language. SQL merupakan bahasa terstruktur yang digunakan untuk mengolah database.

MySQL dapat digunakan untuk membuat dan mengelola data di dalam database. Fungsi dari MySQL adalah untuk menambahkan mengubah, dan menghapus data di dalam database. MySQL berawal dari proyek yang dimulai oleh kedua orang developer, yakni Michael Widenius dan David Axmark di tahun 1994. Proyek ini didasari karena ingin membuat suatu sistem database yang murah, meskipun ketika itu ada database yang power full yakni oracle, namun database ini bersifat komersial yang harganya mahal, dan begitu menguasai pasar.

Menurut Rahmasari, T.(2019:414) “Definisi MySQL merupakan software RDBMS (*Relational Database Management System*) yang dapat mengelola database dengan sangat cepat, dapat menampung data dalam jumlah sangat besar, dapat diakses oleh banyak pengguna dan dapat melakukan suatu proses secara sinkron atau bersamaan”.

Menurut Sitinjak, dkk (2020) “MySQL adalah sebuah software database. MySQL merupakan tipe data relasional yang artinya MySQL menyimpan datanya dalam bentuk table-tabel yang saling berhubungan. Keuntungan menyimpan data di database adalah kemudahannya dalam penyimpanan dan menampilkan data karena dalam bentuk tabel”

2.8 Tinjauan Literatur

Tabel 2.2 - Paper penelitian terkait.

No	Penulis	Tahun	Judul	Metode	Publikasi
1.	Sari, W.K., Wibowo, S., & Amalia, R.	2021	Analisis Pola Pembelian Pelanggan Menggunakan Algoritma <i>FP- Growth</i> dan Segmentasi Pelanggan dengan <i>K- Means</i> untuk Strategi Pemasaran yang Efektif	Penelitian ini menggunakan <i>FP- Growth</i> untuk mengidentifikasi pola pembelian produk yang sering terjadi dan <i>K-Means</i> untuk mengelompokkan pelanggan berdasarkan perilaku pembelian mereka.	Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK), Vol. 8, No. 5, pp. 1031- 1038.
<p>Hasil:</p> <p>memberikan wawasan untuk strategi pemasaran yang disesuaikan dengan segmen pelanggan yang berbeda.</p>					
<p>Kelemahan Penelitian:</p>					

	<p>Keterbatasan Data Penelitian ini mungkin hanya menggunakan data transaksi penjualan dan informasi pelanggan yang terbatas. Tidak adanya data yang lebih komprehensif seperti data demografi, preferensi, atau perilaku pelanggan dapat mempengaruhi akurasi dalam mengidentifikasi pola pembelian dan segmentasi pelanggan.</p>				
No	Penulis	Tahun	Judul	Metode	Publikasi
2.	Hidayat, N., Nugroho, L.E., & Santosa, P.I.	2021	Penerapan Algoritma <i>FP-Growth</i> dan <i>K-Means</i> <i>Clustering</i> untuk Rekomendasi Produk di Toko Ritel	<i>FP-Growth</i>	Jurnal Sistem Informasi (JSI)
<p>Hasil: Sistem rekomendasi produk yang diusulkan berhasil meningkatkan penjualan di toko ritel.</p>					
<p>Keterbatasan Penelitian: Tidak ada penjelasan rinci tentang sumber data dan preprocessing yang dilakukan sebelum analisis.</p>					
No	Penulis	Tahun	Judul	Metode	Publikasi

3.	Rahman, A.F., Supriyanti, R., & Ardiansyah, R.	2021	Implementasi <i>FP-Growth</i> dan <i>K-Means</i> untuk Analisis Pola Pembelian dan Segmentasi Pelanggan di Toko Online	fp-Growth,k- means	Jurnal Teknologi Informasi dan Komunikasi (JUTIK),
<p>Hasil:</p> <p>Kombinasi <i>FP-Growth</i> dan <i>K-Means</i> memungkinkan rekomendasi produk yang lebih relevan dan personalisasi pengalaman belanja online.</p>					
<p>Keterbatasan Penelitian:</p> <p>Kurangnya informasi tentang pemilihan parameter yang optimal untuk algoritma <i>FP-Growth</i> dan K-Means.</p>					