

Jutisi Dewa Ayu Blind

by Eka Ariawan

Submission date: 23-Aug-2024 06:29AM (UTC+0530)

Submission ID: 2419049357

File name: JUTISNEW24fixxRev_-_Blind.doc (1.75M)

Word count: 3535

Character count: 23707

Analisis Keranjang Belanja pada Data Ritel Non-Toko menggunakan Algoritma *FP-Growth*

Abstract

An analysis of sales transaction data is needed to uncover relationships between items in the shopping cart, thus providing strategic insights for business decision-making. This research conducted an association analysis of non-store retail sales transaction data using the CRISP-DM methodology. The analysis is processed through Google Colab with the Python programming language and uses the FP-Growth algorithm with a comparison of various minimum support values. The results obtained a model with a minimum support of 1.5% (0.015) that achieved the best balance. The model generated 3-itemsets in 4.12 seconds with 17 KB of storage and formed 194 association rules. Evaluation of the association rules revealed a maximum confidence value of 89.45% and a minimum of 14.16%, with a lift value > 1, indicating a strong pattern of item combinations and frequent purchases together. The results of these rules can be used for online recommendations as well as the placement of items in the warehouse.

Keywords: *FP-Growth Algorithm; Market Basket Analysis; CRISP-DM; Retail Data; Data Mining*

Abstrak

Analisis data transaksi penjualan diperlukan untuk mengungkap hubungan antar barang dalam keranjang belanja, sehingga memberikan wawasan strategis bagi pengambilan keputusan bisnis. Penelitian ini melakukan analisis asosiasi terhadap data transaksi penjualan ritel non-toko menggunakan metodologi CRISP-DM. Analisis diproses melalui Google Colab dengan bahasa pemrograman Python dan menggunakan algoritma *FP-Growth* dengan perbandingan berbagai nilai *minimum support*. Hasil penelitian memperoleh model dengan *minimum support* 1,5% (0,015) yang mencapai keseimbangan terbaik. Model ini menghasilkan 3-itemset dalam waktu 4,12 detik dengan penyimpanan sebesar 17 KB dan membentuk 194 aturan asosiasi. Evaluasi aturan asosiasi mengungkap nilai *confidence* maksimum 89,45% dan minimum 14,16%, dengan nilai *lift* > 1, yang menunjukkan pola kombinasi barang kuat dan sering dibeli bersamaan. Hasil aturan ini dapat digunakan untuk rekomendasi *online* serta penempatan barang di gudang.

Kata kunci: *Algoritma FP-Growth; Analisis Keranjang Belanja; CRISP-DM; Data Ritel; Penambangan Data*

1. Pendahuluan

Dengan perkembangan teknologi dan komunikasi dalam dunia yang semakin terhubung secara digital [1], data menjadi semakin heterogen, dan volumenya meningkat secara eksponensial [2]. Di tengah perubahan ini, konsep otoritas telah bergeser, jika dahulu manusia percaya bahwa otoritas tertinggi adalah penguasa. Kini, di zaman teknologi, otoritas tertinggi adalah big data dan algoritma. Inilah era *dataisme* [3], di mana volume data yang dihasilkan dari berbagai sumber, seperti layanan kesehatan, pendidikan, media sosial, dan penggunaan internet meningkat secara signifikan [2]. Kehidupan tidak pernah begitu tercatat, namun sekarang terekam jelas dan warga internet telah menjadi kontributor sekaligus konsumen data [4]. Salah satu sektor yang sangat dipengaruhi oleh kemajuan ini adalah perdagangan elektronik (*E-commerce*). Dengan keberadaan internet, transaksi secara elektronik telah menjadi pilihan utama bagi organisasi dan individu. Melalui jaringan global ini, bisnis dapat dengan mudah mempromosikan produk barang dan jasa yang ditawarkan melalui situs web dan menyediakan fasilitas transaksi *online* yang efisien, serta membuka peluang baru untuk menjangkau pasar yang lebih luas. Konsumen juga dimudahkan dengan kemampuan untuk bertransaksi melalui komputer atau telepon pintar tanpa harus bertemu langsung dengan

penjual [5]. Pada tahun 2023, penjualan *e-commerce* ritel global mencapai sekitar 5,8 triliun dolar AS, dengan proyeksi pertumbuhan mencapai 39% dalam beberapa tahun ke depan, diperkirakan akan melampaui delapan triliun dolar AS pada tahun 2027 [6]. Pertumbuhan ini didorong oleh peningkatan minat dan kebutuhan konsumen terhadap berbagai barang yang ditawarkan oleh perusahaan ritel [7]. Akibatnya, kegiatan bisnis ini menghasilkan kumpulan data raksasa, termasuk transaksi penjualan, deskripsi barang, promosi penjualan, profil perusahaan, dan performanya, serta umpan balik pelanggan [8].

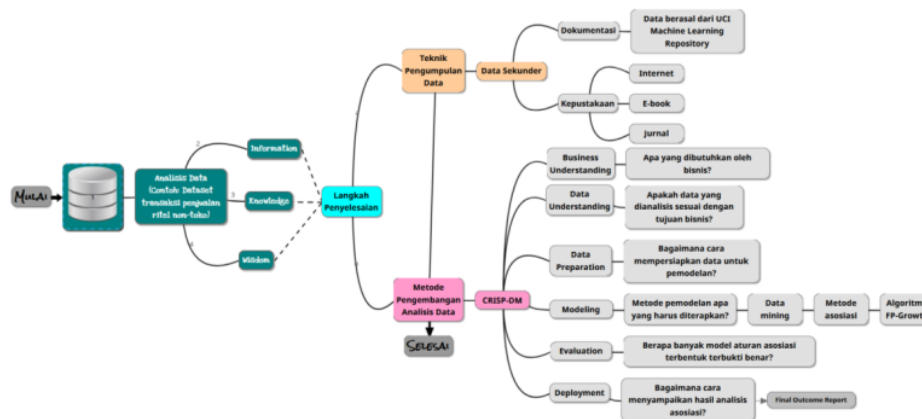
Dalam bukunya yang diterbitkan pada tahun 1982 berjudul *Megatrends*, John Naisbitt menulis, '*We are drowning in information but starved for knowledge*.' Meskipun telah berlalu lebih dari 30 tahun, pernyataan tersebut masih relevan hingga kini. Namun, Eric D. Brown, seorang ilmuwan data, menyesuaikan pernyataan tersebut dengan mengatakan bahwa saat ini kita tenggelam dalam data dan kelaparan akan pengetahuan [9]. Penting dipahami bahwa pada prinsipnya data memang perlu disimpan. Namun, lebih penting lagi adalah transformasi data menjadi informasi yang berguna. Proses ini melibatkan serangkaian langkah, mulai dari pengumpulan, pembersihan, integrasi, transformasi, analisis, hingga visualisasi data [5]. Menurut penelitian Jia Xu, *et al* [10] kemampuan untuk mengelola dan memanfaatkan data telah menjadi keunggulan bersaing masa kini, di mana penggunaan data menjadi kunci keberhasilan. Untuk itu, penting bagi organisasi untuk mengelola data dengan cermat karena data dianggap sebagai aset yang berharga [11]. Dalam menghasilkan pengetahuan baru, diperlukan perhatian terhadap atribut-atribut yang penting pada data tersebut [7]. Oleh karena itu, sebagai hasil evolusi alami teknologi informasi, munculah *data mining* [12], yaitu proses penggunaan teknik analisis dan algoritma untuk mengekstraksi informasi penting dari data besar [13], [14]. Dalam proses ini, data dirangkum dengan cara yang lebih mudah dipahami dan berguna bagi pemilik data [15], sehingga dapat membantu dalam pengambilan keputusan [16]. Menurut penelitian Iffa Musdalifah dan Arief Jananto [17] perusahaan harus merancang berbagai strategi, termasuk memanfaatkan data transaksi untuk meningkatkan penjualan. Salah satu strategi yang dapat dilakukan, menurut penelitian Fandi Achmad, *et al* [18] adalah *cross-selling* berdasarkan analisis pembelian konsumen dengan melihat keterkaitan antara produk yang sering dibeli. Penelitian ini menggunakan *dataset* di perusahaan ritel CV Harmoni Medicine Indonesia yang bergerak dalam penjualan produk kesehatan dan kecantikan. Penelitian ini, juga menggunakan perbandingan nilai *minimum support* dalam pembentukan *itemset*, di mana *itemset* terbanyak terbentuk pada nilai *support* sebesar 0,005, menghasilkan 1-*itemset* sebanyak 34 dan 2-*itemset* sebanyak 28, tanpa adanya pembentuk 3-*itemset*. Dan berhasil membentuk 9 aturan asosiasi, dengan hasil pengujian nilai *minimum confidence* = 0.1 dan nilai *minimum lift* = 1.0. Pola asosiasi ini terbentuk menggunakan algoritma *FP-Growth*, yang memiliki kinerja dan akurasi lebih baik dalam menghasilkan aturan asosiasi dibandingkan dengan algoritma *Apriori*, serta tercepat dalam eksekusi, diikuti oleh algoritma *Apriori* dan *Eclat* [19], [20], [21]. Penelitian lainnya oleh Putra Christianto Purba dan Elisafina Siswanto [22] menyatakan untuk meningkatkan penjualan, baik dari segi finansial maupun brand produk dapat dilakukan penentuan *product bundling* berdasarkan analisis pola pembelian konsumen. Penelitian ini menggunakan *dataset* transaksi penjualan makanan dan minuman, serta mengembangkan aplikasi untuk penentuan *product bundling* berdasarkan pola yang didapat, sehingga penentuan *product bundling* menjadi lebih mudah dan cepat, serta dapat memotong waktu yang diperlukan tim pemasaran dalam menganalisis secara manual.

Dari informasi yang disampaikan, penelitian ini juga menganalisis keranjang belanja melalui teknik asosiasi pada *data mining* untuk menemukan pola barang yang sering dibeli bersamaan oleh pelanggan. Tujuannya adalah membangun model menggunakan algoritma *FP-Growth* dengan membandingkan berbagai nilai *minimum support* untuk menentukan frekuensi kombinasi *item* dalam *dataset*, sehingga terbentuk aturan asosiasi yang sesuai. Perbedaannya adalah menggunakan *dataset* transaksi penjualan ritel *non-toko* yang menjual barangnya melalui media selain toko fisik. Pada *dataset* besar, ambang batas dukungan minimum perlu disetel dengan hati-hati untuk menghindari ledakan kombinatorial dalam pembangkitan pola [23]. Oleh karena itu, eksperimen dan penyesuaian parameter dilakukan untuk menemukan keseimbangan antara signifikansi pola dan efisiensi sumber daya, seperti durasi eksekusi dan kapasitas penyimpanan [24], [25], [26]. Setiap aturan akan dievaluasi berdasarkan nilai *confidence* untuk mengukur kekuatan hubungan antar *item* dalam transaksi, serta nilai *lift* untuk menilai validitas aturan. Penelitian ini tidak menetapkan *minimum confidence*, karena aturan dengan nilai *confidence* di bawah 100%, tetapi memiliki nilai *lift* lebih dari 1 masih dianggap

cukup penting [27]. Temuan ini dapat digunakan untuk mengatur tata letak *item* secara berdekatan di tempat penyimpanan, serta membantu merekomendasikan *item* dalam konteks penjualan ritel *non-toko*, dengan harapan mampu meningkatkan pengalaman belanja dan pelayanan pelanggan. Kehadiran integrasi antara tiga komponen utama, yakni *hardware*, *software*, dan *brainware* [28] sangatlah penting untuk memastikan kelancaran analisis.

2. Metodologi

Pada Gambar 1, ditampilkan peta konsep penelitian yang mencakup berbagai aktivitas terkait langkah-langkah penyelesaian analisis keranjang belanja. Data transaksi penjualan yang terkumpul diolah menjadi informasi, yang kemudian menghasilkan pengetahuan berharga untuk pengambilan keputusan. Proses ini mencakup dari pengumpulan data sampai penyampaian hasil analisis.



Gambar 1. Peta Konsep Penelitian

2.1. Teknik Pengumpulan Data

Penelitian ini memanfaatkan data sekunder yang diperoleh dari sumber-sumber yang telah ada sebelumnya. Metode yang digunakan untuk mendapatkan data sekunder adalah sebagai berikut:

1. Dokumentasi:

Metode dokumentasi adalah teknik yang melibatkan pencarian dan pengumpulan informasi dari *dataset* transaksi penjualan ritel *non-toko* yang diambil melalui situs web resmi UCI (*University of California Irvine Machine Learning Repository*), sebuah platform penyimpanan data yang menyediakan akses gratis untuk mengunduh berbagai kumpulan data. Repositori ini menjadi salah satu dari keempat jenis *public dataset* yang sering dimanfaatkan oleh para peneliti dalam bidang *data mining*, karena kumpulan data yang tersedia sangat terstruktur dan dapat dengan mudah difilter untuk menemukan data yang diinginkan [29].

2. Kepustakaan:

Metode kepustakaan adalah teknik yang melibatkan pencarian dan pengumpulan informasi dari berbagai sumber literatur, seperti internet, *e-book*, serta jurnal yang relevan dengan topik penelitian.

2.2. Metode Pengembangan Analisis Data

Penelitian ini menggunakan model metodologi pengembangan CRISP-DM (*Cross-Industry Standard Process for Data Mining*), sebagai standar umum dalam proyek-proyek penambangan data, analitik, dan ilmu data [30]. CRISP-DM melewati enam tahapan pengembangan, yaitu pemahaman bisnis (*business understanding*), pemahaman data (*data understanding*), persiapan data (*data preparation*), pemodelan (*modeling*), evaluasi

(*evaluation*), dan penyebaran (*deployment*). Setiap tahapan yang berurutan bersifat fleksibel, di mana kemajuan ke tahap berikutnya bergantung pada hasil dari tahap sebelumnya [31].

11

3. Hasil dan Pembahasan

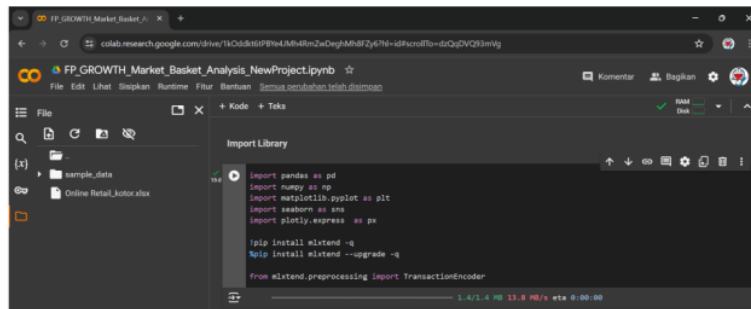
3.1. Pemahaman Bisnis (*Business Understanding*)

Tahap pemahaman bisnis berfokus pada mendefinisikan kebutuhan dan tujuan yang ingin dicapai oleh bisnis. Dalam bisnis penjualan, memberikan pengalaman belanja dan pelayanan terbaik kepada pelanggan merupakan strategi utama untuk meningkatkan penjualan. Dengan mengekstrak informasi dari *dataset* transaksi penjualan, bisnis dapat memperoleh wawasan penting, mengingat aktivitas belanja dilakukan setiap hari. Oleh karena itu, analisis keranjang belanja dilakukan untuk menemukan model aturan asosiasi yang dapat membantu pengambilan keputusan dalam penjualan ritel *non-toko*.

14

3.2. Pemahaman Data (*Data Understanding*)

Tahap pemahaman data adalah proses mempelajari *data* untuk memastikan bahwa *data* yang tersedia sesuai dengan tujuan analisis bisnis. *Dataset* transaksi penjualan ritel *non-toko* ini berlokasi di Inggris. Barang-barang yang dijual, yaitu dekorasi dan aksesoris unik dengan rentang waktu pengumpulan data dari 01/12/2010 hingga 09/12/2011 [32]. Pada tahap awal analisis, seperti yang terlihat pada Gambar 2, *dataset* diimpor ke dalam *tools Google Colab* menggunakan bahasa pemrograman *Python*, memungkinkan pemrosesan *dataset* secara menyeluruh. Selanjutnya, Gambar 3 menunjukkan bahwa *dataset* terdiri dari 541.909 baris data transaksi $\times$ 8 kolom. Jumlah baris transaksi yang ditampilkan masih termasuk transaksi duplikat yang belum dihitung secara unik berdasarkan *InvoiceNo*. Delapan kolom dari *dataset* ini memuat atribut-atribut transaksi yang tidak memiliki hubungan sebab-akibat untuk diklasifikasikan sebagai atribut terikat dan bebas, melainkan sebagai atribut yang mendeskripsikan fakta terkait faktur, seperti yang terlihat pada Tabel 1.



Gambar 2. Memuat *Dataset* ke *Google Colab*

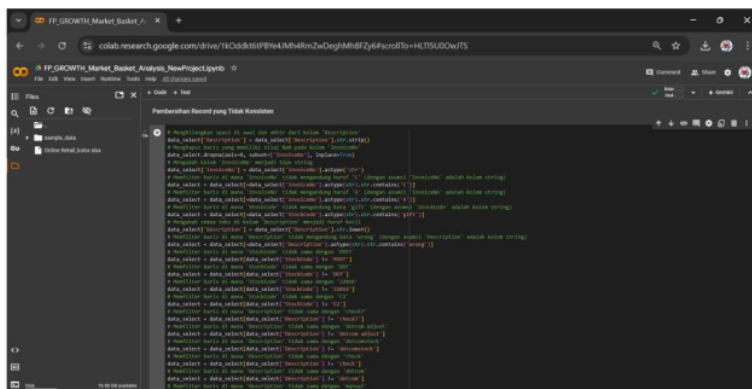
InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country
0	536365	WHITE HANGING HEART T-SHIRT HOLDER	6	2010-12-01 08:26:00	2.55	17850.0	United Kingdom
1	536365	WHITE METAL LANTERN	6	2010-12-01 08:26:00	3.39	17850.0	United Kingdom
2	536365	CIGARETTE CASES	6	2010-12-01 08:26:00	2.75	17850.0	United Kingdom
3	536365	WHITE METAL LANTERN	6	2010-12-01 08:26:00	3.39	17850.0	United Kingdom
4	536365	RED WOOLLY HOTTIE WHITE HEART	6	2010-12-01 08:26:00	3.39	17850.0	United Kingdom
541904	561567	PACK OF 20 SPACEBOY NAPOLEONS	12	2011-12-09 12:50:00	0.85	12080.0	France
541905	561567	CHILDREN'S APRON DOLLY GIRL	6	2011-12-09 12:50:00	2.10	12080.0	France
541906	561567	CHILDREN'S COUTURE DOLLY GIRL	4	2011-12-09 12:50:00	4.15	12080.0	France
541907	561567	CHILDREN'S COUTURE CIRCUS FRAVASE	4	2011-12-09 12:50:00	4.15	12080.0	France
541908	561567	BANKING SET 5 PIECE DETROITPORT	3	2011-12-09 12:50:00	4.95	12080.0	France

Gambar 3. Tampilan *Dataset* Transaksi Penjualan Ritel *Non-Toko*

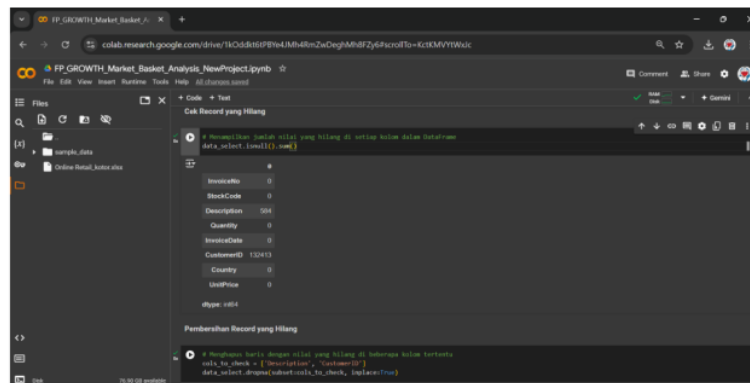
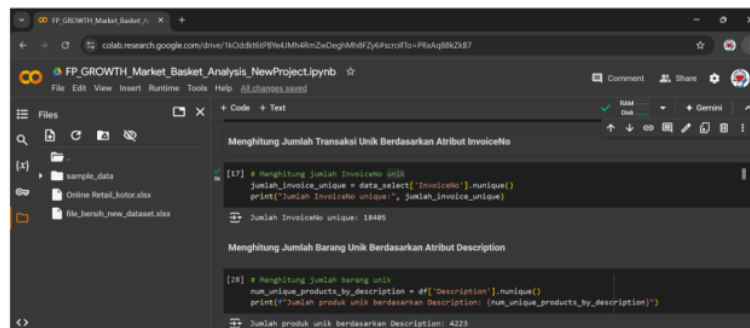
Tabel 1. Atribut-atribut Dalam <i>Dataset</i> Transaksi Penjualan Ritel <i>Non-Toko</i>				
No	Atribut/variabel (Dalam bahasa Inggris)	Atribut/variabel (Dalam bahasa Indonesia)	Deskripsi atribut/variabel	Tipe data
1	<i>InvoiceNo</i>	Nomor Faktur	Merupakan pengenalan unik untuk setiap penjualan	<i>Object</i>
2	<i>StockCode</i>	Kode Stok	Merupakan kode untuk menemukan barang yang dijual	<i>Object</i>
3	<i>Description</i>	Deskripsi	Memberikan informasi tentang nama barang yang dijual	<i>Object</i>
4	<i>Quantity</i>	Kuantitas	Memberikan informasi jumlah barang yang dijual dalam satu faktur	<i>Int64</i>
5	<i>InvoiceDate</i>	Tanggal Faktur	Memberikan informasi kapan transaksi penjualan terjadi, memuat jam, tanggal, bulan, dan tahun	<i>Datetime</i>
6	<i>UnitPrice</i>	Harga Satuan	Merupakan harga per unit barang yang dijual	<i>Float64</i>
7	<i>CustomerID</i>	ID Pelanggan	Merupakan pengenalan unik untuk pelanggan yang membeli barang	<i>Float64</i>
8	<i>Country</i>	Negara	Menunjukkan negara tujuan pengiriman barang kepada pelanggan	<i>Object</i>

3.3. Persiapan Data (*Data Preparation*)

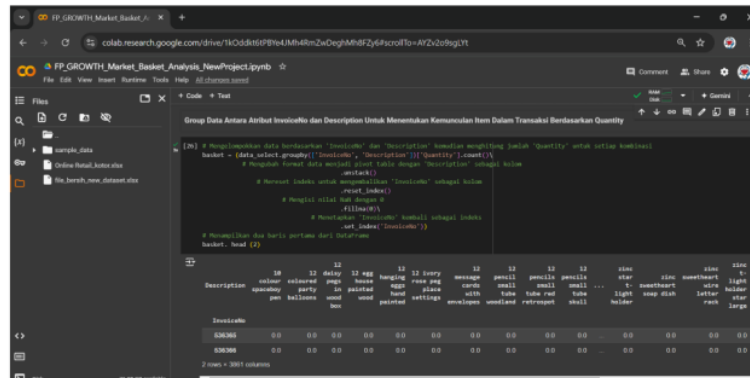
Dalam tahap persiapan data, proses dimulai dari pembersihan pada setiap atribut untuk menghapus nilai *record* yang tidak konsisten, terlihat pada Gambar 4 dan menghapus nilai *record* yang hilang, terlihat pada Gambar 5. Langkah selanjutnya, seperti yang terlihat pada Gambar 6, melibatkan penghitungan secara unik jumlah transaksi berdasarkan atribut *InvoiceNo* dan barang berdasarkan atribut *Description*. Dimana terdapat 18.405 jumlah transaksi dan 4.223 jumlah barang. Lalu, pada Gambar 7 menunjukkan pemilihan atribut nomor faktur (*invoice no*), nama barang (*description*), dan kuantitas (*quantity*) yang kemudian dikelompokkan dan diubah menjadi tabel pivot. Setiap baris dalam tabel pivot mewakili nomor faktur, sementara setiap kolom mewakili nama barang. Jumlah kemunculan barang dalam setiap transaksi dihitung menggunakan kode `.count()` berdasarkan *quantity*. Artinya, terlepas dari jumlah barang (*item*) yang dibeli, baik 1 atau 100, setiap kemunculan barang tetap dihitung sebagai satu kemunculan dengan nilai 1.0 per transaksi. Nilai 1.0 ini muncul karena tipe data kolom adalah *float*, dan dipengaruhi oleh penggunaan kode `.fillna(0)`, yang secara otomatis menggantikan nilai kosong (NaN) dengan 0.0, jika barang tidak ada dalam transaksi. Sedangkan, Gambar 8 menunjukkan proses *one-hot encoding* menggunakan kode `def encode_units(x): if x <= 0: return 0 if x >= 1: return 1`. Kode ini merepresentasikan setiap nilai barang sebagai vektor biner, di mana nilai 1.0 diganti menjadi 1 (barang dibeli) dan nilai 0.0 diganti menjadi 0 (barang tidak dibeli). Melalui proses ini, data kategorikal diubah menjadi format numerik yang lebih sederhana dan dapat dipahami oleh algoritma pembelajaran mesin, sehingga analisis dapat dilakukan dengan lebih baik pada tahap pemodelan.

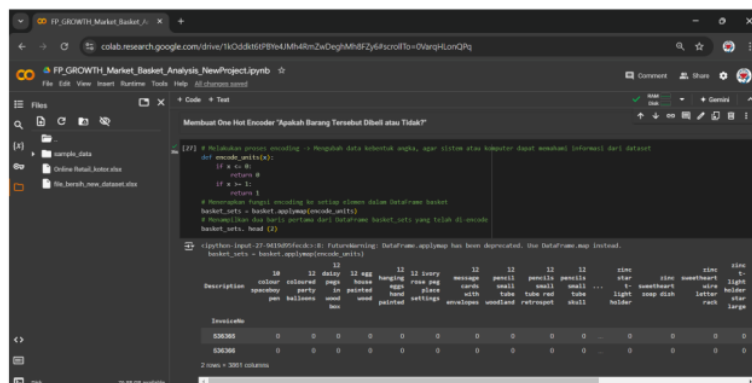


Judul Artikel (Penulis Utama tanpa gelar)

Gambar 4. Penghapusan Nilai *Record* yang InkonsistensiGambar 5. Penghapusan Nilai *Record* yang Hilang

Gambar 6. Menghitung Jumlah Transaksi dan Barang Secara Unik

Gambar 7. Pemilihan Atribut yang Sesuai dari *Dataset*



Gambar 8. Proses One-Hot Encoding

3.4. Pemodelan (Modeling)

Pada tahap ini, model dibangun menggunakan algoritma *FP-Growth* dengan perbandingan berbagai nilai *minimum support*, yaitu 0,09% (0,0009), 0,5% (0,005), 1% (0,01), 1,5% (0,015), 2% (0,02), dan 2,5% (0,025). Interval nilai yang dipilih tidak terlalu jauh maupun terlalu dekat, sehingga analisis dapat dilakukan dengan lebih detail. Model ini menghasilkan beragam jumlah *frequent itemset*, waktu pemrosesan, ukuran penyimpanan, hingga terbentuknya aturan asosiasi, seperti terlihat pada Tabel 2. Masing-masing memiliki kode tertentu untuk menyelesaikannya, antara lain:

1. Kode untuk membangun model dengan berbagai nilai *minimum support*:

```
from mlxtend.frequent_patterns.fpgrowth import fpgrowth frequent_itemsets = fpgrowth(basket_sets, min_support=#sesuaikan nilai minimum support, use_colnames = True) frequent_itemsets
```
2. Kode untuk menghitung jumlah *frequent itemset* yang dihasilkan:

```
combination_counts = frequent_itemsets['itemsets'].apply(lambda x: len(x).value_counts().sort_index() print("Jumlah kombinasi itemset yang terbentuk:") for length, count in combination_counts.items(): print(f"length: {length}-itemset: {count}")
```
3. Kode untuk menampilkan waktu pemrosesan model:

```
from mlxtend.frequent_patterns import fpgrowth import time MIN_SUPPORT = #sesuaikan nilai minimum support start_time = time.time() frequent_itemsets = fpgrowth(basket_sets, min_support=MIN_SUPPORT, use_colnames=True) end_time = time.time() computation_time = end_time - start_time print("Waktu komputasi FP-Growth dengan min_support =", MIN_SUPPORT, ":", computation_time, "detik")
```
4. Kode untuk menyimpan ukuran model:

```
import pickle from google.colab import files with open('frequent_itemsets_#sesuaikan nilai minimum support.pkl', 'wb') as f: pickle.dump(frequent_itemsets, f) files.download('frequent_itemsets_#sesuaikan nilai minimum support.pkl')
```
5. Kode untuk membentuk aturan asosiasi dari *frequent itemset*:

```
from mlxtend.frequent_patterns import association_rules data_asosiasi = association_rules(frequent_itemsets, metric="lift") data_asosiasi=data_asosiasi[['antecedents', 'consequents', 'support', 'confidence', 'lift']]
```
6. Kode untuk menghitung jumlah aturan asosiasi yang terbentuk: 12

```
jumlah_aturan_asosiasi = len(data_asosiasi) print(f"Jumlah aturan asosiasi yang terbentuk: {jumlah_aturan_asosiasi}")
```

Tabel 2. Performa Model

No	Nilai <i>minimum support</i>	Jumlah <i>frequent itemsets</i> yang dihasilkan	Waktu pemrosesan <i>frequent itemsets</i>	Ukuran penyimpanan <i>frequent itemsets</i>	Aturan asosiasi yang terbentuk
1	2,5% (0,025)	1-itemset: 140 2-itemset: 9	2.97 detik	7 kb	18 aturan

2	2% (0,02)	1- itemset: 206 2-itemset: 35 3-itemset: 1	3.95 detik	10 kb	76 aturan
3	1,5% (0,015)	1-itemset: 359 2-itemset: 85 3-itemset: 4	4.12 detik	17 kb	194 aturan
4	1% (0,01)	1-itemset: 621 2-itemset: 304 3-itemset: 53 4-itemset: 2	7.02 detik	33 kb	954 aturan
5	0,5% (0,005)	1-itemset: 1248 2-itemset: 1733 3-itemset: 757 4-itemset: 255 5-itemset: 45 6-itemset: 1	10.98 detik	110 kb	12990 aturan
6	0,09% (0,0009)	1-itemset: 2641 2-itemset: 77613 3-itemset: 88380 4-itemset: 72048 5-itemset: 66609 6-itemset: 49109 7-itemset: 27112 8-itemset: 10712 9-itemset: 2533 10-itemset: 261 11-itemset: 9	53.30 detik	11.383 kb	-

3.5. Evaluasi (Evaluation)

Tahap evaluasi dilakukan setelah fase pemodelan untuk menilai kualitas aturan asosiasi yang terbentuk menggunakan parameter nilai *confidence* dan *lift*, seperti yang terlihat pada Tabel 3. Masing-masing memiliki kode tertentu untuk menyelesaikannya, antara lain:

- Kode untuk menghitung nilai *confidence*:

```
min_confidence = data_asosiasi['confidence'].min() * 100
max_confidence = data_asosiasi['confidence'].max() * 100
print(f"Confidence minimum: {min_confidence}%")
print(f"Confidence maksimum: {max_confidence}%")
```
- Kode untuk menghitung nilai *lift*:

```
data = data_asosiasi['lift']
df = pd.DataFrame(data)
def count_lift(df, criteria):
    return df.query(criteria).shape[0]
lift_1_count = count_lift(df, "lift == 1")
lift_gte_1_count = count_lift(df, "lift >= 1")
lift_lt_1_count = count_lift(df, "lift < 1")
lift_min = df['lift'].min()
lift_max = df['lift'].max()
print("Jumlah data dengan lift == 1:", lift_1_count)
print("Jumlah data dengan lift >= 1:", lift_gte_1_count)
print("Jumlah data dengan lift < 1:", lift_lt_1_count)
print("Nilai lift minimum:", lift_min)
print("Nilai lift maksimum:", lift_max)
```

Tabel 3. Performa Aturan Asosiasi

No	Nilai <i>minimum support</i>	Aturan asosiasi yang terbentuk	Nilai <i>confidence</i>	Nilai <i>lift</i>
1	2,5% (0,025)	18 aturan	Minimum 34.13% Maksimum 78.29%	Minimum 6.75 Maksimum 18.40 Semua aturan asosiasi >1
2	2% (0,02)	76 aturan	Minimum 21.95% Maksimum 89.45%	Minimum 3.80 Maksimum 23.86 Semua aturan asosiasi >1
3	1,5% (0,015)	194 aturan	Minimum 14.16% Maksimum 89.45%	Minimum 1.89 Maksimum 29.29 Semua aturan asosiasi >1
4	1% (0,01)	954 aturan	Minimum 9.39% Maksimum 93.91%	Minimum 1.29 Maksimum 66.38 Semua aturan asosiasi >1
5	0,5% (0,005)	12990 aturan	Minimum 4.72% Maksimum 99.25%	Minimum 0.99 Maksimum 134.43 12986 aturan asosiasi >1 dan 4 aturan asosiasi <1
6	0,09% (0,0009)	-	-	-

3.6. Penyebaran (*Deployment*)

Pada tahap penyebaran, penting untuk memilih model aturan asosiasi berdasarkan performa yang diperoleh dari perbandingan berbagai nilai *minimum support* yang ditentukan. Pemilihan ini dilakukan dengan mempertimbangkan beberapa kriteria, sebagai berikut:

1. Jumlah *frequent itemset*:

Semakin banyak *frequent itemset* yang dihasilkan, semakin baik karena akan menghasilkan lebih banyak aturan. Namun, perlu diperhatikan agar tidak terjadi ledakan kombinatorial yang dapat membuat analisis menjadi kurang bermakna.

2. Kompleksitas komputasi:

Semakin sedikit waktu pemrosesan dan semakin kecil ukuran model yang disimpan, semakin baik karena akan mempercepat waktu eksekusi, sehingga menghindari *downtime* saat aplikasi dijalankan.

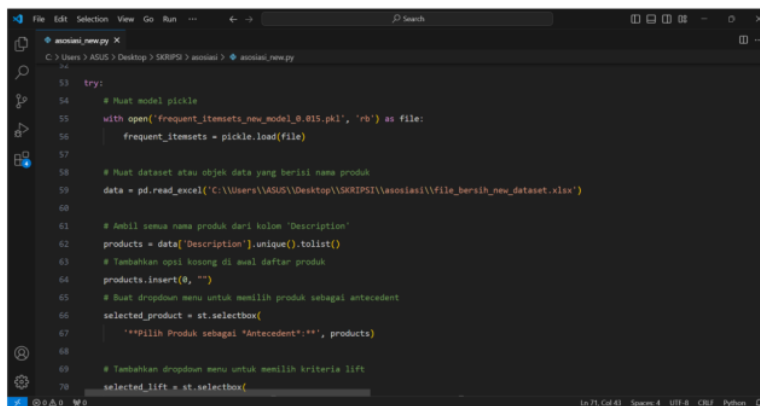
3. Parameter nilai *confidence*:

Nilai *confidence* yang lebih tinggi menunjukkan aturan asosiasi yang lebih kuat, namun, nilai di bawah 100% masih dianggap cukup penting, karena evaluasi aturan asosiasi juga memperhitungkan nilai *lift*.

4. Parameter nilai *lift*:

Nilai *lift* menunjukkan validasi aturan asosiasi. Jika nilai *lift* lebih dari 1, aturan dianggap positif, karena *item-item* tersebut sering dibeli bersamaan dalam transaksi. Sebaliknya, jika nilai *lift* kurang dari 1, aturan dianggap negatif, karena *item-item* tersebut jarang dibeli bersamaan dalam transaksi. Dan nilai *lift* sama dengan 1 menunjukkan bahwa *item-item* tersebut hanya kebetulan dibeli bersamaan dalam transaksi dan tidak memiliki pengaruh khusus.

Dari berbagai pertimbangan tersebut, model ketiga dengan nilai *minimum support* 1,5% (0,015) memberikan keseimbangan terbaik, karena menghasilkan lebih banyak 3-*frequent itemsets* tanpa ledakan kombinatorial. Model pertama dan keenam dieliminasi karena kurangnya *frequent itemsets* dan terjadinya ledakan kombinatorial. Model kedua, keempat, dan kelima menghasilkan banyak *frequent itemsets*, tetapi model kelima dieliminasi karena menghasilkan *frequent itemsets* dengan aturan yang lebih sedikit dibandingkan model ketiga. Model keempat menghasilkan lebih banyak *frequent itemsets* serta memiliki nilai *confidence* dan *lift* yang tinggi, namun memerlukan lebih banyak sumber daya dibandingkan model ketiga. Oleh karena itu, model ketiga digunakan sebagai acuan dalam memberikan pola kombinasi barang (*item*) yang ditemukan dari *dataset* penjualan ritel *non-toko*. Model aturan asosiasi ini akan disampaikan melalui aplikasi analisis asosiasi berbasis web yang mudah digunakan, dengan antarmuka pengguna yang sederhana, intuitif, dan interaktif, yang dibangun menggunakan *library Python Streamlit* versi 1.36.0. Selain *python*, pembaruan ini juga memungkinkan adopsi bahasa pemrograman CSS dan HTML di *Visual Studio Code*. Seperti terlihat pada Gambar 9, kode ini memuat model ketiga dan *dataset* penjualan ritel *non-toko*, saat pengguna memilih barang (*item*) dari *dataset*, fungsi *def analyze_association*, yang terintegrasi langsung dengan model, akan dipanggil untuk melakukan analisis. Setelah itu, aturan asosiasi akan ditampilkan.



```
File Edit Selection View Go Run ... Search
association.py X
C:\Users\ASUS\Desktop>SCRIPSI>association.py
13 try:
14     # Muat model pickle
15     with open('frequent_itemsets_new_model_0.015.pkl', 'rb') as file:
16         frequent_itemsets = pickle.load(file)
17
18     # Muat dataset atau objek data yang berisi nama produk
19     data = pd.read_excel('C:\Users\ASUS\Desktop\SCRIPSI\asosiasi\file_bersih_new_dataset.xlsx')
20
21     # Ambil semua nama produk dari kolom 'Description'
22     products = data['Description'].unique().tolist()
23     # Tambahkan opsi kosong di awal daftar produk
24     products.insert(0, '')
25
26     # Buat dropdown menu untuk memilih produk sebagai antecedent
27     selected_product = st.selectbox(
28         '**Pilih Produk sebagai "Antecedent":**', products)
29
30     # Tambahkan dropdown menu untuk memilih kriteria lift
31     selected_lift = st.selectbox(
```

Gambar 9. Membangun Aplikasi Analisis Asosiasi Menggunakan *Streamlit* di *Visual Studio Code*

Aplikasi ini berhasil di-hosting melalui tools *Streamlit Share* yang terintegrasi dengan repository *GitHub* dan dapat diakses melalui tautan berikut <https://s.id/AplikasiAnalisisAsosiasiFP-Growth> ketika aplikasi digunakan, jika *item A* sebagai *antecedent*, yaitu penyebab yang mengarahkan pada pembelian *item consequent* memiliki aturan asosiasi, maka *item B* akan muncul sebagai *consequent*, yaitu *item* yang dibeli sebagai hasil dari pembelian *antecedent*. Akan tetapi, jika *item A* sebagai *antecedent* yang dipilih tidak memiliki aturan asosiasi, maka muncul pesan bahwa tidak ada aturan asosiasi yang ditemukan.

Penentuan nilai *minimum support* yang tepat sangat krusial dalam menemukan aturan asosiasi. Nilai yang terlalu tinggi dapat menyebabkan terlewatnya aturan penting, sedangkan nilai yang terlalu rendah dapat menghasilkan terlalu banyak kombinasi yang kurang relevan (ledakan kombinatorial). Oleh karena itu, eksperimen dengan berbagai nilai *minimum support* sangat diperlukan dalam analisis keranjang belanja untuk menemukan model yang efisien secara kompleksitas komputasi serta menghasilkan aturan asosiasi yang relevan dan bermakna dengan kualitas evaluasi yang baik.

4. Simpulan

Analisis keranjang belanja dilakukan melalui tahapan CRISP-DM. Dengan membandingkan berbagai nilai *minimum support*, ditemukan bahwa model dengan *minimum support* 1,5% (0,015) memberikan keseimbangan terbaik. Model ini menghasilkan jumlah *frequent itemsets* dengan 1-*itemset* sebanyak 356, 2-*itemset* sebanyak 85, dan 3-*itemset* sebanyak 4, dengan waktu pemrosesan 4,12 detik dan ukuran penyimpanan 17 KB, serta mampu membentuk 194 aturan asosiasi. Evaluasi aturan asosiasi berdasarkan parameter nilai *confidence* menunjukkan nilai maksimum sebesar 89,45% dan minimum sebesar 14,16%. Lalu berdasarkan evaluasi parameter nilai *lift*, semua aturan asosiasi memiliki nilai > 1, dengan maksimum sebesar 29,29 dan minimum sebesar 1,89. Hal ini menunjukkan bahwa semua aturan asosiasi kuat dan sering terjadi, di mana *item B (consequent)* sering dibeli bersamaan dengan *item A (antecedent)* dalam satu keranjang belanja oleh pelanggan. Pengetahuan yang diperoleh dari model aturan asosiasi yang terbentuk memungkinkan rekomendasi *item B (consequent)* ditampilkan secara *online* ketika pelanggan membeli *item A (antecedent)*. Selain itu, karyawan dapat menempatkan barang (*item*) secara berdekatan di gudang penyimpanan, dengan harapan dapat meningkatkan pengalaman belanja dan pelayanan pelanggan yang lebih baik. Untuk memudahkan pemahaman dalam pengambilan keputusan oleh pemilik bisnis, hasil temuan model aturan asosiasi ini disampaikan melalui aplikasi analisis asosiasi berbasis *website* yang mudah digunakan dan telah berhasil di-hosting, sehingga dapat diakses melalui tautan web. Pengembangan selanjutnya dapat menambahkan parameter nilai *leverage* dan *conviction* untuk mengevaluasi lebih lanjut kualitas aturan asosiasi yang terbentuk. Selain itu, uji coba lapangan perlu dilakukan untuk melihat bagaimana analisis berfungsi dalam praktik.

Daftar Referensi

Jutisi Dewa Ayu Blind

ORIGINALITY REPORT

10%

SIMILARITY INDEX

9%

INTERNET SOURCES

4%

PUBLICATIONS

6%

STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Sekolah Tinggi Sandi Negara Student Paper	2%
2	ojs.stmik-banjarbaru.ac.id Internet Source	1%
3	Submitted to UC, Boulder Student Paper	1%
4	www.researchgate.net Internet Source	1%
5	pt.scribd.com Internet Source	<1%
6	eprints.kwikkiangie.ac.id Internet Source	<1%
7	medium.com Internet Source	<1%
8	Submitted to Concordia University Saint Paul Student Paper	<1%
9	Dedy Elisa Limbong. "Rekonsiliasi Data Keuangan dan Karakterisasi Satuan Kerja dengan Teknik Data Mining", Indonesian	<1%

Treasury Review Jurnal Perbendaharaan
Keuangan Negara dan Kebijakan Publik, 2016
Publication

10	Submitted to University of Westminster Student Paper	<1 %
11	es.scribd.com Internet Source	<1 %
12	repository.usd.ac.id Internet Source	<1 %
13	Harianto Harianto, Hadryan Eddy. "Analisa data transaksi penjualan barang menggunakan algoritme Apriori dan FP-Growth", JNANALOKA, 2020 Publication	<1 %
14	digilib.uinsgd.ac.id Internet Source	<1 %
15	ejournal.itn.ac.id Internet Source	<1 %
16	real.mtak.hu Internet Source	<1 %
17	tunasbangsa.ac.id Internet Source	<1 %
18	Agus Fajar Riany, Gusmelia Testiana. "Penerapan Data Mining untuk Klasifikasi Penyakit Stroke Menggunakan Algoritma Naïve Bayes", Jurnal SAINTEKOM, 2023	<1 %

19	repository.unair.ac.id Internet Source	<1 %
20	jasaseomurah-id.blogspot.com Internet Source	<1 %
21	oak.go.kr Internet Source	<1 %
22	ojs.uma.ac.id Internet Source	<1 %
23	www.analyticsvidhya.com Internet Source	<1 %
24	Di Wu. "Data Mining with Python - Theory, Application, and Case Studies", CRC Press, 2024 Publication	<1 %

Exclude quotes Off

Exclude matches Off

Exclude bibliography Off

Jutisi Dewa Ayu Blind

PAGE 1

PAGE 2

PAGE 3

PAGE 4

PAGE 5

PAGE 6

PAGE 7

PAGE 8

PAGE 9

PAGE 10
