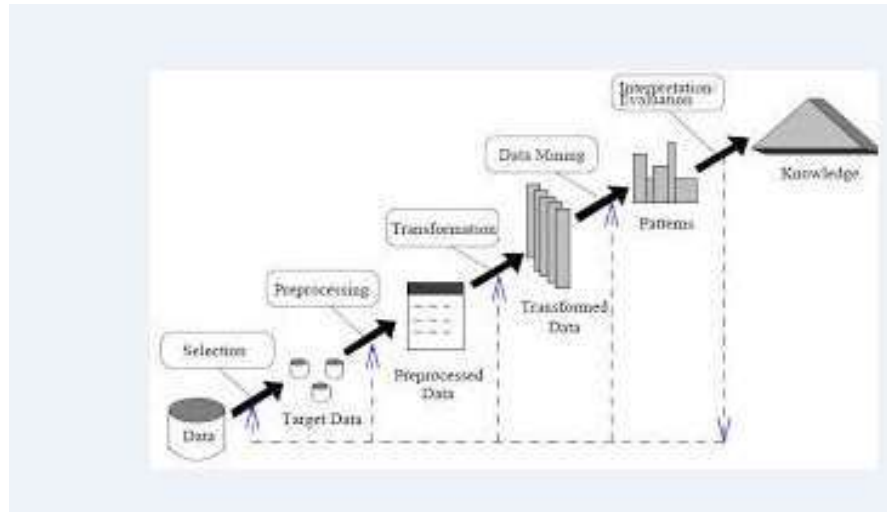


BAB II

LANDASAN TEORI

2.1 *Knowledge Discovery in Database (KDD)*

Menurut (Zulfa et al., 2020), Penambangan informasi sendiri adalah pengelompokan siklus yang mampu mengamati data tambahan yang belum ada dalam kumpulan data dan model yang berarti mengamati model informasi yang dapat diubah menjadi data penting dari hasil pembagian dan penyajian yang bermanfaat atau menarik. Model dari informasi yang terkandung dalam kumpulan data.). KDD menciptakan informasi yang lebih baik dan lebih jelas. KDD khawatir tentang merancang, mencampur dan pengungkapan logis, terjemahan dan persepsi banyak informasi. KDD adalah rangkaian yang dilengkapi untuk pengumpulan informasi, dalam menangani informasi untuk mendapatkan contoh atau koneksi yang diatur secara rutin dalam kumpulan data interior yang sangat besar. Pengungkapan informasi dalam kumpulan data (KDD) pada dasarnya adalah metode yang terlibat dengan melacak informasi yang berguna dalam berbagai informasi.



Gambar 2. 1 Ilustrasi Proses Tahapan KDD

Sumber : Tri Utami Putri, M.Izman H,S.T.,M.M.,PhD , Susan Dian PS M.Kom

1. Seleksi Data (*Data Selection*)

Sekumpulan data yang dilakukan untuk mencari pengolahan informasi dalam KDD. Data hasil seleksi yang digunakan untuk *data mining*, menyimpan data, dan diolah dan akan terpisah dari data yang tidak digunakan. Membuat data target, memilih data, fokus pada satu variabel atau sampel untuk mencari data yang relevan.

2. Pemilihan data (*Preprocessing / Cleaning*)

Untuk membuat proses *data mining*, proses selanjutnya perlu dilakukan pembersihan data yang digunakan pada data KDD. Dimana masih banyak data yang salah dan tidak relevan atau data yang sama (*double*), memeriksa data yang tidak valid atau *real* yang diperlukan pada KDD, seperti data atau informasi eksternal.

3. Transformasi (*Transformation*)

Adalah proses transformasi atau perombakan data yang telah di pilih, dan data harus sesuai dengan fakta atau proses yang dilakukan *data mining*. Dalam proses transformasi diperlukan kreatif tergantung pada pola informasi yang diproses dari database. Fungsi transformasi adalah untuk menjelaskan atau menerangkan data yang relevan atau sudah siap untuk proses *data mining* berdasarkan goal yang ingin di capai.

4. *Data mining*

Proses *data mining* adalah proses untuk mencari dan menggali informasi yang sangat menarik untuk memilih data dengan menggunakan banyak teknik. Pada proses ini akan melakukan pemilihan algoritma untuk pencarian. Pemilihan hasil dari proses KDD adalah seperti *association, regression, forecasting, solution, clustering*. Pemilihan algoritma ini sangat bermanfaat pada goal untuk proses KDD secara menyeluruh.

5. Interpretasi / Evaluasi (*interpretation/evaluation*)

Didalam proses itu pola-pola yang dihasilkan dari *data mining* akan diamati. Pada tahap ini informasi yang sudah diperoleh akan ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang ingin menggunakannya. Proses KDD akan memeriksa semuanya apakah informasi yang dicari berbeda dengan fakta atau hasil pengamatan yang dilakukan sebelumnya.

2.2 *Data mining*

Penambangan informasi adalah informasi yang diperoleh sebagai data yang akan ditangani untuk data yang sangat membantu dalam basis informasi. Penambangan informasi adalah cara paling umum untuk memulihkan data dari berbagai informasi melalui perhitungan dan strategi termasuk pengukuran, mesin, dan kumpulan data. (Sandi Fajar Rodiyansyah, 2015). Penambangan informasi adalah perkembangan siklus untuk mengamati nilai tambah dari informasi sebagai informasi yang belum diketahui hingga saat ini.

Pakar juga memberikan pengalaman tentang penggalian informasi seperti yang diilustrasikan dalam buku yang ditulisnya, yaitu : Menurut (D. Marlina Verawati, A. Achsa, 2021) (Dr. Suyanto,S.T.,2017:1) Data mining adalah cara paling umum untuk menyelidiki informasi dalam kumpulan data untuk mengamati pemikiran baru yang sebelumnya tidak jelas yang dikumpulkan dalam banyak informasi dengan strategi atau teknik yang penting untuk penalaran buatan, AI, pengukuran, dan kumpulan data. Penambangan informasi adalah interaksi atau teknik untuk membuat nilai tambah yang belum diketahui secara fisik dalam kumpulan data (Retno Tri Vulandari, S.Si. 2017:1)

Data mining merupakan suatu hubungan untuk menemukan pola dan kecenderungan untuk mengoreksi dat-data yang berjumlah banyak yang disimpan dalam penyimpanan database dengan menggunakan teknik matematika dan statistika(Saikin & Kusrini, 2019). Berdasarkan pengertian *data mining* yang telah diuraikan maka dapat ditarik kesimpulan bahwa *data mining* itu merupakan suatu proses pencarian dan penggalian data yang selama ini belum diketahui

dalam data yang sangat besar didalam database yang berguna untuk pemilik data kemasa yang akan datang.

2.2.1 Pengelompokan *Data mining*

Menurut (Saikin & Kusrini, 2019), Dalam bukunya *Data mining* dibagi menjadi beberapa kategori berdasarkan tugas yang dapat dilakukan yaitu:

1. Deskripsi

Banyak peneliti secara sederhana ingin mencoba cara atau langkah dalam menggambarkan pola dan kecenderungan yang terdapat dalam data.

2. Estimasi

Model dibuat menggunakan laporan lengkap yang mempunyai nilai dan variable target sebagai nilai dugaan sementara. Estimasi hampir sama dengan klasifikasi.

3. Prediksi

Prediksi hampir sama dengan estimasi dan klasifikasi, bedanya dalam prediksi nilai dari hasil akan ada dimasa yang akan datang.

4. Klasifikasi

Dalam Klasifikasi, ada target variable kategori. Misalnya memperkirakan apakah suatu pengajuan hiptek oleh nasabah merupakan suatu kredit yang baik dan buruk.

5. Pengklusteran

Merupakan pengelompokan laporan, observasi, atau memperhatikan dan membuat tahap objek-objek yang mempunyai kesamaan. Kluster adalah kumpulan laporan yang mempunyai kesamaan antara yang satu dengan

yang lain dan tidak memiliki ketidaksamaan dengan laporan-laporan kluster lainnya. Misalnya dalam pengklusteran dalam bisnis penelitian yaitu untuk tujuan audit akutansi yaitu melakukan pemisahan terhadap perilaku finansial dalam baik dan mencurigakan

6. Asosiasi

Dalam *data mining* bertugas untuk menemukan atribut yang muncul dalam satu tahun. Atau dalam dunia bisnis disebut juga analisis keranjang belanja.

2.2.2 Manfaat Data mining

Menurut (Dr.Suyanto, S.T., 2017:3) kegunaan *data mining* ada 2 yaitu: deskriptif dan prediktif. Deskriptif adalah *data mining* yang digunakan untuk menemukan pola-pola yang dapat dimengerti oleh manusia yang menjelaskan karakteristik data. Prediktif adalah *data mining* untuk membentuk model ilmu pengetahuan yang akan digunakan untuk melakukan perkiraan atau prediksi.

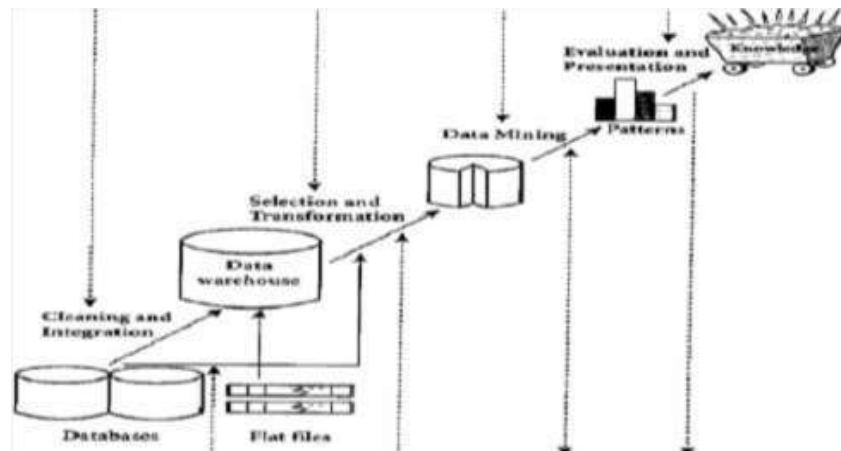
Berdasarkan fungsinya, adapun tugas-tugas *data mining* dibagi menjadi 6 kelompok yaitu:

1. Klasifikasi (*classification*) adalah menguraikan struktur yang diketahui untuk digunakan pada data-data baru.
2. Klasterisasi (*clustering*) adalah mengelompokkan data, yang belum diketahui jenis kelasnya, kedalam beberapa kelompok sesuai dengan kesamaannya.
3. Regresi (*regression*) adalah menemukan fungsi yang memodelkan data (kesalahan prediksi) seminimal mungkin.

4. Deteksi anomali (*anomaly detection*) adalah mengidentifikasi data yang tidak umum, perubahan yang sangat penting dan perlu penyelesaian lebih lanjut.
5. Pembelajaran aturan asosiasi (*association rule learning*) atau pemodelan perbandingan atau mencari relasi antar variabel.
6. Perangkuman (*summarization*) adalah menyediakan representasi data yang lebih sederhana yaitu visualisasi dan pembuatan laporan.

2.2.3 Tahapan Data mining

Salah satu rangkaian dalam proses *data mining* yaitu proses tahapan yang dilakukan. Tahap-tahap tersebut saling berhubungan dengan teknik integrasi dan penemuan ilmiah dan visualisasi dalam pola sejumlah data.



Gambar 2. 2 Gambaran Tahap- Tahap data mining

Sumber: [www:kompasiana.com](http://www.kompasiana.com)

Tahap- tahap *data mining* adalah sebagai berikut:

1. Pembersihan data (*data cleaning*)

Pembersihan data adalah suatu proses untuk membuang data yang tidak valid atau kata yang salah. Data- data yang tidak relevan juga lebih baik dibuang untuk mengurangi data yang akurat nantinya.

2. Integritas Data (*Data Integration*)

Proses penggabungan dari beberapa database kedalam database yang baru.

3. Transformasi data (*Data Transformation*)

Mengubah data dan menentukan kualitas hasil *data mining* yang telah digunakan. Sehingga dapat lebih mudah untuk digunakan.

4. Aplikasi teknik *Data mining*

Masalah satu bagian dalam *data mining*. Proses ekstraksi pola dari data yang sudah ada sebelumnya.

5. Evaluasi pola

Untuk mengevaluasi pola-pola yang ingin dicapai apakah hasil pengamatan memang tercapai. Menjadikan perbaikan untuk proses *data mining* yang dihasilkan.

6. Presentasi pengetahuan (*Knowledge Presentation*)

Mempresentasikan hasil *data mining* dalam bentuk pengetahuan untuk memperoleh informasi yang ingin diperoleh oleh pengguna.

2.3 Tiket

Dalam maskapai, tiket adalah cara untuk menawarkan layanan penerbangan kepada pelancong atau pembeli. Sesuai dengan Keputusan Menteri Perhubungan Nomor 25 Tahun 2008 tentang Penyelenggaraan Angkutan Udara, yang dimaksud

dengan tiket adalah suatu catatan tercetak yang merupakan salah satu bukti adanya pengaturan angkutan udara antara pelancong dan pengangkut. Kemudian, pada saat itu, sesuai Undang-Undang Republik Indonesia Nomor 1 Tahun 2009 tentang Penerbangan, yang dimaksud dengan tiket adalah laporan dalam bentuk tercetak, melalui interaksi elektronik, atau dalam struktur yang berbeda, yang merupakan salah satu konfirmasi adanya persetujuan transportasi udara dengan pelancong.

2.4 Metode Apriori

Dalam proses information mining ada beberapa teknik yang dapat dimanfaatkan. Dalam ulasan ini, penulis menggunakan teknik perhitungan apriori. Perhitungan apriori adalah salah satu perhitungan penting yang digunakan untuk mengubah dari basis informasi yang sangat besar menjadi aturan afiliasi untuk melacak informasi. Susunan afiliasi memutuskan kebutuhan untuk memenuhi kebutuhan dasar harga bantuan dalam basis informasi. Besaran nilai sebuah item diperoleh dengan menggunakan persamaan berikut :

$$\text{Support (A)} = \frac{\text{Jumlah Transaksi mengandung A}}{\text{Total Transaksi}} \times 100\%$$

Sementara nilai support antara 2 item diperoleh dengan menggunakan rumus:

$$\text{Support (A, B)} = \frac{\Sigma \text{Transaksi mengandung A dan B}}{\Sigma \text{Transaksi}} \times 100 \%$$

Rumus 2. 1 Perhitungan Support

Frekuensi *itemset* menunjukkan itemset yang menunjukkan frekuensi kemunculan lebih dari nilai minimum yang ditentukan (). Misalkan =2 Maka frekuensi kemunculanya lebih dari atau sama dengan 2 kali disebut *frequent*.

Himpunan *Frequent k- itemset* dilambang dengan FK. Setelah semua pola frekuensi barulah dicari aturan asosiasi yang memenuhi syarat minimum untuk *confidence* dengan menghitung *confidence* aturan asosiatif A –B.

Menurut (Soma Darmawan, 2015), menyimpulkan bahwa algoritma apriori dapat digunakan dalam metode yang dapat menyelesaikan permasalahan, yang menarik perhatian banyak peneliti menghasilkan algoritma yang efisien.

2.5 Software pendukung



Gambar 2. 3 Logo Tanagra

Sumber : Tanagra. com

Tanagra adalah salah satu software pada *data mining* yang disediakan beberapa metode *data mining* mulai dari perpaduan analisis data, pembelajaran statistik, dan pembelajaran mesin dan database. Tanagra merupakan software pendukung berbasis *open source* dimana semua orang dapat mengakses *source* kodenya dan menambahkan algoritmanya, biasanya banyak digunakan untuk kepentingan akademik dan riset.

Sebagian besar yang menggunakan aplikasi Tanagra ini adalah mahasiswa dan peneliti. Dengan menggunakan software Tanagra memberikan peneliti dan mahasiswa lebih mudah digunakan sesuai dengan aturan yang sudah ada di pengembangan software untuk menganalisis data yang *real*.

2.6 Penelitian Terdahulu

1. Nama Pengarang: (Elisa, 2018), judul : Market Basket Analysis Pada Mini Market Ayu Dengan Algoritma Apriori.

Tahun: Vol. 2 No. 2 (2018) 472 – 478, ISSN : 2580-0760

Pada usaha minimarket Ayu di Kota Batam. Minimarket ini berlokasi dekat dengan rumah penduduk, dimana mempengaruhi berlangsungnya penjualan setiap hari, penjualan yang dilakukan semakin hari membuat data transaksi terus bertambah, membuat penyimpanan data yang besar. Data transaksi penjualan tidak digunakan dengan baik hanya dijadikan arsip. Pada dasarnya informasi yang diperoleh menjadi kumpulan data yang sangat bermanfaat. Metode yang digunakan dalam Analisis keranjang pasar menggunakan Algoritma Apriori, yang bertujuan untuk mencari pola asosiasi melalui pola berbelanja yang dilakukan oleh konsumen, dan apa saja item-item barang yang dibeli dengan bersamaan. Hasil yang diperoleh dari penelitian ini adalah menemukan nilai support dan confidence tertinggi ada pada minyak dan susu, dengan nilai support 42,85% sedangkan confidence 85,71%.

2. **Nama pengarang:** (Zahrotul Wardah, 2017) , **Judul: Implementasi Data Mining Pada Penjualan Tiket Pesawat Menggunakan Algoritma Apriori (Studi Kasus: Pt. Pesona Ceria Travel).**

Tahun: Vol. 2 No. 1 Januari 2018 ISSN 2548-740X

Semakin banyaknya travel yang saat ini berkembang, membuat pengelola ingin menunjukkan strategi pemasaran yang lebih baik, salah satunya adalah PT. Pesona Ceria Trevel ingin meningkatkan penjualan tiket. Dalam kegiatan sehari-hari PT. Pesona Perjalanan Ceria mempekerjakan staf tiket untuk mengelola penjualan tiket. untuk mengetahui tiket terbaik dijual, PT. Pesona Wisata Ceria membutuhkan aplikasi yang dapat menampilkan laporan penjualan tiket sehingga memudahkan pemilik untuk menentukan promosi pada tiket tersebut. Dalam hal ini, data mining dengan metode Apriori digunakan untuk mencari frekuensi item ditetapkan pada kumpulan data. Dengan algoritma ini, koleksi tiket pesawat yang sering dijual adalah Lion Air, Citilink, Garuda dan Batik Air. Asosiasi ini diperoleh dengan nilai Support minimal 30% dan nilai Percaya diri 80%. Berdasarkan hasil pengujian dapat disimpulkan bahwa aplikasi yang dibangun berhasil sesuai dengan desain dan dapat memberikan masukan kombinasi tiket sering dipesan pada waktu tertentu.

3. Nama pengarang: (Purnia & Warnilah, 2017) , **Judul: Implementasi Data Mining Pada Penjualan Kacamata Menggunakan Algoritma Apriori**

Tahun: Vol.2 No.2, November 2017 ISSN: 2527-449X

Optik Nasional merupakan salahsatu optik yang ada di Tasikmalaya, Pada kegiatannya Optik Nasional ini melakukan proses jual beli Kacamata dengan berbagai merk, Banyak nya Transaksi Penjualan sehari-hari maka data penjualan pun semakin lama akan bertambah semakin banyak. Jika dibiarkan saja, maka data-data transaksi penjualan tersebut hanya menjadi sampah yang tidak berarti. Dengan adanya dukungan perkembangan teknologi, semakin berkembang pula kemampuan dalam mengumpulkan dan mengolah data. Pemanfaatan informasi dan pengetahuan yang terkandung di dalam banyaknya data tersebut, pada saat ini disebut dengan data mining.

4. Nama pengarang: (Irene Ananda & Salamah, 2020), **Judul: Determination Of Sales Data Patterns Using The Association Rules Apriori Method**

Tahun: 06 Agustus 2020 eISSN : 2598-246X

Dalam persaingan dalam dunia bisnis perlu dicari strategi yang tepat yang dapat digunakan dalam optimasi penjualan. Faktor yang mempengaruhi kebutuhan analisis pasar adalah tingkat frekuensi konsumen dalam membeli suatu barang. Karena itu diperlukan solusi untuk menemukan pola penjualan dengan website menjadi lebih efektif dan efisien. NS data

yang dibutuhkan diambil dari data transaksi penjualan untuk periode tertentu dan diproses untuk menghasilkan asosiasi aturan barang dan transaksi. Selain bisa cari pola yang sering muncul di antara banyak transaksi, hal ini dapat memudahkan perusahaan untuk meningkatkan omzet penjualan. Pembuatan aplikasi ini menggunakan HTML sebagai pengembangan halaman web, PHP sebagai situs web pengembangan, dan MySQL sebagai manajemen database. Di dalam tahap pengujian, aplikasi ini dimulai dari login hingga mendapatkan hasil analisis asosiasi berjalan dengan baik.

5. Nama pengarang: (Alfianzah et al., 2020), **Judul: Implementation of Apriori Algorithm Data Mining for Increase Sales**

Tahun: Volume 5, Number 1, October 2020 e-ISSN : 2541-2019

Setiap perusahaan atau organisasi yang ingin bertahan perlu menentukan strategi bisnis yang tepat. Data penjualan produk yang dilakukan oleh Lakoe Dessert Pondok Kacang pada akhirnya akan menghasilkan data yang menumpuk, sehingga sangat disayangkan jika tidak dianalisis kembali. Produk yang ditawarkan bervariasi dengan berbagai macam produk sebanyak 45 produk, untuk mengetahui produk dengan penjualan terbanyak dan keterkaitan antara produk yang satu dengan produk yang lain diperlukan salah satu algoritma dalam algoritma data mining yaitu apriori algoritma untuk mengetahuinya, dan dengan bantuan aplikasi Rapidminer 5, dengan nilai dukungan 2,4% dan nilai kepercayaan 50%, produk yang sering dibeli atau diminati pelanggan dapat ditemukan.

Penelitian ini menggunakan data penjualan bulan Maret 2020 yang berjumlah 209 data transaksi. Dari penelitian diketahui bahwa item dengan nama Pudding Strawberry dan Pudding Vanilla merupakan produk yang paling banyak dibeli oleh konsumen. Dengan mengetahui produk yang paling banyak terjual dan pola pembelian barang yang dilakukan oleh konsumen, Lakoe Dessert Pondok Kacang dapat mengembangkan strategi pemasaran untuk memasarkan produk lain dengan menganalisis keuntungan dari penjualan produk yang paling banyak terjual dan mengantisipasi kehabisan atau kosongnya stok atau bahan pada suatu saat.

- 6. Nama pengarang: (Firmansyah & Merlina, 2020) , Judul: Prediksi pola Penjualan Tiket Kapal PT.Pelni cabang Makassar Menggunakan Metode Algoritma Apriori ,Tahun: VOL. 5. NO. 2 FEBRUARI 2020 E-ISSN: 2527-4864**

Tujuan dari penelitian ini adalah untuk memprediksikan penjualan tiket pesawat yang sering dilakukan dengan melihat pola pembelian tiket sebelumnya. Metode yang digunakan dalam penelitian ini menggunakan Algoritma Apriori, dalam hal ini penulis mengambil studi kasus Cabang Makassar. Algoritma Apriori menghasilkan aturan asosiasi antar item pada bulan Januari 2018 sampai dengan Desember 2018 diketahui pola penjualan tiket kapal bahwa jika membeli tiket KM Lambelu maka akan membeli tiket KM Bukit Siguntang secara bersamaan dengan nilai support 75% dan nilai confidence 90%.

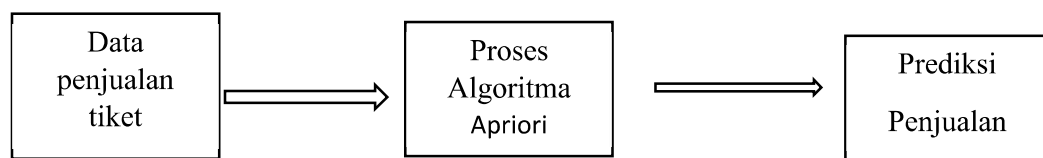
7. **Nama pengarang:** (Anas, 2020) **Judul: Implementasi Algoritma Apriori Untuk Menentukan Strategi Promosi STIE-Graha Karya Muara Bulian Tahun: Vol. 14, No. 1, April 2020 ISSN: 1978-8126**

Tujuan dari penelitian ini adalah untuk mengimplementasikan algoritma apriori untuk mendapatkan gambaran terhadap data mahasiswa baru STIE-GK Muara Bulian. Penulis menggunakan algoritma apriori untuk menganalisis data mahasiswa baru STIE-GK Muara Bulian tahun akademik 2018/2019. Algoritma apriori dipilih karena kemampuannya dalam menganalisis data yang muncul secara bersamaan dan berulang dan data yang telah menumpuk lama, sehingga perlu kiranya digunakan penambangan data, untuk mendapatkan pengetahuan. Data yang penulis gunakan adalah data mahasiswa baru STIE-GK Muara Bulian secara random sebanyak 100 orang. Metode penelitian yang penulis gunakan adalah dengan menghitung data menggunakan algoritma apriori kemudian dilakukan uji coba terhadap software data mining Weka. Adapun hasil penelitian ini menunjukkan, secara parsial Variabel SMK dan jurusan TKJ adalah penyumbang mahasiswa baru terbanyak dengan tingkat confidence mencapai 13%.

2.7 Kerangka Pemikiran

Untuk melakukan prediksi terhadap penjualan tiket pesawat maka perlu dilakukan analisis data. Metode algoritma apriori adalah metode yang sering digunakan untuk melakukan prediksi penjualan ataupun analisis data dengan keakuratan yang tinggi. Data yang akan dianalisis adalah data penjualan tiket

pesawat yang domestik. Data penjualana tersebut akan diolah menggunakan aplikasi Tanagra untuk mendapatkan prediksi yang akurat. Dari data penjualan yang didapat an ukan di susun dan diolah dalam klasifikasi perbulan untuk dijadikan data olahan yang akan di masukkan ke dalam aplikasi Tanagra. Dari ahasil aplikasi Tanagra akan didapatkan tingakt akurasi dengan persentase paling tertinggi. Dalam proses pengolahan data akan ditentukan tingkat akurasi atau persentasi yang akan digunakan sebagai tola ukur. Dengan demikian akan menghasilkan output sebagai nilai prediksi yang diprediksikan aakn naik atau turun di bulan tahun berikutnya Dari uraian latar belakang dan metode yang dipakai maka kerangka pemikiran dari penelitian ini adalah:



Gambar 2. 4 Kerangka pemikiran

Sumber : Penelitian 2021