BAB II

LANDASAN TEORI

2.1 Penelitian Terdahulu

Tabel 2.1 Artikel Jurnal Penelitian Terdahulu

Artikel Jurnal 1	
Nama Penulis	I Made Dwi, Gusti Made, Ni Kadek Dwi
Judul	Analisa Pola Belanja Konsumen serta Prediksi Stok Barang Berbasis Web
Nama Jurnal	JEPIN (Jurnal Edukasi dan Penelitian Informatika)
Tahun	Desember 2023
Metode Penelitian	Asosiasi dengan algoritma FP-Growth dan Apriori, serta prediksi menggunakan algoritma Regresi Linear dan SVR
Hasil Pembahasan	Algoritma FP-Growth dan Apriori berhasil mengidentifikasi 11 aturan asosiasi dari 2568 transaksi dengan nilai support minimum 0,02 dan confidence minimum 0,3, yang dapat dimanfaatkan untuk strategi bisnis seperti pengelompokan barang dan promosi. Pada prediksi stok barang, metode SVR memiliki akurasi lebih baik dibandingkan regresi linier, dengan rata-rata tingkat kesalahan (MAPE) sebesar 11,51% untuk SVR dan 12,09% untuk regresi linier. Hasil asosiasi dan prediksi ini mendukung pengambilan keputusan yang lebih efektif untuk pengelolaan stok dan pengoptimalan bisnis di masa mendatang.

Artikel Jurnal 2	
Nama Penulis	Muliati Badaruddin, Santoso
Judul	Prediksi Persediaan Perlengkapan Hewan Peliharaan Pada Toko Poopy Cat Store Menggunakan Algoritma Apriori
Nama Jurnal	Jurnal Media Informatika Budidarma
Tahun	Juli 2021
Metode Penelitian	Metode data mining dengan algoritma Apriori untuk menganalisis data transaksi dan menemukan pola pembelian barang yang sering dibeli bersama-sama.
Hasil Pembahasan	Algoritma Apriori mampu mengidentifikasi tiga item yang sering dibeli secara bersamaan, yaitu *Bolt 10 gr*, kandang, dan mangkok, dengan nilai confidence tertinggi sebesar 63%. Hasil analisis ini dapat membantu toko dalam pengelolaan stok barang, memastikan ketersediaan barang yang paling diminati, dan meningkatkan strategi tata letak barang untuk menarik minat pembeli. Penelitian ini menunjukkan manfaat algoritma Apriori dalam mendukung keputusan bisnis secara cepat dan tepat.

Artikel Jurnal 3	
Nama Penulis	Rosanti Daeli, Eka Rahayu. Edrian Hadinata
Judul	Analisis Prediksi Persediaan Stok Barang Pada Toko Santi Fotokopi Menggunakan Algoritma Apriori Berbasis Website
Nama Jurnal	Indonesia Journal Computer Science
Tahun	2023
Metode Penelitian	Algoritma Apriori, dengan nilai minimum support sebesar 10% dan

	minimum confidence sebesar 60%.
Hasil Pembahasan	Hasil analisis menghasilkan 6 aturan asosiasi dengan nilai confidence tertinggi sebesar 71,43%, yaitu pada kombinasi *Kertas Binder* dan *Binder*. Dengan menggunakan algoritma Apriori, penelitian ini membantu toko Santi Fotokopi memprediksi stok barang yang sering terjual, sehingga dapat mengurangi risiko penumpukan barang yang jarang dibeli dan meningkatkan efisiensi manajemen stok.

Artikel Jurnal 4	
Nama Penulis	Kristoko Dwi Hartomo, Sri Yulianto, Rahmat Abadi Suharjo.
Judul	Prediksi Stok dan Pengaturan Tata Letak Barang Menggunakan Kombinasi Algoritma Triple Exponential Smoothing dan FP-Growth.
Nama Jurnal	Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK).
Tahun	2020
Metode Penelitian	- Algoritma FP-Growth: Digunakan untuk menemukan pola perilaku konsumen dan aturan asosiasi dengan nilai lift ratio tertinggi. - Triple Exponential Smoothing: Digunakan untuk prediksi stok barang.
Hasil Pembahasan	Penelitian ini menghasilkan 12 aturan asosiasi menggunakan algoritma FP-Growth, dengan aturan asosiasi antara teh dan gula memiliki lift ratio tertinggi sebesar 6,131, menunjukkan hubungan kuat antar barang. Prediksi stoke menggunakan algoritma Triple Exponential Smoothing menunjukkan akurasi tinggi dengan nilai MAPE sebesar 11,7% (akurasi 88,3%). Hasil penelitian digunakan untuk menyusun tata letak barang secara strategis dan memprediksi kebutuhan stok barang, yang

	diharapkan dapat meningkatkan efisiensi pengelolaan dan penjualan.

Artikel Jurnal 5	
Nama Penulis	Andre Setiawan, Farica Perdana Putri
Judul	Implementasi Algoritma Apriori untuk Rekomendasi Kombinasi Produk Penjualan
Nama Jurnal	ULTIMATICS
Tahun	2020
Metode Penelitian	Market Basket Analysis, Association Rules, Algoritma Apriori
Hasil Pembahasan	Sistem rekomendasi yang dibangun berhasil mengidentifikasi aturan asosiasi kombinasi produk dengan nilai lift ratio sebesar 1.18, yang menunjukkan kekuatan hubungan antar item yang valid. Evaluasi usability menggunakan USE questionnaire menunjukkan hasil yang sangat baik (usefulness 90,83%, ease of use 89,09%, ease of learning 95%, dan satisfaction 90,94%), sehingga sistem ini layak digunakan untuk mendukung strategi pemasaran toko daring.

UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

Artikel Jurnal 6	
Nama Penulis	Mayland Trifena, Khoirunnisa Hamidah , Yuyun Umaidah , Apriade Voutama
Judul	Implementasi Algoritma Apriori untuk Menentukan Paket Bundel dalam Penjualan Toko Swalayan XYZ
Nama Jurnal	Journal Sensi
Tahun	2024
Metode Penelitian	Algoritma Apriori
Hasil Pembahasan	Analisis menunjukkan bahwa Air Mineral adalah item yang paling sering dibeli pelanggan, dan mayoritas pelanggan yang membeli produk makanan juga menambahkan Air Mineral ke keranjang belanja mereka . Dari temuan ini, direkomendasikan agar Toko Swalayan XYZ merancang paket bundel yang menggabungkan berbagai produk makanan dengan Air Mineral—seperti paket roti isi + air mineral—sebagai strategi promosi. Dengan memanfaatkan pola pembelian ini, diharapkan penjualan meningkat dan pelanggan memperoleh nilai tambah melalui harga bundel yang lebih kompetitif.

Artikel Jurnal 7	
Nama Penulis	Thedorus Agung Gumilang, Pipin Farida Ariyani
Judul	Algoritma Apriori Dalam Implementasi Data Mining Untuk Pembuatan Paket Penjualan di Mesha Petshop
Nama Jurnal	SENAFTI(Seminar NAsional Mahasiswa Fakultas Teknologi Informasi)
Tahun	2023
Metode Penelitian	Association Rule Mining dengan Algoritma Apriori

Hasil Pembahasan	Analisis menggunakan algoritma Apriori menghasilkan aturan asosiasi yang dapat dijadikan dasar rekomendasi bundling produk. Aturan-aturan yang terbentuk, yang diukur dengan nilai support, confidence, dan lift ratio (nilai lift ratio minimal > 1, contohnya 1.18), menunjukkan bahwa kombinasi produk yang dihasilkan memiliki hubungan yang kuat dan dapat diandalkan. Selain itu, evaluasi usability menunjukkan bahwa sistem memiliki skor yang sangat baik (usefulness 87,5%, ease of use 89,45%, ease of learning 91%, dan satisfaction 88,57%), sehingga sistem tersebut dianggap layak untuk membantu toko daring dalam menyusun paket bundling produk guna meningkatkan penjualan dan kepuasan pelanggan.

Artikel Jurnal 8	
Nama Penulis	Ping Zhanghttps, Yiqiao Jia, Youlin Shang
Judul	Research and application of XGBoost in imbalanced data
Nama Jurnal	SAGE journals
Tahun	2022
Metode Penelitian	- pendekatan data level : SVM-SMOTE - Pendekatan Algoritma level : XGBoost - Optimasi Parameter : Bayesian Optimization - Evaluasi Kinerja : AUC & G-mean
Hasil Pembahasan	Penelitian ini menggabungkan teknik over-sampling (SVM–SMOTE) dan under-sampling (EasyEnsemble) dengan model XGBoost yang dioptimasi menggunakan Bayesian Optimization untuk mengatasi masalah data yang tidak seimbang. Dengan pendekatan ini, model mampu meningkatkan performa klasifikasi secara

	signifikan—ditunjukkan oleh peningkatan nilai AUC dan G-mean dibandingkan model XGBoost standar dan metode lain seperti RUSBoost, CatBoost, serta LightGBM. Keunggulan penggunaan XGBoost di sini antara lain: ● **Kemampuan Adaptasi:** XGBoost mampu menangani kompleksitas data dan menangkap pola non-linear secara efektif. ● **Efisiensi dan Skalabilitas:** Algoritma ini didesain untuk efisiensi komputasi dan dapat diskalakan untuk dataset besar. ● **Regularisasi yang Kuat:** Dengan pengaturan regularisasi (L1/L2) yang mencegah overfitting, model dapat mempertahankan performa meski dalam kondisi data yang tidak seimbang. ● **Integrasi yang Fleksibel:** XGBoost dapat dengan mudah digabungkan dengan teknik sampling dan optimasi parameter, sehingga menghasilkan model yang lebih robust dan akurat.

Artikel Jurnal 9	
Nama Penulis	Zeravan Arif Ali, Ziyad H. Abduljabbar, Hanan A. Tahir, Amira Bibo Sallow, Saman M. Almufti
Judul	Exploring the Power of eXtreme Gradient Boosting Algorithm in Machine Learning: a Review
Nama Jurnal	Academic Journal of Nawroz University (AJNU)
Tahun	2023
Metode Penelitian	XGBoost
Hasil Pembahasan	Pembahasan dari penelitian tersebut, bahwa XGBoost berkat optimasi paralelisasi, regularisasi (L1/L2), dan teknik pruning, terbukti sangat

	cepat, akurat, dan scalable untuk menangani masalah data kompleks. Dengan kemampuannya yang fleksibel dan integrasi dalam ensemble learning, XGBoost menjadi algoritma unggulan di berbagai aplikasi machine learning—baik untuk prediksi, klasifikasi, maupun optimasi sistem—dengan interpretabilitas yang tinggi untuk memahami kontribusi fitur secara mendalam.

Artikel Jurnal 10	
Nama Penulis	Candice Bentéjac, Anna Csörgő, Gonzalo Martínez-Muñoz
Judul	A comparative analysis of gradient boosting algorithms
Nama Jurnal	Artificial Intelligence Review
Tahun	2020
Metode Penelitian	XGBoost, LightGBM (dengan mode GOSS), CatBoost (dengan mode Ordered), Gradient Boosting Standar, dan Random Forest.
Hasil Pembahasan	Penelitian ini menyatukan evaluasi kinerja beberapa algoritma ensemble (XGBoost, LightGBM, CatBoost, gradient boosting standar, dan random forest) pada 28 dataset UCI dengan menggunakan stratified 10-fold cross-validation serta optimasi hyper-parameter melalui grid search. Secara keseluruhan, metode tuning umumnya meningkatkan akurasi dan AUC dibandingkan dengan konfigurasi default; misalnya, tuned Ordered CatBoost dan tuned XGBoost menduduki peringkat teratas dalam rata-rata performa. Khusus untuk XGBoost, penelitian menunjukkan bahwa meskipun LightGBM lebih cepat, XGBoost memberikan keseimbangan yang sangat baik antara akurasi dan kecepatan pelatihan (sekitar 1,6 kali lebih lambat daripada LightGBM) dengan keunggulan dalam hal regularisasi, sparsity-aware split finding, dan fleksibilitas yang memungkinkan model untuk menangani kompleksitas data secara efektif

	sehingga menghasilkan generalisasi yang tinggi. Dengan demikian, XGBoost terbukti menjadi pilihan yang andal dan kompetitif, terutama bila dikombinasikan dengan proses tuning yang tepat untuk meningkatkan performa prediksi.

2.2 Teori Mengenai Topik Skripsi

2.2.1 Chewee

Chewee merupakan platform marketplace online yang menyediakan berbagai macam kebutuhan hewan peliharaan, yang menjual sekitar 4000+ produk seperti makanan, aksesoris, produk kesehatan, yang terbagi berdasarkan 7 kategori hewan peliharaan dan terdapat lebih dari puluhan *brand pet supplies* yang telah melakukan kerjasama dengan Chewee. Selain membeli produk atau memesan layanan di Chewee, para pelanggan juga diperbolehkan untuk membuka toko dan menjual produk-produk yang *pet supplies* di Chewee. Seluruh aktivitas yang ingin dilakukan dalam Chewee, bisa diakses melalui chewee.co.id [9]

2.2.2 Analisis Asosiasi

Analisis asosiasi atau yang dikenal dengan *Association Rules Mining* merupakan teknik analisis dengan membentuk aturan asosiasi hubungan antara satu sama item. Kemampuan tersebut sangat membantu misalkan seperti untuk *market basket analysis* , yang memiliki tujuan untuk mengidentifikasi pola pembelian produk yang dilakukan oleh konsumen. Misalkan, pelanggan yang sering membeli roti maka akan membeli selai juga. Pola tersebut dikenal sebagai aturan asosiasi, yang memberikan wawasan yang dapat ditindaklanjuti untuk meningkatkan manajemen inventaris, meningkatkan kepuasan pelanggan, maupun mendorong penjualan.

Sejarah teknik aturan asosiasi tersebut dimulai pada awal tahun 1990-an ketika Agrawal, Imielinski, dan Swami (1993) memperkenalkan konsep

penambangan asosiasi antar item dalam basis data yang besar. Teknik tersebut menghasilkan pengembangan algoritma Apriori oleh Agrawal dan Srikant pada tahun 1994, yang menjadi landasan karena efisiennya dalam menghasilkan *frequent itemset*. Lalu disusul dengan adanya algoritma FP-Growth oleh Han et al. (2000) dan algoritma ECLAT oleh Zaki (2000), yang semakin menyempurnakan proses tersebut untuk membuat analisis lebih terstruktur dan efisien pada data yang lebih besar[12].

2.2.3 Analisis Prediksi

Analisis prediksi merupakan analisis yang dilakukan untuk menemukan pola yang menarik dan bermakna dalam data, pengenalan pola, statistik, pembelajaran mesin, kecerdasan buatan, dan penggalian data. Perbedaan analisis prediksi dengan analisis lainnya yaitu analisis prediksi digerakkan oleh data, yang berarti algoritma yang mendapatkan karakteristik utama model dari data itu sendiri jadi bentuknya bukan dari asumsi yang dibuat oleh analis. Maka dari itu, algoritma berbasis menginduksi model dari data. Proses induksi tersebut mencakup identifikasi variabel yang akan dimasukkan dalam model, parameter yang mendefinisikan model, bobot atau koefisien dalam model.

Algoritma analisis prediksi juga akan mengotomasikan proses menemukan pola dari data. Algoritma induksi yang kuat tidak hanya menemukan koefisien atau bobot untuk model, namun juga membentuk model tersebut. Misalnya algoritma *decision tree*, yang mempelajari kandidat input yang paling baik dalam memprediksikan variabel target selain mengidentifikasi nilai variabel mana yang akan digunakan dalam membangun prediksi. Algoritma prediksi juga mengotomatiskan proses transformasi variabel input, sehingga dapat digunakan secara efektif dalam model prediktif. Misalnya, jika ada seratus variabel yang merupakan kandidat input untuk model yang dapat/harus ditransformasikan untuk menghilangkan kemiringan data, maka dapat melakukan dengan beberapa *software* analisis prediksi [13].

2.2.4 Sistem Rekomendasi

Sistem rekomendasi adalah jenis sistem penyaringan informasi yang dirancang untuk memprediksi dan merekomendasikan item seperti produk, film, mudik, dan lainnya. Prediksi tersebut didasarkan pada perilaku konsumen maupun preferensi. Tujuan utama sistem rekomendasi untuk meningkatkan pengalaman pengguna, meningkatkan keterlibatan dan, memfasilitasi proses pengambilan keputusan. Hal tersebut dapat diterapkan di berbagai bidang seperti *e-commerce.* hiburan, maupun media sosial, maka dari itu sistem rekomendasi memiliki peran penting dalam penelitian teoritis(akademis) dan aplikasi praktis (industri). Sistem rekomendasi telah berkembang dengan pesat juga, dengan munculnya data besar dan kemajuan kecerdasan buatan. Pengguna berinteraksi dalam platform digital, mereka menghasilkan sejumlah besar data yang dapat dimanfaatkan untuk membuat rekomendasi yang tepat dan personal. Kemampuan tersebut selain untuk meningkatkan kepuasan pengguna, juga dapat meningkatkan kemungkinan pengguna dapat menemukan konten yang baru dan relevan. Selain itu, sistem rekomendasi juga disintegrasi de dalam bidang-bidang yang sedang berkembang seperti pendidikan, yang membantu menyesuaikan pengalaman belajar dengan kebutuhan siswa secara individu. Pengembangan LLM*(Large Language Models*) semakin meningkatkan sistem rekomendasi untuk memahami dan memproses data dalam jumlah besar, sehingga dapat membentuk rekomendasi yang lebih canggih.

Sistem rekomendasi dirintis dari Ellen Rich tahun 1979 yang menyarankan buku berdasarkan preferensi pengguna, lalu ada Jussi Karlgren yang mengonseptualisasikan rak buku digital pada tahun 1990 yang kemudian ide-ide tersebut dikembangkan oleh peneliti SICA, MIT, dan Bellcore. Metode tradisional sistem rekomendasi dapat dikategorikan ke dalam penyaringan kolaboratif, penyaringan berbasisi konten, dan pendekatan hibrida yang bertujuan untuk meningkatkan pengalaman pengguna. Penyaringan tersebut didasarkan pada gagasan bahwa pengguna dengan preferensi yang sama akan memiliki selera yang sama juga di masa depan. Ada juga penyaringan berbasis

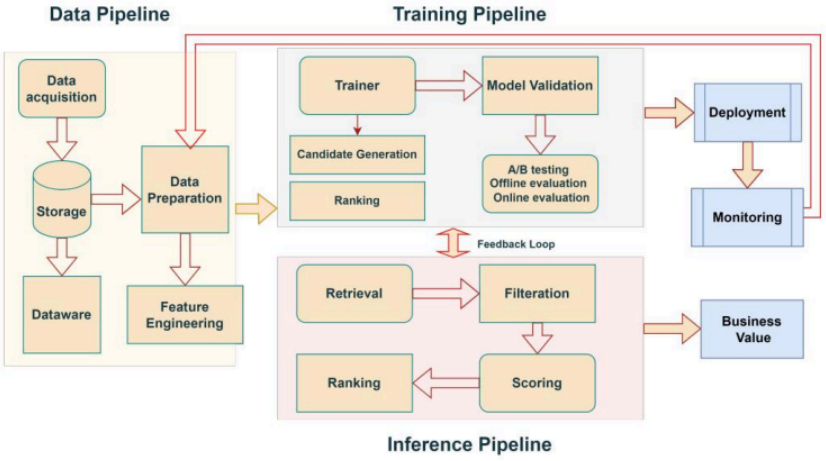
konten yang merekomendasikan item berdasarkan preferensi masa lalu pengguna dan karakteristik item. Dengan perkembangan adanya metode *machine learning* seperti algoritma KNN, *Deep Neural Network,* dan *Natural Language Processing* (NLP) telah meningkatkan sistem rekomendasi dengan memberikan rekomendasi yang lebih tepat.

Secara matematis, sistem rekomendasi direpresentasikan sebagai fungsi f yang memprediksi kegiatan item i bagi pengguna u, yang dilambangkan sebagai $\widehat{r}_{ui}$, memperkirakan seberapa besar pengguna u lebih menyukai item i.

$$\widehat{r}_{ui} = f(u, i;\ \Theta)$$

Rumus 2.1 Rumus Sistem Rekomendasi

Di mana Θ mewakili parameter model, yang dipelajari dari data. Dalam konteks RS, istilah $\widehat{r}_{ui}$ mewakili prediksi peringkat atau utilitas yang akan diberikan pengguna pada suatu item. Prediksi ini digunakan untuk merekomendasikan item yang mungkin menarik bagi pengguna. Kerangka kerja umum RS diilustrasikan dalam gambar 2.1 . Siklus hidup model berbasis data dalam RS dimulai dengan akuisisi data, diikuti oleh penyimpanan dan persiapan. Ini mengarah pada rekayasa fitur, yang membentuk dasar jalur data. [14]



Gambar 2.1 Alur Proses Sistem Rekomendasi [15]

2.3 Teori Mengenai Framework/Algoritma yang Digunakan

2.2.1 CRISP-DM (*Cross-Industry standard Process for Data Mining)*

CRISP-DM merupakan model yang diterbitkan pada tahun 1999 dan merupakan kerangka kerja paling populer untuk menjalankan proyek berbasisi sains data. Kerangka tersebut memberikan deskripsi mengenai siklus hidup sains data (alur kerja dalam proyek yang berfokus pada data). Tahapan CRISP-DM sendiri terbagi menjadi 6 tahapan, yaitu :

1). Business Understanding

Business Understanding merupakan tahapan untuk memahami proyek maupun kebutuhan bisnis terlebih dahulu, juga menentukan bagaimana mendapatkan data dan membangun model, serta mencocokkan model pembelajaran tersebut dengan kebutuhan maupun tujuan proyek/bisnis.

2). Data Understanding

Setelah itu, ada tahap *data understanding*, yaitu mengidentifikasi data yang telah dikumpulkan. Tahap ini memberikan dasar yang kuat untuk memulai proses analitik penelitian dan menemukan potensi masalah dalam data. Tahap ini harus dilakukan secara cermat untuk mendapatkan pemahaman tentang data. Hasil dari ringkasan data dapat membantu memastikan apakah data yang dikumpulkan memenuhi tujuan awal pengolahan model.

3). Data Preparation

Setelah proses pemahaman data selesai, langkah berikutnya adalah *data preparation* Ini termasuk memilih parameter yang akan dianalisis, mengubah parameter tertentu, dan memperbaiki masalah data seperti data null, duplikat, hilang, atau format yang tidak sesuai.

4). Modeling

Data yang telah disiapkan, maka dapat digunakan untuk tahap *modeling*. *Modeling* merupakan tahapan untuk membangun pembelajaran model, sesuai dengan kebutuhan maupun tujuan bisnis yang sudah ditetapkan di awal. Sebagian besar model dibuat bisa secara prediktif maupun deskriptif. Pada

tahap tersebut akan dilakukan seperti metode statistik maupun *machine learning* untuk menentukan teknik *data mining*, lalu algoritma *data mining* yang akan digunakan. Contoh *modeling* yang dapat diterapkan seperti, *classification, ranking, clustering, scoring, relation,* dsb.

5). Evaluation

Pada tahap kelima, model pembelajaran yang telah diterapkan sebelumnya dievaluasi. Tujuan dari tahap ini adalah untuk memastikan bahwa model yang telah dibangun telah mencapai tujuan yang ingin dicapai dan memastikan bahwa model telah berjalan sesuai dengan kebutuhan bisnis.

6). Deployment

Dan tahap terakhir yaitu *deployment*, merupakan tahapan mengimplementasikan model yang sudah dibuat sebelumnya menjadi suatu bentuk yang dapat digunakan *insight* nya untuk perusahaan. Tahap ini merupakan penggabungan antar semua tahap yang telah dibuat sebelumnya dnegan tambahan adanya bentuk nyata yang interaktif [16].

Gambar 2.2 Proses Kerja CRISP-DM [17]

2.2.2 Apriori

Dalam data mining, algoritma apriori menggunakan aturan asosiatif untuk mengumpulkan data dan menentukan hubungan antara kombinasi item . Selain itu, algoritma ini memiliki kemampuan untuk menentukan jenis barang apa yang mungkin dibeli konsumen dalam setiap pembelian mereka dan juga dapat digunakan untuk membuat keputusan tentang stok barang. Aturan asosiasi yang dimaksud dibuat melalui proses perhitungan nilai dukungan dan kepercayaan untuk suatu hubungan item.

Algoritma apriori melakukan dua prosedur umum :

1. Penggabungan (Step Join): Menggabungkan semua set item sampai tidak ada lagi kombinasi vang dapat.
2. Pemangkasan (Prune Step): Memangkas dari masing-masing itemset sesuai dengan nilai minimum support yang ditukar [18].

Pada tahun 1994, Agrawal dan Srikant membuat algoritma Apriori, yang didasarkan pada aturan asosiasi Boolean, untuk menemukan itemset sering dalam dataset (Qoniah & Priandika, 2020). Algoritma ini menggunakan nilai minimum support sebagai kriteria pemilihan kandidat untuk menemukan itemset berikutnya. Proses algoritma Apriori adalah sebagai berikut:

1. Menentukan minimum support
2. Memindai database untuk menemukan kandidat 1-itemset, yang terdiri dari satu itemset, dan menghitung nilai support untuk setiap itemset. Itemset yang tidak memenuhi nilai minimum support akan dihapus, dan itemset yang memenuhi nilai minimum akan dimasukkan ke dalam 1-itemset.
3. Untuk iterasi berikutnya, itemset yang lolos dalam 1-itemset akan digunakan. 1 itemset akan melakukan operasi join antar itemset untuk menghasilkan kandidat 2-itemset, yang terdiri dari dua itemset. Kemudian, nilai support dihitung, dan itemset yang tidak memenuhi nilai minimum akan dieliminasi, sedangkan itemset yang memenuhi nilai support akan dimasukkan ke dalam 2-itemset.

4. Ulangi langkah ini hingga tidak ada lagi k-itemset yang terbentuk.
5. Tentukan nilai kepercayaan minimum
6. Untuk semua k-itemset yang telah terbentuk atau memenuhi nilai kepercayaan minimum, aturan asosiasi akan dibuat dengan menghitung nilai kepercayaan. Item yang tidak memenuhi nilai kepercayaan minimum akan dihapus [19].

Rumus-rumus dalam algoritma Apriori sebagai berikut ini :

1. *Antecedents* : Merupakan kumpulan intemset yang menjadi kondisi pemicu (Misalkan sebagai = A)
2. *Consequents* : Merupakan kumpulan itemset yang diprediksi muncul jika *antecedents* (Misalkan sebagai = B

$$A \Rightarrow B$$

3. *Support :* Mengukur seberapa sering itemset muncul dalam basis data transaksi. Dihitung dengan rumus:

$$S(A) = \frac{jumlah\ transaksi\ yang\ mengandung\ A}{total\ jumlah\ transaksi}$$

$$S(A,B) = \frac{total\ jumlah\ transaksi\ A\ dan\ B}{total\ jumlah\ \ transaksi}$$

Rumus 2.2 Rumus *Suport* Analisis Asosiasi

4. *Confidence :* Mengukur seberapa sering item Y muncul dalam transaksi yang sudah mengandung item X. Dihitung dengan rumus:

$$C(A \rightarrow B) = \frac{jumlah\ transaksi\ yang\ mengandung\ A\ dan\ B}{jumlah\ transaksi\ yang\ mengandung\ A}$$

Rumus 2.3 Rumus *Confidence* Analisis Asosiasi

5. *Lift :* Mengukur seberapa besar peningkatan kemungkinan pembelian item Y ketika item X dibeli. Dihitung dengan rumus:

$$L(A \rightarrow B) = \frac{C(A \rightarrow B)}{S(B)} \text{ [20]}$$

Rumus 2.4 Rumus *Lift*Analisis Asosiasi

2.2.3 XGBoost

Metode XGBoost merupakan pengembangan dari algoritma gradient tree boosting yang berbasis ensemble, dirancang untuk secara efektif menangani kasus pembelajaran mesin berskala besar. Algoritma ini pertama kali diperkenalkan oleh Dr. Tianqi Chen dari University of Washington pada tahun 2014. XGBoost menjadi pilihan utama karena menawarkan sejumlah fitur tambahan yang dapat mempercepat proses komputasi serta meminimalkan risiko overfitting. Metode ini dapat diaplikasikan pada berbagai tugas pembelajaran mesin, termasuk klasifikasi, regresi, dan ranking. Secara teknis, XGBoost merupakan algoritma berbasis ensemble yang menggunakan sekumpulan pohon regresi (CART). Keberhasilan XGBoost sangat bergantung pada kemampuannya untuk beradaptasi dengan berbagai situasi, yang dicapai melalui penyempurnaan algoritma dari metode sebelumnya [21].

Sebagai metode ensemble, XGBoost menggabungkan hasil prediksi dari beberapa pohon regresi untuk menghasilkan prediksi akhir. Rumus prediksi akhir $\widehat{yi}$ untuk data x_i diberikan sebagai berikut:

$$\widehat{yi} = \sum_{t=1}^{T} f_t(x_i), \; f_t \in F$$

Rumus 2.5 Rumus *XGBoost*

Dengan keterangan sebagai berikut :

$\widehat{yi}$: prediksi untuk data ke-i.

$f_t(x_i)$: hasil prediksi dari pohon regresi ke-t untu data x_i

T : total jumlah pohon dalam model XGBoost.

F : kumpulan semua pohon regresi.

Model XGBoost dirancang dengan pendekatan reguler untuk membangun struktur pohon regresi, yang memberikan kinerja lebih optimal sekaligus mengurangi kompleksitas model, sehingga menghindari overfitting. Prediksi akhir dari model XGBoost merupakan hasil penjumlahan output dari setiap pohon regresi [22]. Algoritma berbasis pohon keputusan ini bekerja dengan baik pada data yang memiliki fitur kategorikal dan relatif tidak terpengaruh oleh distribusi kelas yang tidak seimbang. Selain itu, XGBoost mendukung komputasi paralel, yang memungkinkan seluruh inti CPU digunakan secara optimal untuk mempercepat proses pelatihan [23].

Keunggulan-keunggulan tersebut menjadikan XGBoost sebagai salah satu algoritma yang sangat populer di kalangan penggiat dan praktisi pembelajaran mesin. Kecepatan komputasi XGBoost yang tinggi sebagian besar disebabkan oleh kemampuannya memanfaatkan perangkat keras secara efisien melalui paralelisasi. Dalam implementasi algoritma ini, penyesuaian hiperparameter memegang peran penting untuk mengontrol aspek-aspek seperti laju pembelajaran dan pengaruhnya terhadap kinerja model. Oleh karena itu, proses optimasi hiperparameter menjadi langkah esensial dalam membangun model XGBoost yang berkinerja maksimal [24].

2.4 Teori Mengenai *Tools/Software* yang Digunakan

2.4.1 Python

Python adalah bahasa pemrograman serba guna yang dapat digunakan untuk menangani masalah numerik. Meskipun demikian, Python dapat secara efektif menangani masalah numerik dan visualisasi data ketika dipasangkan dengan pustaka seperti NumPy, Seaborn, Matplotlib, dan Pandas [25]. Ini tersedia di berbagai platform termasuk Windows, Linux, dan macOS. Karena sifatnya yang secara inheren mudah diingat bersama dengan fitur berorientasi objek, Python digunakan untuk membuat dan menggambarkan aplikasi dengan cepat.

Python juga berkaitan dengan pemrosesan bahasa alami, pengembang web menggunakan untuk membuat situs web dinamis, dan administrator sistem menggunakan untuk tujuan lain [26]. Bahasa pemrograman Python sendiri diciptakan oleh Guido van Rossum pada akhir tahun 1980-an. Python pun telah menjadi salah satu bahasa pemrograman terjemahan yang paling populer, bersama dengan Perl, Ruby, dan lainnya. Python dan Ruby telah menjadi sangat populer sejak sekitar tahun 2005 untuk membangun situs web menggunakan berbagai kerangka kerja web mereka, seperti Rails (Ruby) dan Django. (Python). Bahasa-bahasa tersebut sering disebut sebagai bahasa skrip, karena dapat digunakan untuk dengan cepat menulis program kecil, atau skrip untuk mengotomatiskan tugas-tugas lain. Dalam dua dekade terakhir, Python telah beralih dari bahasa pemrograman ilmiah yang mutakhir, dan menjadi salah satu bahasa yang paling penting untuk ilmu data, pembelajaran mesin, dan pengembangan perangkat lunak umum di dunia akademis dan industri [27].

Python 2.0 dirilis pada 16 Oktober 2000, dengan banyak fitur baru yang signifikan, termasuk pengumpul sampah yang mendeteksi siklus (selain penghitung referensi) untuk manajemen memori dan dukungan untuk Unicode. Namun, perubahan yang paling penting adalah pada proses pengembangan itu sendiri, dengan peralihan ke proses yang lebih sederhana dan didukung oleh komunitas. Python 3.0, sebuah rilis besar yang tidak kompatibel dengan versi sebelumnya, dirilis pada 3 Desember 2008 setelah melalui periode pengujian yang panjang.. Dalam pembuatan aplikasi di Facebook, Twitter, CERN, Mechanical Light & Enchantment, dan NASA, Python termasuk di dalamnya. Para insinyur perangkat lunak yang berpengalaman dapat melakukan hal-hal menakjubkan dengan Python, tetapi kekuatannya adalah memberdayakan mereka untuk menciptakan inovasi dan menyelesaikan masalah menarik lebih cepat daripada dengan bahasa lain yang lebih kompleks dan memiliki kurva pembelajaran yang lebih curam [28].

2.4.2 Google Collaboratory

Google Collaboratory adalah platform untuk mengimplementasikan dan menjalankan model pembelajaran mesin dengan Python, yang menggabungkan fitur Jupyter Notebooks dengan VM yang dihosting Google dan perangkat keras canggih. Karena Colab Notebook di hosting di Jupyter Notebook, buku catatan tersebut mengikuti alur kerja buku catatan umum yang menggabungkan penjelasan naratif dan demonstrasi interaktif berbasis kode [29].Google Collaboratory menyediakan runtime Python 2 dan 3 yang dikonfigurasikan sebelumnya dengan pustaka pembelajaran mesin dan kecerdasan buatan terkemuka seperti TensorFlow, Matplotlib, dan Keras. Layanan Google ini menawarkan runtime yang dipercepat GPU, yang juga sepenuhnya terdiri dari perangkat lunak yang disebutkan di atas. Infrastruktur Google Collaboratory di hosting di Google Cloud Platform [30].

Ada beberapa manfaat menggunakan Google Collaboratory.

- Pengembang dapat dengan mudah melihat dokumentasi kode sumber dan fitur yang digunakan. Ini akan membantu dalam memahami cara kerja perpustakaan yang diimpor dan memungkinkan untuk melakukan debug lebih cepat.
- Memungkinkan pengembang untuk membagi program menjadi beberapa bagian sementara program terus berjalan. Fitur ini memungkinkan program menjadi lebih terorganisir dan memungkinkan pengembang menjalankan program secara mandiri, memungkinkan mereka dengan cepat mengidentifikasi lokasi kesalahan dalam suatu program.
- Platform ini juga menawarkan GPU berbasis cloud yang membantu pengembang menjalankan kode mereka dengan lebih lancar. Google Collaboratory membantu mencegah masalah penyebab kesalahan kode dengan menyediakan konfigurasi Python berbasis browser yang sama di semua komputer.