

# tode\_Association\_Rule\_Mining\_Untuk\_Menentukan\_Peta\_Pemasaran.pdf

*by*

---

**Submission date:** 18-Nov-2022 09:33AM (UTC+0700)

**Submission ID:** 1957354802

**File name:** tode\_Association\_Rule\_Mining\_Untuk\_Menentukan\_Peta\_Pemasaran.pdf (651.91K)

**Word count:** 5887

**Character count:** 34060

# Implementasi Data Mining dengan Metode Association Rule Mining Untuk Menentukan Peta Pemasaran

**9** Mastuhin  
Program Studi Informatika  
Universitas PGRI Yogyakarta  
Yogyakarta, Indonesia  
mastuhin33@gmail.com

Nurirwan Saputra  
Program Studi Informatika  
Universitas PGRI Yogyakarta  
Yogyakarta, Indonesia  
nurirwan@upy.ac.id

Sunggito Oyama  
Program Studi Informatika  
Universitas PGRI Yogyakarta  
Yogyakarta, Indonesia  
shafa.najla@gmail.com

**2**  
**Abstrak**— Dalam bisnis retail, ada banyak strategi yang bisa digunakan oleh sebuah toko retail agar konsumen tetap loyal. Salah satu contohnya dengan menerapkan *sales promotion* yang tepat sasaran. *Sales promotion* yang tepat hanya bisa dilakukan dengan melihat pola kebiasaan membeli konsumen. Oleh karena itu penulis merancang sebuah sistem implementasi data mining menggunakan metode Association Rule Mining dengan algoritma hash-based yang berbasis website. Sistem ini dibangun dengan menggunakan bahasa pemrograman PHP dan database MySQL. Penelitian ini diharapkan dapat menghasilkan sebuah sistem berbasis website yang mampu memberikan sebuah output berupa gambaran pola membeli konsumen yang di dapat dari olah data transaksi penjualan Toko Matahari sehingga pemilik mendapatkan gambaran dalam menentukan strategi pemasaran yang tepat.

**Kata Kunci:** hash-based, PHP, Retail

## **4** I. PENDAHULUAN

Bisnis retail adalah penjualan barang secara eceran pada berbagai tipe gerai seperti **4** kios, pasar, *department store*, butik dan lain-lain yang umumnya untuk dipergunakan langsung oleh pembeli yang bersangkutan. Bisnis retail di Indonesia dapat dibedakan menjadi 2 kelompok besar, yakni retail tradisional dan retail modern. Retail modern pada dasarnya merupakan pengembangan dari retail tradisional. Format retail ini muncul dan berkembang seiring perkembangan perekonomian, teknologi, dan gaya hidup masyarakat yang membuat masyarakat menuntut kenyamanan yang lebih dalam berbelanja [1].

Dalam 2 dekade terakhir, bisnis retail modern mengalami perkembangan yang cukup signifikan. Salah satu faktor yang memicu perkembangan bisnis retail modern adalah perubahan perilaku konsumen [2]. Meningkatnya pendapatan, tuntutan akan kemudahan, dan kenyamanan adalah beberapa contoh faktor yang mempengaruhi terjadinya perubahan perilaku konsumen [3]. Perkembangan retail modern juga menyebabkan konsumen dihadapkan pada banyaknya pilihan tempat untuk **5** melakukan pembelian.

Masing-masing modern retail dengan berbagai caranya, berusaha menarik konsumen untuk mengunjungi gerainya. Modern retail pun memiliki harapan bahwa setiap konsumen yang mengunjungi gerainya muncul kepuasan. Ada berbagai macam strategi yang dilakukan oleh modern retail untuk mencapai kepuasan konsumen, salah satunya adalah dengan diimplementasikannya bauran pemasaran

(*Marketing Mix*) retail di dalam seluruh kegiatan operasionalnya [4].

Salah satu poin penting dalam *marketing mix* adalah *store atmosphere*. *store atmosphere* atau suasana toko digunakan dalam strategi pemasaran dengan tujuan agar konsumen merasa nyaman sehingga menimbulkan kesan menarik dan nantinya akan meningkatkan keputusan pembelian serta dapat mempengaruhi konsumen untuk melakukan pembelian di luar apa yang direncanakan (*unplanned purchase*) [5].

Pada dasarnya memang *marketing mix* ditujukan untuk mengatasi masalah loyalitas *consumer*. Jumlah retail yang terus bertambah berarti bertambahnya jumlah pesaing. Retailer berlomba-lomba agar *consumer* loyal terhadap retail mereka. Namun jika implementasi *marketing mix* tidak dibarengi dengan strategi lain maka masalah lain juga akan muncul, seperti produk tertentu yang tidak laku terjual, segmentasi pasar tidak sesuai, dan *inventory oversupply*. Masalah-masalah ini sangat umum dialami oleh retail modern dengan omzet yang cukup besar.

Masalah yang sama juga dialami oleh Toko Matahari yang notabene memiliki jumlah *consumer* yang cukup banyak. Hasil observasi menunjukkan bahwa ada beberapa produk yang tidak laku terjual sampai kadaluwarsa, dan beberapa produk kelebihan stok.

Masalah di atas sebenarnya bisa dihindari jika retailer memahami *buying habit* dari *consumer*. Sebagai contoh pada masalah produk yang tidak laku, salah satu solusi yang bisa ditawarkan adalah dengan memberikan diskon pada transaksi yang menyertakan produk tersebut, atau meletakkan produk tersebut di dekat produk lain yang memiliki angka penjualan tinggi. Dua solusi ini tentu melihat status dari produk yang tidak laku, apakah karena memang produk baru sehingga belum dikenal *consumer* atau karena memang daya beli *consumer* yang rendah. *Buying habit* bisa diketahui hanya jika data transaksi penjualan diolah kembali dan ditambang sehingga menjadi informasi yang lebih berguna, dan ini belum ada di Toko Matahari. Pengetahuan yang cukup tentang *buying habit* disamping untuk menghindari masalah-masalah di atas juga secara tidak langsung akan meningkatkan omzet penjualan.

Memahami proses keputusan-untuk-membeli (*decision-to-buy*) konsumen adalah dengan memahami isi “kotak hitam” perilaku konsumen [6]. Salah satu metode yang bisa digunakan untuk mengetahui pola perilaku konsumen adalah dengan menggunakan *Data Mining*.

Data mining adalah sebuah proses untuk mendapatkan informasi yang berguna dari gudang basis data yang besar.

3

Data mining juga dapat diartikan sebagai pengestrakan informasi baru yang diambil dari bongkahan data besar yang membantu dalam pengambilan keputusan. Data mining sering juga disebut dengan *Knowledge Discovery in Database* atau disingkat KDD, adalah kegiatan yang meliputi pengumpulan, pemakaian data histori untuk menelusuri data yang ada untuk membangun sebuah model agar dapat mengenali pola data lain yang berukuran besar [7].

Ada beberapa macam pendekatan yang berbeda yang diklasifikasikan sebagai teknik pencarian informasi/pengetahuan dalam KDD. Ada pendekatan kuantitatif, seperti pendekatan probabilistik seperti logika induktif, pencarian pola, dan analisis pohon keputusan. Pendekatan yang lain meliputi deviasi, analisis kecenderungan, algoritma genetik, jaringan saraf tiruan, dan pendekatan campuran dua atau lebih dari beberapa pendekatan yang ada [8].

Salah satu metode dalam data mining yang biasa digunakan untuk analisis perilaku konsumen adalah metode Asosiasi. Metode asosiasi merupakan suatu bentuk metode dalam data mining yang memberikan informasi hubungan antar *item* data di database. Metode tersebut dapat dimanfaatkan secara luas dalam proses bisnis diantaranya dalam proses penjualan. Data mining asosiasi dapat membantu *retailer* untuk mengetahui *buying habit* konsumen sehingga *retailer* dapat memanfaatkan informasi yang diperoleh untuk mengambil tindakan pemasaran yang sesuai [9].

Algoritma yang umum digunakan dalam pencarian *association rule* adalah Apriori, namun algoritma ini memiliki kelemahan dari sisi performa. Algoritma Apriori akan memakan waktu cukup lama terutama jika melibatkan data yang cukup besar, hal ini karena setiap kali menentukan *k-itemset* algoritma ini akan melakukan scan database. Untuk mengatasi masalah yang timbul dalam algoritma Apriori maka sebagai gantinya akan digunakan algoritma *Hash Based*, yakni data yang telah discan akan dimasukkan ke dalam *bucket* pada tabel hash. Dari *bucket* tersebut akan digunakan untuk mencari *frequent (k+1)-itemset* sehingga scan database hanya dilakukan 1 kali pada iterasi pertama [10].

Banyaknya data penjualan yang dihasilkan selama kurun waktu yang cukup lama ternyata masih bisa dimanfaatkan lagi. Dengan adanya aplikasi ini diharapkan dapat digunakan oleh Toko Matahari untuk mengetahui pola kebiasaan pelanggan dalam memilih item barang yang mereka masukkan ke dalam keranjang belanja, sehingga dapat mengatur strategi pemasaran yang sesuai agar *consumer* merasa nyaman dan loyal.

Penelitian ini dilakukan dengan tujuan untuk membantu *owner* Toko Matahari dalam menentukan arah strategi pemasaran. Strategi yang dimaksud seperti yang telah disebutkan di atas adalah strategi bauran pemasaran dan *sales promotion*. Strategi ini tidak mungkin bisa dilakukan secara efektif tanpa mengetahui *buying habit* dari konsumen. Output dari penelitian ini diharapkan dapat memberikan wawasan tentang *consumer buying habit* sehingga *owner* Toko Matahari dapat menentukan arah strategi pemasaran yang efektif.

Proses pencarian *Buying Habit* dilakukan dengan membuat sebuah sistem berbasis web yang dapat mengolah data transaksi penjualan sehingga didapat informasi berupa item-item yang kurang laku, sangat laku terjual, kombinasi item-item yang sering terjual secara bersamaan dan lain sebagainya. Sistem ini dibangun dengan 3 komponen utama yaitu, Laravel sebagai pengolah data, Vue sebagai tampilan, dan MySQL sebagai media penyimpanan data.

## II. TINJAUAN PUSTAKA

Penelitian ini merupakan pengembangan dari penelitian yang telah ada tentang *Market Basket Analysis* dengan menggunakan algoritma apriori. Oleh karena itu dalam penelitian ini menggunakan algoritma *Hash-Based* yang merupakan pengembangan dari algoritma apriori. Berikut ini adalah beberapa penelitian terdahulu yang nantinya akan digunakan sebagai acuan dalam penelitian ini.

Penelitian yang dilakukan oleh Orisky [11] dengan judul “Analisis Perbandingan Algoritma Apriori Dan Algoritma *Hash Based* Pada *Market Basket Analysis* Di Apotek UAD”, fokus utama penelitian ini adalah melakukan perbandingan antara dua algoritma Asosiasi, dua algoritma ini adalah *Apriori* dan *Hash Based*. Parameter yang digunakan adalah kecepatan masing-masing algoritma dalam menambang data. Hasil penelitian menunjukkan bahwa algoritma *Apriori* lebih cepat dibanding dengan algoritma *Hash Based*, namun dalam penelitian hanya menyertakan data dengan jumlah kecil.

Penelitian yang dilakukan oleh Farha Ramadhan tentang performa algoritma *Hash Based* dibanding dengan algoritma *Apriori* dengan judul penelitian “Implementasi Algoritma *Hash Based* Terhadap Aturan Asosiasi Untuk Menentukan *Frequent Itemset* Study Kasus Rumah Makan Seafood ‘Kita’”. Dari penelitian menunjukkan bahwa Algoritma *Hash Based* menjadi solusi bagi kelemahan algoritma *Apriori* didalam menentukan *frequent itemset*. Tahap dalam menentukan *frequent itemset* merupakan tahap yang paling berpengaruh terhadap performa data mining, terutama jika kandidat *itemset* berjumlah sangat besar. Pada algoritma *Hash Based* hanya dilakukan 1 kali scan database yaitu pada awal iterasi [10].

Penelitian tentang implementasi data mining dengan metode *Association Rule Mining* yang diambil dari skripsi Naimah mahasiswa Universitas Islam Negeri (UIN) Alauddin Makassar jurusan Teknik Informatika 2017 dengan judul “Analisis dan Implementasi Teknik Data Mining dengan Metode *Association Rule Mining* untuk Memprediksi Strategi Pemasaran Produk Unilever pada PT. Tiran Makassar”. Penelitian ini membahas tentang implementasi teknik data mining dengan metode *Association Rule Mining* dan menggunakan algoritma *Apriori*. Penelitian ini bertujuan untuk mengetahui *Buying Habit* terhadap *consumer* PT. Tiran Makassar, dari penelitian ini juga dibangun sebuah sistem yang berguna untuk mempercepat proses penambangan data [12].

Terakhir adalah penelitian yang dilakukan oleh Pancalina Rakhmasari Putri mahasiswa Universitas Airlangga Surabaya program studi Sistem Informasi 2016 dengan judul “*Market Basket Analysis* dengan Menggunakan Algoritma *Hash-Based* Pada Transaksi Penjualan Apotek untuk Menerapkan Konsep *Cross-*



Selling". Dari penelitian ini dapat disimpulkan bahwa penerapan algoritma *Hash-Based* dirasa cukup andal dibandingkan dengan algoritma *Apriori* karena pada algoritma *Hash-Based* hanya dilakukan satu kali scan terhadap database sehingga waktu yang digunakan menjadi lebih efisien. Hasil yang didapatkan pada scan awal disimpan dalam tabel *hash* dalam bentuk *bucket* dan untuk iterasi selanjutnya tidak perlu scan ke database akan tetapi ke tabel *Hash*. Setelah *Association Rule* didapatkan kemudian dilakukan algoritma *Insertion Sort* untuk mendapatkan 10 rekomendasi teratas [13].

Dari beberapa penelitian yang telah diuraikan di atas terdapat beberapa kekurangan, diantaranya adalah dataset yang digunakan masih bersifat statis sehingga kurang cocok jika diterapkan di usaha retail yang datanya terus mengalami perubahan setiap hari, kekurangan lain adalah *user* tidak dapat melakukan *custom request* hasil berdasarkan input minimum *support* dan minimum *confidence* sehingga menghasilkan sedikit aturan. Dalam penelitian ini akan dibangun sebuah sistem dengan menggunakan database sehingga dataset yang digunakan bisa dinamis, dalam sistem ini juga akan disertakan fitur untuk menentukan input parameter berupa *support* secara dinamis agar *user* dapat mendapatkan hasil aturan yang lebih beragam.

### III. METODE PENELITIAN

Objek penelitian dalam penelitian ini adalah Toko Matahari yang merupakan toko retail modern dengan omzet yang cukup besar dengan segmentasi yang beragam. Penelitian ini akan merancang sebuah sistem informasi yang digunakan untuk mengetahui pola belanja konsumen dengan memanfaatkan data transaksi penjualan. Sistem ini diharapkan dapat membantu pihak Toko Matahari dalam menentukan strategi pemasaran yang sesuai kebutuhan.

Langkah pertama dalam penelitian ini adalah proses mengumpulkan data. Data yang digunakan dalam penelitian ini didapatkan melalui beberapa cara, seperti wawancara langsung dengan pemilik toko, observasi ke lokasi, dan studi pustaka.

Data penjualan yang merupakan data utama yang akan digunakan dalam sistem ini didapatkan langsung melalui komputer milik Toko Matahari. Data penjualan yang digunakan yaitu data tahun 2018 sampai dengan tahun 2020, dengan total transaksi sebanyak kurang lebih 85 ribu transaksi.

Data penjualan yang didapatkan dari Toko Matahari merupakan data yang tidak terstruktur dengan banyak atribut yang tidak dibutuhkan pada proses data mining, sehingga sebelum proses data mining dimulai perlu dilakukan *cleaning* terlebih dahulu.

Output dari *cleaning* terhadap data adalah data train siap pakai dalam bentuk vertikal (atas ke bawah). Data dalam bentuk vertikal dimaksudkan agar sistem tidak terlalu terbebani dalam mengolahnya karena tipe data ini lebih sedikit memakan memori dibandingkan dengan data horizontal (menyamping).

id	invoice	product id	product name	created at
1	15491/KSR/UTM/0618	14338054-de0b-4660-b56f-71dca02e122	FF MILKY STRAW 180ML	2018-06-01 08:50:18
2	15491/KSR/UTM/0618	05008b64-4051-4137-944a-03b51c0243c1	ADAM BRI LINK	2018-06-01 08:50:18
3	15491/KSR/UTM/0618	71618400-57d1-426f-a9af-64269869434	PIPSUPPER PLS 120g	2018-06-01 08:53:36
4	15491/KSR/UTM/0618	8b46144a-8615-4c1b-b5a7-9a3863a2517	KISPRAY GOLD POLCH 300ML	2018-06-01 08:53:36
5	15491/KSR/UTM/0618	0ba3d3f6-4e19-4e7b-a788-60ab68f9a489	DOWNY 23 PASSION 6PCS	2018-06-01 08:53:36
6	15491/KSR/UTM/0618	2e59b2d3-d144-4313-9ac7-9c160d92e087	RESIK V FEM HYGIENE 50ML	2018-06-01 08:53:36
7	15491/KSR/UTM/0618	166469e2-5028-4025-9480-924823b39e6d4	GARNIER M C WATER 120ML	2018-06-01 08:53:36
8	15491/KSR/UTM/0618	74680a68-d844-4a84-b7f8-c97eab021a80	MARINIA UV HYDRO COOL 185ML	2018-06-01 08:53:36
9	15491/KSR/UTM/0618	7e2027ca-6ac7-4a77-a27f-dbd1c5a17616	FRESNAT PFM GREEN MIRAGE 100ML	2018-06-01 08:53:36
10	15491/KSR/UTM/0618	e2080ee6-62a1-441b-a064-400a9a5901ac4	GERRY SALLUT 110g COCONUT	2018-06-01 09:02:45
11	15491/KSR/UTM/0618	090e4552-0e6d-4281-b6e6-c23a6787651c	SIDAP SRKABU SPESIAL	2018-06-01 09:02:45
12	15491/KSR/UTM/0618	6aed6f9a-934a-46b4-bb46-1c4e402240dc	SIDAP GR 1818RUK	2018-06-01 09:02:45
13	15491/KSR/UTM/0618	935a3c2c-7541-4317-875f-0f1cf8a9115c	NITTO PEANUT CRACKER 475g	2018-06-01 09:02:45
14	15491/KSR/UTM/0618	e895f67e-8298-408a-87ea-dc13a89c2f9	SUPERSON PINE 450ML	2018-06-01 09:07:23

Gambar. 1. Data Train Siap Pakai

Gambar. 1. merupakan hasil dari *cleaning* terhadap data dan data ini yang nantinya akan digunakan sebagai data train. Hanya 4 atribut saja yang disisakan setelah proses *cleaning* yaitu atribut *invoice*, *product\_id*, *product\_name*, dan *created\_at*.

Dari hasil wawancara didapatkan data bahwa Toko Matahari memiliki beberapa masalah yang belum

Dalam penelitian ini yang menjadi nara sumber adalah istri pemilik Toko Matahari yakni Ibu Tuti. Dari wawancara ini didapatkan data bahwa Toko Matahari memiliki beberapa masalah yang belum ditemukan solusinya secara efektif. Beberapa masalah yang dihadapi antara lain sebagai berikut.

- Permasalahan utama yang sering dihadapi oleh Toko Matahari adalah sebagian produk yang tidak terjual sampai kadaluwarsa. Biasanya merupakan produk baru dengan angka penjualan rendah.
- Belum ada solusi nyata untuk mengatasi masalah ini kecuali sebatas merubah lokasi penempatan produk.
- Dalam beberapa kasus jika barang mendekati masa kadaluwarsa maka diberikan diskon.

### IV. PEMBAHASAN DAN HASIL

Metode data mining yang digunakan dalam penelitian ini adalah metode *Association Rule Mining* dengan menggunakan algoritma *hash-based*. Seperti yang telah dijelaskan bahwa data yang digunakan adalah data transaksi tahun 2018-2020 dengan total transaksi sebanyak 85 ribu transaksi. Sistem data miningnya dibangun dengan menggunakan bahasa pemrograman PHP (*framework* Laravel) sebagai pengolah data dan Vue.js di bagian *view* nya. Sebagai bahan perbandingan dalam penelitian juga akan digunakan bahasa pemrograman Python dengan menggunakan *library* *mlxtend.frequent\_patterns* agar output yang dihasilkan dalam penelitian ini lebih valid.

Agar penelitian ini bisa dipahami oleh banyak pihak penulis akan terlebih dahulu menjelaskan bagaimana cara kerja dari algoritma *Hash-Based* dengan menggunakan contoh dataset. Dalam contoh ini tidak akan dijelaskan keseluruhan proses dari data mining, namun hanya terbatas pada cara kerja algoritma *Hash-Based*.

Tahap pertama dalam algoritma *hash-based* adalah melakukan penelusuran terhadap data untuk mengetahui jumlah *Order of Item* dari data tersebut. *Order of Item* adalah alamat untuk tiap-tiap item yang ada dalam data. *Order of Item* ini nantinya akan dipakai untuk menentukan alamat *bucket* dalam tabel *hash*.

Sebagai contoh terdapat 10 data transaksi penjualan dari sebuah toko retail. Setelah dilakukan *cleaning* hasilnya dapat dilihat pada tabel I.

Tabel I. DATA TRANSAKSI

Id Transaksi	Itemset
--------------	---------

T1	Minyak, Gula
T2	Sabun, Garam, Gula
T3	Sabun, Minyak, Gula
T4	Teh, Sabun, Gula
T5	Minyak, Rokok
T6	Gula, Garam
T7	Beras, Gula
T8	Minyak, Teh, Sabun, Gula
T9	Sabun, Sunlight
T10	Wipol, Rokok

Selanjutnya dilakukan proses *hashing* terhadap kandidat 1-itemset untuk memasukkan tiap itemset ke dalam *bucket* pada tabel *hash*. Proses *hashing* menggunakan rumus [14].

$$h(x) = (\text{order of item } x) \bmod n \quad (1)$$

$h$  = alamat *bucket* pada tabel *hash*  
 $n$  = banyak alamat ( $n = 2m + 1$ ,  $m$  = jumlah keseluruhan item)

Tabel II. ORDER OF ITEM

Item	Order
Minyak	1
Gula	2
Sabun	3
Garam	4
Teh	5
Rokok	6
Beras	7
Sunlight	8
Wipol	9

Tabel II merupakan tabel *Order Of Item* yang berfungsi sebagai nomor urut tiap-tiap item. Setelah tabel *order Of Item* didapat langkah selanjutnya yaitu mencari alamat pada tabel *hash* untuk 1-itemset dengan rumus (1).

$$\begin{aligned} h(\text{Minyak}) &= (1) \bmod 19 = 1 \\ h(\text{Gula}) &= (2) \bmod 19 = 2 \\ h(\text{Sabun}) &= (3) \bmod 19 = 3 \\ h(\text{Garam}) &= (4) \bmod 19 = 4 \\ h(\text{Teh}) &= (5) \bmod 19 = 5 \\ h(\text{Rokok}) &= (6) \bmod 19 = 6 \\ h(\text{Beras}) &= (7) \bmod 19 = 7 \\ h(\text{Sunlight}) &= (8) \bmod 19 = 8 \\ h(\text{Wipol}) &= (9) \bmod 19 = 9 \end{aligned}$$

Tabel III. TABEL HASH 1-ITEMSET

Address	Itemset	Link	Support Count
1	Minyak	T1, T3, T5, T8	4
2	Gula	T1, T2, T3, T4, T6, T7, T8	7
3	Sabun	T2, T3, T4, T8, T9	5
4	Garam	T2, T6	2
5	Teh	T4, T8	2
6	Rokok	T5, T10	2
7	Beras	T7	1
8	Sunlight	T9	1
9	Wipol	T10	1

Tabel III menunjukkan hasil *hashing* untuk 1-itemset. Tiap itemset menempati alamat *hash* sesuai dengan hasil perhitungan. Alamat-alamat yang ditempati akan menjadi node dan membuat *link* yang mengarah kepada transaksi yang mengandung itemset tersebut sehingga terbentuk *link-list*.

Dengan menentukan minimum *support count* = 2 misalnya akan dihasilkan *frequent* 1-itemset seperti pada Tabel IV.

Tabel IV. FREQUENT 1-ITEMSET

Itemset	Support Count
Minyak	4
Gula	7
Sabun	5
Garam	2
Teh	2
Rokok	2

Langkah selanjutnya adalah menentukan semua kemungkinan kandidat 2-itemset berdasarkan *frequent* 1-itemset. Kandidat 2-itemset yang terbentuk dari *frequent* 1-itemset yaitu, (Minyak, Gula), (Minyak, Sabun), (Minyak, Garam), (Minyak, Teh), (Minyak, Rokok), (Gula, Sabun), (Gula, Garam), (Gula, Teh), (Gula, Rokok), (Sabun, Garam), (Sabun, Teh), (Sabun, Rokok), (Garam, Teh), (Garam, Rokok), (Teh, Rokok). Selanjutnya memasukkan kandidat 2-itemset tersebut ke dalam tabel *hash* dengan menggunakan rumus [14]:

$$h(k) = ((\text{order of } x) * 10 + \text{order of } y) \bmod n \quad (2)$$

Mencari alamat pada tabel *hash* untuk 2-itemset dari persamaan (2).

$$\begin{aligned} h(\text{Minyak, Gula}) &= (1 * 10 + 2) \bmod 19 = 12 \\ h(\text{Minyak, Sabun}) &= (1 * 10 + 3) \bmod 19 = 13 \\ h(\text{Minyak, Garam}) &= (1 * 10 + 4) \bmod 19 = 14 \\ h(\text{Minyak, Teh}) &= (1 * 10 + 5) \bmod 19 = 15^* \\ h(\text{Minyak, Rokok}) &= (1 * 10 + 6) \bmod 19 = 16^* \\ h(\text{Gula, Sabun}) &= (2 * 10 + 3) \bmod 19 = 4 \\ h(\text{Gula, Garam}) &= (2 * 10 + 4) \bmod 19 = 5 \\ h(\text{Gula, Teh}) &= (2 * 10 + 5) \bmod 19 = 6 \\ h(\text{Gula, Rokok}) &= (2 * 10 + 6) \bmod 19 = 7^* \\ h(\text{Sabun, Garam}) &= (3 * 10 + 4) \bmod 19 = 15^* \\ h(\text{Sabun, Teh}) &= (3 * 10 + 5) \bmod 19 = 16^* \\ h(\text{Sabun, Rokok}) &= (3 * 10 + 6) \bmod 19 = 17 \\ h(\text{Garam, Teh}) &= (4 * 10 + 5) \bmod 19 = 7^* \\ h(\text{Garam, Rokok}) &= (4 * 10 + 6) \bmod 19 = 8 \\ h(\text{Teh, Rokok}) &= (5 * 10 + 6) \bmod 19 = 18 \end{aligned}$$

Dari perhitungan di atas ditemukan adanya *Collision* (ada lebih dari satu itemset yang menempati alamat yang sama) yaitu pada alamat ke-7 untuk itemset (Gula, Rokok) dan (Garam, Teh), alamat ke-15 untuk itemset (Minyak, Teh) dan (Sabun, Garam), alamat ke-16 untuk itemset (Minyak, Rokok) dan (Sabun, Teh).

Jika terjadi *Collision* maka dilakukan pengecekan pada tabel *hash* untuk mencari alamat yang masih kosong. Jika ternyata separuh dari alamat sudah terisi maka perlu dilakukan *rehashing* dengan jumlah alamat 2 kali lipat dari alamat sebelumnya. Proses *rehashing* menggunakan rumus [14]:

$$h(k) = ((\text{order of } x) * 10 + \text{order of } y) \bmod j \quad (3)$$

$$j = \text{Banyak alamat setelah dilakukan penambahan} \\ (j = 2 * m + 1)$$

$m$  = jumlah alamat pada tabel *hash* sebelum penambahan)

- $h$  (Minyak, Gula) =  $(1 * 10 + 2) \bmod 39 = 12$
- $h$  (Minyak, Sabun) =  $(1 * 10 + 3) \bmod 39 = 13$
- $h$  (Minyak, Garam) =  $(1 * 10 + 4) \bmod 39 = 14$
- $h$  (Minyak, Teh) =  $(1 * 10 + 5) \bmod 39 = 15$
- $h$  (Minyak, Rokok) =  $(1 * 10 + 6) \bmod 39 = 16$
- $h$  (Gula, Sabun) =  $(2 * 10 + 3) \bmod 39 = 23$
- $h$  (Gula, Garam) =  $(2 * 10 + 4) \bmod 39 = 24$
- $h$  (Gula, Teh) =  $(2 * 10 + 5) \bmod 39 = 25$
- $h$  (Gula, Rokok) =  $(2 * 10 + 6) \bmod 39 = 26$
- $h$  (Sabun, Garam) =  $(3 * 10 + 4) \bmod 39 = 34$
- $h$  (Sabun, Teh) =  $(3 * 10 + 5) \bmod 39 = 35$
- $h$  (Sabun, Rokok) =  $(3 * 10 + 6) \bmod 39 = 36$
- $h$  (Garam, Teh) =  $(4 * 10 + 5) \bmod 39 = 6$
- $h$  (Garam, Rokok) =  $(4 * 10 + 6) \bmod 39 = 7$
- $h$  (Teh, Rokok) =  $(5 * 10 + 6) \bmod 39 = 17$

Setelah dilakukan *rehashing* tidak terjadi *Collision* lagi, tiap-tiap itemset menempati alamat yang berbeda. Hasilnya seperti terlihat pada tabel V.

Tabel V. TABEL HASH 2-ITEMSET

Address	Itemset	Link	Support Count
0	-	-	-
...	-	-	-
6	Garam, Teh	-	0
7	Garam, Rokok	-	0
12	Minyak, Gula	T1, T3, T8	3
13	Minyak, Sabun	T3, T8	2
14	Minyak, Garam	-	0
15	Minyak, Teh	T8	1
16	Minyak, Rokok	T5	1
17	Teh, Rokok	-	0
23	Gula, Sabun	T2, T3, T4, T8	4
24	Gula, Garam	T2, T6	2
25	Gula, Teh	T4, T8	2
26	Gula, Rokok	-	0
34	Sabun, Garam	T2	1
35	Sabun, Teh	T4, T8	2
36	Sabun, Rokok	-	0

Dari tabel *hash* tersebut kemudian dilakukan eliminasi itemset dengan support count  $\leq 2$  untuk mendapatkan *frequent 2-itemset*.

Tabel VI. FREQUENT 2-ITEMSET

Itemset	Support Count
Minyak, Gula	3
Minyak, Sabun	2
Gula, Sabun	4
Gula, Garam	2
Gula, Teh	2
Sabun, Teh	2

Tabel VI *frequent 2-itemset* di atas kemudian digunakan untuk membangkitkan kandidat 3-itemset yang mungkin terbentuk yaitu, (Minyak, Gula, Sabun), (Minyak, Gula, Garam), (Minyak, Gula, teh), (Minyak, Sabun, Garam), (Minyak, Sabun, Teh), (Gula, Sabun, Garam), (Gula, Sabun, Teh), (Gula, Garam, Teh). Rumus yang digunakan untuk mencari alamat 3-itemset adalah:

$$h(k) = ((order\ of\ x) * 100 + (order\ of\ y) * 10 + order\ of\ z) \bmod j \quad (4)$$

Mencari alamat pada tabel *hash 3-itemset*:

- $h$  (Minyak, Gula, Sabun) =  $(1 * 100 + 2 * 10 + 3) \bmod 39 = 6$
- $h$  (Minyak, Gula, Garam) =  $(1 * 100 + 2 * 10 + 4) \bmod 39 = 7$
- $h$  (Minyak, Gula, Teh) =  $(1 * 100 + 2 * 10 + 5) \bmod 39 = 8$
- $h$  (Minyak, Sabun, Garam) =  $(1 * 100 + 3 * 10 + 4) \bmod 39 = 17$
- $h$  (Minyak, Sabun, Teh) =  $(1 * 100 + 3 * 10 + 5) \bmod 39 = 18$
- $h$  (Gula, Sabun, Garam) =  $(2 * 100 + 3 * 10 + 4) \bmod 39 = 0$
- $h$  (Gula, Sabun, Teh) =  $(2 * 100 + 3 * 10 + 5) \bmod 39 = 1$
- $h$  (Gula, Garam, Teh) =  $(2 * 100 + 4 * 10 + 5) \bmod 39 = 1$

Tabel *hash 3-Itemset* dapat dilihat di tabel VII.

Tabel VII. TABEL HASH 3-ITEMSET

Address	Itemset	Link	Support Count
0	Gula, Sabun, Garam	T2	1
1	Gula, Sabun, Teh	T4, T8	2
6	Minyak, Gula, Sabun	T3, T8	2
7	Minyak, Gula, Garam	-	0
8	Minyak, Gula, Teh	T8	1
11	Gula, Garam, Teh	-	0
17	Minyak, Sabun, Garam	-	0
18	Minyak, Sabun, Teh	T8	1

Dari Tabel VII kemudian dilakukan eliminasi dengan support count  $\leq 2$  sehingga menghasilkan *frequent 3-itemset* seperti terlihat pada tabel VIII.

Tabel VIII. FREQUENT 3-ITEMSET

Itemset	Support Count
Gula, Sabun, Teh	2
Minyak, Gula, Sabun	2

Setelah dilakukan 3 kali iterasi didapatkan 2 itemset dengan minimum support 2 yaitu (Gula, Sabun, Teh) dan (Minyak, Gula, Sabun). Dari Tabel VIII kemudian digunakan untuk menentukan support dan confidence dengan menggunakan persamaan (5) dan persamaan (6) [15].

$$Support(A, B) = \frac{\sum \text{transaksi mengandung A dan B}}{\sum \text{transaksi}} \times 100\% \quad (5)$$

$$Confidence = P(B|A) = \frac{\sum \text{transaksi mengandung A dan B}}{\sum \text{transaksi mengandung A}} \quad (6)$$

Tabel IX. TABEL SUPPORT DAN CONFIDENCE 3-ITEMSET

Itemset	Support	Confidence
Gula, Sabun, Teh	20%	50%
Minyak, Gula, Sabun	20%	66,6%

Tabel IX menunjukkan bahwa didapat dua itemset dengan nilai support => 20% dan masing-masing itemset memiliki nilai *confidence* yang berbeda. Tabel IX bisa juga ditulis dalam bentuk notasi:

{Gula, sabun} → {Teh} (support = 20%, confidence = 50%)

{Minyak, Gula} → {Sabun} (support = 20%, confidence = 66,6%)

Artinya “50% dari transaksi yang mengandung Gula dan Sabun juga ada Teh di dalamnya, dan 20% dari keseluruhan transaksi memuat Gula, Sabun, dan Teh”. Bisa juga berarti “konsumen yang membeli Gula dan Sabun memiliki kemungkinan 50% untuk juga membeli Teh, dan aturan ini mewakili 20% dalam transaksi yang ada” [16].

Jika menghendaki bisa ditambahkan *Lift Ratio*. *Lift Ratio* adalah nilai yang menunjukkan kekuatan sebuah *rule* yang terbentuk berdasarkan *support* dan *confidence*. *Rule* dikatakan kuat apabila nilai *lift ratio* lebih dari 1. Nilai *lift ratio* dapat diketahui dengan rumus [17].

$$\text{lift ratio} = \frac{\text{confidence}(A,B)}{\text{Benchmark Confidence}(A,B)} \quad (7)$$

Nilai *Benchmark confidence* dapat dihitung dengan rumus:

$$\text{Benchmark Confidence}(A, B) = \frac{\text{jumlah transaksi mengandung B}}{\text{Total transaksi}} \quad (8)$$

Tabel IX menunjukkan data 2-itemset dengan nilai *support* => 20% disertai dengan tambahan nilai *lift ratio* di dalamnya.

Tabel X. LIFT RATIO 3-ITEMSET

Itemset	Support	Confidence	Lift Ratio
Gula, Sabun, Teh	20%	50%	2.5
Minyak, Gula, Sabun	20%	66.6 %	1.3

Setelah diketahui cara kerja data mining dengan menggunakan algoritma *hash-based*, selanjutnya adalah mengimplementasikan algoritma ini ke dalam data riil yang dimiliki.

Terdapat dua jenis output dari implementasi data mining dengan algoritma *hash-based* dengan data penjualan Toko Matahari sebagai *data train*, yaitu dalam bentuk tabel dan dalam bentuk grafik. Terdapat 3 tabel yaitu tabel produk paling tidak laku, tabel kombinasi dua produk paling laku, dan tabel kombinasi tiga produk paling laku.

Tabel XI. TABEL PRODUK PALING TIDAK LAKU

No	Nama Produk	Jumlah Penjualan
1	mie gepeng	1
2	oral-b sikat gg isi 2	1
3	champ nugget coin 200g	1
4	oral-b sikat gg ex soft	1
5	lego 42	1
6	tepung beras padi	1
7	bellagio spray venture 80ml	1
8	hatari banana 90g	1
9	kino pia kijo 75g	1
10	kenji balls 40g	1

Tabel XI memuat data produk dengan nilai penjualan paling rendah. Mengetahui produk dengan nilai penjualan rendah sama pentingnya dengan mengetahui produk dengan penjualan paling bagus, terutama jika dikaitkan dengan toko retail yang bukan waralaba seperti Toko Matahari. Data dari tabel XI nantinya akan digunakan untuk menentukan promosi yang sesuai untuk produk-produk yang terdapat pada tabel tersebut.

Data 2-itemset dengan nilai penjualan paling tinggi dapat dilihat pada tabel XII. Data yang ditampilkan hanya sebagian saja karena hasil proses data mining menunjukkan bahwa hasilnya lebih dari 10 data.

Tabel XII. KOMBINASI DUA PRODUK PALING LAKU

No	Kombinasi Produk	Support	Confidence
1	viva mc 100 green tea → viva fc 100 green tea	0.0209 %	54.55 %
2	viva mc 100 lemon → viva fc 100 lemon	0.0256 %	52.38 %
3	wardah step 2 nc 30g → wardah 1 day cr step 2 30g	0.0291 %	59.52 %
4	wardah lnc step2 20ml → wardah day cream step2 20ml	0.0396 %	42.50 %
5	cerelec mungbean 20g → cerelec 20 b merah	0.0407 %	46.67 %
6	teh celup poci 50 g → gula pasir 1/2	0.0407 %	35.00 %
7	milna bb20sup aywtlb → milna bmp20g sashet	0.0419 %	32.43 %
8	viva mc 100 bengkoang → viva fc 100 bengkuang	0.0419 %	36.36 %
9	sariwangi kotak 50pcs → gulaku premium 1kg	0.0419 %	36.73 %
10	chocolatos 3 in1 matcha 126g → chocolatos 3in1	0.0430 %	36.27 %

Nilai *support* bisa kita artikan sebagai seberapa sering dua produk di tabel XII terjual, semakin besar nilai *support* artinya produk tersebut semakin sering terjual. Sedangkan nilai *confidence* menyatakan hubungan dua produk dalam sebuah transaksi, semakin besar nilainya artinya dua produk tersebut sering terjual secara bersamaan. Untuk data kombinasi tiga produk yang paling laku terjual dapat dilihat pada tabel XIII.

Tabel XIII. TABEL KOMBINASI TIGA PRODUK PALING LAKU

No	Kombinasi Produk	Support	Confidence
1	(milna bb20sup aywtlb - milna bb20 ht ay bym) → milna bmp20g sashet	0.0221 %	42.22 %
2	(forvita margarine 200gr - myk fortune 1 liter) → terigu 1kg	0.0256 %	91.67 %
3	(forvita margarine 200gr - bimoli pouch 1l) → terigu 1kg	0.0268 %	88.46 %
4	(gula pasir 1/2 - terigu 1kg) → terigu 1/2	0.0396 %	44.74 %
5	(forvita margarine 200gr - terigu 1kg) → gula pasir 1/4	0.0535 %	34.33 %
6	(forvita margarine 200gr - terigu 1kg) → terigu 1/2	0.0803 %	51.49 %

Tabel XII dan tabel XIII merupakan tabel yang memuat data produk yang paling laku. Data dalam tabel ini bisa digunakan sebagai acuan dalam penentuan strategi pemasaran seperti promosi dan penataan barang. Salah satu contoh penerapannya misalnya dengan menggunakan



metode *cross selling* yaitu menggabungkan dua jenis produk (yang paling laku dan yang paling tidak laku) dalam suatu paket pembelian. Tujuannya salah satunya adalah agar produk yang tidak laku menjadi lebih dikenal oleh konsumen sehingga angka penjualannya naik.

Setelah data-data dalam bentuk tabel didapat, hal yang seharusnya ada dalam data mining adalah output dalam bentuk grafik. Output dalam bentuk grafik juga perlu disajikan agar gambaran umum hasil data mining dapat dibaca lebih mudah dibandingkan dengan dalam bentuk tabel. Data dalam bentuk grafik lebih memberikan *insight* kepada pembaca tentang gambaran umum model data mining yang dihasilkan.



Gambar. 2. Grafik Support dan Confidence

Gambar. 2. merupakan contoh output dalam bentuk grafik yang dihasilkan oleh sistem yang dikembangkan dengan nilai *support*  $\leq 0,02\%$  dan nilai *confidence*  $\leq 50\%$ . Warna biru menunjukkan data *support* dan warna pink menunjukkan data *confidence* dan bagian bawah menunjukkan label untuk tiap-tiap itemset.

Sistem implementasi data mining dengan menggunakan metode *Association Rule Mining* untuk menentukan peta pemasaran ini dirancang dan dibangun secara mandiri oleh penulis tanpa tambahan tools ataupun *library* pihak ketiga yang terkait dengan data mining, sehingga memungkinkan algoritma yang dibangun tidak sesuai dan output yang dihasilkan tidak tepat. Oleh karena itu perlu dilakukan validasi dan evaluasi terhadap output yang dihasilkan oleh sistem ini dengan menggunakan tools ataupun media lain yang sudah tervalidasi.

Media validasi yang dimaksud di atas yaitu bahasa pemrograman Python dengan menggunakan *library* `mlxtend.frequent_patterns` (*package* `apriori`, `association_rules`). Alasan menggunakan Python yaitu karena bahasa ini merupakan bahasa yang familier digunakan untuk data mining dan tersedia banyak *library* yang sangat mendukung dalam data mining. Data mining menggunakan Python sangat cocok jika alat penelitian yang digunakan kurang mumpuni.

Proses validasi ini menggunakan bantuan Google Colab. Google Colab adalah media yang *cloud-based runtime* sehingga tidak perlu instalasi apapun dan sumber daya yang diberikan cukup mumpuni serta yang paling penting adalah fasilitas ini gratis.

Berikut ini adalah proses validasi dengan Python, `mlxtend.frequent_patterns`, *package* `apriori` `association rules`, Google Colab.

```
#import library
from mlxtend.preprocessing import
TransactionEncoder
from mlxtend.frequent_patterns import apriori
from mlxtend.frequent_patterns import
association_rules
import pandas as pd

#read dataset from google drive
df =
pd.read_csv('/content/drive/MyDrive/TA/DATA/maste
r - transaksi - Copy.csv')

#read first 5 line dataset
df.head()

#check missing value
print(df.isnull().sum().sort_values(ascending=False))

#get unique item
8 transaction_list = []
for i in df['invoice'].unique():
    tlist =
list(set(df[df['invoice']==i]['product_name']))
    if len(tlist)>0:
        transaction_list.append(tlist)
print(len(transaction_list))

#encoding dataset
te = TransactionEncoder()
te_ary=te.fit(transaction_list).transform(transac
tion_list)
df2 = pd.DataFrame(te_ary, columns=te.columns_)

#apriori support 0,002 %
frequent_itemsets=apriori(df,min_support=0.0002,
use_colnames = True, max_len=3)

#add leghth column for filter
frequent_itemsets['length']=frequent_itemsets['it
emsets'].
apply(lambda x: len(x))
frequent_itemsets

#association rules
rules = association_rules(frequent_itemsets,
metric='lift',
min_threshold=1)
rules.sort_values('confidence', ascending=False)

#support 0,05% dan confidence 35%
rules[(rules['support']>=0.0005)&(rules['confiden
ce']>=0.35)]
```

#### Modul 1. Association Rules dengan Python

Modul di atas adalah contoh mencari *association rules* dengan nilai *support* sebesar 0,05 % dan nilai *confidence* sebesar 35%. Hasil dari modul di atas dapat dilihat pada gambar 3.

	antecedents	consequents	antecedent support	consequent support	support	confidence	Lift	leverage	conviction
483	{FARILOVELY FF POAM SASMET}	{FAIR LOVELY 5g}	0.00325	0.00758	0.01396	0.39040	52.53797	0.01389	1.643256
510	{FORVITA MARGARINE 200GR}	{TERIGU 1KG}	0.00327	0.00367	0.01059	0.46851	118.11193	0.01546	1.874115
511	{TERIGU 1KG}	{FORVITA MARGARINE 200GR}	0.00367	0.00327	0.01059	0.30292	118.11193	0.01546	1.641982
743	{TEH POOL 40 SUPERIOR}	{GULA PASIR 1kg}	0.01675	0.02174	0.03926	0.37000	17.23639	0.00882	1.889194
751	{TEH SOGRO 1g}	{GULA PASIR 1kg}	0.00308	0.02174	0.03173	0.30776	16.43734	0.01289	1.822742
753	{TEH TARTLET}	{GULA PASIR 1kg}	0.00213	0.02174	0.03173	0.41720	20.54904	0.01494	1.789544
866	{MUNA BESI HT AY 87MG}	{MUNA BESI GP AYVTLB}	0.00301	0.01291	0.03923	0.30250	436.61843	0.00622	2.292763
867	{MUNA BESI GP AYVTLB}	{MUNA BESI HT AY 87MG}	0.01291	0.00301	0.03923	0.40545	436.61843	0.00622	1.688253
1259	{FORVITA MARGARINE 200GR GULA PASIR 1kg}	{TERIGU 1KG}	0.00065	0.00367	0.00305	0.884615	223.00992	0.00653	6.632267
1260	{TERIGU 1KG GULA PASIR 1kg}	{FORVITA MARGARINE 200GR}	0.00065	0.00327	0.00305	0.884615	285.86981	0.00653	6.637932
1270	{FORVITA MARGARINE 200GR TERIGU 1kg}	{TERIGU 1KG}	0.00019	0.00367	0.00305	0.879418	231.17819	0.00739	7.889662
1271	{FORVITA MARGARINE 200GR TERIGU 1KG}	{TERIGU 1kg}	0.00158	0.00942	0.00305	0.014825	77.52198	0.00762	2.047345
1272	{TERIGU 1kg TERIGU 1KG}	{FORVITA MARGARINE 200GR}	0.01128	0.00327	0.00305	0.71534	213.80489	0.00739	3.452761

Gambar 3. Output Association Rules dengan Python



Setelah dilakukan uji coba terhadap sistem yang dikembangkan dan juga membandingkan dengan media validasi yang lain, dapat ditarik kesimpulan bahwa sistem ini memiliki kelebihan serta kekurangan. Kelebihan yang dimiliki antara lain sebagai berikut.

- Sistem sudah dilengkapi *caching* sehingga tidak perlu melakukan proses yang sama dua kali.
- Format dataset dalam bentuk vertikal sehingga tidak terlalu memakan memori penyimpanan.
- Menggunakan pendekatan SPA (*Single Page Application*) sehingga lebih nyaman digunakan. Adapun kekurangan dari sistem ini adalah sebagai berikut.
- Meskipun berbasis web namun belum *online service* sehingga hanya dapat digunakan oleh satu komputer.
- Agar sistem ini berjalan dengan baik membutuhkan perangkat dengan spesifikasi yang cukup tinggi sehingga sulit diaplikasikan di lapangan.
- Data train yang ada hanya dua tahun saja sehingga hasil yang didapat kurang maksimal.

## V. PENUTUP

2 Berdasarkan uraian yang telah dijelaskan dari implementasi data mining dengan menggunakan metode *Association Rule Mining* dan algoritma *hash-based* ada beberapa hal yang dapat disimpulkan, antara lain sebagai berikut.

- Data mining adalah sebuah metode penggalian informasi yang tersembunyi dalam sebuah data dalam jumlah yang cukup banyak. Data mining dapat diterapkan di berbagai aspek kehidupan. Dalam penelitian ini data mining digunakan untuk menggali informasi yang tersimpan dalam data transaksi penjualan sebuah toko retail modern yaitu Toko Matahari. Informasi yang dicari adalah hubungan asosiatif produk-produk yang terjual di Toko Matahari. Hubungan asosiatif yang dimaksud adalah seberapa sering beberapa produk terjual secara bersamaan.
- Membangun sistem data mining dengan bahasa pemrograman PHP yang dalam penelitian ini menggunakan framework LARAVEL menghasilkan aplikasi yang cukup handal dan powerful, mengingat bahwa PHP merupakan bahasa pemrograman yang tidak familier digunakan dalam data mining.
- Penerapan hasil dari data mining dalam penelitian ini antara lain untuk digunakan dalam strategi *cross selling*, yaitu menjual sebuah produk kemudian menawarkan produk lain yang berbeda. Beberapa contoh penawaran dalam *cross selling* adalah sebagai berikut:
  - 1) Apabila konsumen membeli margarin dan terigu maka kasir menawarkan gula, karena mempunyai kecenderungan 34,32% untuk dibeli;
  - 2) Apabila konsumen membeli teh celup maka kasir akan menawarkan gula, karena memiliki kecenderungan 35%-45% untuk dibeli;
  - 3) membuat paket hemat yang berisi gula, terigu dan produk dengan jumlah penjualan terendah;
  - 4) memberikan diskon untuk produk dengan nilai jual (support) rendah, agar dapat dikenal konsumen;
  - 5) diskon untuk produk tertentu

(produk dengan support rendah) dengan batas minimum transaksi.

Membangun sebuah sistem, terlebih yang digunakan untuk proses data mining tentu akan cukup sulit dan jauh dari kata sempurna. Oleh karena itu ada beberapa hal yang dapat penulis sampaikan sebagai saran agar penelitian selanjutnya lebih baik, antara lain sebagai berikut.

- Penggunaan alat penelitian yang cukup handal. Alat penelitian yang digunakan dalam penelitian ini kurang mampu memenuhi kebutuhan karena dalam proses data mining dengan data yang cukup besar akan membutuhkan sumber daya yang handal.
- Jumlah data yang digunakan cukup ideal. Data yang digunakan dalam penelitian ini adalah data transaksi penjualan tahun 2018-2021 dengan total transaksi 85000. Informasi yang dicari dari penelitian ini adalah pola kebiasaan membeli (*buying habit*) konsumen sehingga jika hanya menggunakan data dalam rentang waktu 2 tahun maka pola yang dihasilkan akan kurang akurat dan aturan yang terbentuk tidak banyak.
- *Frontend* dan *Backend* yang sepenuhnya terpisah, hal ini bertujuan agar dalam pengembangan dan perawatannya lebih mudah.
- Sistem yang dikembangkan belum bisa diaplikasikan di lapangan. Harapannya penelitian selanjutnya dapat membangun sebuah sistem data mining dengan performa yang bagus, ringan dan dapat digunakan dengan baik tanpa harus menggunakan perangkat dengan spesifikasi yang cukup tinggi.

2 *Source code* dari sistem implementasi data mining dengan menggunakan metode *Association Rule Mining* ini dapat dilihat di [18].

## UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih sebanyak-banyaknya kepada semua pihak yang telah ikut andil dalam penelitian ini sehingga penelitian ini dapat berjalan lancar tanpa halangan yang berarti. Secara khusus penulis mengucapkan terima kasih kepada pihak Toko Matahari karena telah mengizinkan penulis untuk melakukan penelitian di Toko Matahari.

## DAFTAR PUSTAKA

- [1] H. Martinus, "Analisis Industri Retail Nasional," *Humaniora*, vol. 2, no. 2, p. 1309, 2011, doi: 10.21512/humaniora.v2i2.3193.
- [2] S. Kartono, *5 Jurus Sukses Berbisnis Sukses di Modern Market*, 1st ed. Jakarta: TransMedia Pustak, 2007.
- [3] A. S. Sujan, *Manajemen Minimarket*, 1st ed. Jakarta: Raih Asa Sukses, 2012.
- [4] A. S. B. Putra, "Pengaruh Retail Marketing Mix Terhadap Loyalitas Konsumen Dengan Kepuasan Konsumen Sebagai Pemoderasi," *JRMB*, Vol. 7, No 2 Desember 2012 berusaha, vol. 1, no. 2, pp. 186-194, 2013.
- [5] E. A. Sukma, A. D. Fauzi, and F. Yaningwati, "Suasana Toko dalam Menciptakan Emosi dan Pengaruhnya Terhadap Keputusan Pembelian," *J. Profit*, vol. 6, no. 1, pp. 60-87, 2012.
- [6] H. Kartajaya et al., *MarkPuls on Strategy*. Gramedia Pustaka Utama, 2005.
- [7] Mastuhin, E. L. Fraehantini, and M. N. Tentua, "Classification of Community Happiness Level Using K-Nearest Neighbor (

- KNN ) Method Klasifikasi Tingkat Kebahagiaan Masyarakat Dengan Metode K- Nearest Neighbor ( KNN ).”
- [8] K. Tampubolon, H. Saragih, and B. Reza, “Implementasi Data Mining Algoritma Apriori Pada Sistem Persediaan Alat-Alat Kesehatan,” *Inf. dan Teknol. Ilm.*, vol. 1, no. 1, pp. 93–106, 2013, doi: 10.1017/S0079497X00014341.
- [9] E. Buulolo, *Data Mining untuk Perguruan Tinggi*, 1st ed. Sleman: Deepublish, 2020.
- [10] F. Ramadhan, “Implementasi Algoritma Hash Based Terhadap Aturan Asosiasi untuk Menentukan Frequent Itemset Study Kasus Rumah Makan Seafood ‘Kita,’” *Semin. Nas. Teknol. Inf. dan Multimed.*, pp. 97–102, 2017.
- [11] O. S. Arifah and E. Wibowo, “Analisis Perbandingan Algoritma Apriori Dan Algoritma Hash Based Pada Market Basket Analysis Di Apotek Uad,” *J. Sarj. Tek. Inform.*, vol. 3, no. 1, pp. 1–10, 2015, doi: 10.12928/jstie.v3i1.2896.
- [12] Naimah, “Analisis dan Implementasi Teknik Data Mining dengan Metode Association Rule Mining untuk Memprediksi Strategi Pemasaran Produk Unilever pada PT . Tiran Makassar,” Universitas Islam Negeri (UIN) Alauddin Makassar, 2017.
- [13] P. R. PUTRI, “Market Basket Analysis Dengan Menggunakan Algoritma Hash-Based Pada Transaksi Penjualan Apotek Untuk Menerapkan Konsep Cross-Selling,” Universitas Airlangga Surabaya, 2016.
- [14] S. Aguru and B. M. Rao, “A Hash Based Frequent Itemset Mining using Rehashing,” *Int. J. Recent Innov. Trends Comput. Commun.*, vol. 2, no. 2, pp. 4198–4204, 2014.
- [15] R. Yanto and H. Di Kesuma, “Pemanfaatan Data Mining Untuk Penempatan Buku Di Perpustakaan Menggunakan Metode Association Rule,” *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 4, no. 1, pp. 1–10, 2017, doi: 10.35957/jatisi.v4i1.83.
- [16] D. Nofriansyah, *Konsep Data Mining vs Sistem Pendukung Keputusan*, 1st ed. Sleman: Deepublish, 2014.
- [17] V. N. Latifah, M. T. Furqon, and N. Santoso, “Implementasi Algoritme Modified-Apriori Untuk Menentukan Pola Penjualan Sebagai Strategi Penempatan Barang Dan Promo,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput. Univ. Brawijaya*, vol. 2, no. 10, pp. 2829–2834, 2018.
- [18] Mastuhin, “Association Rules with LARAVEL,” 2020. [Online]. Available: <https://github.com/tuhinmas/ta-asosiasi.git>. [Accessed: 20-Dec-2020].

# tode\_Association\_Rule\_Mining\_Untuk\_Menentukan\_Peta\_Pe...

## ORIGINALITY REPORT

20%

SIMILARITY INDEX

21%

INTERNET SOURCES

4%

PUBLICATIONS

%

STUDENT PAPERS

## PRIMARY SOURCES

1	<a href="http://ojs.amikom.ac.id">ojs.amikom.ac.id</a> Internet Source	9%
2	<a href="http://senadi.upy.ac.id">senadi.upy.ac.id</a> Internet Source	3%
3	<a href="http://ejurnal.stmik-budidarma.ac.id">ejurnal.stmik-budidarma.ac.id</a> Internet Source	2%
4	<a href="http://www.scribd.com">www.scribd.com</a> Internet Source	1%
5	<a href="http://www.coursehero.com">www.coursehero.com</a> Internet Source	1%
6	<a href="http://sites.google.com">sites.google.com</a> Internet Source	1%
7	<a href="http://repository.uin-suska.ac.id">repository.uin-suska.ac.id</a> Internet Source	1%
8	<a href="http://www.heise.de">www.heise.de</a> Internet Source	1%
9	<a href="http://prosiding.senadi.upy.ac.id">prosiding.senadi.upy.ac.id</a> Internet Source	1%

10

de.scribd.com

Internet Source

1 %

---

11

doku.pub

Internet Source

1 %

---

Exclude quotes  On

Exclude matches  < 30 words

Exclude bibliography  On