Project Machine Learning for Retail with R: Product Packaging

Dormauli Sinta Bella N

# Permasalahan

DQLab.id Fashion adalah 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 tentunya membuat kuatir Pak Agus, manajer DQLab.id Fashion.

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 saya akan membuat paket barang yang inovatif dengan menggunakan R dan algoritma aproriari dari paket arules di sepanjang project ini.

# Data

Data yang digunakan terdiri atas 2 variabel, yaitu Kode Transaksi dan Nama Produk dengan jumlah baris 33,669 baris data (3,450 kode transaksi). Data yang diberikan adalah data transaksi selama 3 bulan.

library(arules)  
transaksi\_tabular <- read.transactions(file = "C:/Users/ASUS/Downloads/transaksi\_dqlab\_retail.tsv", format="single", sep="\t", cols=c(1,2), skip=1)  
transaksi\_tabular

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

write(transaksi\_tabular, file="C:/Users/ASUS/Downloads/test\_project\_retail\_1.txt", sep=",")

# Goals

* Mendapatkan insight top 10 dan bottom 10 dari produk yang terjual.
* Mendapatkan daftar seluruh kombinasi paket produk dengan korelasi yang kuat.
* Mendapatkan daftar seluruh kombinasi paket produk dengan item tertentu.

# Output Awal: Statistik Top 10

Disini Saya akan menampilkan 10 produk dengan jumlah transaksi paling banyak

data\_item <- itemFrequency(transaksi\_tabular, type="absolute")  
data\_item <- sort(data\_item, decreasing = TRUE)  
data\_item <- data\_item[1:10]  
data\_item <- data.frame("Nama Produk"=names(data\_item), "Jumlah"=data\_item, row.names=NULL)  
data\_item

## Nama.Produk Jumlah  
## 1 Shampo Biasa 2075  
## 2 Serum Vitamin 1685  
## 3 Baju Batik Wanita 1312  
## 4 Baju Kemeja Putih 1255  
## 5 Celana Jogger Casual 1136  
## 6 Cover Koper 1086  
## 7 Sepatu Sandal Anak 1062  
## 8 Tali Pinggang Gesper Pria 1003  
## 9 Sepatu Sport merk Z 888  
## 10 Wedges Hitam 849

# Output Awal: Statistik Bottom 10

Disini Saya akan menampilkan 10 produk dengan jumlah transaksi paling sedikit

data\_item <- itemFrequency(transaksi\_tabular, type="absolute")  
data\_item <- sort(data\_item, decreasing = FALSE)  
data\_item <- data\_item[1:10]  
data\_item <- data.frame("Nama Produk"=names(data\_item), "Jumlah"=data\_item, row.names=NULL)  
data\_item

## Nama.Produk Jumlah  
## 1 Celana Jeans Sobek Pria 9  
## 2 Tas Kosmetik 11  
## 3 Stripe Pants 19  
## 4 Pelembab 24  
## 5 Tali Ban Ikat Pinggang 27  
## 6 Baju Renang Pria Anak-anak 32  
## 7 Hair Dye 46  
## 8 Atasan Baju Belang 56  
## 9 Tas Sekolah Anak Perempuan 71  
## 10 Dompet Unisex 75

# Mendapatkan Kombinasi Produk yang Menarik

Saya akan melakukan kombinasi untuk mendapatkan 10 paket kombinasi produk yang paling “menarik”.

Definisi menarik disini sebagai berikut:

* 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.

apriori\_rules <- apriori(transaksi\_tabular, parameter = list(support=10/length(transaksi\_tabular),conf=0.5, minlen=2, maxlen=3))

## 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 FALSE  
##   
## 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 [0.01s].  
## sorting and recoding items ... [68 item(s)] done [0.00s].  
## creating transaction tree ... done [0.00s].  
## checking subsets of size 1 2 3 done [0.03s].  
## writing ... [4637 rule(s)] done [0.02s].  
## creating S4 object ... done [0.01s].

apriori\_rules <- head(sort(apriori\_rules, by='lift', decreasing = T), n=10)  
inspect(apriori\_rules)

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

# Mencari Paket Produk yang bisa dipasangkan dengan Item *Slow-Moving*

Apa itu *Slow-Moving item*? *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.

Misalkan item “Tas Makeup” dan “Baju Renang Pria Anak-anak” adalah item *Slow-Moving*. dan kita akan melakukan hal yang sama seperti sebelumnya dengan kriteria sebagai berikut:

* 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 10 persen.

Masing-masing produk tersebut dikeluarkan 3 rules yang asosiasinya paling kuat, sehingga total ada 6 rules dan kedua item tersebut berada di rhs.

mba<- apriori(transaksi\_tabular, parameter= list(supp=10/length(transaksi\_tabular), conf=0.1, minlen=2, maxlen=3))

## 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 FALSE  
##   
## 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 [0.02s].  
## sorting and recoding items ... [68 item(s)] done [0.01s].  
## creating transaction tree ... done [0.01s].  
## checking subsets of size 1 2 3 done [0.05s].  
## writing ... [39832 rule(s)] done [0.07s].  
## creating S4 object ... done [0.16s].

mba1 <- head(subset(mba, rhs %in% "Tas Makeup"), by = "lift", n = 3L)  
mba2 <- head(subset(mba, rhs %in% "Baju Renang Pria Anak-anak"), by = "lift", n = 3L)  
mba = c(mba1, mba2)  
inspect(mba)

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

# Kesimpulan

Dengan menggunakan algoritma apriori kita mendapatkan 10 kombinasi produk terbaik yang akan diberikan kepada manajer DQLab.id Fashion untuk dipertimbangkan kembali kombinasi apa yang akan nantinya digunakan dalam pasar.