

Project Machine Learning for Retail with R_Product Packaging

Tujuan dari project ini untuk menemukan insight dari kombinasi produk di sebuah toko Fashion.

```
In [1]: install.packages("arules")
```

```
There is a binary version available but the source version is later:
  binary source needs_compilation
arules  1.7-6  1.7-7                TRUE
```

```
Binaries will be installed
package 'arules' successfully unpacked and MD5 sums checked
```

```
The downloaded binary packages are in
  C:\Users\pocad\AppData\Local\Temp\RtmpkRMN3n\downloaded_packages
```

```
In [1]: library(arules)
```

```
# Baca dataset transaksi dari file TSV
transaksi_tabular <- read.transactions(file = "C:\\Users\\pocad\\Downloads\\Project_Produk_Pa

# Tulis dataset transaksi ke dalam file teks dengan delimiter koma (,)
write(transaksi_tabular, file = "C:\\Users\\pocad\\Downloads\\Project_Produk_Packaging\\test_

# Menampilkan dataset transaksi
transaksi_tabular
```

```
Loading required package: Matrix
```

```
Attaching package: 'arules'
```

```
The following objects are masked from 'package:base':
```

```
  abbreviate, write
```

```
transactions in sparse format with
  3450 transactions (rows) and
  69 items (columns)
```

```
In [2]: # Membaca file TSV dengan header
data <- read.delim("C:\\Users\\pocad\\Downloads\\Project_Produk_Packaging\\transaksi_dqlab_re

# Menampilkan beberapa baris data teratas
head(data)

# Menampilkan beberapa baris data terbawah
tail(data)
```

A data.frame: 6 × 2

	Kode.Transaksi	Nama.Barang
	<chr>	<chr>
1	#1	Kaos
2	#1	Shampo Biasa
3	#1	Sepatu Sport merk Z
4	#1	Serum Vitamin
5	#1	Baju Renang Pria Dewasa
6	#1	Baju Renang Wanita Dewasa

A data.frame: 6 × 2

	Kode.Transaksi	Nama.Barang
	<chr>	<chr>
33663	#3450	Wedges Hitam
33664	#3450	Flat Shoes Ballerina
33665	#3450	Sepatu Sandal Anak
33666	#3450	Sepatu Kulit Casual
33667	#3450	Celana Jogger Casual
33668	#3450	Sepatu Sport merk Z

```
In [3]: library(arules)

# Membaca dataset transaksi dari file TSV
data <- read.transactions(file = "C:\\Users\\pocad\\Downloads\\Project_Produk_Packaging\\tran

# Menghitung frekuensi masing-masing item
top10 <- sort(itemFrequency(data, type = "absolute"), decreasing = TRUE)[1:10]

# Membuat data frame dari 10 produk teratas
hasil <- data.frame("Nama Produk" = names(top10), "Jumlah" = as.numeric(top10), row.names = N

# Menyimpan hasil ke dalam file CSV
write.csv(hasil, file = "C:\\Users\\pocad\\Downloads\\Project_Produk_Packaging\\top10.txt", r

# Menampilkan hasil
hasil
```

A data.frame: 10 × 2

Nama.Produk	Jumlah
<chr>	<dbl>
Shampo Biasa	2075
Serum Vitamin	1685
Baju Batik Wanita	1312
Baju Kemeja Putih	1255
Celana Jogger Casual	1136
Cover Koper	1086
Sepatu Sandal Anak	1062
Tali Pinggang Gesper Pria	1003
Sepatu Sport merk Z	888
Wedges Hitam	849

Dalam kumpulan produk dengan frekuensi penjualan tertinggi, terdapat perincian yang menarik. Dari 10 produk teratas, satu produk mampu mencapai angka penjualan melebihi 2000, sementara tujuh produk lainnya memiliki rentang penjualan antara 1000 hingga kurang dari 2000 unit. Sementara itu, dua produk lainnya memperoleh penjualan di bawah angka 1000. Produk yang paling diminati adalah Shampo Biasa, yang mencatat jumlah penjualan tertinggi di atas 2000 unit, khususnya sebanyak 2075 penjualan.

Selanjutnya, tujuh produk lainnya memiliki popularitas yang cukup tinggi, masing-masing dengan angka penjualan antara 1000 hingga kurang dari 2000 unit. Produk-produk ini meliputi Serum Vitamin dengan jumlah penjualan 1685 unit, Baju Batik Wanita (1312 unit), Baju Kemeja Putih (1255 unit), Celana Jogger Casual (1136 unit), Cover Koper (1086 unit), Sepatu Sandal Anak (1062 unit), dan Tali Pinggang Gesper Pria (1003 unit).

Terakhir, terdapat dua produk yang memiliki penjualan di bawah angka 1000, yakni Sepatu Sport merek Z dengan jumlah 888 penjualan dan Wedges Hitam dengan jumlah 849 penjualan.

Dari data tersebut, dapat dilihat bahwa Shampo Biasa menonjol sebagai produk paling diminati dengan penjualan yang signifikan di atas 2000 unit, sementara produk-produk lainnya tetap memiliki kisaran penjualan yang cukup baik, tetapi masih di bawah angka 2000 unit.

```
In [4]: library(arules)

# Membaca dataset transaksi dari file TSV
data <- read.transactions(file = "C:\\Users\\pocad\\Downloads\\Project_Produk_Packaging\\tran

# Menghitung frekuensi masing-masing item
bottom10 <- sort(itemFrequency(data, type = "absolute"), decreasing = FALSE)[1:10]

# Membuat data frame dari 10 produk terbawah
hasil <- data.frame("Nama Produk" = names(bottom10), "Jumlah" = as.numeric(bottom10), row.names =

# Menyimpan hasil ke dalam file CSV
write.csv(hasil, file = "C:\\Users\\pocad\\Downloads\\Project_Produk_Packaging\\bottom10.txt")

# Menampilkan hasil
hasil
```

A data.frame: 10 × 2

Nama.Produk	Jumlah
<chr>	<dbl>
Celana Jeans Sobek Pria	9
Tas Kosmetik	11
Stripe Pants	19
Pelembab	24
Tali Ban Ikat Pinggang	27
Baju Renang Pria Anak-anak	32
Hair Dye	46
Atasan Baju Belang	56
Tas Sekolah Anak Perempuan	71
Dompot Unisex	75

Dalam analisis data penjualan produk, terlihat variasi jumlah penjualan di berbagai produk yang dijual. Celana Jeans Sobek Pria dan Tas Kosmetik masing-masing mencatat penjualan 9 dan 11 unit, sementara Stripe Pants dan pelembab memiliki angka penjualan yang sedikit lebih tinggi dengan masing-masing 19 dan 24 unit.

Terkemuka di antara produk-produk tersebut adalah Tali Ban Ikat Pinggang yang terjual sebanyak 27 unit, diikuti oleh Baju Renang Pria Anak-anak dengan 32 unit penjualan. Di level yang lebih tinggi, Hair Dye mendapatkan popularitas yang cukup besar dengan penjualan sebanyak 46 unit.

Produk Atasan Baju Belang menonjol dengan penjualan mencapai 56 unit, sedangkan Tas Sekolah Anak Perempuan memperoleh jumlah penjualan yang lebih besar lagi, mencapai angka 71 unit. Puncaknya adalah Dompot Unisex yang memimpin dalam penjualan dengan total 75 unit.

```
In [6]: # Mendapatkan Kombinasi Produk yang Menarik
library(arules)

nama_file <- "C:\\Users\\pocad\\Downloads\\Project_Produk_Packaging\\transaksi_dqlab_retail.t
transaksi_tabular <- read.transactions(file=nama_file, format="single", sep="\t", cols=c(1, 2
apriori_rules <- apriori(transaksi_tabular, parameter = list(supp=10/length(transaksi_tabular

Apriori

Parameter specification:
confidence minval smax arem aval originalSupport maxtime      support minlen
      0.5      0.1      1 none FALSE              TRUE        5 0.002898551      2
maxlen target  ext
      3  rules TRUE

Algorithmic control:
filter tree heap memopt load sort verbose
  0.1 TRUE TRUE  FALSE TRUE    2    TRUE

Absolute minimum support count: 10

set item appearances ...[0 item(s)] done [0.00s].
set transactions ...[69 item(s), 3450 transaction(s)] done [0.00s].
sorting and recoding items ... [68 item(s)] done [0.00s].
creating transaction tree ... done [0.00s].
checking subsets of size 1 2 3
```

```
Warning message in apriori(transaksi_tabular, parameter = list(supp = 10/length(transaksi_tabular), :
"Mining stopped (maxlen reached). Only patterns up to a length of 3 returned!"
done [0.00s].
writing ... [4637 rule(s)] done [0.00s].
creating S4 object ... done [0.00s].
```

```
In [7]: apriori_rules <- head(sort(apriori_rules, by="lift", decreasing=T), n=10)
inspect(apriori_rules)
```

lhs	rhs	support	confidence
coverage lift count			
[1] {Tas Makeup, Tas Pinggang Wanita}	=> {Baju Renang Anak Perempuan}	0.010434783	0.8780488 0.0
11884058 24.42958 36			
[2] {Tas Makeup, Tas Travel}	=> {Baju Renang Anak Perempuan}	0.010144928	0.8139535 0.0
12463768 22.64629 35			
[3] {Tas Makeup, Tas Ransel Mini}	=> {Baju Renang Anak Perempuan}	0.011304348	0.7358491 0.0
15362319 20.47322 39			
[4] {Sunblock Cream, Tas Pinggang Wanita}	=> {Kuas Makeup }	0.016231884	0.6913580 0.0
23478261 20.21343 56			
[5] {Baju Renang Anak Perempuan, Tas Pinggang Wanita}	=> {Tas Makeup}	0.010434783	0.8000000 0.0
13043478 19.57447 36			
[6] {Baju Renang Anak Perempuan, Tas Ransel Mini}	=> {Tas Makeup}	0.011304348	0.7959184 0.0
14202899 19.47460 39			
[7] {Baju Renang Anak Perempuan, Celana Pendek Green/Hijau}	=> {Tas Makeup}	0.010144928	0.7777778 0.0
13043478 19.03073 35			
[8] {Tas Makeup, Tas Waist Bag}	=> {Baju Renang Anak Perempuan}	0.004347826	0.6818182 0.0
06376812 18.96994 15			
[9] {Celana Pendek Green/Hijau, Tas Makeup}	=> {Baju Renang Anak Perempuan}	0.010144928	0.6730769 0.0
15072464 18.72674 35			
[10] {Dompot Flip Cover, Sunblock Cream}	=> {Kuas Makeup }	0.016231884	0.6292135 0.0
25797101 18.39650 56			

Pembelian Tas Makeup dan Tas Pinggang Wanita cenderung berhubungan dengan pembelian Baju Renang Anak Perempuan. Dari hasil ini, diperoleh tingkat kepercayaan sekitar 87.8% dan lift sekitar 24.43.

Terdapat asosiasi yang signifikan antara produk Sunblock Cream dan Tas Pinggang Wanita dengan pembelian Kuas Makeup dengan tingkat kepercayaan sekitar 69.1% dan lift sekitar 20.21.

Pola pembelian lainnya menunjukkan bahwa Baju Renang Anak Perempuan berkorelasi dengan produk lain seperti Tas Ransel Mini, Celana Pendek Green/Hijau, Dompot Flip Cover, dan Tas Waist Bag dengan tingkat kepercayaan yang cukup tinggi, berkisar antara 67.3% hingga 79.8%. Hal ini menunjukkan adanya keterkaitan kuat antara pembelian produk tertentu yang mungkin dapat dimanfaatkan dalam strategi pemasaran dan penjualan di masa mendatang.

```
In [9]: # Menulis aturan asosiasi ke dalam file teks
write(apriori_rules, file="C:\\Users\\pocad\\Downloads\\Project_Produk_Packaging\\kombinasi_r
```

Mencari Paket Produk yang bisa dipasangkan dengan item Slow-Moving

Slow-Moving adalah produk yang pergerakan penjualannya lambat dan kurang cepat.

```
In [10]: # Memuat pustaka arules
library(arules)

# Nama file yang berisi transaksi dalam format TSV
nama_file <- "C:\\Users\\pocad\\Downloads\\Project_Produk_Packaging\\transaksi_dqlab_retail.t

# Membaca transaksi dari file TSV dan mengonversinya ke format transaksi arules
transaksi_tabular <- read.transactions(file = nama_file, format = "single", sep = "\t", cols

# Mendapatkan jumlah transaksi
jumlah_transaksi <- length(transaksi_tabular)

# Menentukan jumlah kemunculan minimal yang diinginkan
jumlah_kemunculan_minimal <- 10

# Melakukan analisis aturan asosiasi dengan algoritma Apriori
apriori_rules <- apriori(transaksi_tabular,
                        parameter = list(supp = jumlah_kemunculan_minimal / jumlah_transaksi
                                         conf = 0.1, minlen = 2, maxlen = 3))

# Menampilkan aturan asosiasi yang dihasilkan
inspect(apriori_rules)
```

Apriori

Parameter specification:

confidence	minval	smax	arem	aval	originalSupport	maxtime	support	minlen
0.1	0.1	1	none	FALSE	TRUE	5	0.002898551	2
maxlen	target	ext						
3	rules	TRUE						

Algorithmic control:

filter	tree	heap	memopt	load	sort	verbose
0.1	TRUE	TRUE	FALSE	TRUE	2	TRUE

Absolute minimum support count: 10

```
set item appearances ...[0 item(s)] done [0.00s].
set transactions ...[69 item(s), 3450 transaction(s)] done [0.00s].
sorting and recoding items ... [68 item(s)] done [0.00s].
creating transaction tree ... done [0.00s].
checking subsets of size 1 2 3
```

Warning message in apriori(transaksi_tabular, parameter = list(supp = jumlah_kemunculan_minimal/jumlah_transaksi, :

"Mining stopped (maxlen reached). Only patterns up to a length of 3 returned!"
done [0.00s].

writing ... [39832 rule(s)] done [0.00s].

creating S4 object ... done [0.01s].

IOPub data rate exceeded.

The notebook server will temporarily stop sending output
to the client in order to avoid crashing it.

To change this limit, set the config variable

`--NotebookApp.iopub_data_rate_limit`.

Current values:

NotebookApp.iopub_data_rate_limit=1000000.0 (bytes/sec)

NotebookApp.rate_limit_window=3.0 (secs)

Hasil dari inspect pada algoritma Apriori menunjukkan penggunaan parameter dan kontrol algoritma yang terperinci. Dalam analisis ini, digunakan nilai minimum confidence sebesar 10%, panjang minimum aturan asosiasi sebanyak 2 item, dan panjang maksimum aturan sebanyak 3 item. Proses penambangan aturan asosiasi berjalan dengan pengaturan struktur pohon transaksi dan pemeriksaan subset dari ukuran 1 hingga 3. Meskipun algoritma telah menghasilkan banyak aturan asosiasi (sebanyak 39832),

proses mining dihentikan karena sudah mencapai panjang maksimum yang ditentukan (maxlen reached). Informasi tambahan termasuk statistik terkait data, seperti jumlah item dan transaksi yang diproses, serta pesan peringatan terkait kondisi mining yang dihentikan karena maxlen telah tercapai. Selain itu, output juga memberikan detail mengenai parameter dan kontrol algoritma yang digunakan selama proses penambangan asosiasi.

```
In [11]: # Filter dan mengambil aturan asosiasi yang memenuhi kriteria tertentu
apriori_rules1 <- subset(apriori_rules, lift > 1 & rhs %in% "Tas Makeup")
apriori_rules1 <- apriori_rules1[order(apriori_rules1@quality[, "lift"], decreasing = TRUE)][1]

apriori_rules2 <- subset(apriori_rules, lift > 1 & rhs %in% "Baju Renang Pria Anak-anak")
apriori_rules2 <- apriori_rules2[order(apriori_rules2@quality[, "lift"], decreasing = TRUE)][1]

# Menggabungkan aturan yang telah difilter sebelumnya
apriori_rules <- c(apriori_rules1, apriori_rules2)

# Menampilkan aturan yang telah difilter
inspect(apriori_rules)
```

lhs	rhs	support	confidence	co
verage lift count				
[1] {Baju Renang Anak Perempuan, Tas Pinggang Wanita}	=> {Tas Makeup}	0.010434783	0.8000000	0.01
304348 19.57447 36				
[2] {Baju Renang Anak Perempuan, Tas Ransel Mini}	=> {Tas Makeup}	0.011304348	0.7959184	0.01
420290 19.47460 39				
[3] {Baju Renang Anak Perempuan, Celana Pendek Green/Hijau}	=> {Tas Makeup}	0.010144928	0.7777778	0.01
304348 19.03073 35				
[4] {Gembok Koper, Tas Waist Bag}	=> {Baju Renang Pria Anak-anak}	0.004057971	0.2745098	0.01
478261 29.59559 14				
[5] {Flat Shoes Ballerina, Gembok Koper}	=> {Baju Renang Pria Anak-anak}	0.004057971	0.1866667	0.02
173913 20.12500 14				
[6] {Celana Jeans Sobek Wanita, Jeans Jumbo}	=> {Baju Renang Pria Anak-anak}	0.005507246	0.1210191	0.04
550725 13.04737 19				

Dalam analisis aturan asosiasi menggunakan algoritma Apriori, telah dilakukan filter dan pengambilan aturan yang memenuhi kriteria tertentu berdasarkan lift yang lebih besar dari 1. Kriteria filter melibatkan asosiasi produk-produk tertentu seperti "Tas Makeup" dan "Baju Renang Pria Anak-anak".

Hasilnya menunjukkan beberapa aturan asosiasi yang signifikan:

Ditemukan kombinasi {Baju Renang Anak Perempuan, Tas Pinggang Wanita} yang memiliki lift sebesar 19.57 yang berkorelasi dengan keberadaan "Tas Makeup". Kombinasi {Baju Renang Anak Perempuan, Tas Ransel Mini} juga memiliki korelasi tinggi (lift 19.47) dengan "Tas Makeup". Itemset {Baju Renang Anak Perempuan, Celana Pendek Green/Hijau} menunjukkan korelasi dengan "Tas Makeup" yang cukup signifikan dengan lift 19.03. Selain itu, terdapat beberapa asosiasi antara produk yang berhubungan dengan "Baju Renang Pria Anak-anak" juga ditemukan, seperti kombinasi {Gembok Koper, Tas Waist Bag} yang memiliki korelasi tinggi (lift 29.59) dengan "Baju Renang Pria Anak-anak".

```
In [12]: # Menulis aturan asosiasi ke dalam file teks
write(apriori_rules, file="kombinasi_retail_slow_moving.txt")
```