

## **BAB II**

### **LANDASAN TEORI**

#### **2.1 DATA MINING**

*Data mining*, atau sering disebut *knowledge discovery in database* (KDD), adalah sebuah aktivitas yang melibatkan pengumpulan dan penggunaan data historis, untuk menemukan sebuah aturan, pola, atau hubungan sistem dalam kumpulan data yang berukuran sangat besar. Hasil dari *data mining* dapat digunakan untuk meningkatkan atau memperbaiki dalam pengambilan sebuah keputusan di masa mendatang [7].

Menurut Sigit Arianto & Andung Jati Nugroho dalam Jurnal Cakrawala Ilmiah [8], *data mining* merupakan istilah yang digunakan untuk mencari informasi tersembunyi pada database. *Data mining* juga dapat didefinisikan sebagai proses menemukan korelasi antara pola dan tren yang berguna dan bermanfaat dengan menyaring sejumlah data yang besar (*Big Data*) yang disimpan dalam database.

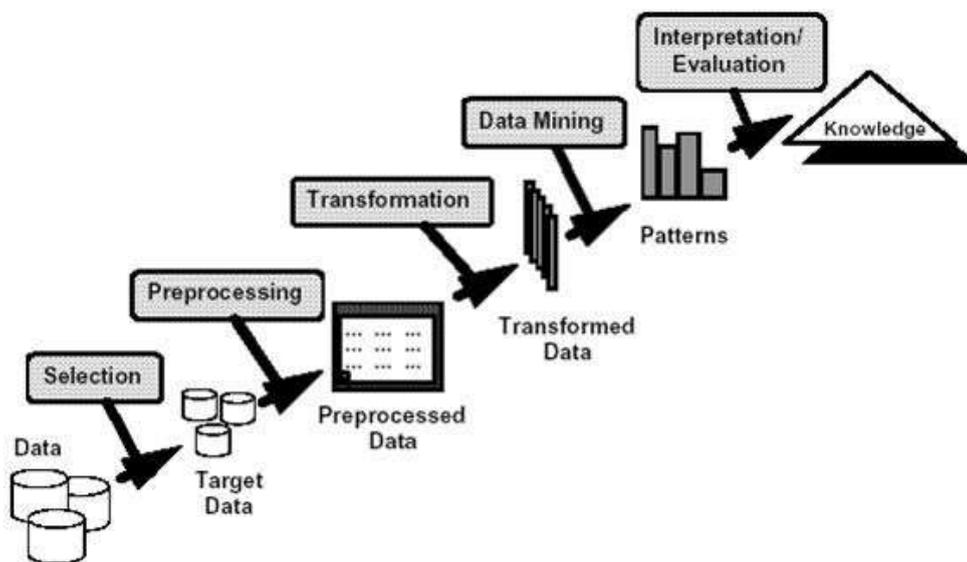
Menurut Ananda Hadi Elyas & J.Prayoga dalam Jurnal *Computer Science and Information Technology* [9], *data mining* adalah sebuah proses yang menggunakan teknik statistik, ilmu matematika, kecerdasan buatan dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang dapat berguna dan relevan dari sebuah database yang besar.

Dari pengertian diatas, maka dapat disimpulkan bahwa *Data mining* merupakan disiplin ilmu yang mempelajari serangkaian proses atau metode untuk mengekstrak atau menemukan pola untuk menghasilkan suatu informasi yang

diinginkan dari dalam suatu basis data atau kumpulan data yang sangat besar. Data yang telah diolah akan menjadi informasi yang nantinya akan menghasilkan sebuah pengetahuan.

### 2.1.1 Tahapan Proses dalam Data Mining

Menurut Ananda Hadi Elyas dkk. [9] dalam penerapan data mining memiliki beberapa tahapan dalam pengolahan data. Adapun tahapan dalam pengolahan data tersebut dapat dijelaskan pada gambar 2.1 sebagai berikut :



Gambar 2.1 Tahapan Proses dalam Data Mining [9]

#### 1. *Data Selection*

*Data Selection* merupakan kumpulan data operasional yang harus dilakukan sebelum tahap penggalian informasi dalam *Knowledge Discover Database* (KDD) dimulai. Data hasil seleksi yang digunakan dalam proses *data mining* disimpan dalam file terpisah dari *database* operasional.

## 2. *Preprocessing / Cleaning*

Sebelum proses data mining dapat dilaksanakan, perlu dilakukan proses cleaning pada data yang menjadi fokus KDD. Proses cleaning mencakup antara lain membuang duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data, seperti kesalahan cetak (tipografi).

## 3. *Transformation*

Pada fase ini yang dilakukan adalah mentransformasi bentuk data yang belum memiliki entitas yang jelas kedalam bentuk data yang valid atau siap untuk dilakukan proses *data mining*.

## 4. *Data Mining*

Pada fase ini yang dilakukan adalah menerapkan algoritma atau metode untuk melakukan pencarian pengetahuan.

## 5. *Interpretation / Evaluation*

Pada fase terakhir ini yang dilakukan adalah proses pembentukan keluaran yang mudah dimengerti yang bersumber pada proses data mining pola informasi.

### **2.1.2 Pengelompokan Data Mining**

Menurut Fitri Rahmawati dkk. [10] *data mining* dibagi menjadi beberapa kelompok berdasarkan tugas yang dapat dilakukan. Diantaranya adalah sebagai berikut :

1. Deskripsi

Merupakan metode pembagian *class* dan mencakup berbagai jenis data. Terkadang penelitian dan analisis hanya ingin menemukan cara untuk menggambarkan pola dan tren dalam data. Deskripsi dari pola kecenderungan sering memberikan penjelasan yang mungkin untuk suatu pola atau tren.

2. Estimasi

Estimasi hampir sama dengan klasifikasi, kecuali variable target estimasi lebih bersifat numerik daripada kategoris. Model dibangun menggunakan baris data (*record*) lengkap yang menyediakan nilai variabel target sebagai nilai prediksi.

3. Prediksi

Prediksi hampir sama dengan klasifikasi dan estimasi, kecuali bahwa dalam prediksi memiliki nilai dari hasil akan ada di masa mendatang. Beberapa metode dan teknik yang digunakan dalam klasifikasi dan estimasi dapat pula digunakan (dalam kondisi yang sesuai) untuk diprediksi.

4. Klasifikasi

Merupakan metode untuk mengklasifikasikan dan memprediksi data untuk mengidentifikasi atribut yang sebelumnya masih belum diketahui.

5. Pengklasteran

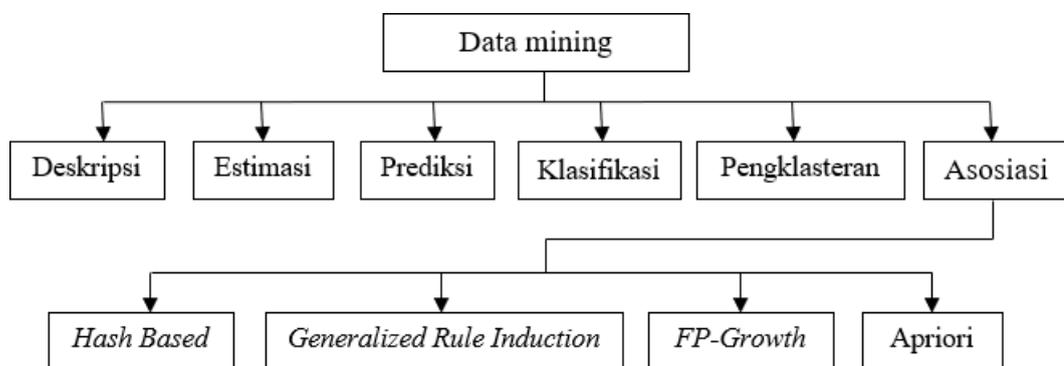
Merupakan metode analisis tanpa label kelas data yang nantinya dapat digunakan untuk menghasilkan label nanti. Pengklasteran adalah pengelompokan record, observasi, dan pembentukan suatu kelas objek yang

memiliki kemiripan satu sama lain dan memiliki record yang berbeda pada klaster lainnya.

#### 6. Asosiasi

Merupakan metode untuk menemukan aturan asosiasi untuk menentukan nilai kondisional dari suatu atribut. Tugas asosiasi dalam *data mining* adalah menemukan atribut yang terjadi secara bersamaan. Salah satu implementasi dari asosiasi tersebut adalah market basket analysis atau analisis keranjang belanja.

Ada beberapa metode untuk menyelesaikan dalam aturan asosiasi yang berdasarkan pada gambar 2.2 yang dapat dijelaskan sebagai berikut :



**Gambar 2.2 Pengelompokkan Data Mining**

##### a. Algoritma *Hash Based*

Algoritma hash based menggunakan teknik hashing untuk menyaring keluar itemset yang tidak penting untuk pembangkitan itemset selanjutnya. Ketika support count untuk kandidat k-itemset dihitung dengan menelusuri database, algoritma hash based mengumpulkan informasi mengenai (k+1)-

itemset dengan cara seluruh kemungkinan  $(k+1)$ -itemset dihash kedalam hash table dengan menggunakan hash (yang menggunakan sebuah bilangan prima untuk operasi modulo).

b. *Algoritma Generalized Rule Induction*

Generalized Rule Induction menggali rule yang memberikan informasi tertinggi berdasarkan support dan confidence. Memiliki batasan jumlah rule yang dipertahankan. GRI dapat mengatasi input data numerik dan kategorikal, namun target harus kategorikal.

c. *Algoritma FP-Growth*

*Frequent Pattern Growth* (FP-Growth) adalah algoritma yang dapat digunakan untuk menentukan himpunan data yang paling sering muncul (frequent itemset) dalam sebuah kumpulan data. Dalam Algoritma Frequent Pattern Growth (FP-Growth) menyimpan informasi mengenai frequent itemset dalam bentuk struktur prefix-tree atau sering di sebut FP-Tree.

d. *Algoritma Apriori*

Algoritma apriori diusulkan oleh R. Agrawal dan R Srikant pada tahun 1994 untuk menggali frequent item set dalam memperoleh association rule yang paling kuat hubungan antar atributnya. Frequent item set adalah himpunan transaksi yang terjadi berdasarkan minimum support. Rule yang kuat adalah yang memenuhi syarat minimum support dan minimum confidence. Algoritma apriori menggunakan pencarian berulang berdasarkan level item set, dimana  $k$ -itemsets digunakan untuk menemukan item set pada  $k+1$ .

Pada gambar 2.2 dapat dilihat bahwa data mining memiliki 6 metode yang terdiri dari Deskripsi, Estimasi, Prediksi, Klasifikasi, Klustering, dan Asosiasi. Metode yang penulis gunakan dalam penelitian ini adalah Asosiasi yang mana dalam metode ini memiliki berbagai macam algoritma, diantaranya adalah Algoritma *Hash Based*, Algoritma *Generalized Rule Induction*, Algoritma *FP-Growth*, Algoritma Apriori. Algoritma yang penulis gunakan dalam penelitian ini yaitu algoritma Apriori.

## 2.2 ATURAN ASOSIASI

Aturan asosiasi atau dikenal sebagai *association rules* adalah salah satu teknik data mining yang menjadi dasar dari beberapa teknik data mining lainnya. Aturan asosiasi adalah salah satu teknik *data mining* untuk menemukan hubungan antara entitas dalam kumpulan data yang besar yang telah ditentukan. Aturan asosiasi berguna untuk menganalisa isi dari keranjang belanja di pasar swalayan, karena itu aturan asosiasi (*association rules*) sering disebut dengan istilah *Market Basket Analysis* [8].

Aturan asosiasi (*association rules*) merupakan salah satu metode yang bertujuan untuk menemukan pola yang sering muncul diantara banyaknya transaksi, dimana setiap transaksi terdiri dari beberapa item, sehingga metode ini mendukung sistem rekomendasi dengan mencari pola diantara transaksi-transaksi yang telah terjadi [4].

Aturan asosiasi adalah suatu proses pada data mining yang mana untuk menentukan semua aturan asosiatif yang memenuhi persyaratan minimum untuk *support* dan *confidence* pada sebuah database. Kedua syarat tersebut akan

digunakan untuk *interesting association rules* dengan dibandingkan dengan batasan yang telah ditentukan [11].

Dari beberapa definisi aturan asosiasi yang telah dijelaskan, maka dapat disimpulkan bahwa aturan asosiasi (*association rules*) adalah teknik *data mining* yang digunakan untuk menemukan aturan asosiatif antara kombinasi item. Karena analisis asosiasi menjadi terkenal karena penerapannya dalam menganalisis isi keranjang belanja di pasar swalayan, maka analisis asosiasi sering disebut sebagai *Market Basket Analysis*. Aturan asosiasi juga merupakan suatu metode yang digunakan untuk mencari pola yang sering terjadi atau muncul pada banyaknya transaksi, dimana setiap transaksi terdiri dari beberapa item atau produk.

Analisis asosiasi dapat didefinisikan sebagai proses menemukan semua aturan asosiasi yang memenuhi persyaratan minimum untuk *support* (nilai penunjang) dan syarat *minimum* untuk *confidence* (nilai kepastian) [3]. Metodologi dasar analisis asosiasi dibagi menjadi dua langkah ,yaitu :

1. Analisa pola frekuensi tinggi

Pada langkah ini mencari kombinasi item yang telah memenuhi persyaratan minimum dari nilai *support* dalam *database*. *Support* (nilai penunjang), adalah suatu ukuran yang menunjukkan seberapa besar tingkat dominasi suatu *item* atau *itemset* dari keseluruhan transaksi. Nilai *support* sebuah item dapat diperoleh yang akan dijelaskan pada persamaan 2.1:

$$Support (A) = \frac{Jumlah\ transaksi\ mengandung\ A}{Total\ transaksi} \dots\dots\dots (2.1)$$

Sementara itu, nilai *support* dari 2 *item* dapat diperoleh dengan menggunakan persamaan 2.2 :

$$\text{Support (A, B)} = P(A \cap B)$$

$$\text{Support (A, B)} = \frac{\Sigma \text{Transaksi mengandung A dan B}}{\Sigma \text{Transaksi}} \dots\dots\dots(2.2)$$

## 2. Pembentukan aturan asosiasi

Setelah semua pola frekuensi tinggi telah ditemukan, barulah dicari aturan asosiasi yang memenuhi persyaratan minimum untuk *confidence* dengan menghitung *confidence* aturan asosiatif dari A ke B. Nilai *confidence* dari aturan A ke B dapat diperoleh dengan persamaan 2.3 :

$$\text{Confidence} = P(B|A) = \frac{\Sigma \text{Transaksi mengandung A dan B}}{\Sigma \text{Transaksi mengandung A}} \dots\dots\dots(2.3)$$

### 2.3 ALGORITMA APRIORI

Algoritma apriori adalah algoritma dasar yang diusulkan oleh *Agrawal* dan *Srikant* pada tahun 1994 untuk menentukan *frequent itemset* untuk aturan asosiasi *boolean*. Algoritma apriori berisi aturan asosiasi yang terdapat dalam *data mining*. Aturan yang menyatakan asosiasi atau *association rule mining* merupakan teknik data mining untuk menemukan aturan suatu kombinasi antara suatu *item*. [12].

Algoritma apriori merupakan salah satu jenis aturan asosiasi yang menjelaskan asosiasi terhadap beberapa atribut, aturan asosiasi sering disebut sebagai *affinity analysis* atau *market basket analysis*. Salah satu langkah analisis

asosiasi pola frekuensi tinggi (*frequent pattern mining*). Pentingnya sebuah asosiasi dapat ditentukan oleh dua kriteria, yaitu *support* dan *confidence*. Nilai *support* adalah persentase kombinasi item-item tersebut dalam database, sedangkan nilai *confidence* adalah kuatnya asosiasi antar item dalam aturan asosiasi [13].

Algoritma apriori merupakan salah satu dari banyak algoritma dalam data mining yang digunakan untuk frequent itemset dan association rule dalam basis data pada data transaksional yang dihasilkan dengan mengidentifikasi setiap item yang ada, dan mengkombinasikan kumpulan item yang lebih besar dengan syarat item tersebut cukup sering muncul dalam database [14]

Dari beberapa definisi algoritma *apriori* yang telah dijelaskan, maka dapat disimpulkan bahwa *algoritma apriori* adalah sebuah aturan yang menyatakan asosiasi atau *association rule mining* yang terdapat pada *data mining* untuk menemukan aturan suatu kombinasi antara suatu item. Salah satu tahap analisis asosiasi pola frekuensi tinggi (*frequent pattern mining*).

## **2.4 PERSEDIAAN**

Secara umum, persediaan adalah bahan atau barang yang akan digunakan untuk memenuhi tujuan tertentu, seperti untuk digunakan dalam proses produksi, untuk dijual kembali, atau sebagai suku cadang untuk peralatan atau mesin/ persediaan dapat berupa bahan baku, barang yang diproses, barang yang telah jadi, atau suku cadang [15].

Persediaan barang atau produk memiliki dua kata dasar yaitu “persediaan” dan “barang”. Persediaan sendiri dapat diartikan sebagai cara, proses, perbuatan untuk menyediakan suatu hal. Persediaan berasal dari kata dasar “sedia” yang dapat

diartikan sudah selesai dibuat, siap, sanggup, ada atau sudah ada. Sedangkan barang dalam kamus bahasa Indonesia adalah benda umum (segala sesuatu yang berwujud atau berjasad). Makna barang sendiri memiliki artian yang sangat luas sesuai dengan bagaimana kita menggunakannya. [16]

Persediaan adalah kegiatan yang berkaitan dengan sebuah perencanaan, pelaksanaan, dan pengawasan penentuan kebutuhan material atau barang lainnya, sehingga kebutuhan operasional dapat terpenuhi dengan tepat waktu. Di lain pihak investasi pdalam persediaan material atau barang dapat ditekan secara optimal [17].

Dari beberapa definisi persediaan yang telah dijelaskan, maka dapat disimpulkan bahwa persediaan adalah sekumpulan barang yang harus disediakan oleh perusahaan-perusahaan ke lokasi tertentu. Artinya tersedianya sejumlah barang yang dipasok oleh beberapa perusahaan untuk memenuhi kebutuhan produksi atau penjualan barang.

## **2.5 WEKA**

WEKA (Waikato Environment for Knowledge Analysis) merupakan aplikasi data mining yang bersifat open source berbasis Java. WEKA pertama kali dikembangkan oleh Universitas Waikato Selandia Baru sebelum menjadi bagian di Pentanho. Weka terdiri dari koleksi algoritma machine learning yang dapat digunakan untuk melakukan generalisasi atau formulasi dari sekumpulan data [18].

Weka merupakan salah satu sistem yang digunakan untuk melakukan mining data. Aplikasi ini berlisensi GNU General Public License sehingga dapat digunakan secara gratis. Tujuan penggunaan aplikasi ini adalah untuk memudahkan pengolahan data mulai dari pengolahan awal hingga pemodelan data [19] .

*WEKA* adalah sebuah perangkat lunak pembelajaran mesin yang ditulis dalam bahasa Java yang dikembangkan di Universitas Waikato, Selandia Baru. Perangkat lunak ini memiliki banyak algoritma pembelajaran mesin untuk keperluan *data mining*. *WEKA* juga memiliki tools untuk pengolahan data, mulai dari preprocessing, klasifikasi, association rules, dan visualization [20].

## **2.6 PENELITIAN SEJENIS**

Penulis memulai penelitian ini dengan terlebih dahulu melakukan studi kepustakaan dari penelitian-penelitian dan sumber-sumber lain. Beberapa penelitian telah banyak dilakukan dengan menggunakan teknik data mining untuk menggali berbagai informasi dari sebuah database transaksi penjualan. Untuk menambah wawasan serta pengetahuan, peneliti melakukan tinjauan penelitian sejenis . Ringkasan tinjauan penelitian sejenis dapat dilihat pada tabel 2.1 :

Tabel 2.1 Penelitian Sejenis

No	Penulis dan Tahun	Masalah	Metode	Hasil
1	Rizal Sidiq Al Amin, Rangga Sanjaya  Tahun 2021 [21].	CV. Sinar Bintang Asia memiliki masalah dalam melakukan persediaan barang, dalam menentukan persediaan barang perusahaan tersebut masih menggunakan cara spontan yaitu dengan melihat barang mana saja yang stoknya sedikit maka setelah melihat stok yang sedikit akan dibuatkan sebagai keputusan persediaan barang yang akan ditambah tidak melihat dari data historis. Maka dari itu perusahaan memerlukan waktu yang cukup lama untuk menentukan persediaan dan jika barang tersebut terjadi penumpukan tidak laku maka terjadi pengeluaran biaya yang kurang efektif	Apriori	Hasil dari penelitian ini yaitu jika membeli hitam maka akan membeli biru navy dengan nilai support dan confidence 50%, jika membeli biru navy maka akan membeli hitam dengan nilai support 50% dan confidence 100%. Hasil ini dapat membantu mengelola persediaan , tata letak maupun penjualan menjadi lebih efektif dan efisien.
2	Elischa Febriyani, Saifullah,	Instalasi farmasi Dr RSUD Djasamen Saragih	Apriori	Hasil yang diperoleh terdapat tiga jenis obat yang sering terjual

	Riki Winanjaya Tahun 2021 [22].	Pematangsiantar sering mengalami permasalahan terhadap stok obat yang akan diberikan kepada pasien. Permasalahan tersebut adalah pihak Instalasi farmasi Dr RSUD Djasamen Saragih Pematangsiantar harus mengganti merek obat dengan kandungan obat yang sama yang telah diresepkan oleh dokter, karena stok obat mengalami kekosongan.		pada tahun 2018 yaitu Isoniazid, dan Neurodex pada Amlodipin 10mg. Hasil pengujian membentuk aturan itemset yang bisa dijadikan item yang saling berkaitan dengan item lain dalam pola penjualan.
3	Sry Pay Tualeka, Faza Alameka, Nariza Wanti Wulan Sari Tahun 2021 [23].	Permasalahan utama yang dihadapi oleh CV Pasti Jaya Houseware adalah sulitnya mengetahui kecenderungan atau pola belanja pembeli terhadap produk-produk yang ada di CV Pasti Jaya Houseware. Dengan menempatkan produk tertentu ditempat yang tepat akan memberikan kemudahan konsumen dalam berbelanja dan dapat mengefisiensi gudang tanpa takut barang menumpuk terlalu lama dan rusak	Apriori	Hasil dari penelitian ini adalah produk piring memiliki keterkaitan yang cukup kuat dengan gelas dibandingkan barang lain yaitu dengan nilai support 60% dan confidence 100%. Barang-barang yang memiliki nilai confidence yang lebih tinggi akan diletakkan bersebelahan karena dengan tingginya nilai confidence antar kedua barang lebih memiliki kesempatan dibeli bersamaan yang tinggi juga.
4	Putri Mai Sarah	Permasalahan yang dialami pada Toko	Apriori	Hasil dari penelitian ini di dapat

	Tarigan, Jaya Tata Hardinata, dkk.  Tahun 2022 [4]	Sinar Harahap adalah persediaan barang atau bahan kebutuhan yang tidak dilakukan secara optimal sehingga menyebabkan kekosongan pada barang atau bahan kebutuhan tersebut		permintaan tertinggi jika terjadi pada pembelian beras dan juga pembelian telur, dengan nilai <i>confidence</i> sebesar 62%.
5	Fitra Panjaitan  Tahun 2021 [24]	Permasalahan yang sering terjadi pada UD. Chandra Jaya Tani adalah kehabisan persediaan stok barang karena tingkat belanja konsumen yang tidak beraturan dan penumpukan barang yang mengakibatkan kerugian pada UD. Chandra Jaya Tani. dimana pihak toko kesulitan dalam pengambilan keputusan menentukan jumlah barang yang akan disediakan.	Apriori	Hasil yang didapatkan dari penelitian ini adalah algoritma apriori terbukti mampu menentukan pola persediaan barang yang sering di minati konsumen dengan minimum support 10% dan <i>confidence</i> sebanyak 20%.

Dari beberapa penelitian yang telah dijelaskan pada tabel 2.1, dapat dijelaskan bahwa algoritma apriori pada teknik data mining sangat efisien dan memiliki tingkat akurasi yang baik dalam mempercepat proses pembentukan kecenderungan pola kombinasi antar itemset. Hasil penelitian dapat membantu pihak perusahaan dalam melakukan penyetokan atau persediaan barang yang berhubungan antar itemnya.

Berdasarkan kesimpulan penelitian sejenis yang telah dijelaskan pada tabel 2.1, dapat dijelaskan bahwa yang membedakan penelitian ini dengan penelitian sebelumnya adalah penulis akan menganalisis data transaksi penjualan barang pada CV. SUARA AGUNG. Penelitian ini berfokus pada bagaimana CV tersebut dapat memanfaatkan data transaksi penjualan sebagaimana mestinya sehingga dapat membantu pihak CV. SUARA AGUNG dalam menentukan strategi dalam persediaan barang.