

Use Case Data Science pada Industri Retail: Market Basket Analysis Menggunakan Association Rule

Oleh Nur Kholifah

Latar Belakang

Pada era perkembangan teknologi dan bertambahnya data yang ada, semua sektor dari kesehatan hingga bisnis mulai memperhatikan data yang mereka punya dan menggunakannya untuk mencapai tujuan yang spesifik. Data science menawarkan menganalisa data yang ada untuk membantu para pihak terkait dalam mengambil sebuah keputusan yang menguntungkan. Pada bidang bisnis, membuat keputusan yang menguntungkan sangat diperlukan untuk terus menjalankan bisnis yang ada dengan mendapatkan profit secara maksimal. Retail adalah bagian dari bisnis yang eksistensinya sering kita temui dan dapat kita lihat perkembangannya. Tentu kita sudah tidak asing dengan *online shop* atau penyedia jasa *entertainment* seperti Netflix. Contoh-contoh tersebut adalah bagian dari industri retail. Perkembangannya terhitung sangat pesat apalagi saat masa pandemi Covid-19. Semua kegiatan diharuskan dilakukan di rumah menjadikan retail yang dilakukan secara daring mengambil peran. Sampai saat ini, retail tersebut masih eksis di kalangan masyarakat. Tidak kalah dengan industri retail daring, retail layaknya minimarket di pinggir jalan masih menjadi primadona. Produk yang terkadang tidak sesuai dengan harapan konsumen pada *online shop*, menjadikan minimarket tempat yang masih sering dikunjungi masyarakat. Lebih lanjut, perkembangan usaha retail ini membuat para *retailers* menjadi lebih sadar akan pentingnya menggunakan data yang ada untuk retail yang mereka punya. Data science juga mengambil peran pada industri retail. Seperti contoh penggunaan data science untuk industri retail antara lain, *price optimization*, *cross selling and upselling*, *fraud detection*, dan *market basket analysis*.

Batasan Masalah

Melihat banyaknya *use case* dari data science yang dapat digunakan untuk sektor retail, maka dalam esai ini hanya akan dibahas mengenai *use case* data science pada sektor retail yakni Market Basket Analysis (MBA) menggunakan *association rule* dengan algoritma apriori untuk promosi produk di minimarket.

Dasar Teori

a. Retail

Retail adalah bagian dari bisnis yang penjualannya untuk konsumen perorangan dengan produknya akan digunakan untuk diri sendiri. Aspek yang dapat menjadikannya sebagai transaksi retail adalah kegiatan bisnis itu berhenti di pembeli (Badole, 2021). Contohnya adalah warung sembako, toko kelontong, minimarket, hingga *online shop* dan Netflix.

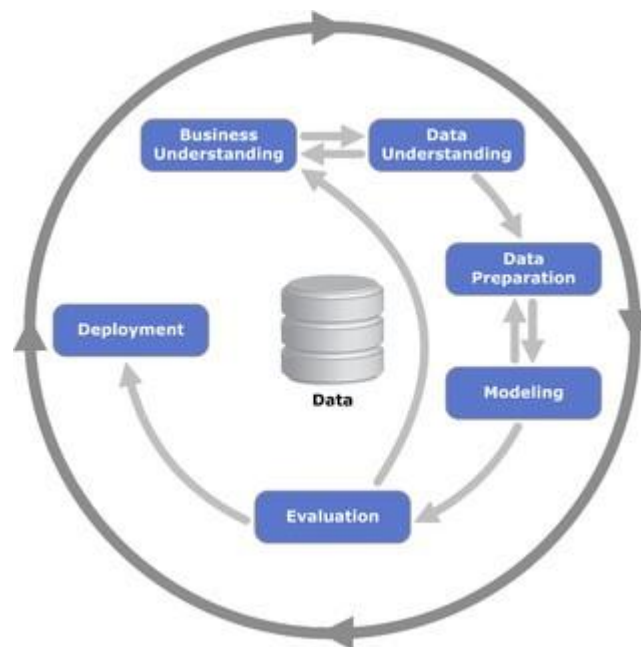
b. Market Basket Analysis

Market Basket Analysis atau MBA adalah teknik untuk mengidentifikasi pola konsumen dengan mengambil data dari *database* transaksi yang ada (Rai, 2022). MBA sering digunakan

pada industri retail. Untuk membuat pola prediksi konsumen, digunakan bermacam-macam *database* yang akan dilihat hubungannya satu sama lain. Hal yang dianggap tidak relevan akan dibuang dalam proses *rule mining*. Pola konsumen yang didapatkan bisa dimanfaatkan untuk mendapatkan banyak *insight*, antara lain cara pengaturan display produk, penentuan harga, dan promosi terbaik sesuai kebutuhan konsumen (Diana, 2020). Semua hal itu tentunya akan menguntungkan apabila digunakan oleh industri retail dengan analisis yang akurat.

c. **Cross industry Standard Proseses for Data Mining (CRISP-DM)**

CRISP DM adalah salah satu standar dalam data mining. Standar tersebut dikembangkan sejak tahun 1996 di Eropa. Proses data mining CRISP DM terdiri dari 6 tahap dengan 3 tahap paling awal dapat *Non Mutually Exclusive* atau dapat dilakukan secara bersama-sama. Berikut ini skema keenam tahapan CRISP DM.



Gambar 1. Skema CRISP DM (Sumber Gambarr: Wikipedia)

Penjabaran dari langkah-langkah tersebut antara lain (Darojatun, 2022).

1. *Business understanding*

Pada tahap ini, *business problem* didefinisikan dengan sederhana dan tepat. Meskipun terkesan sederhana, tahapan ini memerlukan penguasaan pada dua bagian yang berbeda yaitu pemahaman terhadap *business process* termasuk regulasi yang mengaturnya dan pemahaman terhadap cara pengolahannya. Tujuan yang ingin dicapai akan dirumuskan pada tahap ini.

2. *Data understanding*

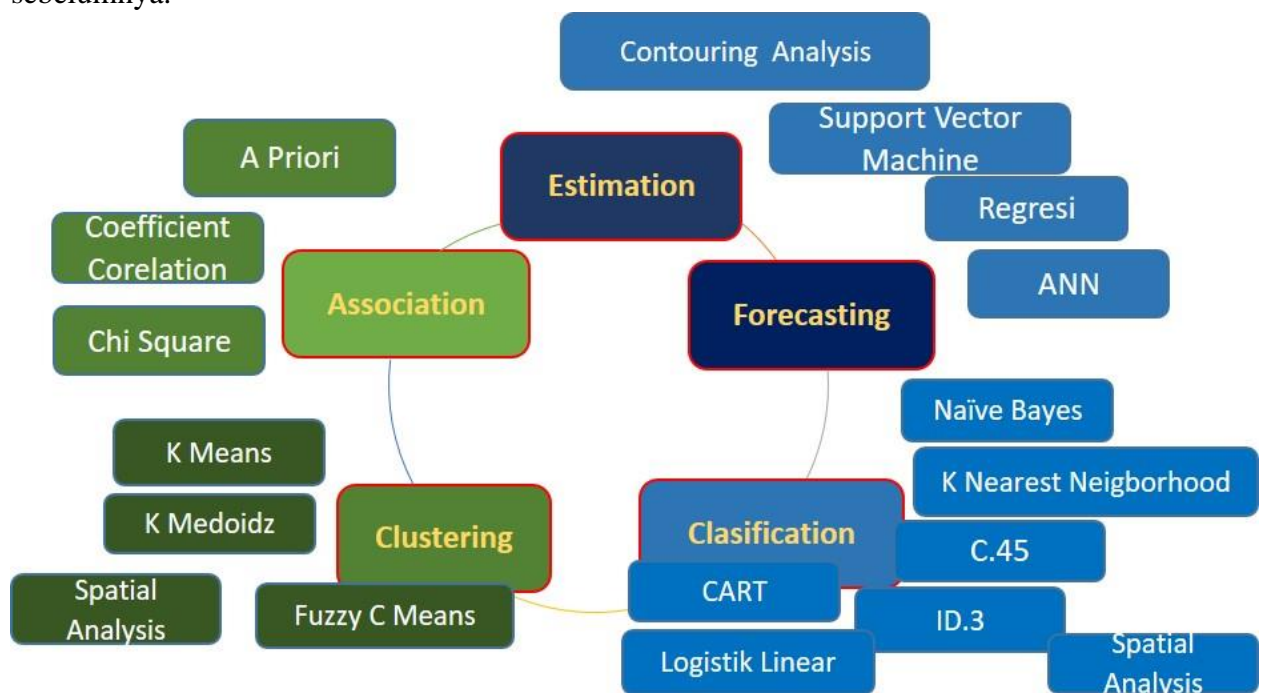
Proses ini akan mempertemukan data yang dimiliki dengan data yang diperlukan. Jenis data akan sangat menentukan jenis algoritma dan tujuan dari data mining yang ingin dicapai. Oleh karena itu, *analyst* perlu menguasai jenis data kuantitatif dan kualitatif.

3. *Data preparation*

Perlu diingat kembali bahwa hasil analisis data bergantung pada data yang digunakan. Semakin baik dan tepat data yang digunakan maka hasilnya akan akurat. Oleh sebab itu, tahap ini perlu perhatian yang lebih. Tahapan umum yang dilakukan pada *data preparation* adalah melakukan pengecekan kembali pada kebenaran data, mengelola data *outlier*, dan mengatasi data *missing* dan inkonsistensi.

4. Modeling

Model yaitu deskripsi yang dibangun oleh sistem yang dihasilkan dari proses kalkulasi. Mayoritas praktisi sepakat dengan La rose bahwa terdapat lima tujuan utama dalam data mining, yaitu estimasi, prediksi atau peramalan, klasifikasi, klastering, dan asosiasi. Hal itu dapat dikombinasikan untuk mencapai tujuan yang lebih detail. Berikut ini adalah algoritma yang dapat digunakan berdasarkan tujuan utama yang telah disebutkan sebelumnya.



Gambar 2. Algoritma dalam data mining

5. Evaluasi

Validasi model yang telah didapatkan dilakukan pada tahap ini. Validasi tersebut berdasarkan parameter-parameter yang relevan. Parameter dapat berupa parameter statistik, parameter ekonomi, dan lain sebagainya. Sebagai contoh pada parameter statistik digunakan parameter uji asumsi klasik atau nilai *R-square*.

6. Deployment

Tahap *deployment* adalah saat para *analyst* dan *engineer* menyusun dan menyampaikan proses dari analisis data yang telah dilakukan. Hal yang perlu dipertimbangkan adalah visualisasi, kemudahan dari penggunaan model, pemeliharaan model di masa yang akan datang, serta payung hukum yang menyertai penggunaan model tersebut. Tidak melanggar

suatu hukum yang ada harus diperhatikan karena modeling akan menuntun kearah pengambilan keputusan yang sarat akan rambu-rambu peraturan.

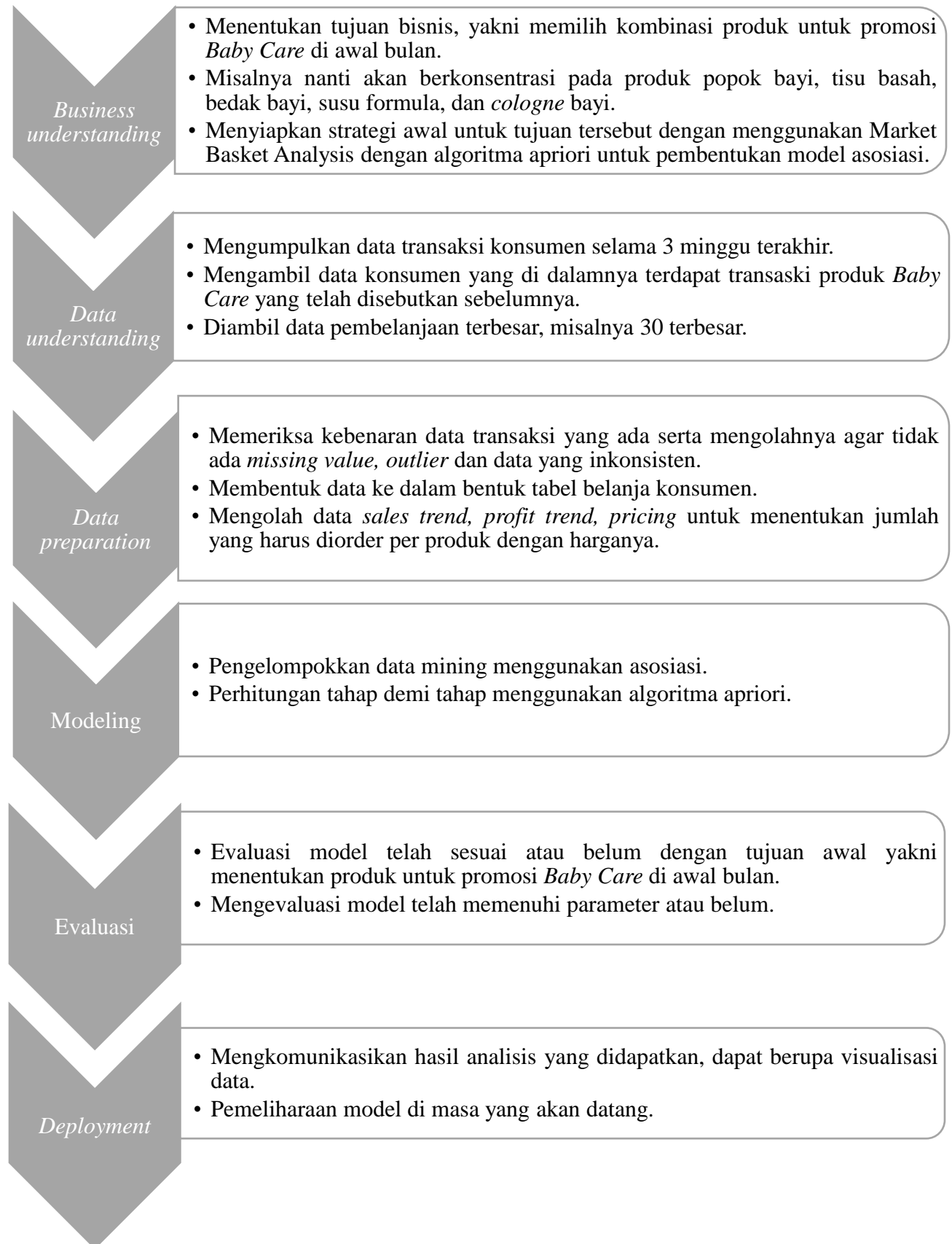
d. *Association Rule*

Salah satu tujuan utama data mining adalah asosiasi. Dengan menggunakan Cross industry Standard Proseses for Data Mining (CRISP-DM) akan dibentuk aturan asosiasi untuk suatu data yang ada. *Association rule* atau aturan asosiasi dapat disebut sebagai hubungan jika-maka yang akan membantu menemukan hubungan antara data *repositories*. Sebagai contoh, orang yang membeli popok bayi akan cenderung juga membeli tisu basah. Hal itu dapat diubah menjadi sebuah pernyataan, **jika orang membeli popok bayi maka mereka membeli tisu basah**. Kebanyakan algoritma *machine learning* untuk dataset numerik dan karena itu cenderung bersifat matematis. Akan tetapi, untuk *association rule* cocok untuk data non numerik, data kategori dan membutuhkan sedikit perhitungan sederhana. *Association rule* memiliki tujuan yakni menemukan pola, korelasi, atau asosiasi yang sering terjadi dari kumpulan data yang satu dengan data lainnya.

Untuk menemukan *association rule* diperlukan terlebih dahulu tahapan yakni mencari *frequent itemset* atau sekumpulan item yang sering muncul secara bersamaan. Setelah semua pola *frequent itemset* ditemukan, akan dicari *association rule* atau keterkaitan dengan syarat yang telah ditentukan. Dalam hal itu, digunakan algoritma Apriori. Algoritma tersebut bekerja dengan cara mengidentifikasi item dari seorang individu yang sering muncul dalam basis data yang diberikan kemudian memperluasnya ke kumpulan item yang lebih besar, dengan memeriksa bahwa kumpulan item tersebut cukup sering muncul di *database*. Hal itu akan membangun asosiasi dan korelasi antara itemset. Akan dihasilkan nilai *support* dan *confidence* yang mengukur kepercayaan dan kegunaan dari pola yang telah ditemukan. Misalnya didapatkan pola konsumen yang jika membeli popok bayi maka membeli tisu basah, nilai *support* 5% menunjukkan bahwa transaksi konsumen yang membeli popok bayi dan tisu basah secara bersamaan adalah sebesar 5%. Kemudian nilai *confidence* 70% menunjukkan nilai kepastian bahwa apabila konsumen membeli popok bayi pasti akan membeli tisu basah adalah sebesar 70% (Irlana & Vydia, 2013).

Pembahasan

Retail menghasilkan data dalam jumlah besar dari data customer. Data science hadir untuk membantu mendapatkan *insight* dari data tersebut mengenai pelanggan dan tren pasar. Salah satu dari *use case* data science pada industri retail adalah Market Basket Analysis (MBA). Apabila sedang membeli celana, maka terdapat kemungkinan kita akan membeli baju untuk melengkapinya. Hal yang sama juga terjadi saat membeli kopi, kita akan membeli gula sebagai pelengkap. *Outcome* dari salah satu *use case* data science yakni MBA dapat digunakan untuk menyarankan pasangan produk yang akan dibeli konsumen, seperti contoh yang disebutkan sebelumnya. Pada *use case* ini, MBA akan digunakan untuk **promosi produk perlengkapan bayi pada industri retail minimarket** di awal bulan. Hal tersebut akan dianalisis menggunakan *association rule* dengan algoritma apriori. Untuk mencapai hal itu akan dilakukan proses data mining yakni Cross industry Standard Proseses for Data Mining (CRISP-DM) dengan tahapan-tahapan sebagai berikut.

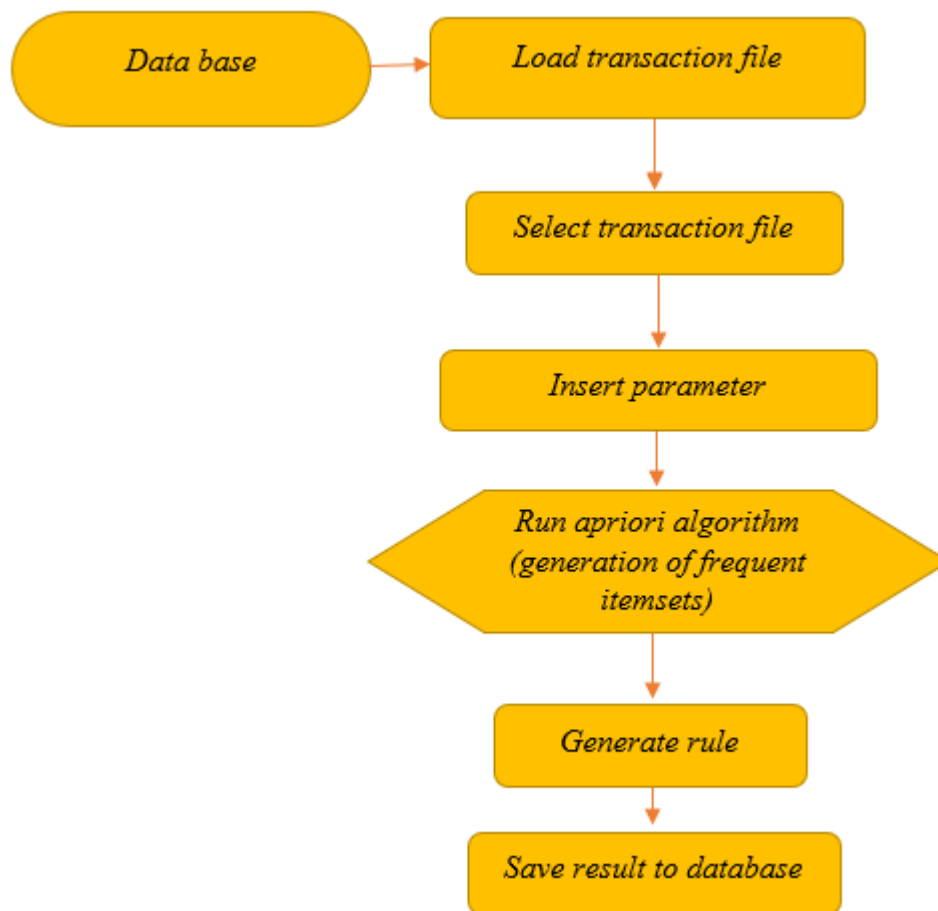


Gambar 3. Deskripsi tahapan CRISP DM

Untuk perincian pada proses modeling menggunakan *association rule* dengan algoritma apriori berikut ini adalah langkah-langkahnya.

1. Menghimpun data transaksi pelanggan.
2. *Upload* data transaksi.
3. Memilih data transaksi yang sesuai.
4. Memasukkan parameter yang digunakan. Parameter yang digunakan adalah *support* dan *confidence*, akan di-*setting* nilai minimal untuk kedua parameter tersebut.
5. Mencari *frequent itemset* menggunakan algoritma apriori.
6. *Generate association rule*.
7. Menyimpan hasil aturan asosiasi pada *database*.

Penggambarannya dalam bentuk diagram adalah sebagai berikut.



Gambar 4. Diagram modeling association rule dengan algoritma apriori

Kesimpulan

Data science telah banyak digunakan pada berbagai sektor, tak kecuali pada bidang retail. Data yang bisa didapatkan dari *retailers* dalam jumlah yang besar adalah data pelanggan. Akan tetapi permasalahannya adalah saat *retailers* belum menggunakan data tersebut dengan baik dan

benar akan kalah dengan kompetitor mereka. Menggunakan data adalah kunci utama untuk semua bisnis sekarang ini, para pebisnis akan dipaksa melihat cara baru untuk menganalisis datanya dan mendapatkan *insight* untuk bisnisnya. Market Basket Analysis adalah salah contoh *use case* data science pada industri retail yang membantu untuk menemukan pola konsumen yang ada menggunakan data transaksi pelanggan. Menggunakan *association rule* dengan algoritma apriori pola dari konsumen yang ada akan dibuat aturan asosiatifnya. Dari hal tersebut, industri retail dapat memanfaatkannya untuk pengaturan display produk, penentuan harga, dan promosi terbaik sesuai kebutuhan konsumen. Pada *use case* kali ini, digunakan MBA dengan algoritma apriori untuk menghasilkan *association rule* yang dimanfaatkan untuk promosi produk *baby care* di minimarket pada awal bulan. Promosi akan menghasilkan kombinasi produk *baby care* yang tepat. sehingga promosi tersebut akan dimaksimalkan untuk mendapatkan *profit* secara maksimal.

Daftar Pustaka

Badole, M. (2021, May 11). *Data Science Use Cases in Retail Industry*.

<https://www.analyticsvidhya.com/blog/2021/05/data-science-use-cases-in-retail-industry/>

Darojatun, H. (2022, June 22). *CRISP DM Sebagai Salah Satu Standard untuk Menghasilkan Data Driven Decision Making yang Berkualitas*.

<https://www.djkn.kemenkeu.go.id/artikel/baca/15134/CRISP-DM-Sebagai-Salah-Satu-Standard-untuk-Menghasilkan-Data-Driven-Decision-Making-yang-Berkualitas.html>

Diana, C. (2020, December 15). *Transformasi Retail dengan Menggunakan Data Science*.

<https://algorit.ma/blog/data-science/tranformasi-retail-data-science/>

Irliana, N., & Vydia, V. (2013). MARKET BASKET ANALYSIS PADA PERUSAHAAN RETAIL MENGGUNAKAN ALGORITMA APRIORI DAN SALES FORECASTING.

Jurnal Transformatika, 11(1), 13. <https://doi.org/10.26623/transformatika.v11i1.92>

Rai, A. (2022, September 29). *An Overview of Association Rule Mining & its Applications*.

<https://www.upgrad.com/blog/association-rule-mining-an-overview-and-its-applications/>

Link GitHub page dan LinkedIn

<https://khlfh.github.io/>

<https://www.linkedin.com/in/nur-kholifah-9581bb215/>