

PROJECT MACHINE LEARNING FOR RETAIL WITH R – PRODUCT PACKAGING

BACKGROUND

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 khawatir 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.

Anda sebagai seorang *data scientist*, akan ditugaskan membantu Pak Agus untuk mengidentifikasi paket produk yang menarik untuk dipaketkan sehingga akhirnya bisa meningkatkan keuntungan dan loyalitas para pelanggan DQLab.id Fashion. Dan untuk wewujudkan ini, Anda akan menggunakan R dan algoritma apriori dari paket *arules* di sepanjang project ini.

Untuk memulai project ini, Pak Agus meminta Pak Charlie, data engineer dari DQLab.id Fashion memberikan data transaksi 3 bulan kepada Anda dalam bentuk format TSV (*Tab Separated Value*) dengan nama transaksi_dqlab_retail.tsv dengan jumlah baris 33,669 baris data (3,450 kode transaksi).

Berikut adalah tampilan datanya :

```
Kode Transaksi  Nama Barang
#1  Kaos
#1  Shampo Biasa
#1  Sepatu Sport merk Z
#1  Serum Vitamin
#1  Baju Renang Pria Dewasa
#1  Baju Renang Wanita Dewasa
#1  Baju Kaos Olahraga
#1  Celana Jogger Casual
#1  Dompot Card Holder
#2  Celana Jeans Sobek Wanita
#2  Blouse Denim
#2  Baju Batik Wanita
#2  Hair and Scalp
#2  Shampo Biasa
#2  Minyak Rambut
#2  Wedges Hitam
#2  Sepatu Sandal Anak
#2  Tas Sekolah Anak Perempuan
#2  Baju Kemeja Putih
#2  Dompot Flip Cover
#3  Shampo Biasa
#3  Hair Tonic
```

Data transaksi di atas telah dirapikan dan hanya mengandung dua variabel, yaitu:

- Kode Transaksi
- Nama Barang

Variabel lain seperti harga, tanggal, jumlah pembelian, dan lain-lain – tidak dimasukkan sesuai permintaan pihak DQLab.id Academy kepada DQLab.id Fashion dengan alasan dua variabel tersebut sudah cukup.

READ THE DATA

```
1 library(arules)
2 transaksi_tabular <- read.transactions(file="https://storage.googleapis.com/dqlab-dataset/transaksi_dqlab_retail.tsv", format="single", sep="\t", cols=c(1,2), skip=1)
3 write(transaksi_tabular, file="test_project_retail_1.txt", sep=",")
```

SHOW ALL DATA

```
1 all_item <- itemFrequency(transaksi_tabular, type = "absolute")
2 names(all_item)
```

Output after input syntax above:

```
## [1] "Atasan Baju Belang" "Atasan Kaos Putih"
## [3] "Baju Batik Wanita" "Baju Kaos Anak - Karakter Kartun"
## [5] "Baju Kaos Anak - Superheroes" "Baju Kaos Olahraga"
## [7] "Baju Kemeja Putih" "Baju Renang Anak Perempuan"
## [9] "Baju Renang Pria Anak-anak" "Baju Renang Pria Dewasa"
## [11] "Baju Renang Wanita Dewasa" "Blouse Denim"
## [13] "Celana Jeans Sobek Pria" "Celana Jeans Sobek Wanita"
## [15] "Celana Jogger Casual" "Celana Panjang Format Hitam"
## [17] "Celana Pendek Casual" "Celana Pendek Green/Hijau"
## [19] "Celana Pendek Jeans" "Celana Tactical "
## [21] "Cover Koper" "Cream Whitening"
## [23] "Dompot Card Holder" "Dompot Flip Cover"
## [25] "Dompot Kulit Pria" "Dompot STNK Gantungan"
## [27] "Dompot Unisex" "Flat Shoes Ballerina"
## [29] "Gembok Koper" "Hair and Scalp"
## [31] "Hair Dryer" "Hair Dye"
## [33] "Hair Tonic" "Jeans Jumbo"
## [35] "Kaos" "Koper Fiber"
## [37] "Kuas Makeup " "Mascara"
## [39] "Minyak Rambut" "Obat Penumbuh Rambut"
## [41] "Pelembab" "Sepatu Kulit Casual"
## [43] "Sepatu Sandal Anak" "Sepatu Sekolah Hitam W"
```

```
## [45] "Sepatu Sport merk Y"          "Sepatu Sport merk Z"
## [47] "Serum Vitamin"                "Shampo Anti Dandruff"
## [49] "Shampo Biasa"                 "Stripe Pants"
## [51] "Sunblock Cream"               "Sweater Top Panjang"
## [53] "Tali Ban Ikat Pinggang"       "Tali Pinggang Anak"
## [55] "Tali Pinggang Gesper Pria"    "Tank Top"
## [57] "Tas Kosmetik"                 "Tas Kulit Selempang"
## [59] "Tas Makeup"                   "Tas Multifungsi"
## [61] "Tas Pinggang Wanita"         "Tas Ransel Mini"
## [63] "Tas Sekolah Anak Laki-laki"   "Tas Sekolah Anak Perempuan"
## [65] "Tas Tangan"                   "Tas Travel"
## [67] "Tas Waist Bag"                "Wedges Hitam"
## [69] "Woman Ripped Jeans "
```

Top 10 Products

```
1 library(arules)
2 transaksi_tabular <-
  read.transactions(file="transaksi_dqlab_retail.tsv",format="single",sep="\t",cols=c(1,
  2),skip=1)
3 top10 <- itemFrequency(transaksi_tabular, type="absolute")
4 top10 <- sort(top10,decreasing=TRUE)[1:10]
5 top10 <- data.frame("Nama.Produk"=names(top10), "Jumlah"=top10,row.names=NULL)
6 write.csv(top10,file="top10_item_retail.txt",sep=",")
```

Output after input syntax above:

```
"", "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
```

SHOWING THE BOTTOM 10 PRODUCTS

```
1 library(arules)
2 transaksi_tabular <- read.transactions(file="transaksi_dqlab_retail.tsv",
   format="single", sep="\t", cols= c(1,2), skip=1)
3 data_bottom <- itemFrequency(transaksi_tabular, type="absolute")
4 data_bottom <- sort(data_bottom, decreasing = FALSE)[1:10]
5 data_bottom <- data.frame ("Nama.Produk" = names(data_bottom), "Jumlah" =
   data_bottom, row.names=NULL)
6 write.csv(data_bottom, file="bottom10_item_retail.txt")
```

Output after input syntax above:

```
"", "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", "Dompot Unisex", 75
```

GET ATTRACTIVE COMBINATION PRODUCTS

Definisi kombinasi produk yang menarik adalah sebagai berikut:

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 50 persen.

```
1 library(arules)
2 transaksi_tabular <- read.transactions(file="transaksi_dqlab_retail.tsv",
   format="single", sep="\t", cols= c(1,2), skip=1)
3 data_bottom <- itemFrequency(transaksi_tabular, type="absolute")
4 data_bottom <- sort(data_bottom, decreasing = FALSE)[1:10]
5 data_bottom <- data.frame ("Nama.Produk" = names(data_bottom), "Jumlah" =
   data_bottom, row.names=NULL)
6 write.csv(data_bottom, file="bottom10_item_retail.txt")
```

Output after input syntax above:

```
"rules" "support" "confidence" "lift"
"797" "{Tas Makeup,Tas Pinggang Wanita} => {Baju Renang Anak Perempuan}" 0.0104347826086957
0.878048780487805 24.4295830055075
"807" "{Tas Makeup,Tas Travel} => {Baju Renang Anak Perempuan}" 0.0101449275362319 0.8139534
88372093 22.6462865716429
"802" "{Tas Makeup,Tas Ransel Mini} => {Baju Renang Anak Perempuan}" 0.011304347826087 0.735
849056603774 20.4732197200243
"649" "{Sunblock Cream,Tas Pinggang Wanita} => {Kuas Makeup}" 0.016231884057971 0.691358024
691358 20.2134337727558
"796" "{Baju Renang Anak Perempuan,Tas Pinggang Wanita} => {Tas Makeup}" 0.0104347826086957
0.8 19.5744680851064
"801" "{Baju Renang Anak Perempuan,Tas Ransel Mini} => {Tas Makeup}" 0.011304347826087 0.795
918367346939 19.4745983499783
"791" "{Baju Renang Anak Perempuan,Celana Pendek Green/Hijau} => {Tas Makeup}" 0.01014492753
62319 0.777777777777778 19.0307328605201
"799" "{Tas Makeup,Tas Waist Bag} => {Baju Renang Anak Perempuan}" 0.00434782608695652 0.681
818181818182 18.9699413489736
"792" "{Celana Pendek Green/Hijau,Tas Makeup} => {Baju Renang Anak Perempuan}" 0.01014492753
62319 0.673076923076923 18.7267369727047
"656" "{Dompot Flip Cover,Sunblock Cream} => {Kuas Makeup}" 0.016231884057971 0.62921348314
6067 18.3964959055418
```

LOOKING FOR PRODUCT PACKAGES THAT CAN BE PAIRED WITH SLOW-MOVING ITEMS

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.

Pak Agus juga meyakini hal ini, dan ingin agar Anda membantu mengidentifikasi dua item produk yang menurut dia stoknya masih banyak dan perlu dicari pasangan item untuk pemaketannya.

```
1 library(arules)
2 transaksi_tabular <- read.transactions(file="transaksi_dqlab_retail.tsv",
  format="single", sep="\t", cols=c(1,2), skip=1)
3 kombinasi <- apriori(transaksi_tabular,
  parameter=list(supp=10/length(transaksi_tabular), conf=0.1, minlen=2, maxlen=3))
4 rules1 <- subset(kombinasi, rhs %in% "Tas Makeup")
5 rules2 <- subset(kombinasi, rhs %in% "Baju Renang Pria Anak-anak")
6 combineRules1 <- sort(rules1, decreasing=TRUE, by="lift")[1:3]
7 combineRules2 <- sort(rules2, decreasing=TRUE, by="lift")[1:3]
8 rules <- c(combineRules1,combineRules2)
9 write(rules, file="kombinasi_retail_slow_moving.txt")
```

Output after input syntax above:

```
> library(arules)

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

> kombinasi <- 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 [0.00s].
sorting and recoding items ... [68 item(s)] done [0.00s].
creating transaction tree ... done [0.00s].
checking subsets of size 1 2 3 done [0.00s].
writing ... [39832 rule(s)] done [0.00s].
creating S4 object ... done [0.01s].

> rules1 <- subset(kombinasi, rhs %in% "Tas Makeup")
> rules2 <- subset(kombinasi, rhs %in% "Baju Renang Pria Anak-anak")
> combineRules1 <- sort(rules1, decreasing=TRUE, by="lift")[1:3]
> combineRules2 <- sort(rules2, decreasing=TRUE, by="lift")[1:3]
> rules <- c(combineRules1,combineRules2)
> write(rules, file="kombinasi_retail_slow_moving.txt")
```

Dua item produk tersebut adalah "**Tas Makeup**" dan "**Baju Renang Pria Anak-anak**". Pak Agus ingin meminta kombinasi yang bisa dipaketkan dengan kedua produk tersebut.

CLOSING

That's all for this project and thanks for watching.

GET IN TOUCH :

LinkedIn [linkedin.com/in/vicky-jodie](https://www.linkedin.com/in/vicky-jodie)

GitHub github.com/vickyjodie