

Pengembangan Gudang Data Pendukung Analisis Tren Penyewaan Peralatan Katering dengan Algoritma Apriori

Ardana Putra, Hapnes Toba

Program Studi Teknik Informatika, Fakultas Teknologi Informasi, Universitas Kristen Maranatha,
Jl.Surya Sumantri No.65, Bandung, 40164, Indonesia

E-mail: 1572041@student.it.maranatha.edu, hapnestoba@it.maranatha.edu

Abstract

PT. XYZ is a company engaged in rental business of catering equipments located in South Jakarta. PT. XYZ rents out various types of catering equipment such as plates, spoons, cutlery, glasses and others. At this time, PT. XYZ needs to analyze which products are the most popular or most frequently borrowed by customers at a specific timeline. This would be beneficial for PT. XYZ to decide the availability of new supplies in the future. PT. XYZ could also adjusted the items owned to the interests or taste of their customers. The data analytics process is carried out in a data warehouse by utilizing apriori algorithm with Pandas software. Pandas is an easy to use package and has a high-performance in Python programming language. By doing the analytics process, we find out which itemsets are often borrowed and also the rental activities trending patterns' from each customer. Based on the analytics results, PT. XYZ prepares some strategic plans to ensure a successful business process re-engineering.

Keywords: apriori algorithm, data analytics, data warehouse, rental management, visualization

Abstrak

PT. XYZ merupakan sebuah perusahaan yang bergerak di bidang penyewaan peralatan katering di Jakarta Selatan. PT. XYZ menyewakan berbagai macam kebutuhan peralatan katering seperti piring, sendok, alat makan, gelas dan lain-lainnya. Saat ini, PT. XYZ ingin menganalisis produk apa saja yang paling diminati atau paling sering dipinjam oleh pelanggan dalam berbagai periode waktu. Hal ini berpengaruh bagi pihak PT. XYZ untuk merencanakan pengadaan barang baru di masa depan agar dapat menyesuaikan dengan minat ataupun selera dari pelanggan mereka. Untuk memenuhi kebutuhan tersebut, dilakukanlah pengembangan gudang data dilanjutkan proses analisis memanfaatkan algoritma apriori dalam bahasa pemrograman Python dengan bantuan pustaka Pandas. Proses pengembangan gudang data dimulai dengan ekstraksi data *invoice* penyewaan dari sistem akunting, dilanjutkan dengan transformasi data, dan diakhiri dengan penyimpanan data dalam tabel dimensi dan fakta. Proses analisis lebih dalam melalui OLAP dilengkapi pula dengan visualisasi grafik penyewaan barang dalam periode-periode waktu yang telah disepakati. Dengan dilakukannya proses analisis dan visualisasi tersebut, dapat diketahui pola minat dari setiap pelanggan dan kombinasi kategori penyewaan dalam periode waktu tertentu. Berpijak pada hasil riset, pihak PT. XYZ telah menyiapkan beberapa langkah strategis untuk menjamin ketersediaan barang sewaan sebagai bagian dari *business process re-engineering* (BPR). Salah satu hal yang direncanakan oleh PT. XYZ sebagai kelanjutan dari hasil riset ini adalah dengan memberikan *teaser* (uji coba) pemakaian peralatan-peralatan baru, terutama yang terkesan mewah dengan berbagai asesorisnya. Melalui proses BPR ini diupayakan terjadinya peningkatan jumlah barang yang disewa serta semakin memperkenalkan PT. XYZ secara lebih luas dalam dunia bisnis katering.

Katakunci: algoritma apriori, analisis data, gudang data, pengelolaan sewa menyewa, visualisasi

Copyright © Journal of Information System and Technology. All rights reserved

I. PENDAHULUAN

Dunia teknologi informasi saat ini berkembang dengan sangat cepat. Banyak perusahaan yang memanfaatkan teknologi informasi untuk meningkatkan kinerja bisnisnya sehingga dapat meraih pangsa pasar yang lebih luas. *Data warehouse* (gudang data) merupakan salah satu pendekatan dalam bidang teknologi informasi yang dapat mendukung kegiatan serta proses pengambilan keputusan di sebuah perusahaan. Keberadaan gudang data dapat membantu perusahaan dalam penentuan kebijakan agar dapat menghasilkan keputusan yang cepat dan tepat berdasarkan hasil analisis data dari fakta yang tersedia [1]. Gudang data juga dapat membantu pihak perusahaan dalam membuat sistem informasi eksekutif sebagai alat analisis bagi para petinggi perusahaan dalam menentukan suatu keputusan. Dengan demikian, salah satu keuntungan utama tersedianya sebuah gudang data adalah agar pihak perusahaan dapat menghimpun data historis dalam volume yang besar serta konsisten untuk mendukung analisis bisnis secara cepat [2].

PT.XYZ merupakan sebuah perusahaan di Jakarta Selatan yang bergerak dalam bisnis penyewaan peralatan catering. Jenis peralatan catering yang disewa oleh para pelanggan, dalam hal ini adalah perusahaan penyedia jasa catering, sangat mungkin untuk berubah-ubah seiring perubahan waktu. Pengembangan gudang data dalam riset ini bertujuan untuk mengolah dan menganalisis data dari setiap transaksi penyewaan yang ada sehingga menghasilkan informasi yang akurat.

Dengan adanya gudang data, pihak manajemen dapat lebih mudah mengetahui jumlah penggunaan masing-masing barang yang disewa oleh pelanggan. Di samping itu, pihak perusahaan juga dapat melihat pertumbuhan penyewaan masing-masing barang dalam rentang periode tertentu yang diinginkan, misalnya: per bulan, kuartal atau tahunan, sehingga dapat merencanakan ketersediaan barang sesuai animo dari pelanggan. Pada bagian-bagian selanjutnya dalam makalah ini akan disampaikan pembahasan mengenai literatur pendukung riset, kemudian dilanjutkan dengan perancangan metode kerja. Setelah itu disampaikan hasil implementasi yang dilakukan, diikuti dengan analisis hasil tercapai. Pada akhir makalah akan dituliskan pula beberapa strategi yang direncanakan PT. XYZ sebagai tindak lanjut dari riset ini. Makalah akan ditutup dengan

kesimpulan berdasarkan pada hasil-hasil tercapai dan analisisnya.

II. KAJIAN LITERATUR

Pada bagian ini dituliskan kajian literatur terkait dengan metode kerja dan teknologi yang digunakan selama riset dilaksanakan. Penulisan dibagi ke dalam tiga topik utama, yaitu: karakteristik gudang data dan proses-proses pendukung di dalamnya, perangkat teknologi yang digunakan, serta algoritma yang digunakan untuk analisis data.

A. Karakteristik Gudang Data

Secara umum, sebuah gudang data dapat didefinisikan sebagai sekumpulan data yang memiliki enam buah karakteristik utama, yaitu: berorientasi subyek, terintegrasi, berorientasi kepada proses, terikat dengan waktu, tidak lekang, dan cepat diakses [1, 2]. Data ditampilkan berdasarkan kebutuhan subjek. Subjek yang dimaksudkan dalam hal ini adalah topik, area, kebutuhan, atau bidang peminatan tertentu yang menjadikan gudang data menjadi lebih spesifik. Data yang dikelola hanyalah data menurut subjek, yaitu yang diperlukan untuk proses pengambilan keputusan. Gudang data dibangun dari proses integrasi berbagai sumber data yang berasal dari berbagai macam aplikasi, menjadi satu kesatuan yang utuh. Contohnya seperti sumber-sumber data yang beragam dan berasal dari berbagai macam basis data yang merupakan aset penting sebuah organisasi. Oleh karena itu gudang data memerlukan adanya kesinambungan proses, dan akan sangat memerlukan dukungan perangkat teknologi informasi [3].

Perlu dipastikan pula bahwa sebuah gudang data dapat mengidentifikasi data dengan mengacu pada periode waktu penyimpanannya. Informasi yang disajikan di dalam gudang data pada umumnya dilihat dari sudut pandang riwayat penyimpanan [2]. Perubahan sekecil apapun tidak boleh dilakukan terhadap data operasional yang telah masuk ke dalam gudang data dan harus diolah secara otentik [4]. Karakteristik berikutnya yang sangat perlu diperhatikan terkait dengan kemudahan dan kecepatan akses oleh pengguna [4]. Pengguna perlu mendapatkan jaminan terkait dengan kecepatan dan ketepatan perolehan data, sesuai dengan kebutuhan dan hak akses yang diberikan oleh sistem. Jaminan ini sangat penting terutama untuk mendukung proses pengambilan

keputusan, misalnya untuk perencanaan *Business Process Re-engineering* [2, 3].

B. *Extraction, Transformation, Loading*

Proses *Extraction, Transformation* dan *Loading* (ETL) terdiri atas suatu urutan langkah saat pemrosesan data dengan melibatkan berbagai tahap [6]. Proses dimulai dengan pengekstrasian data dari sumbernya. Data-data tersebut bisa berasal dari beberapa basis data yang terpisah. Perlu diupayakan agar kualitas data yang disajikan di dalam gudang data tetap konsisten. Hal ini dapat dilakukan dengan menerapkan standarisasi penyajian melalui transformasi, dan pada akhirnya dialirkan atau diteruskan ke dalam gudang data. Hasil dari proses akhir inilah yang digunakan sebagai kebutuhan untuk melakukan analisis data ataupun informasi.

C. *On Line Analytical Processing*

On Line Analytical Processing (OLAP) merupakan sebuah *tools* dalam bentuk perangkat lunak dan sekaligus sebagai sebuah pendekatan untuk pemrosesan pada basis data. Melalui OLAP dilakukan proses analisis data dari dalam gudang data, sehingga mendapatkan informasi serta menyediakan navigasinya kepada pengguna awam. OLAP memiliki tiga operasi analisis dasar yang dapat dimanfaatkan, yaitu [2, 4]: konsolidasi, *drill down*, serta *slice and dice*. Konsolidasi melibatkan agregasi data yang dapat diakumulasikan dan dihitung dalam satu atau lebih dimensi. Misalnya, menghitung total penjualan dari semua kantor cabang untuk mengantisipasi tren kebutuhan pelanggan. Operasi *drill down* merupakan sebuah teknik yang memungkinkan pengguna untuk menavigasi informasi secara lebih terperinci. Misalnya, sebuah organisasi dipisahkan ke dalam departemen atau bagian sehingga lebih mudah untuk menganalisis penjualan produk tertentu dalam suatu area pemasaran.

Slice and dice merupakan sebuah fitur yang memungkinkan pengguna dapat mengambil satu himpunan data tertentu dari kubus OLAP dan melihat irisan dalam sudut pandang yang berbeda. Sudut pandang ini disebut juga sebagai dimensi, misalkan untuk melihat jumlah penjualan sebuah produk berdasarkan periode-periode waktu ataupun berdasarkan pelanggan [4].

D. *Data Multidimensi*

Pemodelan data multidimensi merupakan upaya untuk mendesain sebuah skema, yang di dalamnya melibatkan atribut pengukuran dan atribut dimensi [2, 4]. Atribut-atribut tersebut akan memberikan hubungan antara fakta dan dimensi-dimensi terkait

sehingga dapat disesuaikan dengan proses bisnis dalam sebuah perusahaan [7]. Teknik umum yang digunakan dalam membentuk data multidimensi ini adalah dengan mentransformasikan data ke dalam sebuah tabel relasional sehingga terfokus pada luaran (hasil) proses bisnis pada sebuah organisasi.

Pemodelan data multidimensi dalam gudang data dapat mengambil rupa dalam beberapa bentuk, yaitu: *cube*, *facts*, dan *dimension*. *Cube* merupakan basis data multidimensi yang dioptimalkan untuk aplikasi gudang data dan pemrosesan OLAP. *Facts* merupakan tabel dari suatu kejadian yang berisi luaran (hasil) suatu proses bisnis. Tabel fakta menggabungkan berbagai entitas yang diidentifikasi dalam *logical data* model pada basis data operasional. *Dimension* merupakan tabel yang lebih sederhana yang di dalamnya terdapat *primary key* yang berhubungan dengan salah satu *composite key* yang ada pada tabel fakta [4, 5].

E. *Pustaka Pandas*

Pandas adalah sebuah pustaka berlisensi BSD dan bersumber terbuka yang menyediakan struktur khusus untuk melakukan analisis data dalam bahasa pemrograman Python [8]. Pustaka Pandas dapat digunakan untuk membersihkan data mentah ke dalam bentuk yang cocok untuk dianalisis.

Pustaka Pandas dapat pula melakukan tugas fundamental, misalnya melakukan penyesuaian data untuk membandingkan atau pun menggabungkan berbagai basis data. Proses untuk menangani data yang hilang (*corrupt*) telah menjadi salah satu fitur utama dalam pustaka tersebut [9]. Struktur dasar untuk pengolahan data pada Pandas disebut sebagai *dataframe*, yaitu sebuah kolom yang berurutan dengan nama dan tipe data tertentu. Hal ini dapat dianalogikan seperti halnya sebuah tabel dalam basis data, yaitu sebuah baris tunggal mewakili sebuah contoh tunggal (*record*) dan kolom mewakili atribut tertentu [10].

F. *Algoritma Apriori*

Algoritma apriori, merupakan sebuah algoritma dalam *data mining* yang banyak dimanfaatkan untuk membentuk sejumlah aturan asosiasi (*association rules*). Algoritma ini menggunakan teknik perhitungan berbasis kemunculan (frekuensi) untuk menemukan keterkaitan dalam suatu kombinasi yang terdiri atas beberapa *items* [11]. Contoh sederhana untuk pemanfaatan algoritma apriori adalah untuk menganalisis pembelian di suatu *supermarket*. Misalnya, pihak *supermarket* dapat mengetahui berapa besar kemungkinan seseorang membeli sebuah roti bersamaan dengan

susu. Dengan pengetahuan tersebut pihak *supermarket* dapat mengatur penempatan barang yang akan dijual atau menentukan diskon untuk setiap barang. Oleh karena itu algoritma apriori dan penerapannya dalam analisis asosiasi juga sering disebut sebagai *market basket analysis* [12].

III. METODOLOGI PENELITIAN

Pelaksanaan penelitian ini dilakukan dalam beberapa tahap, yaitu: (1) Identifikasi permasalahan; (2) Pengumpulan data penelitian; (3) *Data Pre-processing* (ETL); (4) Perancangan dan pengisian skema gudang data; (5) Proses penentuan pola data dengan algoritma apriori; dan (6) Proses OLAP dilengkapi dengan visualisasi dalam periode waktu yang sudah ditentukan.

A. Pengumpulan Data Penelitian

Pengumpulan data dilakukan dengan dua macam pendekatan, melalui observasi dan wawancara. Observasi digunakan untuk menghimpun berbagai sumber data dalam pembuatan struktur gudang data. Proses wawancara dilakukan dengan *supervisor* di departemen penjualan (*sales*) PT. XYZ melalui pertanyaan-pertanyaan yang terkait dengan perancangan skema gudang data.

B. Data Pre-Processing melalui ETL

Proses ETL dilakukan sebagai persiapan awal yang dilakukan untuk memastikan data dapat dimanfaatkan dalam gudang data. Pada saat proses ekstraksi dilakukan, terdapat beberapa tahapan, yaitu: mengakses sumber data, pemilihan *grain*, pemilihan dimensi, serta pemilihan fakta.

Tabel 1. Atribut dalam Sumber Data

No.	Nama Kolom	Keterangan
1	<i>Invoice Number</i>	Informasi yang menunjukkan nomor <i>invoice</i> atau faktur penjualan.
2	<i>Bill To</i>	Informasi yang menunjukkan nama pelanggan
3	<i>Ship To</i>	Informasi yang menunjukkan alamat tujuan
4	<i>Terms</i>	Informasi yang menunjukkan status pembayaran.
5	<i>Date</i>	Informasi yang menunjukkan waktu pemesanan
6	<i>Qty</i>	Informasi yang menunjukkan jumlah barang yang dipesan
7	<i>Item Number</i>	Informasi yang menunjukkan kode barang

No.	Nama Kolom	Keterangan
8	<i>Description</i>	Informasi yang menunjukkan nama dari masing masing jenis barang
9	<i>Price</i>	Informasi yang menunjukkan nama dari masing masing jenis barang
10	<i>Unit</i>	Informasi yang menunjukkan satuan barang
11	<i>Disc</i>	Informasi yang menunjukkan nilai diskon dari sebuah barang
12	<i>Sale Amount</i>	Total harga barang

Sumber data yang diolah untuk kebutuhan perancangan gudang data pada PT. XYZ ini adalah data penjualan berbentuk berkas *excel*. Data yang digunakan merupakan data historis penjualan dalam 3 tahun, yaitu: dari bulan Januari tahun 2016 sampai dengan bulan September tahun 2018. Tabel 1 memberikan deskripsi untuk atribut-atribut utama terkait proses penyewaan barang yang berhasil dihimpun dari PT. XYZ.

Pemilihan *grain* atau granularitas dilakukan untuk menentukan tingkat kedetilan informasi dalam sebuah gudang data. Contoh tingkat granularitas, misalnya terkait pengelompokan waktu, yaitu: dalam satuan hari, minggu, bulan, dan tahun. Proses pemilihan fakta dapat dilakukan dengan memilih fakta berupa data yang dapat dihitung (bertipe numerik) sehingga dapat ditampilkan dalam bentuk laporan atau divisualisasikan dengan grafik [2, 4].

Proses transformasi dilakukan pertama-tama dengan mengubah beberapa atribut, yaitu: harga, jumlah pesanan, dan kategori barang ke dalam beberapa kategori (diskretisasi). Sesuai kesepakatan dengan PT. XYZ, terdapat tiga kategori harga yang ditentukan, yaitu: "Murah", "Sedang", dan "Mahal". Untuk kategori "Murah" rentang harga yang ditentukan adalah dari harga Rp 200,00 sampai dengan Rp 1.100,00. Untuk kategori "Sedang" rentang harga yang ditentukan adalah dari harga Rp 1.200,00 sampai dengan Rp 5.000,00. Untuk kategori "Mahal" rentang harga Rp 7.000,00 sampai dengan Rp 50.000,00.

Atribut jumlah pesanan berisi tiga kategori, yaitu: "sedikit", "cukup", dan "banyak". Untuk Kategori "sedikit" rentang jumlah pemesanan berjumlah 1 sampai dengan 350 buah. Untuk kategori "cukup" rentang jumlah pemesanan berjumlah 351 sampai dengan 700 buah. Untuk jumlah 700 ke atas masuk dalam kategori "banyak". Pada atribut "Kategori Barang", ditentukan kategori dari setiap barang yang disewakan, yaitu: "Piring", "Gelas", "Alat

Makan”, “*Serving Spoon*”, “Wadah *Serving*”, “Mangkok”, dan “Peralatan”. Penentuan kategori barang ini disesuaikan dengan informasi keseluruhan barang yang dimiliki dan disewakan oleh PT. XYZ.

Pemilihan dimensi dilakukan dengan analisis deskriptif melalui pemberian informasi yang berhubungan dengan setiap kolom pada Tabel 1. Dimensi barang dipilih untuk menganalisis kategori dan nama barang yang telah disewakan. Dimensi waktu untuk memastikan jangka waktu penyewaan yang telah dilakukan. Dimensi pelanggan untuk memastikan nama pelanggan dan status pelanggan yang telah melakukan penyewaan barang.

Setelah dimensi terbentuk, dilakukanlah proses identifikasi dan penyesuaian. Hal ini dilakukan dengan menentukan granularitas data yang dipresentasikan dalam tabel fakta. Granularitas yang digunakan untuk membangun gudang data didasarkan pada fakta barang yang disewakan.

Tabel 2. Dimensi Terpilih

Dimensi	Atribut	Keterangan
Waktu	<i>Date</i>	Menyimpan informasi
	<i>Month</i>	tentang keterangan waktu
	<i>Year</i>	Menyimpan informasi tentang data pelanggan
Pelanggan	Nama Pelanggan	Menyimpan informasi tentang data pelanggan
	Status Pelanggan	Menyimpan informasi tentang jumlah pesanan barang
Pemesanan	Jumlah pesanan	Menyimpan informasi tentang jumlah pesanan barang
	<i>Qty</i> Barang	Menyimpan informasi data barang
Barang	Nama Barang	Menyimpan informasi data barang
Harga	<i>Id</i> Barang	Menyimpan informasi dan kategori harga
	Harga Barang	Menyimpan informasi dan kategori harga
	Kategori Harga	

Kebutuhan informasi yang dianalisis adalah jumlah masing-masing barang yang disewakan pada tahun 2016 sampai tahun 2018, waktu dari setiap pemesanan penyewaan (berdasarkan hari, bulan, dan tahun), nama pelanggan, serta status pelanggan yang melakukan penyewaan barang. Dalam Tabel 3 diberikan identifikasi atribut yang menjadi kandidat untuk pengisian tabel fakta.

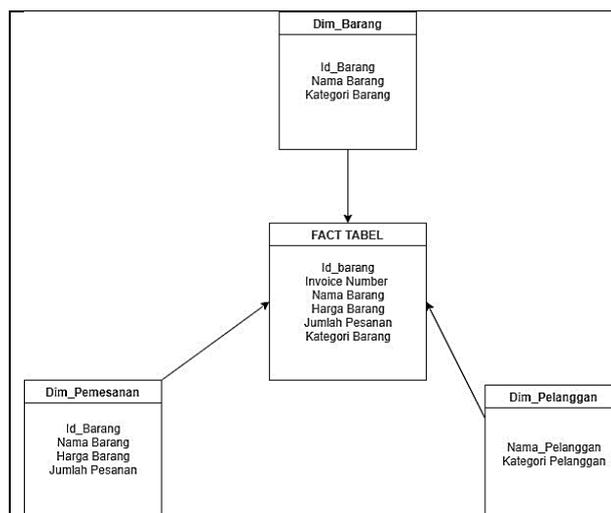
Tabel 3. Kandidat Atribut dalam Tabel Fakta

Atribut	Keterangan
<i>Id</i> Barang	Menyimpan informasi tentang <i>Id</i> dari setiap barang
<i>Invoice Number</i>	Menyimpan informasi tentang nomor pemesanan.
Pelanggan	Menyimpan informasi mengenai nama pelanggan dari PT. XYZ
Jumlah Pesanan	Menyimpan informasi tentang jumlah barang yang dipesan oleh pelanggan
Harga Barang	Menyimpan informasi tentang kategori harga dari setiap barang
Barang	Menyimpan informasi nama produk-produk yang dimiliki PT. XYZ
Waktu Pemesanan	Menyimpan informasi waktu pemesanan dari setiap barang

Saat proses *loading* dilakukan, ada beberapa tahapan dengan tujuan utama untuk memastikan data dialirkan dari sumber ke dalam gudang data secara konsisten. Data dalam suatu tabel tidak selamanya memiliki nilai yang tetap, maka dari itu setiap perubahan nilai perlu untuk disesuaikan terhadap data yang telah tersimpan di dalam gudang data. Misalnya ada sebuah data yang memiliki *id* yang sama tetapi *id* tersebut digunakan pada lebih dari satu nama yang sejenis, maka perubahan data tersebut harus ditelusuri untuk memastikan makna data yang sesungguhnya.

C. Perancangan Skema Gudang Data

Dalam Gambar 1, disampaikan rancangan *star schema* pada PT. XYZ. Dalam perancangan *star schema* ini ada beberapa informasi yang dipilih untuk menjadi suatu dimensi sehingga dapat mendukung proses analisis yang akan dilakukan. Tabel fakta digunakan sebagai tabel utama, dan direlasikan dengan tabel dimensi lainnya.



Gambar 1. Rancangan Star Schema

Tabel fakta yang digunakan dalam perancangan *star schema* ini berisikan: *Id_Barang*, *Invoice Number*, Pelanggan, Jumlah Pesanan, Waktu, dan Harga. Tabel fakta dikelilingi oleh beberapa dimensi yaitu: dimensi pelanggan, dimensi jumlah pesanan, dimensi barang, dimensi harga, dan dimensi waktu (lihat juga dalam Tabel 2). Masing-masing dimensi ini direlasikan dengan tabel fakta, hal ini dimaksud agar tabel fakta dapat diisi dengan informasi penting yang terdapat di dalam masing-masing dimensi.

IV. IMPLEMENTASI

Pada bagian ini dibahas mengenai implementasi setelah dibuatnya rancangan gudang data yang disesuaikan dengan kebutuhan. Pihak PT. XYZ membutuhkan analisis data historis penjualan untuk menentukan pengambilan keputusan di masa depan, baik itu terkait pembelian alat ataupun pemasaran barang. Proses implementasi diawali dengan pengumpulan data dari PT. XYZ kemudian data tersebut dianalisis untuk dilakukan pemilihan data apa saja yang dibutuhkan melalui proses ETL. Pada akhirnya, dilakukanlah penentuan pola, berupa kumpulan *itemsets* dan tren dari masing-masing alat catering yang telah disewakan.

A. Identifikasi Kebutuhan Analisis

Setelah melakukan analisis terhadap sumber data sebagaimana disampaikan dalam Tabel 1, tahap selanjutnya adalah melakukan identifikasi pemilihan *record* yang akan digunakan dalam pembuatan gudang data. Pada Tabel 4 disampaikan informasi statistik terhadap data-data yang digunakan dalam proses analisis hingga dihasilkannya pola.

Tabel 4. Ketersediaan Data untuk Analisis

No.	Atribut	Keterangan	Jumlah
1	<i>Invoice Number</i>	Informasi nomor <i>invoice</i> atau nomor faktur penjualan	1.309
2	<i>Date</i>	Informasi Tentang waktu pemesanan barang	1.309
3	<i>Qty</i>	Informasi tentang jumlah barang yang dipesan	22.875
4	<i>Item No</i>	Informasi tentang kode barang	22.875
5	<i>Description</i>	Informasi tentang nama setiap barang yang ada	22.875
6	<i>Price</i>	Informasi yang harga sewa per satuan untuk setiap barang	22.875

B. Impelementasi *Extraction, Transform, dan Loading (ETL)*

Proses *Extraction*

Data histori penyewaan diambil menggunakan *software accounting MYOB*, yang kemudian diekspor data ke dalam format Excel. Data yang diambil tersebut berasal dari *invoice* penjualan yang sudah dilakukan dari tahun 2016 sampai dengan tahun 2018.

Proses *Transformation Cleaning and Replace Data*

Setelah data disatukan dan menjadi sebuah tabel fakta, dilakukan proses *cleaning* dan *replace data*. Proses *cleaning* dilakukan agar konsistensi data tetap terjaga karena berasal dari banyak sumber.

Proses *replace* data dilakukan jika terdeteksi duplikasi data, seperti nama produk yang berbeda namun memiliki kode *Id_barang* yang sama ataupun sebaliknya (misalnya dalam Gambar 2 pada kolom kedua). Proses transformasi dilakukan menggunakan perangkat *Microsoft Excel* dengan menyamakan data master barang yang didapat dari PT.XYZ.

10	A-021bB	Rice Ball / mangkuk sambal	Rp. 1.100,00	buah
14	A-021bB	mangkuk samabl	Rp. 1.100,00	buah

Gambar 2. Contoh barang yang memiliki kode barang yang sama

Dalam Gambar 2, dicontohkan data yang memiliki kesamaan kode barang (A-021bB), namun memiliki nama produk yang berbeda dengan adanya pula kesalahan ketik (yaitu: 'Rice Ball / mangkuk sambal' dan 'mangkuk samabl'). Setelah proses *replace data* pada bagian nama produk, akan dilakukan proses *replace data* pada bagian *Id_barang*. Proses ini dilakukan karena didalam beberapa *invoice* terdapat perbedaan *Id_barang*.

Penambahan Kolom didalam Tabel Fakta

Setelah data *Id_barang* diubah, kemudian ditambahkan beberapa kolom di dalam tabel fakta. Kolom yang ditambahkan adalah: "Status Pelanggan", "Kategori Harga", "Jumlah Pesanan", dan "Kategori Barang". Kolom "Status Pelanggan" berisikan status dari pelanggan yang melakukan pemesanan barang apakah pelanggan ini berstatus "Kontrak" atau "Non Kontrak".

Untuk Kolom "Kategori Harga" berisikan data harga barang yang sudah di konversi dari nilai numerik menjadi nominal. Kolom "Jumlah Pesanan" berisikan data jumlah *quantity* yang sudah di konversi dari nilai numerik menjadi nilai nominal. Untuk kolom "Kategori Barang" berisikan kategori barang sesuai ketentuan dari PT.XYZ.

Proses Loading

Tabel fakta yang sudah dipersiapkan akan dianalisis menggunakan bahasa pemrograman Python dengan memanfaatkan Jupyter Notebook. Proses *loading* dilakukan dengan memanggil berkas *invoice* yang telah ditransformasi saat tahap persiapan. Contoh cuplikan data yang berhasil diambil dapat dilihat pada Tabel 5.

Tabel 5. Contoh Hasil *Loading Data*

Qty	Item No.	Desc.	Price	Date	Invoice number	Month	Customer
215	a-002b	Piring Ceper 10'	400	2017-06-01	A5470117	Januari	Yvones
50	a-003	Plate Snack	300	2017-06-01	A5470117	Januari	Yvones
175	a-005	Gelas Juice	300	2017-06-01	A5470117	Januari	Yvones
200	a-006	Gelas Minum	300	2017-06-01	A5470117	Januari	Yvones

C. Implementasi Algoritma Apriori

Setelah skema gudang data terbentuk, selanjutnya dilakukan proses analisis data dari tabel fakta. Teknik yang digunakan adalah menggunakan algoritma apriori. Proses pertama yang dilakukan dengan membuat sebuah *basket* atau keranjang belanja.

Basket ini berfungsi untuk mengambil data yang terdapat pada tabel fakta berdasarkan dimensinya, yaitu: dimensi waktu, pelanggan, jumlah pesanan, barang, dan harga. Setelah proses pembuatan *basket*, dilanjutkan dengan proses denormalisasi tabel. Contoh hasil denormalisasi dapat dilihat dalam Tabel 6. Jika proses denormalisasi telah dilakukan [5], yaitu untuk memastikan kemunculan *item* dalam sebuah *invoice* dan memastikan tidak terjadi kesalahan dalam pemrosesan data, barulah proses penemuan *frequent itemsets* dapat dilakukan.

Tabel 6. Contoh Hasil Proses Denormalisasi Data

Invoice Number	Garpu Kecil	Sendok Garpu	Sendok Garpu VIP	Sendok Pudding	Sendok Soup Makan	Sendok Teh
A5472117	1	0	1	0	1	1
A5472118	1	1	0	0	1	1
A5472119	1	0	1	0	1	1
A5472120	1	1	0	0	1	1
A5472121	1	0	1	0	1	1

Gambar 3 memperlihatkan lima nilai *frequent itemsets* tertinggi sebagai hasil dari algoritma apriori dari hasil denormalisasi data, yaitu: Garpu Kecil, Sendok Garpu, Sendok Garpu VIP, Sendok *Pudding*, dan Sendok *Soup Makan*.

Contoh isi tabel dimensi pemesanan dapat dilihat pada Gambar 4. Pengisian tabel dimensi pemesanan ini dimanfaatkan sebagai dasar untuk melakukan analisis kategori barang dalam tabel fakta. Dalam tabel fakta inilah analisis visual secara OLAP akan dilakukan sesuai kategori barang dan satuan periode peminjaman.

```
In [19]: frequent_itemsets = apriori(basket_sets, min_support=0.07, use_colnames=True)
In [20]: frequent_itemsets.head()
Out[20]:
```

	support	Itemsets
0	0.840989	(Garpu Kecil)
1	0.522968	(Sendok Garpu)
2	0.487633	(Sendok Garpu Vip)
3	0.717314	(Sendok Pudding)
4	0.876325	(Sendok Soup Makan)

Gambar 3. Lima Nilai Frequent Itemsets Tertinggi

ITEM NO	DESCRIPTION	PRICE	HARGA	QTY	JUMLAH PEMESANAN
A-003	Plate Snack	275	Murah	200	Sedikit
A-004	Gelas Goblet	1250	Sedang	50	Sedikit
A-005	Gelas Juice	250	Murah	400	Cukup
A-036	Gelas Ice Cream	500	Murah	50	Sedikit
A-006	Gelas Minum	250	Murah	1500	Banyak
A-007	Garpu Kecil	250	Murah	300	Sedikit
A-008	Sendok Garpu	300	Murah	400	Cukup
A-009	Sendok Soup Makan	300	Murah	700	Banyak
A-011B	Sendok Teh	250	Murah	300	Sedikit
A-012	Sendok Sayur	1250	Sedang	18	Cukup
A-013	Sendok Nasi	1250	Sedang	6	Sedikit
A-014	Sendok Soup Pendek	1250	Sedang	2	Sedikit
A-015	Sendok Soup Panjang	1250	Sedang	5	Sedikit
A-016	Capit Kue	1250	Sedang	4	Sedikit
A-017	Taart Schaap	1250	Sedang	6	Sedikit
A-018A	Mangkok Pudding	250	Murah	400	Cukup
A-021	Mangkok Soup Vip	1100	Murah	200	Sedikit
A-021B	Rice Ball	1100	Murah	6	Sedikit
A-020	Ringkok 6 Inch	350	Murah	500	Cukup
A-022	Nampan Biasa	5000	Sedang	6	Sedikit
A-023	Cambung	2000	Sedang	2	Sedikit
A-024	Pitcher	2000	Sedang	4	Sedikit
A-032	Meja Persegi Panjang	7000	Mahal	30	Banyak
A-032A	Meja Bulat	8000	Mahal	10	Sedikit
A-033	Container	5000	Sedang	10	Sedikit

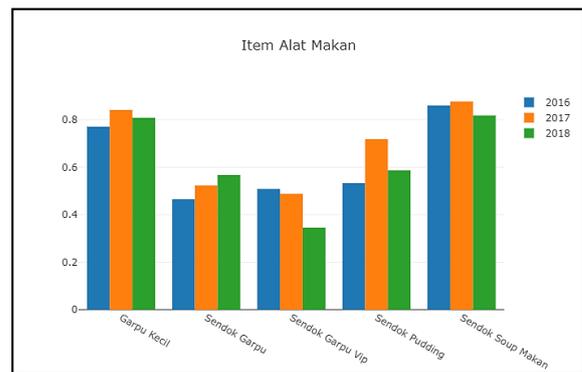
Gambar 4. Cuplikan Dimensi Pemesanan Barang

V. Pembahasan Visualisasi OLAP

A. Analisis Kategori “Alat Makan”

Gambar 5 menunjukkan grafik OLAP terhadap nilai *frequent itemsets* untuk kategori “Alat Makan” dari tahun 2016 sampai tahun 2018. Grafik OLAP dibuat dengan menggunakan pustaka Dash pada Python. Proses pembuatan grafik ini dilakukan dengan memanfaatkan nilai *frequent itemsets* pada rentang periode waktu tertentu, misalnya dalam satuan tahun yang dicontohkan dalam makalah ini.

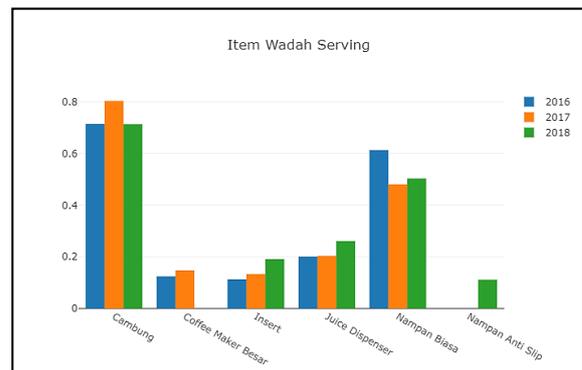
Dari grafik pada Gambar 5, dapat dilihat ada terjadinya kenaikan dan penurunan pada beberapa *itemsets* dalam kategori “Alat Makan”. *Item* Garpu Kecil mengalami kenaikan pada tahun 2017 dan mengalami penurunan pada tahun 2018. *Item* Sendok Garpu mengalami kenaikan pada tahun 2017 dan 2018. *Item* Sendok Garpu VIP mengalami penurunan pada tahun 2017 dan 2018. *Item* Sendok *Pudding* mengalami kenaikan pada tahun 2017 dan mengalami penurunan pada tahun 2018. Sedangkan untuk *item* Sendok *Soup* Makan cenderung stabil namun ada terjadi sedikit penurunan pada tahun 2018.



Gambar 5. Grafik Kategori “Alat Makan”

B. Analisis Kategori “Wadah Serving”

Gambar 6 menunjukkan grafik OLAP terhadap nilai *frequent item sets* untuk kategori “Wadah Serving” dari tahun 2016 sampai tahun 2018.



Gambar 6. Grafik Kategori “Wadah Serving”

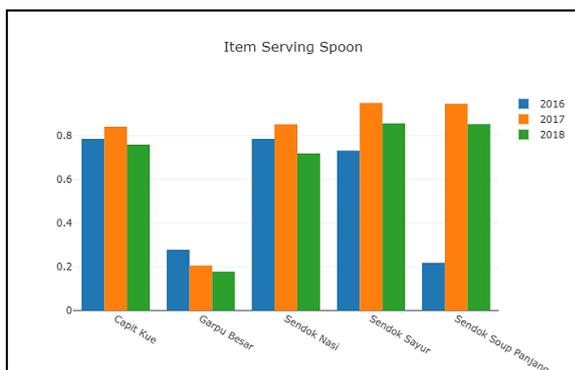
Pada grafik ini dapat dilihat ada terjadi kenaikan dan penurunan pada beberapa *item* peralatan “Wadah Serving”. *Item* Cambung mengalami kenaikan pada tahun 2017 dan mengalami penurunan pada tahun 2018.

Item Coffee Maker mengalami kenaikan pada tahun 2017 dan pada tahun 2018. Namun demikian, *Item* Coffee Maker tidak termasuk kedalam kategori *frequent item sets* sesuai hasil algoritma apriori, sehingga digantikan oleh *item* Nampan Anti Slip. *Item* Juice Dispenser tidak mengalami kenaikan pada tahun 2017 dan mengalami sedikit kenaikan pada tahun 2018. *Item* Nampan Biasa mengalami penurunan pada tahun 2017, namun mengalami sedikit kenaikan pada tahun 2018.

C. Analisis Kategori “Serving Spoon”

Gambar 7 menunjukkan grafik OLAP terhadap nilai *frequent item sets* untuk kategori “Serving Spoon” dari tahun 2016 sampai tahun 2018. Proses pembuatan grafik ini dengan memasukan nilai *frequent item sets* dari setiap *item* yang sudah di

analisis setiap tahunnya. Dari grafik pada Gambar 7, dapat dilihat ada kenaikan dan penurunan pada beberapa *item* "Serving Spoon".



Gambar 7. Grafik Kategori "Serving Spoon"

Item Capit Kue mengalami kenaikan pada tahun 2017 dan mengalami penurunan pada tahun 2018. *Item* Garpu Besar mengalami penurunan pada tahun 2017 dan pada tahun 2018, *item* Sendok nasi mengalami kenaikan pada tahun 2017 dan mengalami sedikit penurunan pada tahun 2018. *Item* Sendok Sayur mengalami kenaikan pada tahun 2017 dan mengalami sedikit penurunan pada tahun 2018.

D. Analisis Umum Tren Penyewaan

Dari hasil eksekusi algoritma apriori dalam gudang data dan analisis visual yang telah dilakukan, terlihat beberapa arahan untuk menjawab animo para penyewa berdasarkan pada kategori barang. Dalam kategori "Alat Makan", *item* seperti Sendok dan Garpu Biasa, hampir selalu diminati oleh konsumen dilihat dari jumlah pemesanan yang meningkat setiap periodenya. Untuk *item* seperti Sendok Garpu VIP mungkin dapat diperbaharui dengan model yang lebih kekinian, dengan mempertimbangkan jumlah pemesanan yang turun setiap tahunnya.

Dalam kategori "Wadah *servings*", *item* seperti Nampan Biasa disarankan untuk dilakukan pembaharuan model, seperti pada tahun 2018. Terdapat pula *item* baru yang bernama Nampan Anti Slip yang mulai diminati oleh konsumen. Analisis yang dilakukan dalam kategori "Serving Spoon", memperlihatkan bahwa *item* seperti garpu besar dan sendok nasi kemungkinan besar harus diperbaharui dengan model yang lebih bagus dilihat dari jumlah pemesanan yang mengalami penurunan.

Berdasarkan pada analisis kategori barang pada Gambar 5, 6 dan 7, pihak PT. XYZ merencanakan beberapa langkah strategis untuk pengadaan ketersediaan barang sewaan sebagai bagian dari

Business Process Re-engineering (BPR) [13, 14]. Salah satu hal yang direncanakan adalah untuk mengadakan survei kepuasan secara langsung di lapangan pada saat sebuah *event* diadakan, misalnya saat pesta pernikahan, pelantikan pejabat ataupun ulang tahun. PT. XYZ tidak lagi bergantung pada pihak penyewa (perusahaan catering) untuk penyebaran survei, tapi langsung datang ke lapangan untuk memantau kepuasan dan usulan dari para pengguna peralatan.

Di samping itu, PT. XYZ juga berencana untuk memberikan *teaser* (uji coba) pemakaian peralatan-peralatan baru, terutama yang terkesan mewah dengan berbagai asesornya, pada *event* dengan calon pengguna di atas 300 orang. Dengan upaya BPR ini diharapkan terjadi peningkatan jumlah barang yang disewa, dan memperkenalkan PT. XYZ secara lebih luas dalam dunia bisnis catering.

VI. KESIMPULAN

Berdasarkan hasil evaluasi dapat disampaikan bahwa implementasi gudang data telah berhasil dilakukan dengan menggunakan data historis penjualan selama tiga tahun terakhir. Gudang data pada PT. XYZ telah dapat digunakan untuk mendukung pengambilan keputusan yang berasal dari berbagai dimensi. Mekanisme ETL yang telah dihasilkan melalui riset ini akan terus disempurnakan dan ditindak lanjuti untuk periode-periode penyewaan selanjutnya. Melalui penerapan algoritma apriori dapat diketahui pula pola dan minat dari setiap pelanggan. Hal ini dapat dilakukan melalui pengamatan terhadap seberapa seringnya sebuah *item* disewa dalam satuan waktu tertentu ataupun kombinasi *item* apa saja yang dipesan oleh pelanggan. Dengan cara seperti ini pihak perusahaan dapat terbantu dalam menentukan strategi pengadaan barang di masa mendatang.

REFERENSI

- [1] Redman, P., Building the Unstructured Data Warehouse: Architecture, Analysis and Design, New Jersey: Technics Publications, 2006.
- [2] W.H. Inmon, Daniel Linstedt., Data architecture: a primer for the data scientist: big data, data warehouse and data vault, Waltham: Morgan Kaufmann, Elsevier, 2014.
- [3] Darudiato, S., "Perancangan data warehouse penjualan untuk mendukung kebutuhan informasi eksekutif cemerlang skin care", Seminar Nasional Informatika (SEMNASIF), 1 (5). Yogyakarta, 14 November 2015.
- [4] Kimball, R., Ross. M., The Data Warehouse Toolkit: The Definitive Guide To Dimensional

- Modeling, Indianapolis: John Wiley & Sons, 2013.
- [5] D. M. Kroenke dan D. J. Auer, Database Processing: Fundamentals, Design, and Implementation, Upper Saddle River, New Jersey: Pearson Education, Inc, 2012.
- [6] Amanpartap Singh, P. A. L. L., & Khaira, J. S., A "Comparative Review of Extraction, Transformation, and Loading Tools", Database Systems Journal BOARD 42, 2013.
- [7] Object Management Group, Business Process Model and Notation Resource Page, 2014. <http://www.omg.org/bpmn/index.htm>. [diakses 27 April 2020].
- [8] Fabio Nelli, Python Data Analytics: with Pandas, Numpy, and Matplotli, New York: Springer Science+Business Media, 2018.
- [9] Wes Mckinney., Python for Data Analysis: Data Wrangling with Pandas, Numpy, lpython, Sebastopol: O'Reilly Media, 2012.
- [10] Ivan Idris, Python Data Analysis, Birmingham: Packt Publishing, 2014.
- [11] Gunadi, G., & Sensuse, D. I., "Penerapan metode data mining market basket analysis terhadap data penjualan produk buku dengan menggunakan algoritma apriori dan frequent pattern growth (fp-growth): studi kasus percetakan PT. Gramedia", Telematika MKOM, 4(1), 118-132, 2016.
- [12] Elisa, E.,. "Market Basket Analysis Pada Mini Market Ayu Dengan Algoritma Apriori", Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi), 2(2), 472-478, 2018.
- [13] Bhaskar, H. L., "A critical analysis of information technology and business process reengineering", International Journal of Productivity and Quality Management, 19(1), 98-115, 2016.
- [14] Luo, Y., "Innovative Research on Catering Management", International Workshop on Advances in Social Sciences. Hongkong, China, 12-13 Desember 2018.