Nama : St Dhiah Raniah Napian

NIM : 09031181621011

**ANALISIS PENERAPAN *KNOWLEDGE DISCOVERY IN DATABASES* (KDD) PADA SUPERMARKET**

# PENDAHULUAN

Produksi data dewasa ini sangat banyak, namun masih sedikit informasi yang dapat diperoleh. Agar data yang kita peroleh dapat menjadi informasi yang bermanfaat, kita bisa melakukan *Knowledge Discovery in Databases* (KDD). KDD merupakan sekumpulan proses untuk menemukan pengetahuan yang bermanfaat dari data. Kumpulan proses tersebut berupa:

1. *Data cleaning*

*Data cleaning* proses pada KKD yang menghilangkan *noise*, membuang duplikasi data, memeriksa data yang inkonsisten dan memperbaiki kesalahan pada data seperti kesalahan cetak. Selain itu *data cleaning* juga melakukan proses penambahan data yang sudah ada dengan data yang relevan dan diperlukan.

1. *Data integration*

*Data integration* merupakan proses menggabungkan atau menyatukan dua atau lebih data dari berbagai sumber *database* yang berbeda kedalam sebuah penyimpanan.

1. *Data selection*

*Data selection* merupakan proses pengambilan data yang relevan dengan tugas analisis dari basis data. Proses pemilihan data dari sekumpulan data operasional yang perlu dilakukan sebelum tahap penggalian informasi dalam KDD.

1. *Data transformation*

*Data transformation* merupakan proses mentransformasi atau menggabungkan data ke dalam bentuk yang sesuai untuk penggalian lewat operasi *summary* atau *aggregation*. Perubahan bentuk atau transformasi dari data yang belum memiliki entitas yang jelas ke dalam bentuk data yang valid.

1. *Data mining*

*Data mining* merupakan proses dimana diterapkannya algoritma atau metode pencarian pengetahuan. *Data mining*  merupakan langkah penting karena pada fase inilah teknik kecerdasan diterapkan untuk mengekstrak pola informasi yang memiliki potensi berguna dari data yang dipilih.

1. *Pattern evaluation*

*Pattern evaluation* merupakan proses identifikasi pola yang menarik dan merepresentasikan pengetahuan berdasarkan *interestingness measures*.

1. *Knowledge presentation*

*Knowledge presentation* merupakan proses penyajian pengetahuan yang digali kepada pengguna dengan menggunakan visualisasi dan teknik representasi pengetahuan.

Berdasarkan penjelasan di atas, dapat kita ketahui bahwa *data mining* merupakan salah satu proses yang ada pada KDD. *Data mining* merupakan proses ekstraksi data menjadi informasi dan pengetahuan yang berguna secara otomatis dari beberapa data dalam jumlah besar. *Data mining* memiliki peran yang sangat penting dalam menemukan pola-pola yang berguna yang bersumber dari beberapa data yang besar dan data tersebut bisa disimpan di dalam basis data.

# DATASET PADA SITUS KDD

Dataset merupakan kumpulan data yang saling tekait dan memiliki kesatuan dan bersifat spesifik terhadap suatu kasus. Data set dapat direpresentasikan ke dalam berbagai bentuk tabel seperti basis data, matriks, teks, dan lain sebagainya. Dataset dapat dikumpulkan olah satu orang, sekelompok orang, atau bahkan suatu organisasi dan dipublikasikan secara *online*. Salah satu contoh situs yang berisi berbagai macam dataset adalah www.kaggle.com.

Dataset digunakan sebagai referensi data yang valid untuk suatu penelitian seperti untuk referensi data dalam sistem pembelajaran cerdas dan untuk data dalam pengujian sistem otomatis. *Dataset* yang baik memiliki ciri memiliki data yang lengkap, selalu *up to date*, bersifat konsisten dalam representasi datanya, jumlah variabelnya jelas, tidak mengandung *noise*, menarik, dan mudah dipahami.

Pada analisis penerapan KDD ini, dataset KDD penulis peroleh situs Kaggle (www.kaggle.com) tentang Penelitian terhadap informasi penjualan dan stok produk pada supermarket. Dataset tersebut terdiri dari 1568 baris, dan 63 kolom. Pada kolom dataset tersebut terdiri dari Share, Category, Sector, RM, Up, Last Traded Price, Percentage Change, High Price, Low Price, % High Movt, % Low Movt, Yearly Gainner, 27th Dec, 1st Feb, 1st March, 1st April, TB, RH, Corr, PeRatio, New Pe, W52\_High, Corre, Latest, Annual\_Pat, Cum PAT 3 Quarter, Pat Jump, Annual Growth, Year End, Mar\_17ReportedPAT, Dec\_16ReportedPAT, Sep\_16ReportedPAT, Jun\_16ReportedPAT , Mar\_16ReportedPAT, Dec\_15ReportedPAT, Y oY PAT, LQ>0.5(4q’s), PAT Growth, PeRatio, ResultDAte, Mar\_17\_Inst\_no\_of\_shares, Dec\_16\_Inst\_no\_of\_shares, Int/ Incr/ Decr, Pledge, DividendYield, DERatio, FB, Mar\_17\_Eps\_Before, Dec\_16\_Eps\_Before, Sep\_16\_Eps\_Before, Jun\_16\_Eps\_Before, Mar\_16\_Eps\_Before, Dec\_15\_Eps\_Before, Annual\_Eps, Volume, MarketCap, Annual\_Pat, EquityCapital, FaceValue, New Listing, PE High Price High, Future, Cap Category.

# JURNAL PENDUKUNG INFORMASI

1. **JURNAL 1**: PENERAPAN DATA MINING UNTUK MENGANALISIS PENJUALAN BARANG DENGAN MENGGUNAKAN METODE APRIORI PADA SUPERMARKET SEJAHTERA LHOKSEUMAWE

 Berdasarkan penelitian yang dilakukan oleh Dewi Astika Nurdin pada Supermarket Sejahtera Lhokseumawe, dapat diketahui bahwa sistem informasi penjualan yang diterapkan masih kurang efektif dan efisisen padahal perkembangan dunia bisnis sangat pesat, persaingan pun kian ketat. Oleh sebab itu dibutuhkannya suatu pola yang dapat meningkatkan penjualan dan pemasaran barang supermarket terkhusus Supermarket Sejahtera Lhokseumawe. Salah satu usahanya yaitu dengan memanfaatkan data transaksi. Pada sistem informasi penjualan yang sudah diterapkan pemanfaatan data transaksi terdahulu sangatlah kurang, atau bahkan bisa dikatakan tidak digunakan. Selama ini data transaksi yang terdahulu hanya disimpan namun tidak pernah diolah sama sekali.

 Dengan adanya kegiatan operasional sehari-hari data semakin lama akan semakin bertambah banyak. Jumlah data yang begitu besar justru bisa menjadi masalah bagi supermarket tersebut jika tidak bisa dimanfaatkan sebaik mungkin. Semakin banyak data, maka supermarket tersebut semakin memerlukan usaha untuk mengolah data-data agar dapat dijadikan informasi yang berguna di masa depan. Oleh karena itu diperlukan sebuah aplikasi yang bisa memanfaatkan kumpulan data yang besar, agar dapat diperoleh informasi yang berguna bagi pengguna di masa depan.

 Pemanfaatan data yang ada di dalam sistem informasi untuk menunjang kegiatan pengambilan keputusan, tidak cukup hanya mengandalkan data operasional saja, diperlukan suatu analisis data untuk menggali potensi-potensi informasi yang ada. Para pengambil keputusan berusaha untuk memanfaatkan gudang data yang sudah dimiliki untuk menggali informasi yang berguna membantu mengambil keputusan, hal ini mendorong munculnya cabang ilmu baru untuk mengatasi masalah penggalian informasi atau pola yang penting atau menarik dari data dalam jumlah besar yaitu dengan menggunakan *data mining*.

 Teknik analisa keranjang pasar merupakan teknik yang mengadaptasi ilmu data mining. Teknik ini digunakan untuk merancang suatu strategi penjualan dan pemasaran barang melalui proses pencarian asosiasi atau hubungan antar item data dari suatu basis data relasional. Pencarian asosiasi berawal dari pengolahan data transaksi pembelian barang dari setiap pembeli, kemudian dicari hubungan antar barang-barang yang dibeli. Pencarian informasi ini hampir sama dengan mencari peluang kemunculan barang yang dibeli sesuai dengan kebiasaan berbelanja masyarakat dan jumlah transaksi yang ada. Proses pencarian asosiasi ini menggunakan algoritma apriori, yang berfungsi untuk membentuk kandidat kombinasi item yang mungkin, lalu diuji apakah kombinasi tersebut memenuhi parameter *support* dan *confidence* minimum yang merupakan nilai ambang yang diberikan oleh user. (Nurdin, 2015)

1. **JURNAL 2**: A DATA MINING FRAMEWORK FOR OPTIMAL PRODUCT SELECTION IN RETAIL SUPERMARKET DATA: THE GENERALIZED PROFSET MODEL

 Pada penelitian yang dilakukan oleh Tom Brijs dan kawan-kawan, dapat kita ketahui bahwasanya penulis sebelumnya telah memperkenalkan sebuah model pemilihan produk yang disebut profset. Model profset ini memilih produk yang paling menarik dari berbagai produk berdasarkan potensi *cross-selling* meningat setiap produk dari pengecer pasti memiliki batasan.

 Pemanfaatan data transaksi pada penelitian ini dilihat untuk memperoleh peningkatan penjualan menggunakan strategi *cross-selling*. Dalam beberapa tahun terakhir, peneliti penambangan data telah mengembangkan algoritma aturan asosiasi yang efisien untuk analisis keranjang pasar ritel. Namun, masih banyak pengecer yang belum memahami dan belum bisa mengimplemenasikan algoritma aturan asosiasi tersebut ke dalam bisnis mereka dan masih banyak perbedaan data pada data supermarket.

 Oleh karena itu pada jurnal ini, penulis menyajikan generalisasi penting dari model yang ada untuk membuatnya cocok untuk data supermarket juga, dan untuk memungkinkan pengecer menambahkan pembatasan kategori ke model. Eksperimen pada data dunia nyata yang diperoleh dari jaringan supermarket Belgia menghasilkan hasil yang sangat menjanjikan dan menunjukkan efektivitas model profset umum.(Brijs & Goethals, 2001)

1. **JURNAL 3**: PERANCANGAN APLIKASI DATA MINING DENGAN ALGORITMA APRIORI UNTUK FREKUENSI ANALISIS KERANJANG BELANJA PADA DATA TRANSAKSI PENJUALAN

 Penelitian ini berupaya mengembangkan strategi bisnis peletakkan barang belanjaan yang disesuaikan dengan pola konsumsi konsumen di Swalayan KPRI Universitas Brawijaya Malang. Metode yang digunakan adalah Market Basket Analysis (MBA) dengan memanfaatkan data transaksi penjualan selama bulan Februari 2013. Hasil peneltian ini adalah sebuah prototipe aplikasi MBA. Pengujian prototipe dilakukan dengan batasan minimum transaksi (support) sebesar 7 transaksi dan minimum confidence sebesar 5%. Dengan batasan tersebut, aplikasi MBA membentuk 11 aturan asosiasi. Salah satu aturan asosiasi yang terbentuk adalah jika membeli gula pasir lokal 1kg, indofood bmb.racik sayur sop 20gr 9117, maka membeli indofood bmb.racik sy.asem 20gr rsah.463 dengan nilai support = 0,52% dan nilai confidence = 90,91% yang merupakan aturan dengan nilai confidence tertinggi.

 Proses selanjutnya adalah mengkategorikan item yang digunakan sebagai acuan perbaikan layout, sehingga mendapatkan rekomendasi perbaikan layout yang menyatakan bahwa gula didekatkan dengan telur, bumbu masak jadi, dan minyak goreng; minyak goreng didekatkan dengan bumbu masak jadi; telur didekatkan dengan beras dan mie instant serta minuman didekatkan dengan roti. Dengan demikian, penataan barang dagangan bisa disesuaikan dengan aturan asosiasi agar sesuai dengan pola konsumsi konsumen.

 Pada jurnal ini, kita ketahui bahwa, pemanfaatan data menggunakan proses *data mining* yaitu penggalian data transaksi sebagai penentu keputusan penempatan rak atau barang berdasarkan pola kebiasaan konsumen dilihat dari periode tertentu. (Dewantara, Santosa, & Setyanto, 2013)

1. **JURNAL 4**: ANALISIS CLUSTERING MENGGUNAKAN METODE K-MEANS DALAM PENGELOMPOKKAN PENJUALAN PRODUK PADA SWALAYAN FADHILA

 Pada jurnal ini, menjelaskan tentang aplikasi perangkat lunak *data mining* yang dapat digunakan untuk mengakses beberapa metode *data mining* yang ada. Dalam melaksanakan pengujian algoritma ini, data yang dipakai adalah data barang di swalayan Fadhila Bengkulu. Dalam segi penerapan, penerapan yang digunakan adalah penerapan *clustering* dengan menggunakan algoritma *K-means*.

 Dari data yang diolah dengan sampel data yang diambil di Swalayan Fadhila Bengkulu, maka menghasilkan dua jenis kelompok data, yaitu data penjualan rendah dan data penjualan tinggi. Sehingga dengan adanya pengelompokkan data ini, pihak Swalayan Fadhila dapat mengetahui jenis barang yang laris terjual dan tidak agar barang yang ada di gudang tidak menumpuk.

 Pemanfaatan penggalian data transaksi pada Swalayan Fadhila Bengkulu ini berupa pemanfaatan data transaksi sebagai alat pendukung keputusan dalam melakukan penambahan atau pengurangan stok barang. Agar tidak terjadi kekurangan dan penumpukan barang di gudang. Pemanfaatan tersebut dapat menghemat biaya distribusi barang dan efisiensi kuota barang.(Metisen & Sari, 2015)

1. **JURNAL 5**: ALGORITMA APRIORI PADA DATA PENJUALAN DI SUPERMARKET

 Proses kegiatan penjualan pada supermarket berjalan terus dan begitu juga data yang dihasilkan semakin lama maka akan semakin bertambah. Data-data penjualan yang semakin lama maka akan semakin besar tidak akan berguna dan bermanfaat jika dibiarkan begitu saja. Supaya data tersebut data berguna maka maka perlu di olah dengan suatu algoritma tertentu.

 Algoritma apriori merupakan bagian dari data mining yaitu kegiatan pengumpulan data dan pemakaian data yang lama untuk menemukan keteraturan, pola atau hubungan dalam suatu data. Keluaran dari algoritma ini adalah bisa membantu dalam memperbaikin pengambilan keputusan dimasa yang akan datang.Salah satu manfaat dari pengambilan keputusan ini adalah penyusunan katalog produk pada supermarket seperti produk yang paling banyak terjual diletakkan ditempat yang mudah dicari dan begitu juga dengan produk yang sering diterjual secara bersamaan maka produk tersebut perlu diletakkan pada tempat yang sama.

 Pemanfaatan penggalian data transaksi pada supermarket ini berupa pemanfaatan data transaksi sebagai alat pendukung keputusan dalam melakukan penyusunan katalog produk. Penyusunan tersebut seperti penempatan produk dengan jumlah penjualan tertinggi atau terlaris diletakkan di bagian paling depan supermarket (Buulolo, 2017)

# KESIMPULAN

Berdasarkan beberapa jurnal di atas, dapat disimpulkan bahwa data-data transaksi yang pernah dilakukan di supermarket dapat dimanfaatkan untuk melihat pola-pola pelanggan. Pola-pola tersebut dapat terbagi menjadi banyak klsaifikasi contohnya pemanfaatan data transaksi untuk mempelajari pola kebiasaan yang dilakukan oleh orang-orang yang berusia produktif. Seperti strategi yang dilakukan oleh Walmart dalam menentukan penempatan tata letak rak barang yakni bir, *cookies*, dan popok dilihat dari pola konsumen yang berusia produktif di mana mereka melihat data transaksi yang menunjukkan bahwa orang-orang yang berusia produktif biasa menghilangkan stress dengan minum bir dan *cookies* sebagai cemilan serta popok yang selalu digunakan anak-anak mereka.

Contoh lainnya yaitu pemanfaatan data transaksi untuk mempelajari pola-pola kebiasaan konsumen untuk mendukung pengambilan keputusan penempatan produk berdasarkan jumlah penjualan tertinggi, pemanfaatan data transaksi untuk mempelajari pola-pola produksi agar tIdak terjadi kekurangan dan penumpukkan di gudang, pemanfaatan data transaksi untuk mempelajari pola-pola *supplier*, pemanfaatan data transaksi dengan mempelajari pola-pola kebiasaan konsumen sebagai alat pendukung keputusan penentuan strategi seperti strategi *cross-*selling contohnya membaca pola-pola kebiasaan konsumen seperti ibu rumah tangga untuk menhemat pengeluaran keluarga ibu memilih membeli sabun cuci baju yang mendapat bonus piring ataupun sabun pembersih lantai dan lain sebagainya.

# DAFTAR PUSAKA

Brijs, T., & Goethals, B. (2001). A Data Mining Framework for Optimal Product Selection in Retail Supermarket Data : The Generalized PROFSET Model, 1–13.

Buulolo, E. (2017). Algoritma apriori pada data penjualan di supermarket, (September 2015), 4–7.

Dewantara, H., Santosa, P. B., & Setyanto, N. W. (2013). PERANCANGAN APLIKASI DATA MINING DENGAN ALGORITMA APRIORI UNTUK FREKUENSI ANALISIS KERANJANG BELANJA PADA DATA TRANSAKSI PENJUALAN ( Studi Kasus di Swalayan KPRI Universitas Brawijaya ) DESIGNING APPLICATIONS DATA MINING WITH THE APRIORI ALGORITHM TO FREQ, 415–426.

Metisen, B. M., & Sari, H. L. (2015). ANALISIS CLUSTERING MENGGUNAKAN METODE K-MEANS DALAM PENGELOMPOKKAN PENJUALAN PRODUK PADA SWALAYAN FADHILA, *11*(2), 110–118.

Nurdin, D. A. (2015). PENERAPAN DATA MINING UNTUK MENGANALISIS PENJUALAN BARANG DENGAN PADA SUPERMARKET SEJAHTERA.