**Aplikasi Market Basket Analysis Pada Transaksi Barang di Google Store Tahun 2017**

TUGAS UTS DATA MINING



Deta Putri Prakoso

1706047315

1 November 2019

**LATAR BELAKANG**

Google merupakan sebuah perusahaan multinasional Amerika Serikat yang berkekhususan pada jasa dan produk Internet. Produk-produk tersebut meliputi teknologi pencarian, komputasi web, perangkat lunak, dan periklanan daring. Sejak didirikan, pertumbuhan perusahaan yang cepat telah menghasilkan berbagai produk dan layanan. Salah satunya adalah *e-commerce*, Google Store, yang menjual produk fisik dari Google seperti *merchandise* resmi dan alat-alat elektronik.

Seperti pada *e-commerce* lainnya, tentu dibutuhkan rekomendasi bagi pengunjung/pelanggan ataupun tim pemasaran Google guna menaikkan pendapatan dan kepuasan pelanggan. *Market Basket Analysis* dapat diterapkan untuk mengetahui mana saja produk yang sering dibeli bersamaan ataupun produk yang dapat direkomendasikan untuk di *bundle*.

1. **Data**

Data diambil dari Kaggle, dengan tautan sebagai berikut: <https://www.kaggle.com/jpallard/google-store-ecommerce-data-fake-retail-data>

File “Online.csv” berisi data pesanan/transaksi aktual yang diimpor secara manual dari akses publik Google Store (<https://store.google.com/us/>) melalui Google Analytics. Data berupa jejak transaksi selama satu tahun, yaitu dari 1 Januari 2017 hingga 31 Desember 2017.

Beberapa observasi pertama dari data transaksi Google Store:

A screenshot of a computer

Description automatically generated

1. **Tujuan Analisis Data**

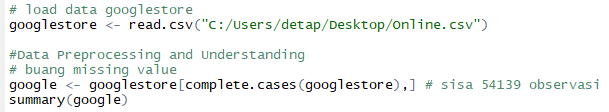
* Mencari tahu produk apa yang sering dibeli bersamaan.
* Menentukan produk apa yang sebaiknya direkomendasikan, jika diketahui pelanggan telah membeli produk tertentu.

1. **Langkah-langkah Analisis Data**

Untuk dokumentasi lengkap, dapat diakses pada file “uts data mining\_deta\_MBA google store.R” yang terlampir pada *e-mail*.

Langkah yang dilakukan dalam mengolah data adalah sebagai berikut.

**Langkah 1. *Data Preprocessing***



Dataset diimpor ke lingkungan R. Kemudian, didefinisikan variabel baru yaitu **google** yang membuang *missing value* pada data awal. Digunakan fungsi *summary* untuk melihat statistika deskriptif dari data yang dipakai.

A close up of text on a white background

Description automatically generated

Untuk memastikan tipe data yang akan kita gunakan sesuai dengan yang diinginkan didefinisikan:A picture containing indoor

Description automatically generatedA screenshot of a cell phone

Description automatically generated

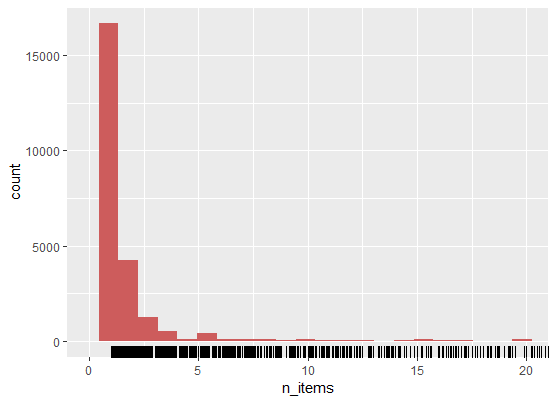
Setelah *preprocessing*, dataset memuat 52.924 observasi dengan 10 variabel.

**Langkah 2. *Data Understanding***

Ingin diketahui berapa banyak barang yang pelanggan beli di Google Store.

A picture containing indoor

Description automatically generated

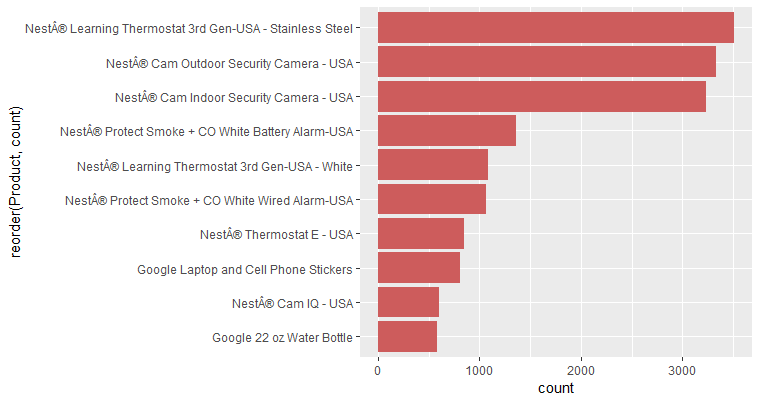


Artinya, kebanyakan pelanggan Google Store di tahun 2017 hanya membeli kurang dari 5 barang per transaksinya.

Ingin diketahui apa saja barang pada Google Store yang paling laku dibeli oleh pelanggan.

A picture containing indoor

Description automatically generated



*Best-selling products* pada Google Store di tahun 2017 adalah *“Nest Learning Thermostat 3rd Gen-USA – Stainless Stell.”*

**Langkah 3. Menyiapkan Data Transaksi**

Sebelum menggunakan algoritma aturan (*rules)* mana pun, kita perlu mengubah data dari format *data frame*, menjadi data transaksi. Sehingga kita memiliki semua barang yang dibeli bersama dalam satu baris yang dipisahkan koma (entri tiap baris berupa himpunan berisi barang yang dibeli 1 pelanggan pada 1 kali transaksi).

A screenshot of a cell phone

Description automatically generated

A screenshot of a cell phone

Description automatically generated

* Terdapat 25.221 transaksi dengan 29.709 barang (*items*). *Items* adalah nilai unik dari variabel **Product** dalam dataset asli (**googlestore**). Transaksi di sini adalah koleksi atau himpunan bagian dari 29.709 item ini.
* Dari fungsi kepadatan probabilitas yang tertera di atas, kita dapat menghitung berapa barang yang terjual pada tahun 2017 di Google Store yaitu barang.

**Langkah 4. Association Rules.**

Digunakan Algoritma Apriori pada **arules** library untuk memperoleh himpunan yang sering muncul (*frequent itemsets*) dan aturan asosiasi jika-maka (*association rules*).

* *Output* dari algoritma apriori adalah *set of rules* (*left hand side (lhs) right hand side (rhs)*)
* *Support Minimum (support*: seberapa sering kumpulan item di *lhs* dan *rhs* muncul bersama dari keseluruhan dataset) : 0.0002 minimal muncul 0.2% dari data. Nilai ini telah disesuaikan dengan banyak items terjual, dan banyaknya barang yang dibeli pelanggan per transaksinya.
* *Minimum Confidence (confidence*: seberapa yakin kita yakin pelanggan akan membeli barang di *rhs*, jika sudah memiliki barang di *lhs*) : 0.6 minimal 60% yakin

A picture containing indoor, wall

Description automatically generatedA screenshot of text

Description automatically generated

Terdapat 135 aturan dengan barang terbeli maksimum yaitu 4, dan lebih banyak aturan yang g*enerate* 3 barang. Akan dilihat 30 aturan teratas saja.

A screenshot of a cell phone

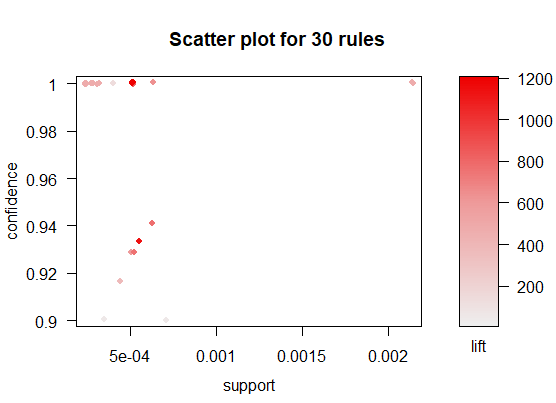
Description automatically generated

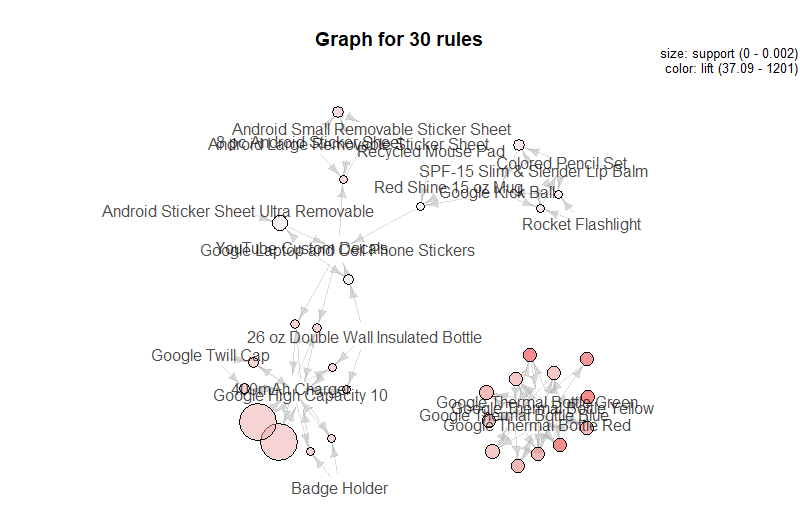
A screenshot of text

Description automatically generated

Beberapa interpretasi dari *output* di atas:

* 100% pelanggan yang membeli *“400mAh Charger”* juga membeli *“Google High Capacity 10”,* dan berlaku sebaliknya.
* 100% pelanggan yang membeli *“Google Thermal Bottle Red”* dan *“Google Thermal Bottle Yellow”* juga membeli *“Google Thermal Bottle Green”*





Dari grafik di atas, dapat dilihat bahwa:

* *“Google Thermal Bottle”* dalam berbagai warna mengelompok sendiri. Berarti, pelanggan yang membeli botol ini tidak berniat untuk membeli produk jenis lain.
* Pelanggan 100% membeli *“Google High Capacity 10”* berbarengan dengan *“400mAh Charge”* dapat dilihat dari lingkarannya yang besar dibandingkan dengan yang lain.
* Barang aksesori/hiasan seperti *stickers* mengelompok. Bisa diindikasikan bahwa pelanggan cenderung membeli *sticker* jenis yang lain lagi saat sudah membeli *sticker* tertentu.

1. **Kesimpulan**

Dengan data yang ada, bisa disimpulkan bahwa:

* *“Google High Capacity 10”* dan *“400mAh Charge”* paling sering dibeli bersamaan.
* Pelanggan cenderung membeli *Google Thermal Bottle* dalam berbagai warna (tidak hanya 1 saja), dan tidak membeli produk lainnya. Rekomendasi: dipasangkan iklan/rekomendasi botol dengan lebih banyak warna.
* Pelanggan yang membeli *stickers* cenderung membeli *stickers* jenis lain. Rekomendasi: dipasangkan rekomendasi produk dengan *bundle free sticker* sehingga untuk menambah penjualan produk lain yang masih berhubungan/dalam aturan asosiasi dengan produk *stickers*.

1. **Rekomendasi**

Rekomendasi penelitian selanjutnya:

* Untuk menggunakan data yang lebih up to date, sesuai dengan tahun terkini, agar dapat merekomendasi untuk strategi pemasaran yang lebih relevan.
* Untuk rekomendasi yang lebih akurat, lebih baik menambahkan variabel tambahan, seperti apakah website sedang ada diskon atau *seasonal sale*.
* Menggunakan sistem rekomendasi lain yang memperhitungkan kemiripan jenis produk dan harga produk.