



Project Machine Learning for Retail with R: Product Packaging



Dataset Retail Toko Fashion

Created by: Huan Wendy Ariono
Last Update: 28 November 2022

Program language



knowledge

Capaian

Dapat menerapkan bahasa pemograman R untuk menyelesaikan persoalan untuk analisis data.



Dapat mengolah dataset skala kecil hingga besar



Dapat menerapkan exploratory data analysis (EDA) untuk persoalan bisnis



Dapat menerapkan algoritma Arules



mengidentifikasi paket produk yang menarik

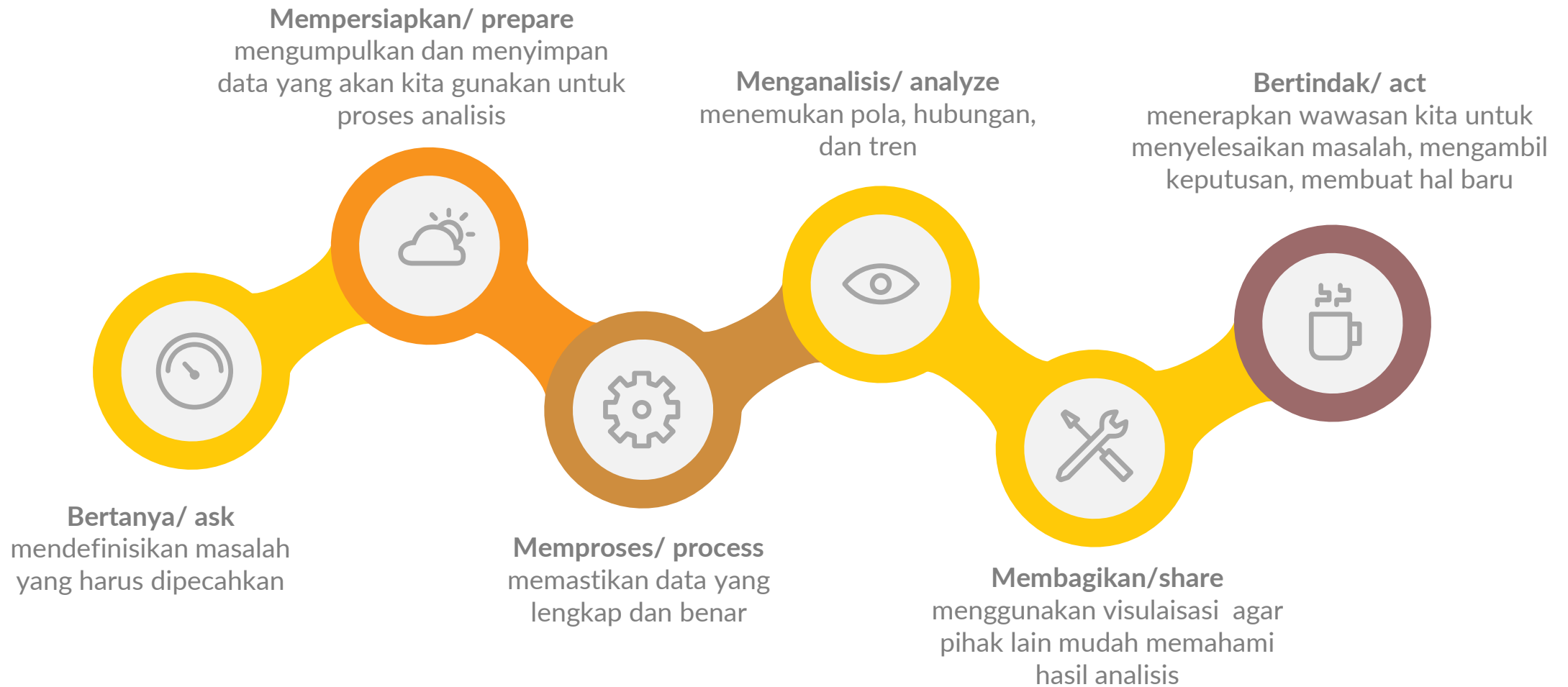


Dapat menerapkan teknik investigasi data-data yang memiliki anomali



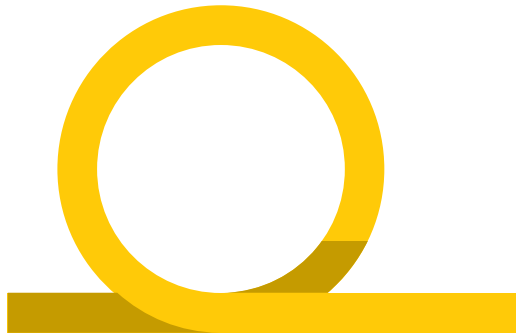
Kunjungi [medium/huans502](https://medium.com/huans502) untuk penjelasan lebih detail

6 FASE ANALISIS DATA



Python

Library Yang Digunakan



Arules

Algoritma apriori dari paket arules digunakan agar computer dapat mempelajari aturan asosiasi, mencari pola hubungan antar satu atau lebih item dalam suatu dataset



Dplyr

Digunakan untuk manipulasi data

Study Case



Sebuah toko fashion yang menjual berbagai produk seperti jeans, kemeja, kosmetik, dan lain-lain. Walaupun cukup berkembang, namun dengan semakin banyaknya kompetitor dan banyak produk yang stoknya masih banyak membuat manajer toko khawatir. Salah satu solusi adalah membuat paket yang inovatif. Dimana produk yang sebelumnya tidak terlalu laku tapi punya pangsa pasar malah bisa dipaketkan dan laku. Disini manajer toko menyuruh kita untuk mengidentifikasi paket produk yang menarik untuk dipaketkan sehingga akhirnya bisa meningkatkan keuntungan dan loyalitas para pelanggan toko fashion. Dan untuk mewujudkan ini, saya menggunakan R dan algoritma apriori dari paket arules di sepanjang project ini

Untuk coding lebih jelas silahkan kunjungi <https://rpubs.com/Worstone57/961353>

Bertanya/ ask



Dari study case diatas dapat kita ketahui pemangku kepentingannya adalah manajer toko. Harapan dari manajer tersebut adalah agar kita mengidentifikasi paket produk yang menarik

Mempersiapkan/ prepare



Dalam case ini kita sudah disediakan data oleh pihak toko yang mana berisi 2 column saja yaitu kode_transaksi dan Nama_barang. Data yang disediakan dalam format tsv yang memiliki 33.669 baris dan 3.450 kode transaksi

Memproses/ process

Seperti yang sudah disinggung diatas disini saya menggunakan R dan algoritma aproriari dari paket arules. Algoritma aproriari merupakan salah satu algoritma klasik untuk mencari frequent item/itemset pada transactional database. Algoritma Apriori digunakan agar computer dapat mempelajari aturan asosiasi, mencari pola hubungan antar satu atau lebih item dalam suatu dataset. Dalam study case ini kita memiliki data transaksi/market basket sehingga kita dapat melakukan analisa terhadap kemungkinan membeli product yang lainnya.

```
#Menyajikan informasi tiap kolom dataset dalam format yang compact
str(dataset_raw)
```

```
## 'data.frame': 33668 obs. of 2 variables:
## $ Kode.Transaksi: chr "#1" "#1" "#1" "#1" ...
## $ Nama.Barang : chr "Kaos" "Shampo Biasa" "Sepatu Sport merk Z" "Serum Vitamin" ...
```

```
#tampilkan 6 data teratas
head(dataset_raw)
```

##	Kode.Transaksi	Nama.Barang
## 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

- Cek tipe data sebelum diproses, Dari data diatas dapat kita ketahui jumlah baris data ada 33.668 baris yang terdiri dari 2 column yaitu Kode.Transaksi dan Nama.Barang dengan tipe data nya chr

Cek 6 data teratas sebelum di proses

Lanjutan

Setelah kita pastikan data sesuai kita lanjutkan untuk menggunakan arules

```
#cek tipe datanya  
class(transaksi_tabular)
```

```
## [1] "transactions"  
## attr(,"package")  
## [1] "arules"
```

```
#Cek nilai null  
is.null(dataset_raw)
```

```
## [1] FALSE
```

```
#rubah bentuk data frame
```

```
transaksi_tabular <- read.transactions(file="transaksi_dqlab_retail.tsv", format="single", sep="\t", cols=c(1,2), skip=1)
```

```
#simpan hasil konversi ke dalam file test_project_retail_1.txt
```

```
write(transaksi_tabular, file="test_project_retail_1.txt", sep=",")
```

```
#cek data transaksi_tabular
```

```
transaksi_tabular
```

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

Memastikan dataset bebas dari null value

Pertama kita ubah format dari dataset transaksi_dqlab_retail.tsv menjadi format yang dapat digunakan oleh paket arules

Lanjutan

Setelah kita pastikan data sesuai kita lanjutkan untuk merubah bentuk data agar dapat menggunakan arules

```
#cek tipe datanya  
class(transaksi_tabular)
```

```
## [1] "transactions"  
## attr(,"package")  
## [1] "arules"
```

```
#Cek nilai null  
is.null(dataset_raw)
```

```
## [1] FALSE
```

```
#rubah bentuk data frame
```

```
transaksi_tabular <- read.transactions(file="transaksi_dqlab_retail.tsv", format="single", sep="\t", cols=c(1,2), skip=1)
```

```
#simpan hasil konversi ke dalam file test_project_retail_1.txt
```

```
write(transaksi_tabular, file="test_project_retail_1.txt", sep=",")
```

```
#cek data transaksi_tabular
```

```
transaksi_tabular
```

```
## transactions in sparse format with
```

```
## 3450 transactions (rows) and
```

```
## 69 items (columns)
```

Memastikan dataset bebas dari null value


Pertama kita ubah format dari dataset transaksi_dqlab_retail.tsv menjadi format yang dapat digunakan oleh paket arules

Menganalisis/ analyze & Membagikan/share




1

Mendapatkan insight top 10 dan bottom 10 dari produk yang terjual.




2

Mendapatkan daftar seluruh kombinasi paket produk dengan korelasi yang kuat.



3

Mendapatkan daftar seluruh kombinasi paket produk dengan item tertentu.



Output Awal: Statistik Top 10

```
#rubah bentuk data frame ke bentuk tabular seperti variable transaksi_tabular diatas
data <- read.transactions(file = "transaksi_dqlab_retail.tsv", format = "single", sep = "\t", cols = c(1,2), skip=1)

#select top 10
#decreasing true artinya diurutkan dari besar ke kecil, false dari kecil ke besar
top_10 <- sort(itemFrequency(data, type="absolute"), decreasing = TRUE)[1:10] #menggunakan cara slicing
#tampilkan isinya
top_10
```

```
##           Shampo Biasa           Serum Vitamin           Baju Batik Wanita
##           2075           1685           1312
##           Baju Kemeja Putih           Celana Jogger Casual           Cover Koper
##           1255           1136           1086
##           Sepatu Sandal Anak Tali Pinggang Gesper Pria           Sepatu Sport merk Z
##           1062           1003           888
##           Wedges Hitam
##           849
```

Output Awal: Statistik Bottom 10

```
#rubah bentuk data frame kebentuk tabular seperti variable transaksi_tabular diatas
data <- read.transactions(file ="transaksi_dqlab_retail.tsv", format = "single", sep = "\t", cols = c(1,2), skip=1)

#select top 10
#decreasing true artinya diurutkan dari besar ke kecil, false dari kecil ke besar
bottom_10 <- sort(itemFrequency(data, type="absolute"), decreasing = FALSE)[1:10] #menggunakan cara slicing
#tampilkan isinya
bottom_10
```

```
##      Celana Jeans Sobek Pria          Tas Kosmetik
##                9                11
##      Stripe Pants          Pelembab
##                19                24
##      Tali Ban Ikat Pinggang Baju Renang Pria Anak-anak
##                27                32
##      Hair Dye          Atasan Baju Belang
##                46                56
## Tas Sekolah Anak Perempuan          Dompet Unisex
##                71                75
```

Penjelasan Algoritma Apriori

Apa itu Algoritma Apriori?

Merupakan salah satu algoritma klasik untuk mencari frequent item/itemset pada transactional database. Algoritma Apriori digunakan agar computer dapat mempelajari aturan asosiasi, mencari pola hubungan antar satu atau lebih item dalam suatu dataset. Dengan memiliki data transaksi/market basket, Kita dapat melakukan Analisa terhadap kemungkinan membeli product yang lainnya. Berbagai algorithm yang digunakan dalam 'association rule mining' meliputi algorithm Apriori yang sangat terkenal (dimana sekumpulan item yang sering muncul diidentifikasi).

Apa itu Association?

Pengelompokkan hal yang biasanya terjadi bersamaan, misalnya 10 orang membeli susu kental manis, 5 orang dari 10 orang tersebut membeli keju lalu disimpulkan bahwa jika pelanggan membeli susu kental manis ia juga akan membeli keju.

Apa itu Confidence?

Paramenter yang digunakan, nilai confidence (nilai kepastian) yaitu kuatnya hubungan antar item dalam aturan assosiatif.

Apa itu Lift Case?

Peningkatan rasio penjualan suatu produk Ketika produk lain dijual

Mendapatkan Kombinasi Produk yang menarik

```
#n = 10 paket kombinasi produk yang memenuhi syarat
#Kemudian kita sort berdasarkan nilai lift terbesar
apriori_rules <- head(sort(apriori_rules, by='lift', decreasing = T),n=10)
print(apriori_rules)
```

```
## set of 10 rules
```

Kemudian kita sort berdasarkan 10 nilai lift terbesar. Kita memilih nilai lift karena Lift adalah peningkatan rasio penjualan suatu produk saat kita menjual produk lain

```
# menampilkan hasilnya
print(inspect(apriori_rules))
```

##	lhs	rhs	support	confidence	coverage	lift	count
## [1]	{Tas Makeup,	=> {Baju Renang Anak Perempuan}	0.010434783	0.8780488	0.011884058	24.42958	36
##	Tas Pinggang Wanita}						
## [2]	{Tas Makeup,	=> {Baju Renang Anak Perempuan}	0.010144928	0.8139535	0.012463768	22.64629	35
##	Tas Travel}						
## [3]	{Tas Makeup,	=> {Baju Renang Anak Perempuan}	0.011304348	0.7358491	0.015362319	20.47322	39
##	Tas Ransel Mini}						
## [4]	{Sunblock Cream,	=> {Kuas Makeup }	0.016231884	0.6913580	0.023478261	20.21343	56
##	Tas Pinggang Wanita}						
## [5]	{Baju Renang Anak Perempuan,	=> {Tas Makeup}	0.010434783	0.8000000	0.013043478	19.57447	36
##	Tas Pinggang Wanita}						
## [6]	{Baju Renang Anak Perempuan,	=> {Tas Makeup}	0.011304348	0.7959184	0.014202899	19.47460	39
##	Tas Ransel Mini}						
## [7]	{Baju Renang Anak Perempuan,	=> {Tas Makeup}	0.010144928	0.7777778	0.013043478	19.03073	35
##	Celana Pendek Green/Hijau}						
## [8]	{Tas Makeup,	=> {Baju Renang Anak Perempuan}	0.004347826	0.6818182	0.006376812	18.96994	15
##	Tas Waist Bag}						
## [9]	{Celana Pendek Green/Hijau,	=> {Baju Renang Anak Perempuan}	0.010144928	0.6730769	0.015072464	18.72674	35
##	Tas Makeup}						
## [10]	{Dompot Flip Cover,	=> {Kuas Makeup }	0.016231884	0.6292135	0.025797101	18.39650	56
##	Sunblock Cream}						
##	NULL						

```
#Hasilnya disimpan ke file kombinasi_retail.txt
write(apriori_rules, file="kombinasi_retail.txt")
```

Detail

Pada tahap ini kita akan membuat daftar 10 paket kombinasi produk yang memenuhi syarat dibawah ini:

- Memiliki asosiasi atau hubungan erat.
- Kombinasi produk minimal 2 item, dan maksimum 3 item.
- Kombinasi produk itu muncul setidaknya 10 dari dari seluruh transaksi.
- Memiliki tingkat confidence minimal 50 persen.

Mencari Paket Produk yang bisa dipasangkan dengan Item Slow-Moving

Slow-moving item adalah produk yang pergerakan penjualannya lambat atau kurang cepat. Ini akan bermasalah apabila item produk tersebut masih menumpuk. Kadang kala item ini belum tentu tidak laku, hanya saja mungkin harganya tidak bagus dan jarang dibutuhkan jika dijual satuan. Nah, jika tidak dijual satuan kita perlu cari asosiasi kuat dari item produk ini dengan produk lain sehingga jika dipaketkan akan menjadi lebih menarik.

Menurut manajer toko ada 2 produk slow-moving di tokonya adalah “Tas Makeup” dan “Baju Renang Pria Anak-anak”. Manajer toko ingin meminta kombinasi yang bisa dipaketkan dengan kedua produk tersebut

Masing-masing produk tersebut dikeluarkan 3 rules yang asosiasinya paling kuat, sehingga total ada 6 rules (kenapa ada 6 hal itu karena minnya 2 maxnya 3 itu dikalikan hasilnya 6 untuk kemungkinannya). Persyaratan-persyaratan asosiasi kuat ini masih sama dengan yang telah disebutkan manajer toko sebelumnya, kecuali tingkat confidence dicoba pada tingkat minimal 0.1.

Jika lupa bisa cek dibawah ini untuk syaratnya

1. Memiliki asosiasi atau hubungan erat.
2. Kombinasi produk minimal 2 item, dan maksimum 3 item.
3. Kombinasi produk itu muncul setidaknya 10 dari dari seluruh transaksi.
4. Memiliki tingkat confidence minimal 10 persen.

Lanjutan

```
#Tampilkan rulesnya
inspect(apriori_rules)
```

##	lhs	rhs	support	confidence	coverage	lift	count
## [1]	{Baju Renang Anak Perempuan,						
##	Tas Pinggang Wanita}	=> {Tas Makeup}	0.010434783	0.8000000	0.01304348	19.57447	36
## [2]	{Baju Renang Anak Perempuan,						
##	Tas Ransel Mini}	=> {Tas Makeup}	0.011304348	0.7959184	0.01420290	19.47460	39
## [3]	{Baju Renang Anak Perempuan,						
##	Celana Pendek Green/Hijau}	=> {Tas Makeup}	0.010144928	0.7777778	0.01304348	19.03073	35
## [4]	{Gembok Koper,						
##	Tas Waist Bag}	=> {Baju Renang Pria Anak-anak}	0.004057971	0.2745098	0.01478261	29.59559	14
## [5]	{Flat Shoes Ballerina,						
##	Gembok Koper}	=> {Baju Renang Pria Anak-anak}	0.004057971	0.1866667	0.02173913	20.12500	14
## [6]	{Celana Jeans Sobek Wanita,						
##	Jeans Jumbo}	=> {Baju Renang Pria Anak-anak}	0.005507246	0.1210191	0.04550725	13.04737	19

Detail

Dari data didamping dapat kita lihat kombinasi yang bisa dipaketkan dengan kedua produk slow-moving di toko yaitu produk “Tas Makeup” dan “Baju Renang Pria Anak-anak”

#DQLABPRJCTWBGVEP

CERTIFICATE OF COMPLETION

This certificate is proudly presented to

Huan Wendy Ariono

Has Completed in

Project Machine Learning for Retail with R: Product Packaging

Oct 20, 2022



A yellow horizontal line spanning the width of the image, with a small yellow dot at each end, positioned behind the word Selesai.

Selesai

THANKS FOR WATCHING

Huan Wendy Ariono

Link Portofolio

Github

<https://github.com/huwea>



Rpubs

<https://rpubs.com/Worstone57>



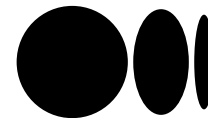
Tableau

<https://public.tableau.com/app/profile/huan.wendy.ariono>



Medium

<https://medium.com/@huans502>





About

My name is Huan Wendy Ariono, I am Fresh Graduate of Informatics Engineering at University Muhammadiyah Surakarta. Currently I focus on data analysis. I am also active in adding new knowledge in the field of data by attending courses, workshops, reading articles and writing articles related to data in the medium.

My experience in data field is being able to use python, R, SQL, excel, googlesheet and tableau as well as other tools to analyze data and get valuable input to solve problems. I got these skills through lectures, independent projects and taking courses related to data.



LINKEDIN

<https://www.linkedin.com/in/huanwendyariono/>

