

**IMPLEMENTASI ALGORITMA NAIVE BAYES
DALAM KLASIFIKASI PRODUK AIR CONDITIONER
TERLARIS PADA PT DAMAI LESTARI SEJAHTERA**

SKRIPSI



Oleh:

Ester Devi Br Nadapdap

181510070

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS TEKNIK DAN KOMPUTER
UNIVERSITAS PUTERA BATAM**

2023

**IMPLEMENTASI ALGORITMA NAIVE BAYES
DALAM KLASIFIKASI PRODUK AIR CONDITIONER
TERLARIS PADA PT DAMAI LESTARI SEJAHTERA**

SKRIPSI

**Untuk memenuhi salah satu syarat
guna memperoleh gelar Sarjana**



Oleh:

Ester Devi Br Nadapdap

181510070

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS TEKNIK DAN KOMPUTER
UNIVERSITAS PUTERA BATAM**

2023

SURAT PERNYATAAN ORISINALITAS

Yang bertanda tangan di bawah ini saya:

Nama : Ester Devi
NPM : 181510070
Fakultas : Teknik dan Komputer
Program Studi : Sistem Informasi

Menyatakan bahwa “Skripsi” yang saya buat dengan judul:

IMPLEMENTASI ALGORITMA NAIVE BAYES DALAM KLASIFIKASI PRODUK AIR CONDITIONER TERLARIS PADA PT DAMAI LESTARI SEJAHTERA

Adalah hasil karya sendiri dan bukan “duplikasi” dari karya orang lain. Sepengetahuan saya, didalam naskah Skripsi ini tidak terdapat karya ilmiah atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis dikutip didalam naskah ini dan disebutkan dalam sumber kutipan dan daftar pustaka.

Apabila ternyata di dalam naskah Skripsi ini dapat dibuktikan terdapat unsur-unsur PLAGIASI, saya bersedia naskah Skripsi ini digugurkan dan gelar akademik yang saya peroleh dibatalkan, serta diproses sesuai dengan peraturan perundang-undangan yang berlaku.

Demikian pernyataan ini saya buat dengan sebenarnya tanpa ada paksaan dari siapapun.

Batam, 31 Januari 2023



Ester Devi Br Nadandap
181510070

**IMPLEMENTASI ALGORITMA NAIVE BAYES DALAM KLASIFIKASI
PRODUK AIR CONDITIONER TERLARIS PADA PT DAMAI LESTARI
SEJAHTERA**

SKRIPSI

**Untuk memenuhi salah satu
syarat memperoleh gelar sarjana**

Oleh

Ester Devi Br Nadapdap

181510070

Telah disetujui oleh Pembimbing pada tanggal

Seperti tertera di bawah ini

Batam, 31 Januari 2023



**Sasa Ani Arnomo, S.Kom., M.Si., Ph.D
Pembimbing**

ABSTRAK

PT Damai Lestari Sejahtera merupakan salah satu produser *Air Conditioning (AC)* ternama dunia yang berada di Batam. Dalam membantu perusahaan agar semakin maju diperlukan sebuah sistem yang akan membantu kemajuan perusahaan dalam memaksimalkan penjualan produk mereka dari barang yang sering laris sehingga dapat memuaskan pelanggan. Algoritma Metode Naive Bayes dapat membantu dalam menentukan produk yang laris dan tidak laris pada penjualan AC yaitu dapat menghasilkan data yang memiliki tingkat akurasi yang dapat dimanfaatkan untuk memprediksi kelas dari kasus yang baru. Sehingga dapat diketahui tingkat akurasi data. Penelitian ini dibuat untuk membantu dalam mengklasifikasi produk yang laris dijual dalam penjualan AC sehingga pemilik dapat menentukan stok produk yang banyak diminati masyarakat agar setiap kali ada pembeli barang selalu ada untuk dijual. Teknik Naive Bayes sesuai untuk informasi distribusi frekuensi. Tingkat pengukuran semacam ini memiliki elemen dengan kuantitas dalam arti sebenarnya yang dapat dikategorikan tetapi dapat dievaluasi berdasarkan panjangnya atau dijumlahkan. Manfaat dari algoritma ini adalah sangat baik dalam memperkirakan informasi dengan perubahan yang disebabkan dan membuat perkiraan untuk minggu berikutnya. Hasilnya, teknik Naive Bayes merupakan pendekatan yang tepat untuk memproyeksikan pangsa pasar AC di perusahaan yang saat ini sedang berkembang di pasar oleh konsumen di bulan berikutnya.

Kata Kunci : Penjualan, Data Mining, Algoritma Naive Bayes

ABSTRACT

PT Damai Lestari Sejahtera is one of the world's leading Air Conditioning (AC) producers in Batam. In helping the company to be more advanced, a system is needed that will help the company's progress in maximizing the sales of their products from items that are often in demand so that they can satisfy customers. The Naive Bayes Method algorithm can help in determining products that are in demand and not in demand in AC sales, which can produce data that has an accuracy level that can be used to predict the class of a new case. So that the accuracy of the data can be known. This research was made to assist in classifying products that are in demand in the sale of air conditioners so that the owner can determine the stock of products that are in great demand by the community so that every time there is a buyer the goods are always there for sale. Naive Bayes technique is suitable for frequency distribution information. This kind of measurement level has elements with quantities in the true sense that can be categorized but can be evaluated based on their length or summed. The benefit of this algorithm is that it is very good at estimating information with induced changes and making forecasts for the following week. As a result, the Naive Bayes technique is an appropriate approach to project the market share of air conditioners in the company that are currently growing in the market by consumers in the next month.

Keywords: Sales, Data Mining, Naive Bayes Algorithm

KATA PENGANTAR

Puji syukur penulis panjatkan kepada Tuhan Yang Maha Esa atas segala berkat dan anugerahNya, sehingga penulis dapat menyelesaikan skripsi ini yang merupakan salah satu persyaratan untuk menyelesaikan program studi strata satu (S1) pada Program Studi Sistem Informasi di Universitas Putera Batam.

Dengan segala keterbatasan, penulis juga menyadari bahwa skripsi ini takkan terwujud tanpa bantuan, bimbingan, dan dorongan dari berbagai pihak. Untuk itu, dengan segala kerendahan hati, penulis menyampaikan ucapan terima kasih kepada:

1. Rektor Universitas Putera Batam Ibu Dr. Nur Elfi Husda, S.kom., M.SI.,
2. Bapak Welly Sugianto, S.T., M.M. selaku Dekan Fakultas Teknik dan Komputer di Universitas Putera Batam
3. Ketua Program Studi Sistem Informasi Bapak Muhammat Rasid Ridho, S.Kom., M.SI. di Universitas Putera Batam.
4. Ibu Erlin Elisa, S.Kom., M.Kom. selaku Pembimbing Akademik pada Program Studi Sistem Informasi di Universitas Putera Batam,
5. Bapak Sasa Ani Arnomo, S.Kom., M.Si., Ph.D. selaku Pembimbing Skripsi pada Program Studi Sistem Informasi di Universitas Putera Batam.
6. Dosen dan staff Universitas Putera Batam, yang sudah banyak memberikan pengetahuan selama perkuliahan berlangsung,

7. Kedua orang tua penulis serta keluarag yang selalu berdoa, memberikan kasih sayang dan juga menyemangati penulis untuk menyelesaikan perkuliahan S1 ini.
8. Wahyudi yang sudah memberikan segala dukungannya dalam berbagai bentuk selama perkuliahan dan penulisan skripsi ini berlangsung
9. Seluruh pihak yang telah banyak membantu dalam menyelesaikan penelitian ini yang tidak dapat penulis sebutkan satu persatu.

Semoga Tuhan membalas kebaikan dan selalu mencurahkan berkat yang melimpah. Penulis berharap skripsi ini dapat menjadi ilmu yang berharga bagi para pembaca di masa yang akan datang. Penulis juga berharap kritik dan saran yang bisa membangun dari para pembaca.

Batam, 31 Januari 2023

Ester Devi Br Nadapdap

DAFTAR ISI

HALAMAN COVER	i
HALAMAN JUDUL	ii
ABSTRAK	v
ABSTRACT	vi
KATA PENGANTAR	vii
DAFTAR ISI	ix
DAFTAR GAMBAR	xi
DAFTAR TABEL	xii
BAB I	1
PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Identifikasi Masalah	6
1.2 Batasan Masalah	6
1.4 Rumusan Masalah	6
1.5 Tujuan Penelitian	7
1.6 Manfaat Penelitian	7
BAB II	9
TINJAUAN PUSTAKA	9
2.1 Tinjauan Teori Dasar	9
2.1.1 Tinjauan Teori Umum	9
2.1.2 Tinjauan Teori Khusus	18
2.2 Kerangka Pemikiran	22
2.3 Penelitian Terdahulu	22
BAB III	28
METODE PENELITIAN	28
3.1 Desain Penelitian	28
3.2 Objek Penelitian	29
3.2.1 Lokasi Penelitian	29
3.3 Populasi dan Sampel	30
3.3.1 Populasi Penelitian	30
3.3.2 Sampel	30

3.4	Variabel Penelitian	30
3.5	Pengumpulan Data	35
3.6	Model Penelitian	36
BAB IV		39
HASIL DAN PEMBAHASAN		39
4.1	Hasil Penelitian	39
4.2	Pembahasan	39
4.3	Analisis Metode Naïve Bayes	41
4.3.1	Menghitung Probabilitas Kelas	41
4.3.2	Menghitung probabilitas masing-masing atribut	42
4.3.3	Menghitung probabilitas akhir untuk setiap kelas	61
4.3.5	Pengujian menggunakan <i>Software WEKA</i>	63
BAB V		69
KESIMPULAN DAN SARAN		69
5.1	Kesimpulan	69
5.2	Saran	69
DAFTAR PUSTAKA		lxx
LAMPIRAN		lxxii
	DATA PENELITIAN	lxxii
	DATA PENJUALAN	xcvi
	SURAT IZIN KAMPUS	cxviii
	SURAT BALASAN PENELITIAN	cxix
	DAFTAR RIWAYAT HIDUP	Error! Bookmark not defined.
	TURNITIN SKRIPSI	cxxi
	TURNITIN JURNAL	cxxii
	LOA JURNAL	cxxiii

DAFTAR GAMBAR

Gambar 2. 1 Tahapan Proses KDD	12
Gambar 2. 2 Kerangka Penelitian	22
Gambar 3. 1 Design penelitian	28
Gambar 3. 2 Flowchart Algoritma Naive Bayes	37
Gambar 4. 1 Format Data Microsoft Excel	64
Gambar 4. 2 Software WEKA	64
Gambar 4. 3 Data Pada Menu Explorer	65
Gambar 4. 4 Memilih Tabel yg akan di Input	66
Gambar 4. 5 Data pada Menu Classify	66
Gambar 4. 6 Classify Output pada WEKA	67
Gambar 4. 7 Output tes	68

DAFTAR TABEL

Tabel 2. 1 Penelitian Terdahulu	Error! Bookmark not defined.
Tabel 3. 1 Atribut dalam data training	30
Tabel 3. 2 Penjualan AC Laris September	32
Tabel 3. 3 Penjualan AC Laris Oktober	34
Tabel 3. 4 Penjualan AC Laris November	34
Tabel 3. 5 Penjualan AC Laris Desember	35
Tabel 3. 5 Penjualan AC Laris Januari	35
Tabel 4. 1 Probabilitas Kelas	41
Tabel 4. 2 Hasil Probabilitas Keputusan	42
Tabel 4. 3 Merk	42
Tabel 4. 4 Hasil Perhitungan Atribut Merk	44
Tabel 4. 5 Type	44
Tabel 4. 6 Hasil Perhitungan Atribut Type	46
Tabel 4. 7 Daya	47
Tabel 4. 8 Hasil Perhitungan Atribut Daya	48
Tabel 4. 9 PK	49
Tabel 4. 10 Hasil Perhitungan Atribut PK	50
Tabel 4. 11 Hasil Perhitungan Atribut Harga	52
Tabel 4. 12 Jenis	52
Tabel 4. 13 Hasil Perhitungan Atribut Jenis	53
Tabel 4. 14 Alamat	54

Tabel 4. 15 Hasil Perhitungan Atribut Alamat	57
Tabel 4. 16 Hasil Perhitungan Atribut Bulan	60
Tabel 4. 17 Data Testing	61

BAB I

PENDAHULUAN

1.1 Latar Belakang

Penemuan, khususnya di bidang pengembangan perangkat lunak, merupakan salah satu elemen yang menyebabkan modifikasi dalam cara orang belajar untuk menarik kesimpulan dengan cepat dan efektif. Pada prinsipnya, intelijen bisnis digunakan untuk menangani data dalam jumlah besar, seperti menyimpan, memperoleh, memodifikasi, dan menampilkan statistik untuk memberikan informasi yang sangat baik dan dapat dipercaya. Pentingnya terobosan teknis memang terkait erat dengan memahami dan mengidentifikasi keadaan yang mungkin terjadi, yang sangat penting dalam organisasi pabrik dalam hal antara output dan pemasaran. Pendekatan prediksi atau estimasi merupakan salah satu strategi yang dibutuhkan untuk menyelesaikan tantangan ini secara akurat. Prediksi atau analisis pasar adalah teknik yang dapat menilai sebuah komponen atau beberapa komponen sebelumnya untuk mengidentifikasi anomali menggunakan hasil sebelumnya (Saputra and Herdiansyah 2022).

Peperangan dalam dunia pemasaran dan penjualan semakin meningkat seiring dengan keadaan. Sehingga para kapitalis berjuang untuk meningkatkan bisnis yang sudah mapan dengan memanfaatkan perangkat teknologi, untuk memastikan bahwa bisnisnya terus berkembang dalam menghadapi daya saing yang cepat. Untuk memenuhi harapan pelanggan, pengusaha harus meningkatkan kepuasan pelanggan dan menambahkan jenis barang yang disukai pelanggan dan lebih sering dibeli. Untuk meningkatkan pendapatan toko di tengah persaingan

yang kompetitif, para pemangku kepentingan terkait harus menentukan rencana bisnis untuk penjualan yang dilakukan. Para eksekutif perusahaan juga menginginkan evaluasi untuk mengantisipasi hal-hal yang sedang populer. Para direktur untuk memastikan harus mampu membantu dan memfasilitasi staf toko dalam mengidentifikasi barang-barang yang harganya lebih rendah dengan mengantisipasi riwayat pembelian (Pransiska, Mirza, and Andri. 2019).

Praktik dalam profesi untuk menghasilkan pendapatan dikenal sebagai pemasaran. Perusahaan memberikan kendali atas produk tertentu kepada pelanggan melalui tenaga penjual dengan tujuan untuk memuaskan permintaan pelanggan dan dapat memberikan keuntungan kepada pemasok untuk kontrak berjangka. Ada beberapa statistik yang berhubungan dengan pemasaran dan pertukaran barang atau jasa dalam perusahaan pemasaran. Informasi ini harus dilakukan untuk memastikan kemajuan perusahaan. Sebagian besar perusahaan sekarang sadar akan penggunaan informasi dalam mengembangkan bisnis masing-masing. Sayangnya, banyak organisasi yang tidak menyadari apa yang harus dilakukan untuk mendapatkan informasi yang benar, dan yang lainnya tidak yakin dapat memanfaatkan statistik dengan baik. Ekonomi pengetahuan semakin cepat. Teknik dan aplikasi ini memudahkan untuk menjalankan perintah. Implementasi ekstraksi informasi adalah salah satu metode pengolahan data. Penggalan data adalah kombinasi dari alat dan metodologi informasi dan teknik klasifikasi untuk menganalisis informasi dalam jumlah besar (Widodo and Sa'adah 2019).

Data mining adalah jenis ekstraksi informasi di mana keahlian diekstraksi menggunakan data dalam jumlah yang sangat besar. Metode untuk menganalisis angka pendapatan adalah dengan melakukan otomatisasi untuk memastikan temuan kategorisasi dari data (Ismasari Nawangsih and Setyaningsih 2020).

Teknik Naive Bayes adalah pendekatan kategorisasi dan percabangan otomatisasi. Sistem AI adalah kapasitas program untuk secara akurat membaca rangsangan sensorik, beradaptasi melalui informasi, dan menggunakan pemahaman tersebut untuk menyelesaikan berbagai hal. Hasilnya, banyak merek dagang dalam analisis tersebut akan diklasifikasikan sebagai sesuai permintaan atau tidak sesuai keinginan, sehingga pembeli, produsen, dan cendekiawan dapat menentukan perusahaan mana yang lebih sesuai dengan kriteria untuk dievaluasi dalam kategori dan produksi (Romli, Pusnawati, and Bangsa 2019).

Hasil penelitian terdahulu dari (Ade Izyuddin and Setyawan Wibisono 2020) menyatakan type item AC paling banyak terjual berdasarkan brand item ialah Sharp. Algoritma C4.5 memprediksi penjualan produk terlaris dengan menggunakan metode decision tree yang bisa di kelompokkan informasi sesuai jumlah kepastian menggunakan analisa value dari pendapatan info hingga nilainya tidak dapat dihitung (0). Maka didapatkan Harga murah, merk sharp dengan jenis inverter dengan nilai entropy 0,721928 dan nilai gain 0.170951 merupakan barang terlaris pada toko tersebut. Selanjutnya penelitian dari (Sri yanti 2018) Metodologi posterior dapat digunakan untuk mencari pendapatan dari barang kipas angin listrik yang paling banyak dibeli di PT Anugerah Cipta Parmana dengan memeriksa barang yang memenuhi pengaruh lemah dan cukup diambil. Barang yang sering dibeli antara lain LG, Dainkin, Panasonic, dan Sharp. Menurut

algoritma apriori pamungkas, jika Anda membeli Bright, maka juga akan membeli Philips dengan dukungan tidak lebih dari 58,33% serta kepercayaan 77,78%. Jika Anda membeli Panasonic, Anda akan mendapatkan Sharpe dengan dukungan 58,33% dan kepastian 87,5%. Lumix memiliki dukungan 41,67% dan kepastian 83,33% sementara pelanggan membeli Smart dan juga LG. LG memiliki 41,67% dukungan dan 83,33% kepastian jika pelanggan membeli Sharper dan Lumix.

PT Damai Lestari Sejahtera atau PT DLS merupakan Salah satu produser *Air Conditioning (AC)* ternama dunia yang berada di Batam. officially, kantor perwakilan Daikin ialah PT DLS mulai beroperasi pada Sabtu tanggal 01 Maret 2014. Perusahaan ini melayani penjualan unit *Daikin Group (Daikin, McQuay, AAF)* yaitu *Split Wall mounted, Cassette, Ceiling concealed atau split duct, Floor Standing, VRV, Chiller, Marine Product dan Filtration*. Perusahaan ini juga melakukan pemasangan seluruh *type AC Daikin Group* tersebut diatas, memberikan Sistem pemanas dan pendingin, konstruksi lokasi AC, termasuk konstruksi AC (untuk pengaturan perumahan dan komersial) yang menggunakan berbagai jenis pembersih udara mulai dari panel pemisah, kartrid, dan mekanisme tabung semuanya tersedia. Organisasi ini menekankan kepuasan pelanggan melalui pemeriksaan kualitas, manajemen tenaga kerja, kecepatan kerja, dan juga konsisten yang baik. Dibantu oleh bantuan purna jual dalam bentuk pemeliharaan terjadwal, latensi tercepat sejauh memberikan umpan balik dan melaporkan hasil organisasi sesuai dengan spesifikasi. Kemudian, mulai dari ide hingga pengembangan tata letak hingga implementasi keahlian, menawarkan bantuan. Bisnis ini juga menyediakan material yang berhubungan dengan pemasangan AC seperti : *Panel PU ducting, Accessories Ducting, Pipa Inaba, Kabel, damper,*

Grille, dll. Melakukan Penjualan spare Part untuk *AC Daikin Group* serta perbaikan. Produk layanan ini melaksanakan *system maintenance* atau peremajaan AC dengan konstan.

PT. Damai Lestari Sejahtera lebih ingin menyenangkan pembeli dengan meningkatkan nilai dari hal-hal yang ditawarkan perusahaan. Oleh karena itu, diperlukan suatu metode yang dapat membantu perusahaan ini untuk terus berkembang dalam meningkatkan brand awareness dari barang-barang yang sering diminati oleh para pelanggan. Pendekatan Metodologi Naive Bayes dapat membantu dalam mendeteksi barang siapa yang membutuhkan dan yang tidak mendukung dalam pemasaran AC, menghasilkan informasi dengan presisi yang dapat digunakan untuk meramalkan klasifikasi contoh baru. Untuk menentukan keandalan informasi dapat ditentukan. Penelitian ini dilakukan untuk membantu menentukan barang yang saat ini memiliki persediaan terbatas dalam pembelian sistem pemanas sehingga konsumen dapat membuat keputusan pembelian yang tepat.

Teknik Naive Bayes sesuai untuk informasi distribusi frekuensi. Tingkat pengukuran semacam ini memiliki elemen dengan kuantitas dalam arti sebenarnya yang dapat dikategorikan tetapi dapat dievaluasi berdasarkan panjangnya atau dijumlahkan. Regresi Logistik adalah strategi untuk menangani peramalan variabel dan pola di kuartal mendatang. Manfaat dari algoritma ini adalah sangat baik dalam memperkirakan informasi dengan perubahan yang disebabkan dan membuat perkiraan untuk minggu berikutnya. Hasilnya, teknik Naive Bayes merupakan pendekatan yang tepat untuk memproyeksikan pangsa pasar AC di perusahaan yang saat ini sedang berkembang di pasar oleh konsumen di bulan

berikutnya. Infrastruktur jaringan perusahaan masih belum menentukan item mana yang paling baik dibeli, dilelang, atau bahkan tidak dilelang pada periode sebelumnya dalam upaya untuk menentukan apakah item tersebut akan dipasok pada bulan berikutnya.

Bergantung pada konteksnya, penelitian ini dimaksudkan untuk melakukan studi terkait **“IMPLEMENTASI ALGORITMA NAIVE BAYES DALAM KLASIFIKASI PRODUK AIR CONDITIONER TERLARIS PADA PT DAMAI LESTARI SEJAHTERA”**

1.2 Identifikasi Masalah

Adapun identifikasi masalah yang didapatkan dari latar belakang diatas, yaitu:

1. Belum diketahui produk AC mana yang paling diminati konsumen
2. Tidak ada akurasi yang tepat dalam menentukan ramalan penjualan

1.2 Batasan Masalah

Agar penelitian ini lebih fokus, lengkap dan mendalam, batasan masalah dalam penelitian ini adalah:

1. Penelitian ini hanya dilakukan di PT Damai Lestari Sejahtera
2. Penelitian ini menggunakan metode Naive Bayes untuk memprediksi produk terlaris.

1.4 Rumusan Masalah

Dengan latar belakang permasalahan di atas, maka permasalahan utama dari penelitian ini adalah:

1. Bagaimana klasifikasi produk untuk mengetahui kecenderungan pembelian produk terlaris berdasarkan riwayat transaksi September 2022 – Januari 2023
2. Keakuratan apa yang dapat membantu wiraniaga membuat keputusan untuk memprediksi produk mana yang akan sangat dicari (menjual dan tidak menjual) oleh pelanggan?

1.5 Tujuan Penelitian

Sesuai rumusan masalah yang telah diuraikan di atas, maka tujuan dari penelitian ini adalah sebagai berikut:

1. Untuk mengetahui klasifikasi produk kecenderungan pembelian produk terlaris berdasarkan riwayat transaksi September 2022 – Januari 2023
2. Untuk mengetahui tingkat akurasi dalam memprediksi produk mana yang banyak diminati pelanggan (laris dan tidak laris) ?

1.6 Manfaat Penelitian

Setiap karya memiliki kelebihan yang dapat diambil dari penelitian ini, juga dari sudut pandang teoritis dan praktis. Di bawah ini adalah manfaat masing-masing dari penelitian ini:

1. Aspek Teoritis.

Investigasi ini dapat digunakan untuk mengadaptasi konsep-konsep yang dipelajari di universitas dan mengintegrasikannya ke dalam situasi saat ini, serta memperluas keahlian dan pemahaman editor mengenai tantangan perusahaan, khususnya analisis data menggunakan teknik Nave Bayes.

2. Aspek Praktis.

Para ilmuwan berteori tentang bagaimana penelitian ini akan mampu menerapkan pengetahuan pada bidang teknis ini:

- a. Para akademisi akan memanfaatkan penelitian ini karena akan meningkatkan pemahaman mereka tentang pencarian informasi menggunakan pendekatan Nave Bayes.
- b. Untuk bisnis tersebut: Temuan dari penelitian ini dimaksudkan untuk memberikan data dan saran yang dapat digunakan oleh bisnis untuk meningkatkan bisnis.
- c. Untuk universitas: Penelitian ini dimaksudkan untuk membantu individu di masa depan yang ingin melakukan penelitian lebih lanjut, terutama yang terkait dengan pengumpulan data menggunakan teknik Nave Bayes, dengan menyediakan data penting.

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*

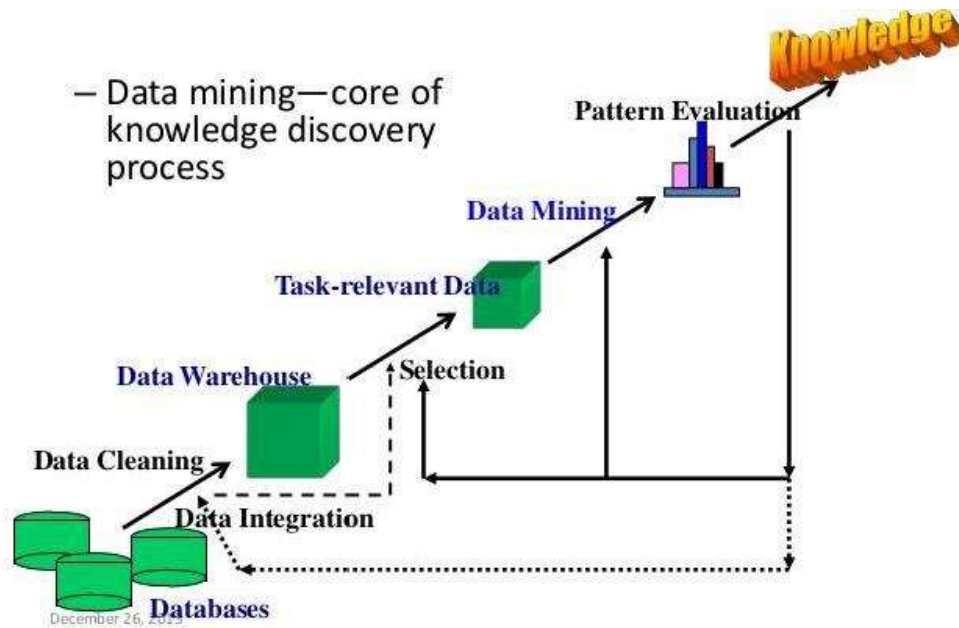
Ruang untuk penyesuaian adalah teknik yang digunakan untuk mengungkap wawasan berguna yang terkubur dalam statistik.

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) Penambahan 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 Universitas 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 dikarnakan 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 dikarnakan 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

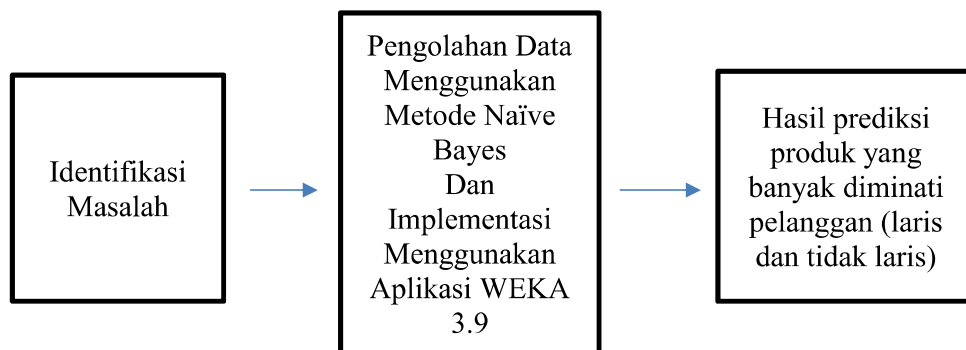
9. Anti bocor pada unit indoor

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 0.945 (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.

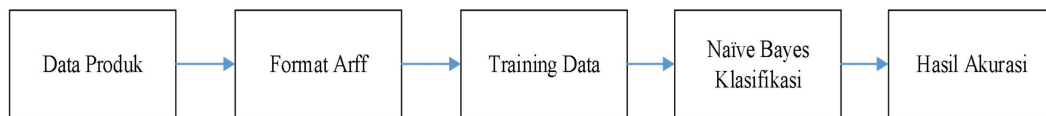
Penelitian kesepuluh oleh Muhammad Juanda Saputra, Muhammad Izman Herdiansyah (2022) berjudul Penerapan Naive Bayes Dalam Memprediksi Penjualan Dan Persediaan Kain Jumputan Pada Toko Batiq Colet Tuan Kentang Palembang. Menunjukkan Algoritma Naïve Bayes telah berhasil diterapkan untuk menganalisis prediksi 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%.

BAB III

METODE PENELITIAN

3.1 Desain Penelitian

Pada desain penelitian, peneliti memaparkan hasil dari yang telah dibuat dengan cara yang terstruktur agar dapat diambil sebagai acuan dalam melakukan penelitian dan dapat membantu peneliti pada saat proses penelitian, desain penelitian ini juga dapat berperan sebagai pedoman bagi peneliti.



Gambar 3. 1 Design penelitian

Adapun desain penelitian pada penelitian ini dijelaskan sebagai berikut:

1. Data Produk

Data ini didapat dari tempat penelitian langsung. Terdapat atribut yang digunakan yaitu Merk, Type, Daya, PK, Harga, Jenis, Alamat, Bulan, Status Laris dan tidak laris.

2. Format Afr

WEKA mendukung beberapa format file dalam inputnya diantaranya adalah format arff, Data produk tersebut di buat menggunakan NotePad++ dengan format arff, karena WEKA merupakan tipe file text yang berisi berbagai instance data yang berhubungan dengan suatu set

attribute data. Kemudian data-data ini disusun dengan menggunakan tanda koma (,) sebagai pemisah data peratributnya.

3. Training Data

Pada tahap ini dilakukan proses training oleh algoritma klasifikasi naïve bayes untuk membentuk sebuah model classifier pada data produk yang sudah ada. Model ini merupakan representasi pengetahuan yang akan digunakan untuk prediksi kelas data baru yang belum ada.

4. Naïve Bayes klasifikasi

Proses pengklasifikasi data menggunakan tools WEKA terhadap data produk yang sudah di testing menggunakan model classifier yang sudah dibentuk.

5. Hasil Akurasi

Hasil penelitian yang telah diuji melalui perhitungan teknik data mining dengan metode Naïve Bayes serta diuji menggunakan aplikasi WEKA, yang menunjukkan kedekatan hasil pengukuran dengan nilai sesungguhnya.

3.2 Objek Penelitian

3.2.1 Lokasi Penelitian

Penelitian ini berlokasi di Daikin Pro Shop Batam Tlk. Tering, Kec. Batam Kota, Kota Batam, Kepulauan Riau 29444.

3.3 Populasi dan Sampel

3.3.1 Populasi Penelitian

Populasi ialah wilayah generalisasi (golongan dari banyak wilayah) yang terdiri dari objek atau subjek dengan beberapa karakteristik yang ditetapkan oleh pengkaji buat pengkajian ini lalu dapat ditarik kesimpulannya (Anggraini et al., 2020).

Dalam pengkajian ini yang dijadikan populasi ialah semua pembeli yang melakukan transaksi pada bulan September 2022 hingga Januari 2023 di PT DLS Batam.

3.3.2 Sampel

Dalam penelitian ini, teknik pengambilan sampel yang digunakan adalah probability sampling, yaitu teknik pengambilan sampel yang memberikan kesempatan yang sama bagi setiap elemen atau anggota populasi untuk dipilih sebagai sampel. Pemeriksa melakukan pengambilan sampel dengan menggunakan easy random sampling, artinya pengambilan sampel secara acak.

3.4 Variabel Penelitian

Operasional variabel yang digunakan dalam penelitian ini adalah data penjualan di PT DLS. Untuk mendukung penelitian ini, peneliti mengambil hasil data penjualan produk AC yang ada di perusahaan tersebut.

Tabel 3. 1 Atribut dalam data training

Nama Atribut	Nilai Atribut
Merk	Daikin Thailand =1

	Daikin Malaysia =2
Type	FTK =1 FTV = 2 FTC = 3 FTX = 4
Daya	Low Watt = 1 (400 Watt) Standart Watt Middle = 2 (600 Watt) High Watt = 3 (800 Watt)
PK	1/2 - 1 PK = 1 2 - 2.5 PK = 2
Harga	Murah = 1 Dikatakan murah jika harga di dibawah 10 juta Mahal = 2 Dikatakan mahal jika harga diatas 10 juta
Jenis	Inverter = 1 Splitt = 2
Alamat	Batu Aji = 1 Sagulung = 2 Batam Centre = 3 Bengkong = 4 Tiban = 5

	Sekupang = 6
Bulan	September = 1 Oktober =2 November =3 Desember =4 Januari =5
Status	Laris = 1 Tidak Laris = 2 Data diambil dari hasil penjualan Rata-Rata perbulanya.

Adapun jenis AC yang tergolong laris pada penjualan dari bulan september sampai bulan januari yakni :

Tabel 3. 2 Penjualan AC Laris September

Merk	Type	Daya	PK	Harga	Jenis	Jumlah	status
Daikin Thailand	FTK	High Watt	1/2 - 1 PK	Murah	Splitt	8	Laris
Daikin Thailand	FTK	Standart Watt Middle	1/2 - 1 PK	Murah	Inverter	25	Laris
Daikin Thailand	FTK	Low Watt	2 - 2.5 PK	Murah	Inverter	46	Laris
Daikin Thailand	FTV	High Watt	1/2 - 1 PK	Murah	Splitt	2	Tidak Laris
Daikin Malaysia	FTV	High Watt	1/2 - 1 PK	Murah	Splitt	1	Tidak Laris
Daikin Malaysia	FTV	Low Watt	2 - 2.5 PK	Murah	Inverter	1	Tidak Laris
Daikin Thailand	FTV	Low Watt	1/2 - 1 PK	Murah	Inverter	1	Tidak Laris
Daikin Thailand	FTK	High Watt	2 - 2.5 PK	Murah	Inverter	2	Tidak Laris

Daikin Malaysia	FTK	High Watt	2 - 2.5 PK	Murah	Splitt	1	Tidak Laris
Daikin Thailand	FTK	High Watt	1/2 - 1 PK	Murah	Inverter	10	Laris
Daikin Malaysia	FTK	Standart Watt Middle	1/2 - 1 PK	Murah	Inverter	1	Tidak Laris
Daikin Thailand	FTK	Low Watt	2 - 2.5 PK	Mahal	Splitt	1	Tidak Laris
Daikin Malaysia	FTK	High Watt	1/2 - 1 PK	Murah	Inverter	4	Tidak Laris
Daikin Thailand	FTX	Low Watt	1/2 - 1 PK	Murah	Inverter	1	Tidak Laris
Daikin Thailand	FTX	High Watt	1/2 - 1 PK	Murah	Inverter	1	Tidak Laris
Daikin Malaysia	FTX	High Watt	1/2 - 1 PK	Murah	Inverter	1	Tidak Laris
Daikin Thailand	FTC	High Watt	1/2 - 1 PK	Murah	Inverter	5	Tidak Laris
Daikin Malaysia	FTC	Low Watt	2 - 2.5 PK	Murah	Splitt	1	Tidak Laris
Daikin Thailand	FTC	Standart Watt Middle	1/2 - 1 PK	Murah	Inverter	1	Tidak Laris
Daikin Malaysia	FTK	Standart Watt Middle	2 - 2.5 PK	Murah	Splitt	1	Tidak Laris
Daikin Malaysia	FTK	Low Watt	2 - 2.5 PK	Murah	Splitt	1	Tidak Laris
Daikin Malaysia	FTV	High Watt	1/2 - 1 PK	Murah	Inverter	3	Tidak Laris
Daikin Thailand	FTV	Standart Watt Middle	1/2 - 1 PK	Murah	Inverter	4	Tidak Laris
Daikin Thailand	FTV	High Watt	2 - 2.5 PK	Murah	Inverter	1	Tidak Laris
Daikin Thailand	FTK	Low Watt	1/2 - 1 PK	Murah	Inverter	1	Tidak Laris
Daikin Malaysia	FTV	Low Watt	2 - 2.5 PK	Murah	Splitt	1	Tidak Laris
Total						125	
Rata-rata						8.33	

Diketahui pada bulan september jenis ac yang tergolong laris berdasarkan nilai rata-rata adalah Daikin Thailand, FTK, Low watt, 2-2,5 PK, Murah, Inverter dengan jumlah penjualan 46.

Tabel 3. 3 Penjualan AC Laris Oktober

Merk	Type	Daya	PK	Harga	Jenis	Jumlah	Status
Daikin Thailand	FTK	Standart Watt Middle	1/2 - 1 PK	Murah	Inverter	17	Laris
Daikin Thailand	FTK	High Watt	2 - 2.5 PK	Mahal	Splitt	18	Laris
Daikin Thailand	FTV	High Watt	1/2 - 1 PK	Murah	Splitt	29	Laris
Total						64	
Rata-Rata						4.2666 67	

Diketahui pada bulan oktober jenis ac yang tergolong laris berdasarkan nilai rata-rata adalah Daikin Thailand, FTV, High watt, 1/2-1 PK, Murah, splitt dengan jumlah penjualan 29.

Tabel 3. 4 Penjualan AC Laris November

Merk	Type	Daya	PK	Harga	Jenis	Jumlah	Status
Daikin Thailand	FTV	High Watt	1/2 - 1 PK	Murah	Splitt	10	Laris
Daikin Thailand	FTX	Low Watt	1/2 - 1 PK	Murah	Inverter	21	Laris
Total						31	
Rata-rata						2.06667	

Diketahui pada bulan november jenis ac yang tergolong laris berdasarkan nilai rata-rata adalah Daikin Thailand, FTX, Low watt, 1/2-1 PK, Murah, inverter dengan jumlah penjualan 21.

Tabel 3. 5 Penjualan AC Laris Desember

Merk	Type	Daya	PK	Harga	Jenis	Jumlah	Status
Daikin Thailand	FTX	Low Watt	1/2 - 1 PK	Murah	Inverter	6	Tidak Laris
Total						6	
Rata-rata						0.4	

Diketahui pada bulan desember jenis ac yang tergolong tidak laris berdasarkan nilai rata-rata adalah Daikin Thailand, FTX, Low watt, 1/2-1 PK, Murah, inverter dengan jumlah penjualan 6.

Tabel 3. 6 Penjualan AC Laris Januari

Merk	Type	Daya	PK	Harga	Jenis	Jumlah	Status
Daikin Thailand	FTX	Low Watt	1/2 - 1 PK	Murah	Inverter	100	Laris
Daikin Thailand	FTK	Standart Watt Middle	2 - 2.5 PK	Mahal	Inverter	12	Laris
Total						112	
Rata-rata						7.46667	

Diketahui pada bulan januari jenis ac yang tergolong laris berdasarkan nilai rata-rata adalah Daikin Thailand, FTX, Low watt, 1/2-1 PK, Murah, inverter dengan jumlah penjualan 100

3.5 Pengumpulan Data

Pengumpulan data dilakukan dengan tujuan buat mendapatkan berita informasi yang bisa mempermudah dan melancarkan dalam mencapai tujuan pengkaji. Adapun metode yang digunakan di pengkajian ini menjadi berikut :

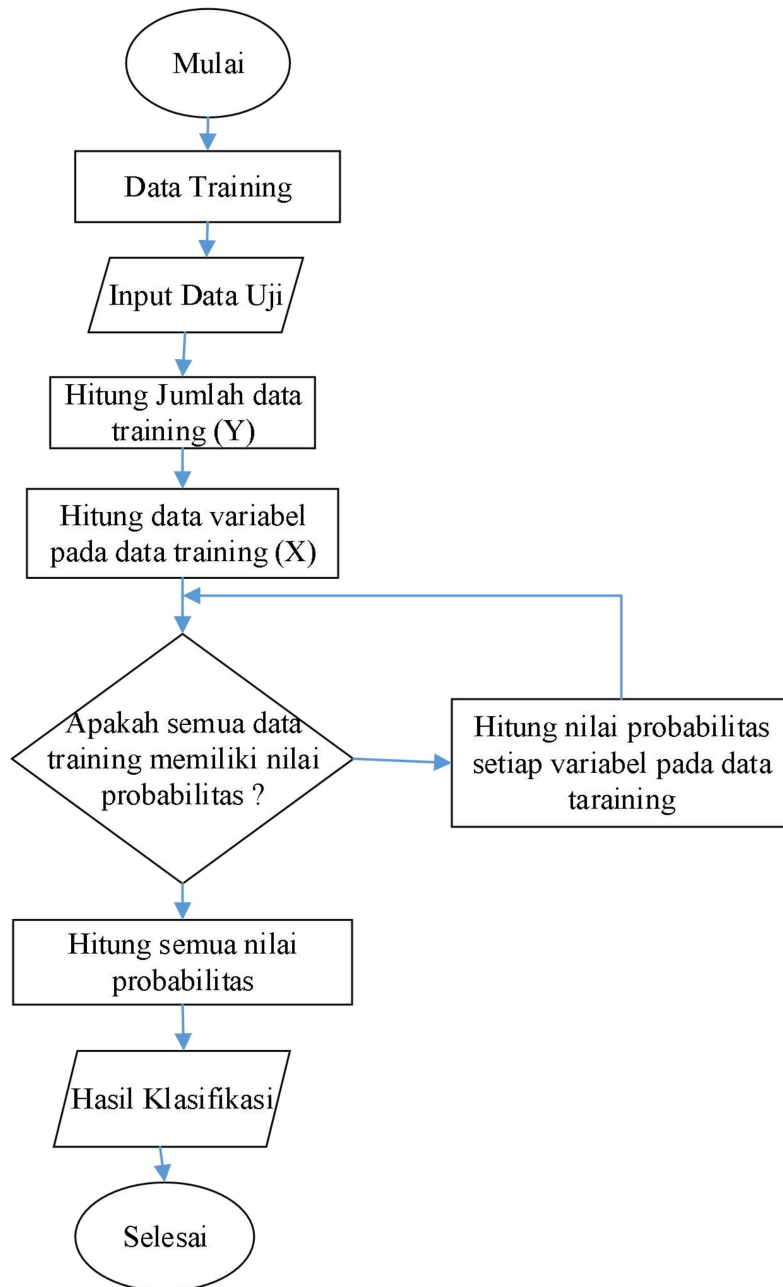
1. Teknik Observasi di termin observasi pengkaji melakukan pengamatan secara eksklusif buat tahu insiden di PT DLS, pada hal ini pengkaji

mengambil data yang akan terjadi transaksi penjualan yang ada di database sesuai menggunakan data yg diperlukan dalam menyelesaikan persoalan di pengkajian ini.

2. Studi Pustaka merupakan pengkaji mengkaji memakai mencari informasi melalui asal-asal tertulis seperti kitab dan jurnal yang mempunyai korelasi menggunakan pengkajian ini.

3.6 Model Penelitian

Pada penelitian ini model yang digunakan bisa dilihat pada *flowchart* dibawah ini :



Gambar 3. 2 Flowchart Algoritma Naive Bayes

Berdasarkan model diatas dapat jelaskan bahwa proses awal yan dilakukan pada penelitian ini yaitu memisahkan terlebih dahulu antara data numerik dan tidak. Kemudian akan dicari nilai setiap probabilitas setiap atribut yang tedapat pada data training. Setelah ditemukan hasil probabilitas setiap atribut maka data

set akan dihitung dengan mencari probabilitas laris dan tidak laris pada setiap data set. Setelah diketahui setiap probabilitas akan dibandingkan mana yang terlaris maka itulah hasil dari prediksi algoritma Naïve Bayes.