

**IMPLEMENTASI *DATA MINING* UNTUK MENCARI KETERHUBUNGAN ANTAR *ITEM*  
PADA DATA TRANSAKSI PENJUALAN MENGGUNAKAN  
ALGORITMA *APRIORI*  
(STUDI KASUS : THE CANOPY KAFE TANJUNGPINANG)**

Daeng Ajis Susanto<sup>1</sup>, Tekad Matulatan<sup>2</sup>, Dwi Amalia Purnamasari<sup>3</sup>  
daengajis@gmail.com

Program Studi Teknik Informatika, Fakultas Teknik, Universitas Maritim Raja Ali Haji

**Abstract**

*The Canopy cafe has a lot of sales transaction data from month to month. This data is only used for daily and monthly reports by The Canopy cafe. But after that the sales transaction data is no longer used and stored in the database. The researcher then builds a system or application by utilizing data mining techniques using a priori algorithms to obtain information in the form of interrelationships between items or association rules from sales data of The Canopy cafe. System or application test results by inputting 0.07 for minimum support and 0.20 for minimum confidence, namely the formation of 7 association rules with the strongest rule being Fruit Tea Classic Drink (if consumers buy Fruit Tea then buy Classic Drink too) with a support value of 0.08 and the confidence value is 0.33 followed by the lift value of 1.256 and the cosine value of 0.317 and conviction of 0.796. These results can be used by the management of The Canopy cafe to determine its business strategy such as making product recommendations, discounts, packages, and others.*

Keywords: *Data, Mining, Apriori, Association.*

**I. Pendahuluan**

**1.1 Latar Belakang**

Kafe adalah salah satu bentuk usaha bisnis restoran yang mengutamakan pada penyajian tempat yang nyaman untuk bersantai, beristirahat, dan berbincang-bincang sambil menikmati minuman dan makanan seperti kopi serta hidangan-hidangan ringan lainnya (Santoso dan Danisa, 2019). Salah satu kafe yang berada di kota Tanjungpinang adalah *The canopy* kafe. *The canopy* kafe memiliki data historis transaksi penjualan dari bulan ke bulan. Data-data transaksi tersebut biasanya disimpan dalam sebuah *database* untuk keperluan laporan perhari ataupun perbulan. Namun setelah itu data-data tersebut tidak digunakan lagi dan dibiarkan begitu saja dalam *database*.

Data-data yang transaksi penjualan yang telah terkumpul dan tersimpan dapat memberikan manfaat bagi manajemen kafe dalam melakukan kegiatan-kegiatan yang terkait penjualan seperti membuat rekomendasi, paket, ataupun diskon pada produk yang dijual. Dalam hal menggali informasi yang terkandung dalam data transaksi penjualan tersebut dapat menggunakan suatu teknik yang biasa disebut *data mining*. *Data mining* adalah proses yang menggunakan teknik statistic, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai *database* besar (Kusrini dan Luthfi, 2009).

Berdasarkan masalah yang dijabarkan di atas, maka peneliti ingin membantu pihak kafe untuk menemukan informasi yang terkandung di dalam data transaksi penjualan dengan melakukan penelitian dengan judul implementasi *data mining* untuk mencari keterhubungan antar *item* pada data transaksi penjualan *The Canopy Kafe* menggunakan algoritma *apriori*.

## II. Metode Penelitian

### 2.1 Data Mining

*Data mining* adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam *database*. *Data mining* adalah proses yang menggunakan teknik statistic, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai *database* besar (Kusrini dan Luthfi, 2009). Istilah *data mining* dan *knowledge discovery in database* (KDD) sering kali digunakan secara bergantian untuk menjelaskan proses penggalian informasi dalam suatu basis data yang besar, sebenarnya kedua istilah tersebut memiliki konsep yang berbeda, tetapi berkaitan dengan satu sama lain. Salah satu tahapan dalam keseluruhan proses KDD adalah *data mining*. Proses KDD secara garis besar yaitu *data selection, cleaning, transformation, data mining, dan evaluation* (Nofriansyah, 2014).

### 2.2 Algoritma Apriori

Algoritma *apriori* adalah suatu algoritma dasar yang diusulkan oleh Agrawal dan Srikant pada tahun 1994 untuk menentukan *Frequent itemsets* untuk aturan asosiasi *Boolean*. Proses utama yang dilakukan pada algoritma *apriori* untuk mendapatkan *frequent itemset* yaitu:

1. *Join* (penggabungan), Proses ini dilakukan dengan cara mengkombinasikan item dengan item lainnya hingga tidak dapat terbentuk kombinasi lagi.
2. *Prune* (pemangkasan), Proses pemangkasan yaitu hasil dari item yang telah dikombinasikan kemudian dipangkas dengan memakai minimum *support* yang telah ditentukan.

Penting tidaknya aturan asosiatif dapat diketahui melalui *interestingness measure*. Berikut beberapa *interesting measure* (Manimaran dan Velmurugan, 2015):

- a) *Support* adalah presentase kombinasi item dalam *database*. Nilai *support* didapat dari rumus 2.1.

$$Support(A) = \frac{\text{Jumlah Transaksi mengandung A}}{\text{Jumlah Transaksi}} \quad (2.1)$$

Sedangkan nilai *support* dari dua item diperoleh dari rumus 2.2

$$Support(A, B) = \frac{\text{Jumlah Transaksi mengandung A dan B}}{\text{Jumlah Transaksi}} \quad (2.2)$$

- b) *Confidence* adalah tingkat kepercayaan atau kuatnya hubungan antar item. Nilai *confidence* didapat menggunakan rumus 2.3.

$$Confidence(A \Rightarrow B) = \frac{\text{Jumlah Transaksi mengandung A dan B}}{\text{Jumlah Transaksi mengandung A}} \quad (2.3)$$

- c) *Lift* digunakan untuk mengukur seberapa penting aturan yang telah terbentuk berdasarkan nilai *support* dan *confidence*. Nilai *lift* yang lebih besar dari 1 menunjukkan bahwa A dan B muncul lebih sering bersamaan. Apabila nilai *lift* lebih besar dari 1, maka aturan tersebut dapat disimpulkan memberi manfaat. Untuk menghitung *lift* menggunakan rumus 2.4.

$$Lift(A \Rightarrow B) = \frac{confidence(A \Rightarrow B)}{support(B)} \quad (2.4)$$

- d) *Conviction* digunakan untuk mengukur tingkat implikasi. *Conviction* menyatakan bahwa  $conviction(A \Rightarrow B) \neq conviction(B \Rightarrow A)$ . nilai *conviction* didapat dari rumus 2.5.

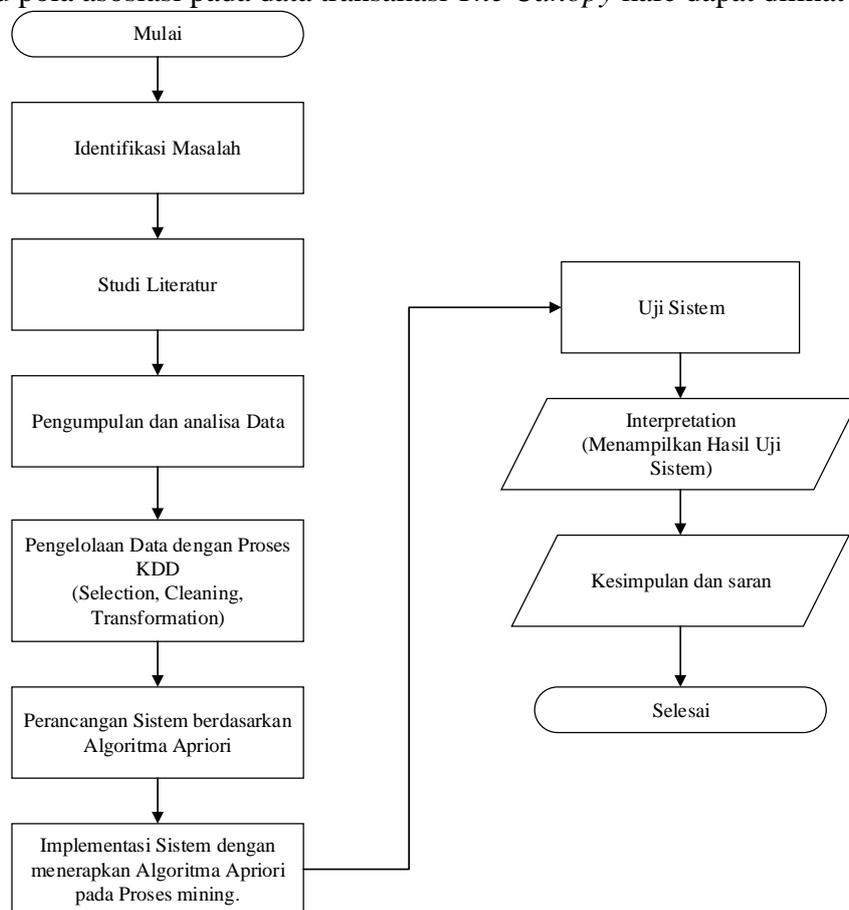
$$Conviction(A \Rightarrow B) = \frac{1 - support\ B}{1 - confidence(A \Rightarrow B)} \quad (2.5)$$

- e) *Cosine* Ini digunakan untuk mengukur kesamaan antara LHS dan RHS. Semakin dekat *cosine* ( $A \Rightarrow B$ ) ke 1, semakin banyak transaksi yang mengandung item A juga mengandung item B. Sebaliknya, semakin dekat *cosine* ( $A \Rightarrow B$ ) ke 0, semakin banyak transaksi mengandung item A tanpa mengandung item B. Nilai *cosine* didapat dari rumus 2.6.

$$cosine = \frac{P(A, B)}{\sqrt{P(A)P(B)}} \quad (2.6)$$

## 2.3 Tahapan Penelitian

Tahapan Penelitian dalam mengimplementasikan teknik *data mining* dalam mencari keterhubungan atau pola asosiasi pada data transaksi *The Canopy* kafe dapat dilihat pada gambar 1.



**Gambar 1.** Tahapan Penelitian

## 2.4 Pengolaan Data

### 2.4.1 Data Selection

Pada penelitian ini data yang digunakan adalah data transaksi penjualan dari *The Canopy* kafe Tanjungpinang. Sebelumnya dilakukan seleksi data pada data transaksi penjualan tersebut karena tidak semua atribut memiliki keterhubungan dan dapat digunakan untuk menemukan informasi yang dibutuhkan. Atribut yang di seleksi adalah kode dan item. Hasil seleksi data dapat dilihat pada tabel 1.

**Tabel 1.** Hasil Seleksi Data

No	Kode	Items
1	TS9OTO	Green Tea Latte HOT x 2
2	TS9OTN	Lemon Tea COLD
3	I7IEVL	Lemon Tea COLD x 2, CHICKEN RICE BOWL Spicy Sweet x 2, Naughty Fries, Mineral Water BIASA
4	TS9OTM	Carlsberg Pint x 2, Carlsberg Pint, Cappucino COLD, ROTI BAKAR Cheese
...	...	...
...	...	...
1265	I7IF27	Lemon Tea COLD, Garlic French Fries Garlic Saos
1266	I7IF26	Hazelnut Latte COLD, Flat White HOT, Vanilla Latte HOT

### 2.4.2 Data Cleaning

Setelah data di seleksi kemudian data di *cleaning* atau dibersihkan terlebih dahulu sebelum di proses. Proses *cleaning* atau membersihkan data mencakup beberapa hal seperti membuang data yang sama persis atau duplikat serta memeriksa apabila ada kesalahan pada data misalnya salah ketik, *custom order*, *refund*, dan sebagainya. Data transaksi yang didapat dari *The Canopy* Kafe sebanyak 1266 *record*. Setelah dibersihkan jumlah data yang dapat digunakan sebanyak 886 *record*.

### 2.4.3 Data Transformation

Selanjutnya data ditransformasikan ke bentuk yang mudah digunakan sebab ada beberapa data yang mengandung informasi yang sama. Atribut yang ditransformasikan adalah atribut Item menjadi kategori. Kategori merupakan pengelompokan dari beberapa item yang memiliki jenis yang sama baik itu varian maupun ukuran. Kemudian data tersebut disimpan dengan format CSV (*Comma Separated Values*) agar lebih mudah untuk digunakan pada *platform* berbeda. Hasil transformasi dapat dilihat pada tabel 2.

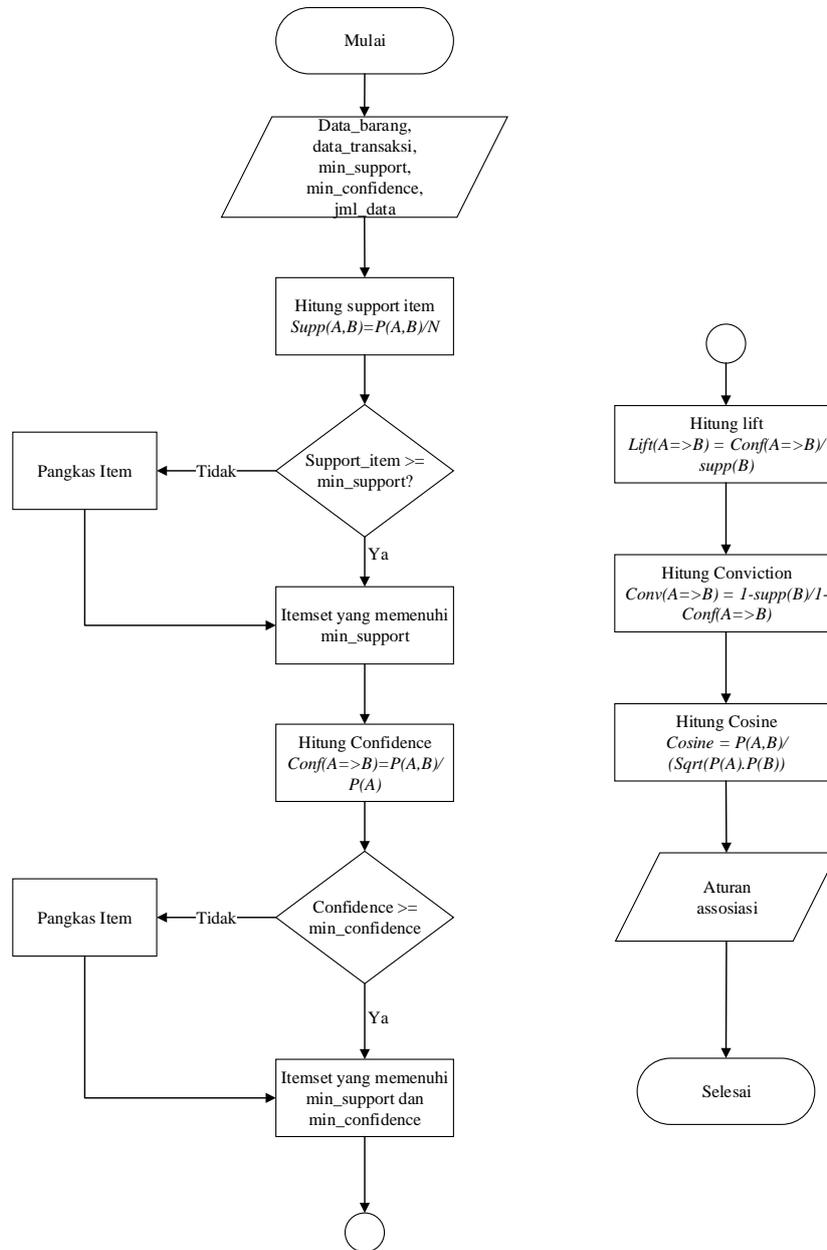
**Tabel 2.** Hasil Transformasi Data

No	Nama Item	Kategori Item
1	Lipton Tea Cold	Lipton Tea
2	Lipton Tea Hot	
3	Lipton Milk Tea Cold	
4	Lipton Milk Tea Hot	
5	Orange Julius	Yogurt Smoothie
6	Strawberry Yogurt Smoothie	
7	Mango Yogurt Smoothie	
8	Grape Yogurt Smoothie	
9	Chocolate cold	Chocolate
10	Chocolate Hot	
11	Chocolate Milkshake	
12	Chocomint	
...	...	...
...	...	...
132	Green Tea Latte Cold	Greentea

## 2.5 Perancangan

### 2.5.1 Perancangan Sistem

Adapun Perancangan sistem aplikasi digambarkan dalam bentuk flowchart yang dapat dilihat pada gambar 2.



Gambar 2. Flowchart Sistem Aplikasi dengan Algoritma Apriori

### 2.5.2 Perancangan Basis Data

Setelah mengumpulkan dan menganalisa data yang akan dibutuhkan termasuk bagian-bagian yang akan berinteraksi dengan basis data, penulis merancang kebutuhan tersebut dalam bentuk struktur tabel yang dapat dilihat pada tabel basis data dibawah. Tabel transaksi dapat dilihat pada tabel 3.

Tabel 3. Tabel transaksi

No	Atribut	Tipe data
1	Id_transaksi	Int (11)
2	Kode	Varchar (10)

Tabel transaksi digunakan untuk menyimpan data yang nantinya diinputkan atau di import kedalam sistem.

**Tabel 4.** Tabel Item

No	Atribut	Tipe data
1	Id_item	Int (11)
2	Kode	Varchar (10)
3	Nama	Varchar (50)

Tabel 4 yaitu tabel item digunakan untuk menyimpan data barang atau item yang digunakan. Tabel ini juga akan digunakan untuk proses lainnya sebagai acuan.

**Tabel 5.** Tabel Transaksi Detail

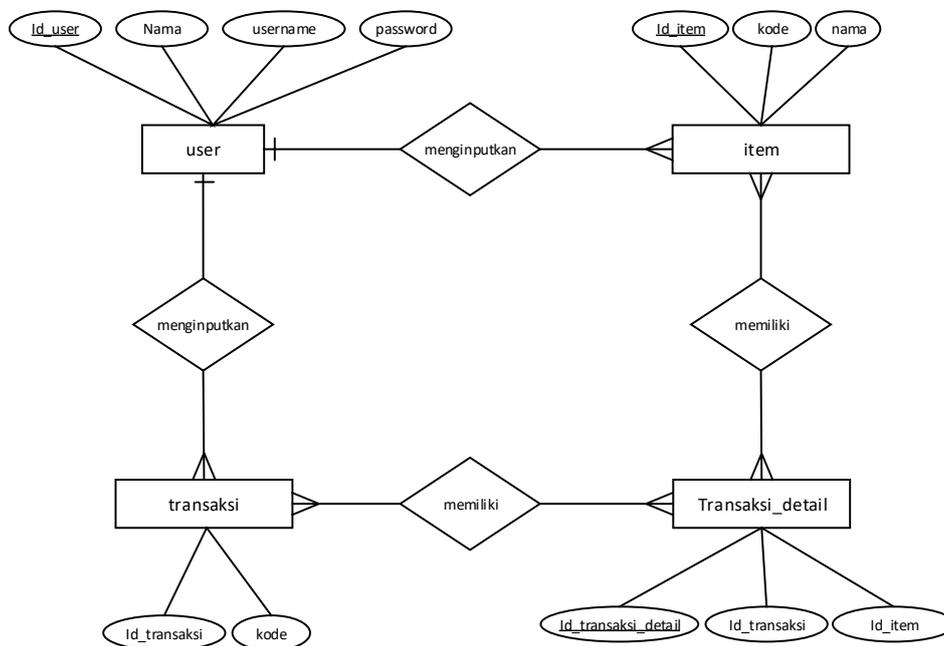
No	Atribut	Tipe data
1	Id_transaksi_detail	Int (11)
2	Id_transaksi	Int (11)
3	Id_item	Int (11)

Tabel 5 yaitu Tabel Transaksi Detail digunakan untuk menyimpan detail transaksi saat proses perhitungan menggunakan algoritma *apriori*.

**Tabel 6.** Tabel User

No	Atribut	Tipe data
1	Id_user	Int (11)
2	Nama	Varchar (50)
3	Username	Varchar (50)
4	Password	Varchar (50)

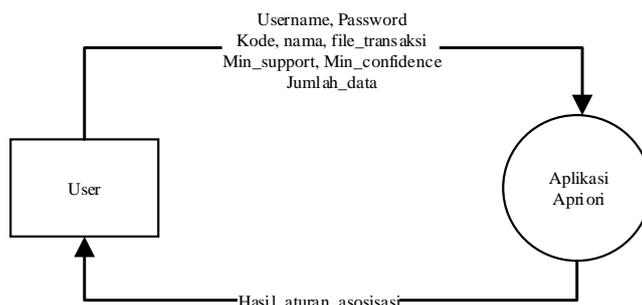
Tabel 6 yaitu Tabel *user* digunakan untuk menyimpan data pengguna. Relasi antar tabel dapat dilihat pada gambar 3.



**Gambar 3.** Entity Relationship Diagram

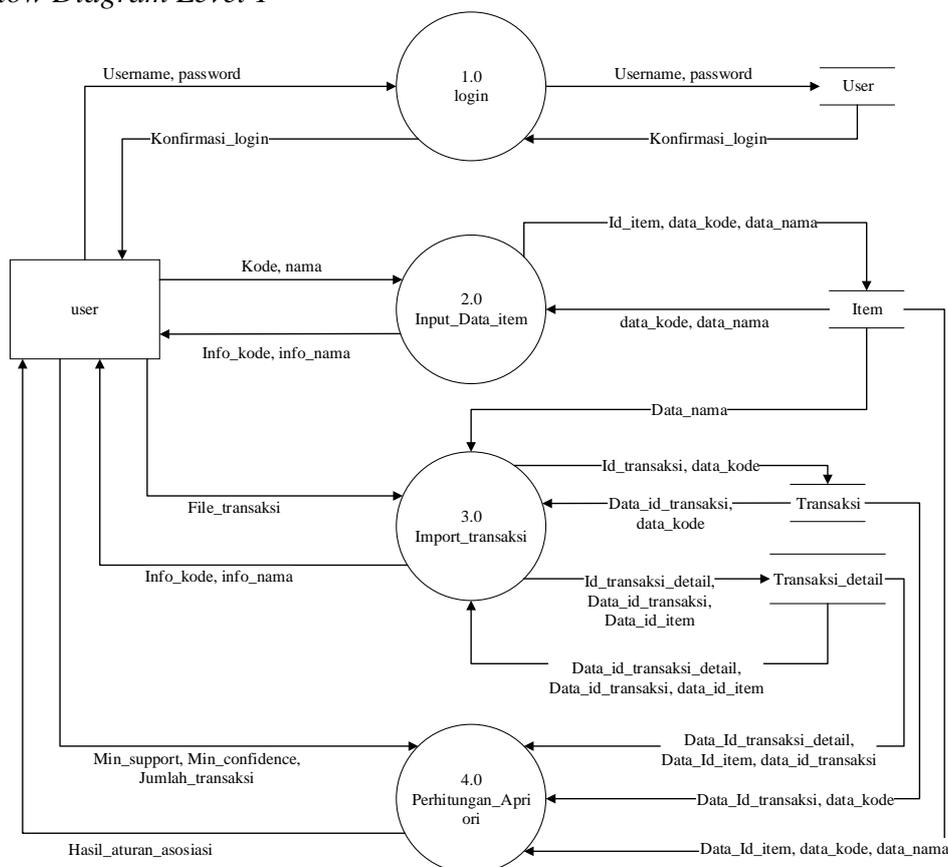
## 2.5.3 Perancangan Data Flow Diagram

### a. Context Diagram



Gambar 4. Context Diagram

### b. Data Flow Diagram Level 1



Gambar 5. Data Flow Diagram Level 1

## III. Hasil dan Pembahasan

### 3.1 Hasil Proses Apriori

Berikut ini merupakan hasil dari proses uji sistem aplikasi yang telah dibangun menggunakan algoritma *apriori* untuk menemukan keterhubungan antar item atau aturan asosiasi. Pada proses ini dilakukan percobaan menggunakan 886 data transaksi yang telah di *cleaning* atau di bersihkan. Jumlah data ini merupakan data selama 3 bulan dimulai dari November - Desember 2019 hingga Januari 2020. Pengujian dilakukan beberapa kali dengan kombinasi minimum *support* dan minimum *confidence*. Untuk minimum *support* digunakan 5 nilai yang berbeda yaitu 0.03, 0.05, 0.07, 0.09 dan 0.10. Sedangkan untuk minimum *confidence* digunakan 5 nilai yang berbeda yaitu 0.10, 0.15, 0.20, 0.25 dan 0.30.

Dapat diamati dari jumlah aturan asosiasi yang terbentuk (Tabel 7), kombinasi minimum *support* yang rendah dan minimum *confidence* yang juga rendah akan menghasilkan aturan yang banyak. Sedangkan jika minimum *support* tinggi dan minimum *confidence* tinggi, maka jumlah aturan yang terbentuk akan menjadi sedikit, bahkan tidak ada. Nilai minimum *support* dan minimum *confidence* yang ideal untuk data transaksi ini adalah 0.07 untuk *support* dan 0.20 untuk *confidence*. Pada percobaan menggunakan minimum *support* 0.07 dan minimum *confidence* 0.20 didapat 7 aturan asosiasi yang dapat dilihat pada tabel 8.

**Tabel 7.** Kombinasi minimum *support* dan minimum *confidence*

No	Minimum Support	Minimum Confidence	1-itemset	2-itemset	Aturan
1	0.03	0.10	32	45	45
2	0.05	0.15	26	17	17
3	0.07	0.20	17	7	7
4	0.09	0.25	14	1	1
5	0.10	0.30	0	0	0

**Tabel 8.** Hasil Aturan Asosiasi

No	Aturan Asosiasi	Support	Confidence	Lift	Cosine	Conviction
1	Jika konsumen membeli Classic Drink, maka membeli French Fries	0.08	0.28	0.886	0.266	1.129
2	Jika konsumen membeli Classic Drink, maka membeli Milkshake	0.07	0.26	0.896	0.25	1.116
3	Jika konsumen membeli French Fries, maka membeli Coffee Latte	0.07	0.21	0.918	0.253	1.09
4	Jika konsumen membeli Fruit Tea, maka membeli Classic Drink	0.08	0.33	1.256	0.317	0.796
5	Jika konsumen membeli Fruit Tea, maka membeli French Fries	0.08	0.33	1.067	0.292	0.937
6	Jika konsumen membeli Milkshake, maka membeli Coffee Latte	0.07	0.24	1.061	0.272	0.943
7	Jika konsumen membeli Milkshake, maka membeli French Fries	0.09	0.32	0.979	0.297	1.021

Berdasarkan tabel 8, dari 7 aturan asosiasi yang terbentuk dapat dilihat bahwa aturan yang memiliki keterhubungan yang baik adalah *Fruit Tea*  $\Rightarrow$  *Classic Drink* (Jika konsumen membeli *Fruit Tea*, maka membeli *Classic Drink*) dengan *support* 0.08 artinya artinya *Fruit Tea* dan *Classic Drink* dibeli secara bersamaan sebesar 0.08 atau 8% dari seluruh transaksi. *Confidence* sebesar 0.33 atau 33% konsumen yang membeli *Fruit Tea* akan membeli *Classic Drink* juga. Nilai *lift* sebesar 1.256 menyatakan bahwa aturan yang terbentuk kuat. Selanjutnya nilai *cosine* sebesar 0.317, pada nilai *cosine* yang mendekati 0 menyatakan bahwa banyak nya konsumen membeli *Fruit Tea* tetapi tidak membeli *Classic Drink*. Serta nilai *conviction* sebesar 0.796.

#### IV. Kesimpulan

Berdasarkan uraian pembahasan sebelumnya maka dapat ditarik kesimpulan bahwa proses pencarian informasi berupa keterhubungan antar item pada data transaksi penjualan *The Canopy* kafe dapat dilakukan dengan membangun sistem aplikasi yang menggunakan algoritma *apriori*. Sistem kemudian diuji dengan data transaksi penjualan *The Canopy* kafe yang telah melalui proses *selection*, *cleaning*, dan *transformation*. Uji sistem dilakukan dengan kombinasi minimum *support* dan minimum *confidence* yang kecil menghasilkan aturan yang banyak. Minimum *support* dan minimum *confidence* yang baik untuk data transaksi penjualan *The Canopy* kafe adalah 0.07 dan 0.20 yang menghasilkan 7 aturan asosiasi. Informasi keterhubungan antar item atau aturan asosiasi yang

dihasilkan sistem aplikasi ini dapat digunakan oleh *management The Canopy* kafe untuk menentukan strategi bisnisnya seperti rekomendasi produk, diskon, paket dan lainnya

## V. Daftar Pustaka

- Anggraeni, R. M. (2014). Perbandingan Algoritma Apriori dan Algoritma FP-Growth untuk Rekomendasi Pada Transaksi Peminjaman Buku di Perpustakaan Universitas Dian Nuswantoro. *Teknik Informatika*, 1–6.
- Fauzy, M., Saleh W, K. R., & Asror, I. (2016). Penerapan Metode Association Rule Menggunakan. *Jurnal Ilmiah Teknologi Informasi Terapan*, II(2), 221–227. <https://doi.org/2407-3911>
- Hermawati, F. A. (2013). *DATA MINING*. Yogyakarta: ANDI.
- Ikhwan, A., Sriani, & Nofriansyah, D. (2015). Penerapan Algoritma Apriori Untuk Menganalisa Transaksi Penjualan Untuk Promo Produk Furniture Jepara. *Konferensi Nasional Pengembangan Teknologi Informasi Dan Komunikasi*, 19–24.
- Kusrini, & Luthfi, E. T. (2009). *Algoritma Data Mining*. Yogyakarta: ANDI.
- Manimaran, J., & Velmurugan, T. (2015). Analysing the quality of association rules by computing an interestingness measures. *Indian Journal of Science and Technology*, 8(15), 1–12. <https://doi.org/10.17485/ijst/2015/v8i15/76693>
- Nofriansyah, D. (2014). *Konsep Data Mining Vs Sistem Pendukung Keputusan*. Yogyakarta: Deepublish.
- Riszky, A. R., & Sadikin, M. (2019). Data Mining Menggunakan Algoritma Apriori untuk Rekomendasi Produk bagi Pelanggan. *Jurnal Teknologi Dan Sistem Komputer*, 7(3), 103. <https://doi.org/10.14710/jtsiskom.7.3.2019.103-108>
- Santoso, B., & Danisa, P. (2019). *Penerapan algoritma apriori sebagai rekomendasi menu itemsets di trotoar steak kafe*. 5(2), 203–210.
- Tan, P., Steinbach, M., & Kumar, V. (2005). *Introduction to Data Mining*. PEARSON.

## Ucapan Terimakasih

Penulis mengucapkan terima kasih yang sebesar-besarnya dan penghargaan yang mulia kepada :

1. Bapak Sapta Nugraha, S.T, M.Eng, selaku Dekan Fakultas Teknik Universitas Maritim Raja Ali Haji
2. Bapak Tekad Matulatan, S.Sos., S.Kom., M.Inf.Tech. dan Ibu Nola Ritha, S.T., Mcs selaku dosen Pembimbing Akademik yang telah memberikan semangat dan arahan.
3. Bapak M. Radzi Rathomi, S.Kom., M.Cs., selaku Kepala Jurusan Program Studi Teknik Informatika.
4. Bapak Tekad Matulatan, S.Sos., S.Kom., M.Inf.Tech. selaku Pembimbing I yang telah menyediakan waktu, tenaga dan pikiran serta membimbing dalam penyusunan skripsi.
5. Ibu Dwi Amalia Purnamasari, S.T., Mcs. selaku Pembimbing II yang telah membimbing penulis dan memberikan masukan serta mengarahkan dalam penyelesaian skripsi.
6. Para Dosen Fakultas Teknik yang telah banyak membantu membekali ilmu dan berbagi pengalaman selama masa perkuliahan, serta staff yang selalu dengan senang hati melayani dalam administrasi perkuliahan.