

## 2. TEORI PENUNJANG

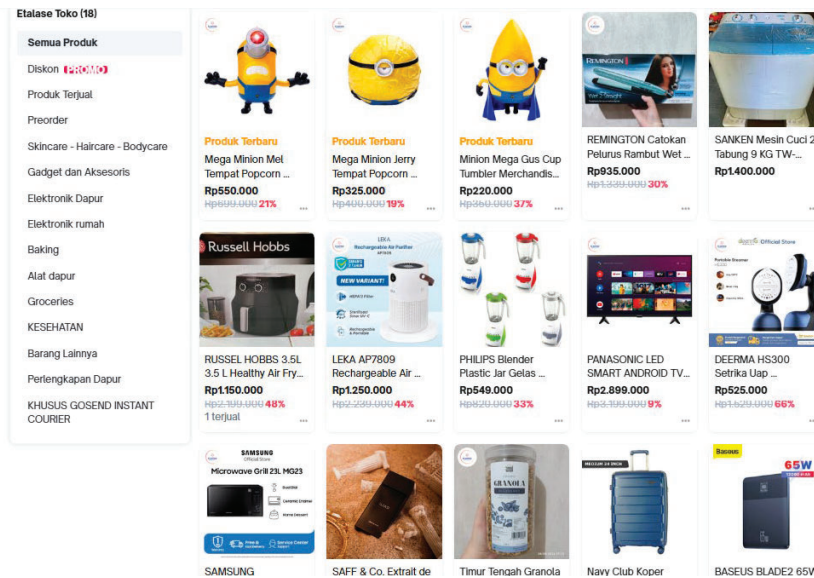
### 2.1. Tinjauan Pustaka

#### 2.1.1. Tokopedia

Tokopedia adalah sebuah perusahaan *e-commerce* terbesar di Indonesia yang didirikan pada tahun 2009. Tokopedia memberikan platform bagi penjual dan pembeli untuk bertransaksi secara daring. Tokopedia menyediakan berbagai macam produk, termasuk barang-barang elektronik, pakaian, aksesoris, makanan, dan lain sebagainya. Layanan *e-commerce* Tokopedia dapat diakses melalui berbagai sistem operasi. Aplikasi *mobile* tersedia untuk pengguna *smartphone* dengan sistem operasi iOS maupun Android, sedangkan pengguna komputer dapat mengakses layanan ini melalui situs *web* Tokopedia.

#### 2.1.2. Toko X

Toko X adalah toko yang terdaftar pada *platform e-commerce* Tokopedia sejak tahun 2018 silam. Produk – produk yang dijual oleh Toko X bervariasi mulai dari kebutuhan sehari – hari, kesehatan dan kecantikan hingga berbagai produk gadget dan elektronik lainnya. Toko X memiliki rating toko sebesar 4.9. Berikut adalah gambaran produk - produk yang dijual seperti pada gambar dibawah ini.



### 2.1.3. Data

Data adalah sekumpulan fakta yang berupa angka, sekumpulan teks ataupun lambang dari suatu periode waktu yang dapat diolah / dianalisis untuk menjadi sebuah informasi. Informasi ini dapat memperdalam wawasan seseorang terkait suatu keadaan dan menjadi salah satu faktor penting dalam pengambilan keputusan. Beberapa cara pengolahan data dengan memanfaatkan teknologi adalah *data mining*, *data warehousing* dan *big data processing*. Dalam penelitian ini, analisis data diterapkan terhadap data laporan pesanan dan data laporan pembeli dari Toko X untuk mendapatkan hasil perhitungan yang bermanfaat dalam pengambilan keputusan. Pemilik Toko X menggunakan *library Python* untuk membaca, memproses, mengupload dan menganalisis data seperti *library pandas*, *mlxtend*, *mysql.connector* dan lain sebagainya.

### 2.1.4. Data Mining

*Data mining* adalah metode untuk mengelola serta memanipulasi data guna memperoleh informasi bermanfaat. Proses ini melibatkan penggunaan teknik statistik, matematika, dan *machine learning* untuk mengekstrak serta mengidentifikasi informasi berharga dan pengetahuan yang terkait dari berbagai dataset besar. Salah satu contoh penerapan data mining yang terkenal adalah "*fraud detection*" pada data laporan finansial perusahaan. Teknik *data mining* untuk hal ini terbagi menjadi dua, klasifikasi biner dan deteksi anomali. Setelah didapat cukup sampel data, dilakukan penerapan metode *machine learning* untuk mengklasifikasikan data kedalam kelas *fraud* ataupun bebas *fraud* (Aftabi et al., 2023). Pada proyek ini, prinsip *data mining* yang akan diterapkan oleh penulis adalah aturan asosiasi terhadap data transaksi penjualan toko X di *e-commerce* Tokopedia.

### 2.1.5. Analisis Keranjang Belanja / *Market Basket Analysis*

Analisis Keranjang Belanja (*Market Basket Analysis* / MBA) adalah pola transaksi yang kebetulan menunjukkan bahwa pembelian satu produk bisa berdampak pada pembelian produk lainnya. Pada MBA, terdapat konsep *Association Rule Mining* / ARM yang bertujuan untuk menemukan informasi tentang item-item yang sering muncul bersama dalam bentuk aturan. Tiga parameter utama dalam MBA adalah *support*, *confidence*, dan *lift*. *Support* adalah rasio kejadian B karena kejadian A. *Confidence* adalah rasio probabilitas terjadinya kejadian B karena kejadian A yang saling berkaitan. *Lift* adalah rasio probabilitas terjadinya kejadian B karena kejadian A secara independen (Halim et al., 2019).

Penulis menggunakan salah satu algoritma ARM pada MBA, yaitu algoritma *FP-Growth* yang dapat mengelompokkan set produk yang sering dibeli bersamaan. Selain untuk pengelompokan set produk, analisis keranjang belanja juga dapat dimanfaatkan untuk alokasi stok produk dalam suatu set sehingga semua produk didalam suatu set dapat memiliki stok yang seimbang untuk meningkatkan penjualan.

#### **2.1.6. Algoritma FP-Growth (Frequent Pattern Growth)**

Algoritma *FP-Growth* (*Frequent Pattern Growth*) adalah salah satu algoritma yang paling efisien dalam mengidentifikasi pola dalam sekumpulan data. Perbedaan yang terdapat pada algoritma ini dengan *Apriori* dan *Eclat* adalah sebagai berikut :

1. Algoritma *Apriori* : Membaca set data berulang kali dan menerapkan teknik *Breadth-First Search* (BFS) pada data untuk mencari kandidat set produk yang paling sering dibeli dan kemudian mengeliminasi hasil kandidat dengan support kurang dari nilai minimum. Kompleksitas waktu dan penggunaan memori yang tinggi menyebabkan penerapannya sangat terbatas pada data besar (Srinadh, 2022).
2. Algoritma *FP-Growth* : Langkah pertama adalah set data dikompresi ke dalam struktur pohon *FP-Tree*. Struktur *FP-Tree* dapat diilustrasikan sebagai pohon utama yang menyimpan data transaksi awal. Data transaksi akan dibaca kembali untuk membentuk pohon kedua setelah item yang sering muncul teridentifikasi. *Nodes* dalam pohon diatur berdasarkan urutan nilai *support* dari tinggi ke rendah, sehingga cabang-cabang pohon yang mewakili transaksi dengan item-item yang sama memiliki peluang tumpang tindih yang lebih baik untuk mengurangi ukuran *FP-tree* (Shabtay et al., 2021). Langkah terakhir, algoritma *FP-Growth* akan menganalisis *FP-Tree* untuk mengekstrak set item yang sering muncul bersama dengan nilai *support*, *confidence* dan *lift* dari tiap set item.
3. Algoritma *Eclat* : Lebih cocok diterapkan pada set data kecil dan membutuhkan waktu yang lebih sedikit untuk menganalisis set item yang sering muncul dibanding algoritma *Apriori*. Menerapkan teknik *Depth-First Search* (DFS) untuk menghubungkan setiap set item yang teramati dengan daftar identifikasi transaksi yang mendukungnya. Kemudian, dengan

pendekatan *rekursif*, melakukan penyilangan antara set-set ini untuk mendapatkan daftar untuk kumpulan item yang lebih besar (Bifet et al., 2018).

### **2.1.7. Web Scraping**

*Web scraping* adalah sebuah teknik untuk mengekstrak data dari halaman *website* secara otomatis, *web scraping* telah menjadi hal yang sangat diperlukan dalam sistem pengambilan informasi yang efektif dan efisien. Teknik *web scraping* dapat secara umum dikategorikan menjadi statis dan dinamis. Teknik statis melibatkan ekstraksi data langsung dari halaman *HTML*, sementara teknik dinamis menggunakan program web *browser* untuk mengakses dan mengekstrak data dari situs *web*. (Pichiyan et al., 2023)

Saat ini, metode *web scraping* yang lebih canggih menggunakan perpustakaan seperti *BeautifulSoup* dan *Scrapy* yang dapat ditemukan pada Python. Salah satu contoh penerapan *web scraping* adalah ekstraksi ulasan produk dari situs *web e-commerce* untuk riset pasar dan analisis sentimen konsumen.

### **2.1.8. Web Application**

*Web application* adalah halaman di dalam suatu domain di internet yang dapat diakses melalui *browser*. Tujuan pembuatannya mirip dengan *website*, yaitu untuk menyampaikan suatu informasi kepada pengguna. Namun *web application* memungkinkan adanya interaksi dengan user lewat penggunaan fitur yang disediakan seperti *upload* data. Data akan diproses oleh sistem dan menampilkan hasil analisis secara otomatis. *Web application* dibuat khusus sesuai dengan kebutuhan user dimana sistem atau program dibuat sendiri dengan bahasa pemrograman *PHP*, *Python*, *Ruby* dan lain sebagainya yang kemudian diintegrasikan dengan struktur *web* untuk kemudahan akses bagi pengguna.

## **2.2. Tinjauan Studi**

Pada sub bab ini akan dibahas penelitian yang serupa dengan penelitian ini

### **2.2.1. Penerapan Metode Web Scraping Menggunakan Scrapy untuk Pembuatan Konten pada Aplikasi Pencari Lowongan Kerja (Lorence, 2023)**

- Masalah yang diangkat pada penelitian ini adalah lowongan pekerjaan yang banyak ditemukan di internet, namun proses pencarian yang seringkali memakan waktu dikarenakan banyaknya situs web dan lowongan kerja yang terdapat di internet.

- Metode yang digunakan dalam penelitian ini adalah *Web Scraping* dengan memanfaatkan *Scrapy* untuk mengambil data lowongan kerja dari berbagai situs penyedia lowongan kerja dan disatukan dalam suatu wadah.
- Hasil uji coba dari penelitian ini adalah pengambilan data lowongan pekerjaan pada 9 dari 10 situs lowongan kerja berhasil dilakukan.
- Perbedaan penelitian yang dilakukan dengan skripsi ini adalah penggunaan *web scraping* akan dilakukan pada *e-commerce* Tokopedia dan targetnya adalah daftar harga dari hasil pencarian suatu produk berdasarkan *keyword* pencarian.

### **2.2.2. Product Bundling for Online Supermarkets by Frequent Itemset Mining and Optimization Approach (Fang et al., 2022)**

- Masalah yang diangkat pada penelitian ini adalah bagaimana cara menghasilkan *itemset* yang menarik dan menguntungkan sesuai dengan karakter konsumen di berbagai daerah.
- Metode yang digunakan pada penelitian ini adalah identifikasi pola pembelian melalui penambangan *frequent itemset* dengan algoritma Apriori pada data penjualan.
- Hasil uji coba dari penelitian ini adalah bahwa *itemset* dapat dipilih secara efektif dan penjualan supermarket *online* dapat meningkat melalui penggabungan paket.
- Perbedaan penelitian yang dilakukan dengan skripsi ini adalah penggunaan Algoritma *FP-Growth* untuk mengenal pola transaksi produk yang sering dibeli.

### **2.2.3. Market Basket Analysis of Heterogeneous Data Sources for Recommendation System Improvement (Kutuzova & Melnik, 2018)**

- Masalah yang diangkat pada penelitian ini adalah perkembangan pasar *online* dan teknologi pasar mengarah pada kebutuhan bagi para pemilik toko untuk menganalisis perilaku pelanggan. Sehingga diperlukan sistem rekomendasi yang dimana peningkatan kualitas hasil rekomendasi memungkinkan ketika menganalisis data yang lebih banyak, yang dapat diperoleh dari sumber-sumber heterogen eksternal.
- Metode yang digunakan dalam penelitian ini adalah *Association Rule Mining* dengan algoritma *Apriori* untuk menemukan hubungan antar produk dalam bentuk aturan agar dapat tercipta sistem rekomendasi.
- Hasil uji coba dari penelitian ini adalah meningkatnya kualitas sistem rekomendasi dengan menggunakan sumber data heterogen tambahan.

- Perbedaan penelitian yang dilakukan dengan skripsi ini adalah penggunaan Algoritma *FP-Growth* untuk mengenal pola transaksi produk yang sering dibeli.

#### **2.2.4. Analisis dan Perancangan Sistem Informasi e-Commerce Berbasis Web Pada UMKM**

##### **Batik Rindani Jambi (Setiawan, D., & Lutfi, 2018)**

- Masalah yang diangkat pada penelitian ini adalah UMKM, khususnya Batik Rindiani, sering menghadapi kesulitan dalam mengelola penjualan dan promosi produk secara efektif dan efisien menggunakan sistem informasi, yang berdampak negatif pada biaya, tenaga, dan waktu.
- Metode yang digunakan dalam penelitian ini adalah pembuatan blueprint perancangan, desain dan pengembangan sistem informasi e-commerce berbasis web, serta implementasi sistem ini pada UMKM Batik Rindiani Jambi untuk meningkatkan efektivitas dan efisiensi tata kelola penjualan dan promosi produk.
- Hasil uji coba dari penelitian ini menunjukkan bahwa semua fitur yang dibuat pada sistem informasi e-commerce berbasis web ini dapat berjalan dengan baik dan para stakeholder di UMKM Rindiani memiliki keinginan besar untuk keberhasilan sistem informasi e-commerce yang diharapkan dapat meningkatkan pemasaran dan penjualan produk UMKM Rindiani.
- Perbedaan penelitian yang dilakukan dengan skripsi ini adalah pada tujuan pembuatan web dan metode yang digunakan, dimana penulis menggunakan Python dan library tambahan untuk pengolahan data.