



Plagiarism Checker X - Report

Originality Assessment

7%



Overall Similarity

Date: Aug 23, 2024

Matches: 326 / 4500 words

Sources: 26

Remarks: Low similarity detected, check with your supervisor if changes are required.

Verify Report:

[View Certificate Online](#)

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

I Dewa Ayu Indah Saraswati¹, I Made Agus Oka Gunawan^{2*}, I Made Agus Widiana Putra³

^{1,2,3}Sistem Informasi, Universitas Tabanan, Jl. Wagimin No.8, Kediri, 82121, Tabanan
Bali, Indonesia

dewayuindahsaraswati@gmail.com, agusokagunawan@gmail.com,

imadeagusclass@gmail.com

*e-mail Corresponding Author: agusokagunawan@gmail.com

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 **5 yang lebih mudah dipahami** dan berguna bagi pemilik data [15], sehingga dapat **membantu dalam pengambilan keputusan** [16]. Menurut penelitian Ifa 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 ¹⁶ 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 ¹ 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. Perbedaanannya 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 5 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. 1 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 ¹² 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 ¹² 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, ²² 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].

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.

3.2. Pemahaman Data (Data Understanding)

Tahap pemahaman data adalah proses mempelajari data ⁵ 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 □ 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. ²³ Memuat Dataset ke Google Colab

Gambar 3. Tampilan Dataset Transaksi Penjualan Ritel Non-Toko

Tabel 1. Atribut-atribut Dalam Dataset Transaksi Penjualan Ritel Non-Toko

No

Atribut/variabel

(Dalam bahasa Inggris)

Atribut/variabel

(Dalam bahasa Indonesia)

Deskripsi atribut/variabel

Tipe data

1

InvoiceNo

Nomor Faktur

Merupakan pengenal unik untuk setiap penjualan

Object

2

StockCode

Kode Stok

Merupakan kode untuk menemukan barang yang dijual

Object

3

Description

Deskripsi

Memberikan informasi tentang nama barang yang dijual

Object

4

Quantity

Kuantitas

Memberikan informasi jumlah barang yang dijual dalam satu faktur

Int64

5

InvoiceDate

Tanggal Faktur

Memberikan informasi kapan transaksi penjualan terjadi, memuat jam, tanggal, bulan, dan tahun

Datetime

6

UnitPrice

Harga Satuan

Merupakan harga per unit barang yang dijual

Float64

7

CustomerID

ID Pelanggan

Merupakan pengenalan unik untuk pelanggan yang membeli barang

Float64

8

Country

Negara

Menunjukkan negara tujuan pengiriman barang kepada pelanggan

Object

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.

Gambar 4. Penghapusan Nilai Record yang Inkonsistensi

Gambar 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. **18 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}-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:

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

Tabel 2. Performa Model

No

Nilai minimum support

Jumlah frequent itemsets yang dihasilkan

Waktu pemrosesan frequent itemsets

Ukuran penyimpanan frequent itemsets

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:

1. 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}%")
```

2. 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 minimum support

Aturan asosiasi yang terbentuk

Nilai confidence

Nilai lift

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 ¹ 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 **1** sering dibeli bersamaan dalam transaksi. Sebaliknya, jika **24** 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 aturan dengan **25** nilai lift kurang dari satu. Model kedua 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 **17** 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.

Gambar 9. Membangun Aplikasi Analisis Asosiasi Menggunakan Streamlit 17 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 1 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) ¹ 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

- [1] R. Farmana Putra et al., DATA MINING : Algoritma dan Penerapannya, I. Jambi: PT. Sonpedia Publishing Indonesia, 2023.
- [2] A. Yunita, H. B. Santoso, and Z. A. Hasibuan, “‘Everything ⁷ is data’: towards one big data ecosystem using multiple sources of data on higher education in Indonesia,” J. Big Data, vol. 9, no. 1, 2022, doi: 10.1186/s40537-022-00639-7.
- [3] K. Budi Aryasa and C. Ahmadjayadi, ²⁶ Ainomics - Economic Artificial Intelligence :

Pentingnya Sebuah Data. Jakarta: PT Elex Media Komputindo, 2022.

[4] Situngkir and Hokky, "Indonesia Embraces the Data Science," in Seminar statistics, SEAMS 7th Conference, Universitas Gadjah Mada, Jogjakarta, Indonesia., pp. 1–11, 18 August 2015.

[5] A. Baijuri et al., Analisis Sistem Informasi, I. Padang: CV. Gita Lentera, 2023.

[6] S. Chevalier, "Global retail e-commerce sales 2014-2027", Statista, May 22, 2024.

[Online]. Tersedia: <https://www.statista.com/statistics/379046/worldwide-retail-e-commerce-sales/> [Diakses: 5 Mei 2024].

[7] 2 D. Supriyadi, "Penerapan Association Rule Mining Berbasis Algoritma Frequent Pattern Growth untuk Rekomendasi Penjualan," JATISI (Jurnal Tek. Inform. dan Sist. Informasi), vol. 7, no. 2, pp. 135–148, 2020, doi: 10.35957/jatisi.v7i2.339.

[8] U. Sa'adah, M. Yessi Rochayani, D. Wahyu Lestari, and D. Ayu Lusia, 13 Kupas Tuntas Algoritma Data Mining dan Implementasinya menggunakan R, I. Malang: UB Press, 2021.

[9] E. D. Brown, "Drowning in Data, Starved for Information", Big Data, Featured, October 22, 2014. [Online]. Tersedia: <https://ericbrown.com/drowning-in-data-starved-for-information.htm> [Diakses: 12 Maret 2024].

[10] J. Xu, H. Naseer, S. Maynard, and J. Filippou, "Leveraging 6 Data and Analytics for Digital Business Transformation through DataOps: An Information Processing Perspective," in Seminar Digital Business Transformation through DataOps, ACIS - Australas. Conf. Inf. Syst. Proc., pp. 1–11, 2021.

[11] A. F. Ramadhan, N. I. Jaafar, and F. P. Tajudeen, "Data 14 Governance and its Scientific Outlook in Indonesia: A Literature Review," J. Manag. Inf. Decis. Sci., vol. 24, no. 3, pp. 1–10, 2021.

[12] M. Fitri, L. M. Anastasia, and M. A. Tubagus, Data Mining Konsep Dan Penerapannya, I. Yogyakarta: Deepublish, 2021.

[13] A. T. Y. Al Abd Alazeez, "Data 11 Stream Mining Between Classical and Modern Applications: A Review," J. Educ. Sci., vol. 30, no. 5, pp. 30–43, 2021, doi:

10.33899/edusj.2021.130093.1158.

[14] P. N. Harahap and S. Sulindawaty, "Implementasi 8 Data Mining Dalam Memprediksi Transaksi Penjualan Menggunakan Algoritma Apriori (Studi Kasus PT.Arma Anugerah Abadi Cabang Sei Rampah)," J. Sist. Inf. Kaputama, vol. 4, no. 1, p. 8, 2020.

[15] L. Indah Prahartiwi, S. Informasi, S. Nusa Mandiri, J. Damai No, and W. Jati Barat Jakarta Selatan DKI Jakarta, "Pencarian Frequent Itemset pada Analisis Keranjang Belanja Menggunakan Algoritma FP-Growth," Inf. Syst. Educ. Prof., vol. 2, no. 1, pp. 1–10, 2017.

[16] P. N. Harahap and S. Sulindawaty, "Implementasi Data Mining Dalam Memprediksi Transaksi Penjualan Menggunakan Algoritma Apriori (Studi Kasus PT.Arma Anugerah Abadi Cabang Sei Rampah)," Matics, vol. 11, no. 2, p. 46, 2020, doi: 10.18860/mat.v11i2.7821.

[17] I. Musdalifah and A. Jananto, "Analisis 10 Perbandingan Algoritma Apriori Dan FP-Growth Dalam Pembentukan Pola Asosiasi Keranjang," J. Ilm. Komput., vol. 18, no. 2, p. 10, 2022.

[18] F. Achmad, O. Nurdiawan, and Y. Arie Wijaya, 4 "Analisa Pola Transaksi Pembelian Konsumen Pada Toko Ritel Kesehatan Menggunakan Algoritma Fp-Growth," JATI (Jurnal Mhs. Tek. Inform., vol. 7, no. 1, pp. 168–175, 2023, doi: 10.36040/jati.v7i1.6210.

[19] D. Rachmawati, Y. Cahyana, E. E. Awal, and S. Faisal, "Perbandingan 15 Algoritma Apriori dan Algoritma FP-Growth dalam Menentukan Pola Penjualan Pupuk," vol. 3, no. 1, pp. 21–31, 2024.

[20] A. 3 Anggrawan, M. Mayadi, and C. Satria, "Menentukan Akurasi Tata Letak Barang dengan Menggunakan Algoritma Apriori dan Algoritma FP-Growth," MATRIK J. Manajemen, Tek. Inform. dan Rekayasa Komput., vol. 21, no. 1, pp. 125–138, 2021, doi: 10.30812/matrik.v21i1.1260.

[21] 9 Yusuf Husain, Enny Dwi Oktaviyani, and Sherly Christina, "Analisis Perbandingan Algoritma Apriori, FP-Growth, Dan Eclat dalam Menemukan Pola Pembelian Konsumen," KONSTELASI Konvergensi Teknol. dan Sist. Inf., vol. 3, no. 2, pp. 231–243, 2023, doi:

10.24002/konstelasi.v3i2.7007.

[22] E. S. P. C. Purba, "Implementasi 1 Algoritme Fp-Growth Untuk Market Basket Analysis Dalam Menentukan Product Bundling", ResearchGate, pp. 1–8, February 2018.

[23] J. Ha, M. Kambe, and J. Pe, Data Mining : Concepts and Techniques, III. Amerika: Morgan Kaufmann, 2011.

[24] L. Rokach, O. Maimon, and E. Shmueli, Machine Learning for Data Science Handbook : 21 Data Mining and Knowledge Discovery Handbook, III. Jerman: Springer, 2023.

[25] R. Srikant and R. Agrawal, "Mining 20 quantitative association rules in large relational tables," pp. 1–12, 1996, doi: 10.1145/233269.233311.

[26] V. Karyotis and M. H. . Khouzani, Malware Diffusion Models for Modern Complex Networks : Theory and Applications. Amerika: Morgan Kaufmann, 2016.

[27] Lia Farokhah, "Evaluasi Algoritma Asosiasi Menggunakan Lift Ratio | DATA MINING", YouTube, 15 Desember 2021. [Online]. Tersedia: <https://youtu.be/A2jKP2vke2o?si=X3AkL3nv2qLcnQbq> [Diakses: 25 Maret 2024].

[28] BAKTI, "3 Komponen Sistem Komputer : Hardware, Software dan Brainware", BAKTI KOMINFO, 2024. [Online]. Tersedia: <https://www.baktikominfo.id/id/detail-berita/3-komponen-sistem-komputer-hardware-software-dan-brainware> [Diakses: 25 Maret 2024].

[29] Mustakim, "Empat Sumber Dataset untuk Belajar dan Penelitian Bidang Data Mining", Konten edukasi, Mei 18, 2022. [Online]. Tersedia: <https://mustakim.irpi.or.id/2022/05/18/empat-sumber-dataset-untuk-belajar-dan-penelitian-bidang-data-mining/> [Diakses: 25 Maret 2024].

[30] J. Saltz, I. Shamshurin, and C. Connors, "Predicting data science sociotechnical execution challenges by categorizing data science projects," J. Assoc. Inf. Sci. Technol., vol. 68, no. 12, pp. 2720–2728, Dec. 2017, doi: 10.1002/ASI.23873.

[31] P. Palupiningsih and B. Prayitno, "Implementasi Algoritma Fp-Growth Untuk Penentuan Rekomendasi Produk UMKM berdasarkan Frekuensi Pembelian," J. Teknoinfo\, vol. 17, no. 2, pp. 493–501, 2023.

[32] D. Aha, "Datasets", Home - UCI Machine Learning Repository, 2023. [Online].
Tersedia: <https://archive.ics.uci.edu/> [Diakses: 25 Maret 2024].

Jutisi e-ISSN: 2685-0893 □

□ e-ISSN: 2685-0893

□ 1

Jutisi: 19 Jurnal Ilmiah Teknik Informatika dan Sistem Informasi

Jl. Ahmad Yani, K.M. 33,5 - Kampus STMIK Banjarbaru

Loktabat – Banjarbaru (Tlp. 0511 4782881), e-mail: puslit.stmikbjb@gmail.com

e-ISSN: 2685-0893

p-ISSN: 2089-3787

Judul Artikel (Penulis Utama tanpa gelar)

Jutisi: Vol. , No. , Bulan Tahun : hal.

Judul Artikel (Nama Penulis Utama tanpa gelar)

Sources

1	https://www.academia.edu/26171819/IMPLEMENTASI... INTERNET 1%
2	https://ojs.serambimekkah.ac.id/index.php/jnkti/article/view INTERNET 1%
3	https://jurnal.ulb.ac.id/index.php/informatika/article/view/6128 INTERNET <1%
4	https://repository.uksw.edu/bitstream/123456789... INTERNET <1%
5	pemburukode.com/data-processing-adalah-apa/ INTERNET <1%
6	https://dro.deakin.edu.au/articles/conference... INTERNET <1%
7	https://scholar.ui.ac.id/en/publications/... INTERNET <1%
8	https://www.researchgate.net/publication/338632311... INTERNET <1%
9	https://www.semanticscholar.org/paper/Perbandingan... INTERNET <1%
10	https://garuda.kemdikbud.go.id/documents/detail/2901653 INTERNET <1%
11	https://doaj.org/article/a22ebcbf63454d43b63ac5d9d4af2aa7 INTERNET <1%
12	https://akademia.co.id/teknik-pengumpulan-data... INTERNET <1%
13	https://auth.store.ums.ac.id/product/kupas-tuntas... INTERNET <1%
14	https://eprints.um.edu.my/35570 INTERNET <1%

15	https://www.semanticscholar.org/paper/Analysis-Per... INTERNET <1%
16	https://ejournal.itn.ac.id/index.php/jati/article/... INTERNET <1%
17	https://dewabiz.com/fungsi-visual-studio-code INTERNET <1%
18	bing.com/videos INTERNET <1%
19	https://sinta.kemdikbud.go.id/journals/profile/5073 INTERNET <1%
20	https://scholar.archive.org/work INTERNET <1%
21	https://mmsi.binus.ac.id/2020/09/18/cross- INTERNET <1%
22	http://repository.upi.edu/89841/5/S_MAT_1905589_Chapt... INTERNET <1%
23	https://hamhrmwn.medium.com/belajar-data-mining... INTERNET <1%
24	https://jurnal.untan.ac.id/index.php/justin/... INTERNET <1%
25	https://revou.co/kosakata/algoritma-apriori INTERNET <1%
26	https://play.google.com/store/books/details/... INTERNET <1%

EXCLUDE CUSTOM MATCHES ON

EXCLUDE QUOTES OFF

EXCLUDE BIBLIOGRAPHY OFF