

## **BAB II**

### **TINJAUAN PUSTAKA**

#### **2.1 Tinjauan Teori Dasar**

Pada bab ini, peneliti menjelaskan teori-teori terkait yang digunakan dalam penelitian ini.

##### **2.1.1 Tinjauan Teori Umum**

Argumen dasar adalah argumen yang menyatakan bahwa setiap kali sebuah proposisi akurat, proposisi tersebut benar secara menyeluruh. Pernyataan ini berlaku untuk semua lokasi, periode, kondisi, dan isu. Pernyataan ini disesuaikan dengan studi yang akan dilakukan.

###### **2.1.1.1 Datamining**

Ruang untuk penyesuaian adalah proses pengumpulan dan analisis informasi sebelumnya untuk menemukan simetri, struktur, atau korelasi dalam data yang sangat besar. Hasil dari pengumpulan data tersebut digunakan untuk meningkatkan keputusan dan kebijakan (Tana, Marisa, and Wijaya 2018).

Berdasarkan (Hasugian 2020) Ruang untuk penyesuaian adalah eksplorasi dan pemeriksaan set data yang sangat besar untuk menemukan korelasi yang menarik dengan tujuan memperoleh data dan pemahaman yang andal dan berpotensi efektif dari arsip yang sangat besar.

Lalu (Ikhwan 2018) Ruang untuk penyesuaian adalah metode untuk mengekstraksi wawasan dari data yang heterogen dengan mengidentifikasi pola dan hubungan. Fasilitas berada, juga dikenal sebagai penggalian informasi relasional (KDD), adalah proses yang melibatkan pengumpulan dan analisis informasi sebelumnya untuk mendeteksi proses, tema, atau korelasi dalam data

dalam jumlah besar. Hasil dari ekstraksi data dapat digunakan untuk membantu pengambilan keputusan. Seiring berkembangnya KDD, penggunaan pemikiran analitis berkurang karena sekarang menjadi komponen dari data mining.

#### **2.1.1.2 Knowledge Discovery in Database (KDD)**

berdasarkan (Yanto and Kesuma 2017) Ekstraksi Informasi dalam Basis Data (KDD) tampaknya merupakan operasi yang terdiri dari pengumpulan dan pemanfaatan informasi sebelumnya untuk tujuan mempelajari informasi saat ini, khususnya tujuan untuk menentukan nilai-nilai yang diterapkan untuk mendeteksi pola-pola yang lebih kecil yang bermakna, di samping bentuk data yang sangat besar.

Lalu (Abarca 2021) Kecerdasan bisnis dan Jawaban Pertanyaan dalam Basis Data (KDD) adalah frasa yang sering digunakan secara bersamaan untuk merujuk pada metode menemukan data yang berguna dari basis data yang sangat besar. Secara teknis, kedua istilah tersebut memiliki arti yang berbeda, meskipun keduanya saling berkaitan. Kecerdasan bisnis merupakan salah satu langkah di dalam proses penemuan pengetahuan.

Langkah-langkah berikut ini merupakan contoh dari prosedur penggalian informasi pada prinsipnya (Elisa 2018) :

1. *Data Cleaning*

Pemeliharaan rutin dilakukan pada awal metode untuk meminimalkan ketidakkonsistenan atau kesalahan pengukuran.

2. *Data Integration*

Agregasi informasi adalah proses mengintegrasikan informasi dari banyak sistem ke dalam satu repositori baru.

### 3. *Data Selection*

Tujuan dari pemilihan informasi adalah untuk memilih informasi yang berguna dari internet.

### 4. *Data Transformation*

Segmentasi pasar adalah proses mengubah informasi dengan cara yang dapat diproses oleh Packet sniffing.

### 5. *Datamining*

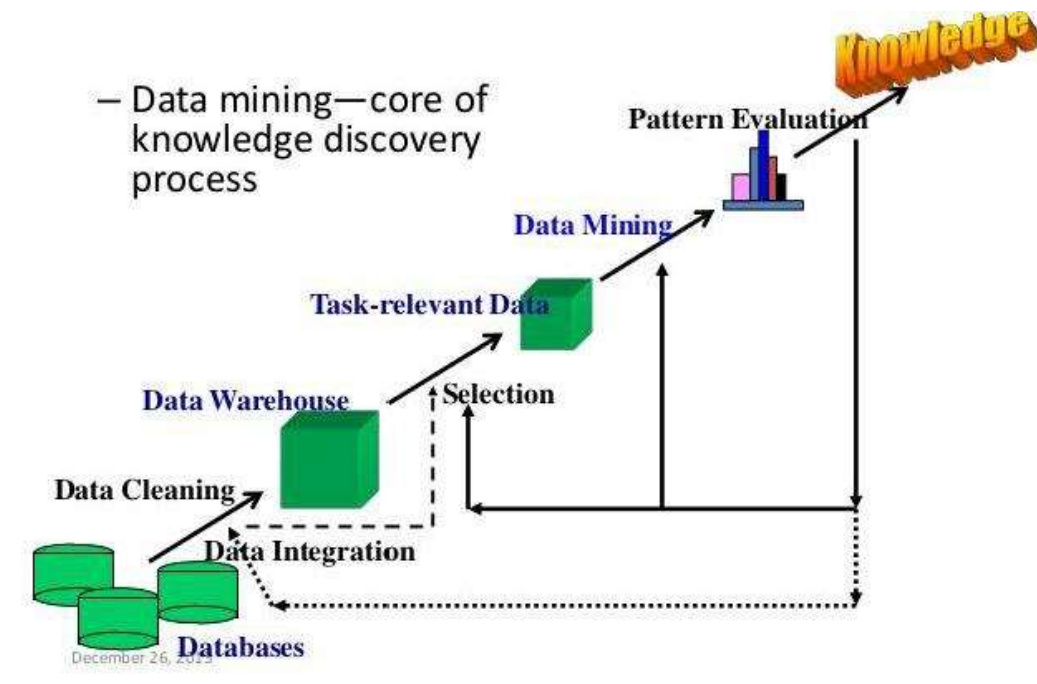
Ruang untuk penyesuaian adalah teknik yang digunakan untuk mengungkap wawasan berguna yang terkubur dalam statistik.

### 6. *Parttern Evaluation*

Penilaian tren digunakan untuk menemukan karakteristik menarik yang akan ditampilkan dalam sistem berbasis pengetahuan.

### 7. *Knowledge Presentation*

Tujuan dari memahami presentasi adalah untuk memahami dan memberikan informasi tentang langkah-langkah yang digunakan untuk mendapatkan pemahaman penerima.



Sumber : (Yanto and Kesuma 2017)

**Gambar 2. 1** Tahapan Proses KDD

Berinteraksi dengan volume data yang sangat besar merupakan salah satu dari enam aspek fundamental dari tujuan bisnis komunikasi dan strategis dalam KDD (Yanto and Kesuma 2017) :

1. Mengerjakan sejumlah besar data.
2. Karena banyaknya informasi, maka diperlukan akurasi.
3. Mengutamakan keandalan.
4. Memerlukan penggunaan terminologi tingkat lanjut.
5. Memanfaatkan beberapa jenis pemahaman otomatis.
6. Menghasilkan temuan yang menarik.

### 2.1.1.3 Pengelompokan Datamining

Berdasarkan (Fajrin and Handoko 2018) Penambangan data dibagi menjadi beberapa kelompok sesuai dengan tugas yang akan dilakukan, misalnya:

### 1. Deskripsi

Terkadang analis keuangan hanya ingin menyelidiki informasi untuk mengidentifikasi korelasi yang mereka temukan. Para pengumpul data, misalnya, dapat berjuang untuk melihat pengetahuan atau kenyataan bahwa individu yang tidak memiliki kualifikasi yang memadai akan memiliki dukungan yang lebih sedikit dalam kampanye federal. Pernyataan berpola dan gerakan sering kali menyertakan alasan yang masuk akal untuk sebuah struktur atau perkembangan.

### 2. Estimasi

Prediksi sebanding dengan identifikasi karena parameter tujuannya adalah numerik, bukan kualitatif. Model yang dihasilkan hanya menggunakan catatan komprehensif yang memprediksi probabilitas posterior. Perkiraan jumlah nilai atribut kemudian dihitung berdasarkan nilai kemungkinan suatu peristiwa dalam penilaian berikutnya. Sebagai referensi, kami akan menghitung detak jantung dan tekanan darah peserta layanan kesehatan tergantung pada durasi perawatan, kebangsaan, indeks tinggi badan, dan konsentrasi garam vena. Dalam proses pembelajaran, hubungan yang melibatkan tekanan darah arteri dan variabel independen yang diproyeksikan akan menghasilkan kerangka kerja estimasi. Metodologi estimasi yang dihasilkan dapat digunakan untuk situasi baru.

### 3. Prediksi.

Peramalan hampir sama dengan klasifikasi dan estimasi, hanya saja dengan peramalan nilai hasil akhirnya akan di masa yang akan datang.

### 4. Klasifikasi

Klasifikasi memiliki variabel target kategorikal. Misalnya, klasifikasi pendapatan dapat dibagi menjadi tiga kategori, yaitu pendapatan tinggi, pendapatan menengah, dan pendapatan rendah.

#### 5. Pengklusteran (*Clustering*)

Enkapsulasi adalah proses menggabungkan data, penilaian, atau pertimbangan yang terkait dan menetapkan pengelompokan item. Botnet adalah sekelompok informasi yang dapat dibandingkan dengan seseorang tetapi tidak dengan informasi yang berbeda dalam berbagai kelompok. Segmentasi berbeda dengan kategorisasi karena tidak memiliki nilai prediksi. Segmentasi tidak melakukan upaya untuk mengkategorikan, menghitung, atau mengantisipasi harga saham yang ditargetkan. Sebaliknya, pengelompokan fuzzy mencoba mempartisi seluruh dataset ke dalam kelompok-kelompok yang kohesif, dengan kemiripan dalam satu kategori yang ditekankan dan kemiripan dengan nilai-nilai di organisasi lain berkurang.

#### 6. Asosiasi

Dalam analisis Heuristik, tujuan koneksi adalah untuk menemukan kualitas yang muncul pada saat yang sama. Analisis ini semakin sering dikenal sebagai pemantauan troli supermarket di dalam industri keuangan.

### **2.1.1.4 Tahapan – Tahapan Datamining**

Berdasarkan (Erwansyah 2019) Solusi yang ditingkatkan adalah bagian dari Ekstraksi Informasi (KDD). KDD difokuskan pada metodologi serta perkembangan ilmu pengetahuan, pemahaman, dan representasi data yang dikumpulkan. Langkah-langkah prosedurnya adalah sebagai berikut:

1. Pemurnian data

Prosedur ini berupaya menghilangkan ketidakkonsistenan informasi dan distorsi.

2. Manajemen informasi mengacu pada metode penggabungan beberapa sumber data secara bersamaan.

3. Konversi data

Segmentasi gambar adalah proses mengubah informasi ke dalam format yang sesuai untuk ekstraksi.

4. Penggunaan teknik penggalian data

Pengambilan pola dari informasi yang ada.

5. Penilaian penemuan pola

Metode untuk mengubah kesamaan menjadi informasi yang dapat digunakan untuk mengambil keputusan.

6. Demonstrasi kompetensi

Metode visualisasi digunakan untuk menyampaikan informasi..

#### **2.1.1.5 Manfaat Datamining**

Penerapan pengumpulan data dilihat dari dua perspektif bisnis dan ilmiah (Erwansyah 2019):

1. Dari sudut pandang bisnis, datamining dapat digunakan untuk mengelola jumlah data yang terus bertambah dengan menggunakan pendekatan untuk memberikan rincian yang diperlukan, sehingga menghasilkan komoditas yang dapat meningkatkan kelangsungan hidup suatu negara.

2. Dari sudut pandang biologis, datamining memiliki potensi untuk memperoleh, mengevaluasi, dan melestarikan sejumlah besar informasi yang tepat waktu.

#### **2.1.1.6 Klasifikasi**

Segmentasi adalah proses menganalisis informasi yang diambil yang menggunakan kerangka kerja untuk menganalisis kategori informasi dari kedua statistik individu berdasarkan sifat-sifatnya. Statistik kategorikal, numerik, numerik, dan nyata dapat digunakan sebagai karakteristik (Ismasari Nawangsih and Setyaningsih 2020).

Analisis adalah strategi statistik yang memiliki potensi untuk membuat kumpulan kategori (fungsi) yang juga dapat menggambarkan dan memisahkan kelas-kelas yang berisi data atau ide, dengan tujuan menggunakan metode untuk menghitung karakteristik kualitas yang pelabelannya tidak tersedia atau untuk meramalkan tren informasi yang potensial. Algoritma pengelompokan berusaha untuk membagi informasi di antara dari sebelum klasifikasi sesuai dengan nilai item detail (Han dan Kamber, 2006).

#### **2.1.1.7 Naïve Bayes Classifier**

Tes adalah teknik sistematis yang memiliki kemampuan untuk menghasilkan sekumpulan klasifikasi (fungsi) yang dapat digunakan untuk mendefinisikan dan memisahkan kelas-kelas basis data yang mencakup atau konsep, juga dengan tujuan menerapkan teknik ini untuk menentukan parameter kinerja yang tidak tersedia atau untuk mengantisipasi bagian yang akan memeriksa tren. Metode klasifikasi bertujuan untuk mempartisi pengetahuan ke



dalam kelompok-kelompok sambil mengkategorikannya berdasarkan pentingnya bagian dataset (Saputra and Herdiansyah 2022).

Segmentasi Bayesian adalah pembelajar probabilistik yang memprediksi kemungkinan masalah klasifikasi biner. Penilaian Bayesian hanya didasarkan pada teorema Bayes yang memberikan kemampuan kategorisasi yang sebanding dengan pohon regresi dan pembelajaran mesin. Ketika digunakan pada set data yang sangat besar, pemilahan Bayesian telah terbukti memiliki kompleksitas komputasi yang tinggi. Analisis Bayes adalah metodologi numerik untuk melakukan penalaran induktif pada klasifikasi multikelas (Wijaya and Dwiasnati 2020).

Naive Bayes adalah pengklasifikasi deterministik dasar yang diawasi berdasarkan Teorema Bayes yang berhasil mengklasifikasikan fakta dengan menggunakan algoritma klasifikasi. Harga parameter keluaran dalam contoh target diasumsikan berbeda dari kuantitas banyak karakteristik lainnya oleh Naive Bayes. Filsuf Inggris Thomas Bayes membuat Rumus Bayes, yang pada dasarnya sepenuhnya mencerminkan kemungkinan yang memanfaatkan pengetahuan sebelumnya, dan dengan demikian namanya (Ismasari Nawangsih and Setyaningsih 2020). Bentuk umum atau persamaan dari teorema Bayes adalah

$$P(H|X) = \frac{P(X|H) \cdot P(H)}{P(X)}$$

Keterangan

X: sampel data dengan klasifikasi yang belum ditentukan (label)

H: anggapan bahwa X mewakili kelas (label)

P(H): kemungkinan asumsi H

P(X|H): kemungkinan data uji X jika premisnya benar.

$P(X)$ : probabilitas set data yang dilaporkan

Tahapan-tahapan Algoritma Naive Bayes adalah sebagai berikut:

1. Tentukan jumlah mata kuliah.
2. Hitung jumlah situasi di setiap kelas.
3. Tambahkan semua parameter kelas bersama-sama.
4. Periksa temuan berdasarkan kategori.

Naive Bayes memiliki manfaat yang signifikan, salah satunya adalah mudah dipahami. Dibutuhkan lebih sedikit kode dan cepat untuk menghitung.

### **2.1.2 Tinjauan Teori Khusus**

Hipotesis khusus adalah hipotesis yang terkait dengan kumpulan data yang sangat kompleks.

#### **2.1.2.1 Weka**

Waikato System for Hypothesis Testing (WEKA), sering dikenal sebagai WEKA, adalah program perangkat lunak visi komputer yang dibuat di Universidad of Waikato di Selandia Baru dan digunakan untuk akademis, instruksi, dan berbagai tujuan lainnya. WEKA adalah program open source, yang berarti dapat diakses oleh siapa saja. WEKA dimaksudkan untuk mengatasi berbagai masalah kehidupan nyata dengan data yang sedikit, seperti kategorisasi, yang melibatkan visi komputer sebagai komponennya. WEKA juga mencakup data sebelum, klasifikasi, analisis, pengelompokan, pola yang sering muncul, dan struktur data. Pengujian WEKA mungkin lebih memuaskan dan mampu menciptakan teknik data mining yang hampir sama dengan perhitungan manusia (Sinaga and Handoko 2021).

### 2.1.2.2 Air Conditioner

Pendingin Udara (AC) adalah bentuk khusus dari teknologi otomatisasi pendinginan. Perangkat ini berfungsi untuk menghantarkan udara dingin atau mengatur jumlah uap yang dibutuhkan oleh tubuh. Penggunaan AC adalah hal yang umum di tempat-tempat tropis yang dikenal dengan iklim yang sangat hangat. Suhu yang tinggi pada musim panas dapat menyebabkan kelelahan pada sekresi tubuh. Sistem pendingin juga digunakan untuk memberikan relaksasi. AC sering digunakan di ruang kelas untuk meningkatkan efisiensi kerja. Karena orang membutuhkan kontrol iklim yang menyenangkan untuk bekerja dengan baik dalam beberapa keadaan, yaitu 20-25 ° C dan kelembaban 40-60%. Pemanasan, kelembapan, mobilitas, dan kemurnian nafas, semuanya berkontribusi pada lingkungan termal ruang (<http://www.sumbarsehat.com>).

Penyejuk udara, pendingin udara, pengkondisi udara, penyaman udara, erkon, atau AC (*air conditioner*) adalah sistem atau mesin yang dirancang untuk menstabilkan suhu udara dan kelembapan suatu area (yang digunakan untuk pendinginan maupun pemanasan tergantung pada sifat udara pada waktu tertentu). Umumnya menggunakan siklus refrigerasi tetapi kadang-kadang menggunakan penguapan, biasanya untuk kenyamanan pendingin di gedung-gedung dan kendaraan bermotor. Konsep pendingin udara diketahui telah diterapkan di Romawi Kuno dan Persia abad pertengahan. Pendingin modern muncul dari kemajuan dalam ilmu kimia selama abad 19, dan pendingin udara skala besar listrik pertama ditemukan dan digunakan pada tahun 1902 oleh Willis Haviland Carrier (wikipedia 2022).

### 2.1.2.3 Penjualan

Penjualan ialah proses sosial manaherial dimana individu dan kelompok mendapatkan apa yang mereka butuhkan dan inginkan, menciptakan, menawarkan dan mempertukarkan produk yang bernilai dengan pihak lain. Penjualan ialah semua kegiatan yang bertujuan untuk melancarkan arus barang dan jasa dari produsen ke konsumen secara paling efisien dengan maksud untuk menciptakan permintaan yang efektif (Barata and Kurniawati 2019).

Berikut dibawah ini indikator penentu AC paling laris :

1. Harga

Harga yang murah menjadi alasan utama konsumen membeli produk AC dikarenakan cukup terjangkau disemua kalangan, walaupun kualitas yang diberikan belum tentu sesuai keinginan.

2. Merek

Merek menjadi sebuah stigma dimana beberapa merk menjadi tolak ukur didalam pemilihan produk AC dikarenakan kredibilitas yang baik.

3. Ukuran Watt

Produk AC yang memiliki penggunaan watt yang kecil lebih cenderung dipilih oleh konsumen disebabkan oleh penggunaan AC yang terus menerus untuk menghindari tagihan listrik yang besar.

4. Spesifikasi

Spesifikasi produk AC yang baik seringkali dijadikan alasan konsumen membeli produk AC sesuai dengan apa yang konsumen inginkan.

5. Peredam suara

AC yang dilengkapi dengan set peredam suara tidak hanya mengurangi kebisingan, tetapi juga efektif mengurangi getaran saat AC digunakan.

6. Garansi

Produk yang memiliki masa garansi yang panjang biasanya mempunyai kualitas yang baik sehingga produk bisa bertahan lama.

7. Lulus standar International

Produk yang telah lulus uji standarisasi biasanya telah memiliki performa yang baik dan teruji

8. AC yang memiliki cover mesin

Bagian yang menghubungkan outdoor AC dengan pipa tembaga terbuat dari struktur logam, dan mudah berkarat jika sering terpapar air hujan. Cover pelindung akan sangat berguna untuk mencegah kerusakan bagian tersebut

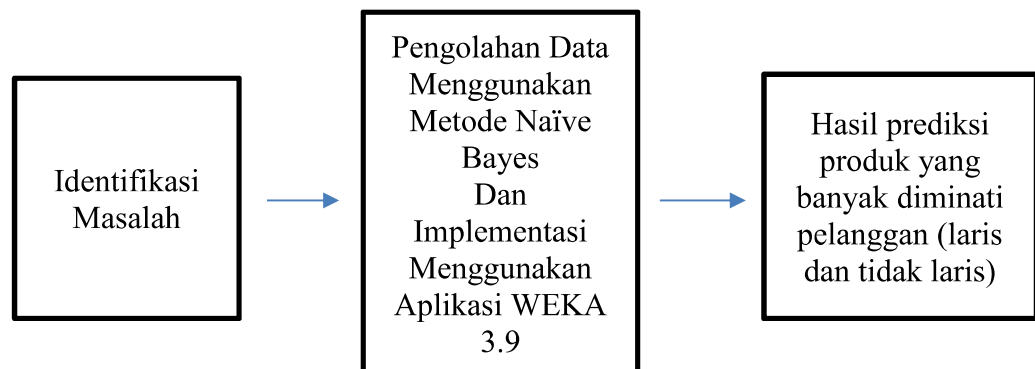
9. Anti bocor pada unit indoor

Pemilihan AC yang memiliki desain gesper pada sambungan pipa airnya dapat mencegah masalah mesin internal lepas yang menyebabkan kebocoran

10. AC dengan panel indikator LED

AC yang memiliki indikator LED lebih mudah dikontrol serta dapat dengan mudah menunjukkan error atau kerusakan pada mesin

## 2.2 Kerangka Pemikiran



Sumber : (Peneliti 2021)

**Gambar 2. 2 Kerangka Penelitian**

Penjelasan dari kerangka pemikiran pada *flowchart* diatas:

1. Identifikasi masalah merupakan proses mendefinisikan problem atau masalah yang ada pada penelitian.
2. Data di proses dengan metode naïve bayes lalu dilakukan pengujian dengan weka untuk mengetahui output berupa hasil prediksi produk yang banyak diminati pelanggan
3. Hasil dari penelitian

## 2.3 Penelitian Terdahulu

Penelitian pertama oleh Ade Izyuddin, Setyawan Wibisono, 2020. Berjudul Aplikasi Prediksi Penjualan AC Menggunakan Decision Tree Dengan Algoritma C4.5. Menunjukkan Dengan kemampuan melakukan prediksi penjualan AC terlaris ini maka manajemen Global Elektronik dapat secara lebih tepat dalam memperhitungkan ketersediaan stok produk AC, sehingga permintaan konsumen pada produk AC dapat terpenuhi, dengan demikian kepuasan konsumen dapat terjaga dan keuntungan yang diperoleh dapat lebih besar.

Algoritma C4.5 dalam penelitian ini digunakan sebagai algoritma alternatif untuk memprediksi penjualan produk terlaris dengan menggunakan metode decision tree yang mampu mengklasifikasi data dengan tingkat nilai keakuratan yang tinggi dengan menghitung nilai information gain hingga nilainya tidak dapat dihitung (0).

Penelitian Kedua oleh Natalia Sri Yanti, 2018. Berjudul Penerapan Data Mining Analisis Terhadap Data Penjualan Air Conditioner Dengan Menggunakan Algoritma Apriori. Menunjukkan Penjualan produk air conditioner paling banyak terjual pada PT Anugerah Cipta Parmana dapat diketahui dengan menggunakan algoritma apriori, dengan melihat produk yang memenuhi minimal support dan minimal confidence, produk yang paling banyak terjual tersebut adalah LG, Dainkin, Panasonic dan Sharp. Dari aturan asosiasi final yang diketahui Jika membeli Sharp, maka akan membeli Panasonic dengan support 58,33% dan confidence 77,78%. Jika membeli Panasonic, maka akan membeli Sharp dengan support 58,33% dan confidence 87,5%. Jika membeli Sharp dan LG, maka akan membeli Panasonic dengan support 41,67% dan confidence 83,33%. Jika membeli Sharp dan Panasonic, maka akan membeli LG dengan support 41,67% dan confidence 83,33%.

Penelitian Ketiga oleh Wahyu Hadikristanto, Asti Setyaningsih (2020, berjudul Penerapan Algoritma Naive Bayes Untuk Menentukan Klasifikasi Produk Terlaris Pada Penjualan Pulsa. Berdasarkan hasil perhitungan yang telah dilakukan oleh peneliti dengan perhitungan menggunakan Algoritma Naive Bayes yang dihitung secara manual, Microsoft Excel, dan tool RapidMiner menghasilkan data yang sama sehingga data mining algoritma metode Naive Bayes sangat tepat digunakan untuk menghitung klasifikasi data penjualan pulsa karena

menghasilkan data yang akurat. Berdasarkan hasil perhitungan menggunakan metode Naive Bayes produk terlaris dalam penjualan pulsa adalah Pulsa Telkomsel. Berdasarkan hasil perhitungan Metode Naive Bayes menggunakan RapidMiner nilai accuracy : 97,50%, Precision : 100,00%, dan Recall : 93,48%

Penelitian Keempat oleh Edy Widodo, Pathyatus Sa'adah (2019) berjudul Klasifikasi Data Penjualan Alat Tulis Kantor (Atk) Terlaris Untuk Optimasi Strategi Pemasaran Di Toko Citramedia Menggunakan Metode Naive Bayes. Menunjukkan Penerapan data mining pada tools Rapid Miner menggunakan algoritma naive bayes sangat cocok diterapkan dalam mengklasifikasi data dengan jumlah yang cukup banyak seperti data penjualan ATK. Karena dapat menentukan produk mana saja yang termasuk dalam kategori Laris maupun Tidak Laris tanpa memerlukan waktu yang lama dan lebih efisien dalam pengolahan data, sehingga owner Citra Media dapat lebih mudah dalam memilah produk yang mampu menunjang dan dapat meningkatkan keuntungan bisnisnya. Perhitungan menggunakan metode naive bayes dalam klasifikasi penjualan ATK terlaris memperoleh hasil yang sangat baik. Dari pengujian yang dilakukan pada tools RapidMiner menghasilkan tingkat akurasi paling tinggi sebesar 99.26% dan akurasi paling rendah sebesar 94.39%. sedangkan evaluasi dengan kurva ROC/AUC hasil dari empat kali pengujian yang telah dilakukan termasuk dalam kategori klasifikasi sangat baik (excellent classification) yakni dengan evaluasi tertinggi sebesar 0.997.

Penelitian Kelima oleh Ismasari Nawangsih, Asti Setyaningsih (2020) berjudul Penerapan Algoritma Naive Bayes Untuk Menentukan Klasifikasi Produk Terlaris Pada Penjualan Pulsa. Berdasarkan hasil perhitungan yang telah



dilakukan oleh peneliti dengan perhitungan menggunakan Algoritma Naive Bayes yang dihitung secara manual, Microsoft Excel, dan tool RapidMiner menghasilkan data yang sama sehingga data mining algoritma metode Naive Bayes sangat tepat digunakan untuk menghitung klasifikasi data penjualan pulsa karena menghasilkan data yang akurat. Berdasarkan hasil perhitungan menggunakan metode Naive Bayes produk terlaris dalam penjualan pulsa adalah Pulsa Telkomsel. Berdasarkan hasil perhitungan Metode Naive Bayes menggunakan RapidMiner nilai accuracy : 97,50%, Precision : 100,00%, dan Recall : 93,48%.

Penelitian Keenam oleh Sidik Rahmatullah, Mukrim, Monica Nindy Pramitha (2019) berjudul Data Mining Untuk Menentukan Produk Terlaris Menggunakan Metode Naive Bayes. Menunjukkan Sistem pengklasifikasian pada produk terlaris dapat menghasilkan aturan klasifikasi dalam setiap kategori berdasarkan 8 Atribut dengan menggunakan metode naïve bayes classifier dengan mencari nilai kategori yang paling maksimum. Pembangunan sistem pengklasifikasian produk terlaris dapat membantu mempermudah Kepala Area Operational Supervisor (AOS) PT. Cipta Niaga Semesta SubBranch Kotabumi dalam melakukan pendataan produk terlaris berdasarkan kategori ke dalam 8 atribut.

Penelitian ketujuh oleh Robi Wariyanto Abdullah, Dwi Hartanti, Hanifah Permatasari, Arif Wicaksono Septyanto, Yuda Abi Bagaskara (2022) berjudul Penerapan Data Mining untuk Memprediksi Jumlah Produk Terlaris Menggunakan Algoritma Naive Bayes Studi Kasus (Toko Prapti). Menunjukkan Teknik Penerapan Algoritma Naïve Bayes Classifier pada teknik Data Mining sangat efisien dan dapat mempercepat proses pembentukan hasil penjualan di

Toko Prapti, yaitu dengan perhitungan peluang dari masing-masing atribut adalah triwulan 1, triwulan 2, triwulan 3 dan triwulan 4. Keterkaitan suatu barang yang dibeli oleh konsumen bisa dihitung dengan teknik algoritma naïve bayes classifier. Keakuratan sistem menggunakan confusion matrix dengan nilai accuracy sebesar 83,3 % , precision sebesar 84,2% dan recall sebesar 88,9 %.

Penelitian kedelapan oleh Ikhsan Romli, Toga Pardamean, Sufajar Butsianto, Tri Ngudi Wiyatno, Effendi bin Mohamad (2021) berjudul *Naive Bayes Algorithm Implementation Based on Particle Swarm Optimization in Analyzing the Defect Product*. Menunjukkan *The results of training data testing conducted by increasing accuracy using the RapidMiner tools produce an accuracy rate of 88.62% (800 data) and 92.22% (372 data), while the results of the evaluation using the ROC curve with an excellent classification accuracy of 0.953 (800 data) and 0.0.945 (372 data). And also increase accuracy by 5.02% (800 data) and 2.10% (372 data). This proves that the more data tested, the results obtained will be better.*

Penelitian kesembilan oleh Novi Pransiska, A.HaidarMirz, Andri (2019) berjudul Penerapan Datamining Prediksi Penjualan Barang Elektronik Terlaris Menggunakan Algoritma Naïve Bayes (Study Kasus : Planet Cash And Credit Cabang Muara Enim). Menunjukkan Analisis data penjualan barang pada perusahaan Planet Cash And Credit ini digunakan untuk menampilkan informasi barang yang laris terjual dan barang kurang laris menggunakan algoritma naive bayes. Algoritma naive bayes sangat cocok diterapkan dalam memprediksi peluang dimasa depan berdasarkan pengalaman dimasa sebelumnya sehingga memudahkan perusahaan untuk memprediksi peminatan masyarakat terhadap

barang elektronik yang dijual. Dengan mengetahui barang yang laris atau kurang laris terjual akan meminimalisir kerugian pada perusahaan dan perusahaan akan lebih selektif dalam menyetok barang elektronik tersebut.

Penelitian kesepuluh oleh Muhammad Juanda Saputra, Muhammad Izman Herdiansyah (2022) berjudul Penerapan Naive Bayes Dalam Memprediksi Penjualan Dan Persediaan Kain Jumputan Pada Toko Batiq Colet Tuan Kentang Palembang. Menunjukkan Algoritma Naïve Bayes telah berhasil diterapkan untuk menganalisis prediksi Penjualan dan Persediaan kain jumputan pada Toko Batiq Colet untuk menjadi acuan dimasa yang akan datang agar tidak terjadinya penumpukan barang di Gudang dan dapat memperkecil kerugian. Penerapan data mining membantu Toko Batiq Colet dalam hal pengambilan keputusan yang melakukan prediksi (forecasting) terhadap penjualan dan persediaan produk sehingga toko dapat mengambil Langkah – Langkah penting dalam menghadapinya. Hasil prediksi menggunakan algoritma Naïve Bayes dengan alat bantu Rapidminer 9.4 terbukti mampu memprediksi penjualan pada Toko Batiq Colet Tuan Kentang Palembang, dengan nilai akurasi sebesar 89.47% dengan class precision yaitu “YA” 100.00%, “TIDAK” 81.82%, dan untuk class recall yaitu “YA” 80.00%, “TIDAK” 100.00%. Hasil prediksi menggunakan algoritma Naïve Bayes dengan alat bantu Rapidminer 9.4 terbukti mampu memprediksi produksi pada Toko Batiq Colet Tuan Kentang Palembang, dengan nilai akurasi sebesar 78.95% dengan class precision yaitu “YA” 77.27%, “TIDAK” 81.25%, dan untuk class recall yaitu “YA” 85.00%, “TIDAK” 72.22%.