BAB II

LANDASAN TEORI

2.1. *Knowledge Discovery in Database* (KDD)

Knowledge Discovery in Database adalah bagian dari ilmu pengetahuan yang mencakup database serta dapat diuraikan sebagai proses pengenalan pola kepada pengetahuan yang memiliki sekelompok besar data dan menghasilkan informasi yang mudah untuk dipahami. KDD memiliki beberapa proses untuk mendapatkan data yang ingin dicapai seperti pemurnian data, integrasi data, dan pemilihan data dan presentasi pengetahuan (Elmayati, 2017).

Data mining adalah akuisisi pengetahuan dalam praktiknya meliputi pembersihan data, integrasi data, pemilihan data, konversi data,analisis data. Pengumpulan data mengacu pada proses penggalian informasi dari sejumlah besar data untuk mendapatkan informasi baru tentang situs tertentu. (Fajrin & Handoko, 2018).

Knowledge Discovery in Database (KDD) merupakan proses untuk menentukan informasi yang bermanfaat seperti berbentuk suatu data yang bersifat baru dan bisa digunakan (Gukguk & Sitohang, 2021). Data informasi yang berukuran besar berguna untuk pola-pola yang ada dalam data dan potensinya yang bermanfaat. Data mining merupakan salah satu langkah dari serangkaian proses iterative KDD. Menurut (Ikhwan, 2015) langkah-langkah proses *Knowledge Discovery in Database* (KDD) yang dilakukan diawali pada gambar 2.1 sebagai berikut:

Gambar 2.1 *Knowledge Discovery in Database* (KDD) (Ikhwan, 2015)

1. Data *selection* (seleksi data)

   Sebelum tahap information *mining* KDD dimulai, data perlu diseleksi dari sekumpulan data operasional. Data yang dipilih akan diproses dan data disimpan dalam file.

2. *Pre-processing* / pembersihan

   Sebelum menjalankan proses data *mining*, harus melalui proses pembersihan data yang menjadi perhatian KDD. Proses pembersihan secara khusus meliputi deduplikasi, pengecekan data yang tidak konsisten, dan perbaikan kesalahan data.

3. Transformasi

   Proses transformasi untuk data yang dipilih. Hal ini membuat data cocok untuk proses data *mining*. Proses pengkodean KDD adalah proses kreatif yang bergantung pada jenis dan pola informasi yang di cari di database.

4. Data *mining*

   Proses menemukan pola atau informasi yang menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Pilihan metode dan

algoritma yang benar sangat berpengaruh pada tujuan proses KDD secara semuanya.

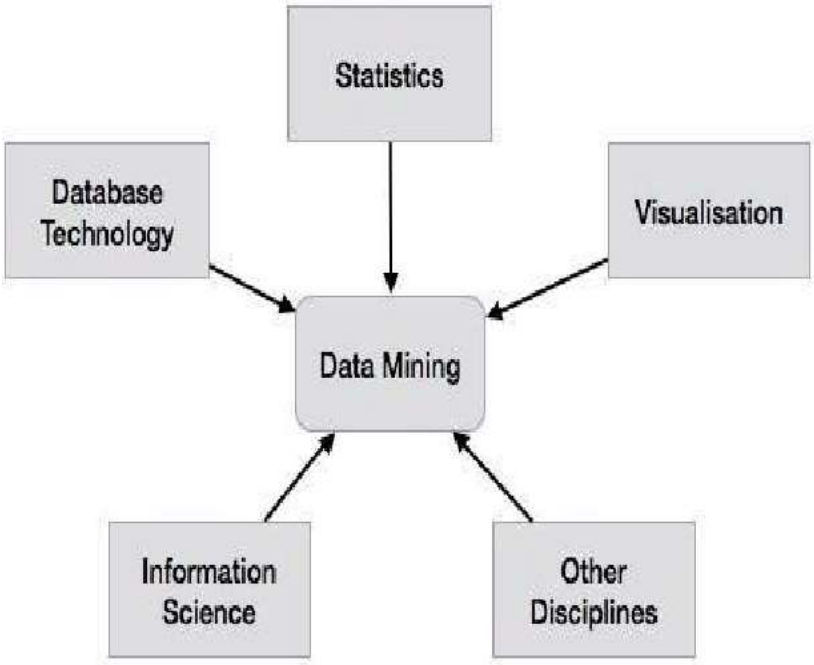
5. Interpretasi /evaluasi

   Pola informasi yang dihasilkan oleh proses data *mining* harus ditampilkan dalam format yang mudah dipahami oleh para pemangku kepentingan. Tahap ini merupakan bagian dari proses KDD dan disebut interpretasi.

2.2. Data Mining

Data mining merupakan kumpulan dari proses Kumpulan sistem suatu informasi penunjang dalam pengambilan keputusan pada data besar disebut juga dengan *data mining* (Fajrin & Handoko, 2018). *Data mining* dapat diukur di tempat yang berbeda dengan data yang berbeda. Menurut Daryl Pregibon, penambangan data adalah kombinasi dari banyak sistem penyimpanan dan pengambilan data yang cerdas. Salah satu metode yang digunakan dalam penambangan data adalah menyempurnakan riset pasar untuk menciptakan strategi pemasaran atau pasar yang efektif menggunakan data pasar yang tersedia untuk perusahaan.

Pengambilan informasi dapat digunakan di berbagai situs, termasuk informasi. Menurut Daryl dalam (Fajrin & Handoko, 2018), pengolahan data merupakan kombinasi dari statistik, keahlian dan pengumpulan data. Salah satu metode yang digunakan dalam pencarian informasi adalah aturan perusahaan yang menganalisis produk untuk menetapkan strategi praktis untuk penjualan atau periklanan dengan menggunakan informasi penjualan yang tersedia di perusahaan.



Gambar 2.2 Ilmu *Data Mining* (Elmayanti, 2017)

Pada *data mining* terdapat bagian yang menjadi fungsi utamanya yaitu fungsi *descriptive* dan fungsi *predictive.*

1. *Descriptive*

   Descriptive berfungsi sebagai informasi dari analisis data untuk mendapatkan pemahaman yang lebih baik dari informasi yang dipantau atau yang telah diamati. Setelah proses, berharap untuk mengetahui sifat data. Dari data yang telah di proses tersebut maka dapat dilihat jenis informasi yang terkait. Dengan menggunakan fungsi penggalian data, dapat menemukan pola lain yang tersembunyi dalam rekaman. Dengan kata lain, ketika model diulang dan sesuai, elemen data dapat diidentifikasi.

2. *Estimation*

   Model ini disusun menggunakan *database* yang lengkap dan didasarkan pada nilai tukar yang diharapkan. Selain itu, dalam analisis lain, *varians* yang diestimasi didasarkan pada *varians* yang diharapkan. Misalnya, kami

mengevaluasi tekanan darah sistolik pada pasien rumah sakit berdasarkan usia pasien, jenis kelamin, berat badan, dan kadar natrium darah. Korelasi antara variabel prognostik dalam tekanan darah sistolik dan proses pembelajaran memungkinkan terciptanya model hasil. Model penilaian yang dihasilkan dapat digunakan untuk masalah baru lainnya (Harahap dkk., 2019).

3. *Predictive*

Prediktif adalah bagian dari bagaimana satu pola menangkap data untuk pola lain. Pola-pola ini dapat diidentifikasi dengan perubahan yang berbeda dalam data. Setelah pola menemukan modelnya, maka pola dapat menggunakan model hasil untuk memprediksi beberapa variasi dalam nilai atau jenis yang tidak diketahui. Oleh karena itu, karya ini dikatakan sebagai karya prediksi dan peramalan. Fungsi ini juga dapat digunakan untuk mengidentifikasi variabel tertentu yang tidak ada dalam data. Tugas ini mudah dan nyaman bagi siapa saja yang membutuhkan informasi akurat untuk memperbaiki masalah penting ini.

Pada *data mining* terdapat fungsi yang perlu diketahui, yaitu klasifikasi, *asosiasi* dan *clauster*.

1. Klasifikasi

Klasifikasi adalah bagian dari aktivitas yang menyusun suatu sistem atau data dalam bentuk berkelompok atau golongan dengan standar yang telah ditetapkan. Metode ini adalah metode yang paling umum digunakan metode ini bertujuan untuk memperkirakan kelas dari suatu objek yang lebel nya belum diketahui. Kelas adalah jenis tertinggi di mana dua kalimat adalah "ya" dan "tidak."(Ulfha & Amin, 2020).

2. Asosiasi

Asosiasi merupakan teknik yang digunakan untuk menemukan suatu hubungan yang menunjukan suatu kondisi di dalam satu set data yang beberapa nilai atribut akan muncul secara bersamaan. Analisis hubungan mencakup studi tentang “apa dan apa”. Contoh: 90% konsumen membeli produk kain dan 60% konsumen membeli produk kain dan pakaian.

Analisis asosiasi ini akan menghasilkan aturan asosiasi *(association rules).* Contoh: 90% orang yang berbelanja di suatu toko yang membeli gamis juga membeli kerudung, dan 60 % dari semua orang yang berbelanja membeli keduanya. Sekumpulan *record,* masing-masing berisi kumpulan elemen tertentu, menciptakan aturan untuk mengamati salah satu dari empat elemen dari sesuatu yang terjadi. Aturan umumnya sering disebut sebagai riset pasar, karena menganalisis data transaksi pelanggan untuk menentukan produk mana yang dijual.(Despitaria dkk., 2016).

3. *Cluster*

Cluster adalah topik sederhana untuk mengklasifikasikan data ke dalam satu kategori objek yang tidak didasarkan pada kumpulan data tertentu. *Cluster* memiliki tujuan untuk mengelompokan suatu kelas ke dalam beberapa segmen berdasarkan atribut tang telah ditentukan. Mengklasifikasikan, memperkirakan, atau memprediksi nilai dari suatu variabel yang diestimasi. Namun, *algoritma clustering* mencoba untuk membagi semua data menjadi segmen-segmen yang identik, di mana data di satu bagian lebih berharga, tetapi di bagian lain datanya kurang penting (Elmayati, 2017).

2.1.1. Tingkatan-tingkatan *Data mining*

Menurut (Efori Buulolo, 2017) ada beberapa tingkatan pada *data mining* yaitu:

1. Pembersihan data adalah metode yang digunakan untuk menghapus data yang tidak akurat dan tidak sesuai.
2. *Integrasi* data dapat digunakan dalam satu *database.*
3. Pilihan data yang akan diperoleh harus konsisten dengan hasil analisis
4. Memperbarui data Pemrosesan data memerlukan format dan konten *Internet* yang tepat.
5. Proses pengiriman.
6. Selama proses peninjauan, coba tunjukkan pola yang menarik kontras dengan informasi tersembunyi.
7. Manifestasi kebijaksanaan. Ini adalah proses yang memberikan informasi dan pengalaman yang membantu pengguna tetap mendapat informasi tentang hasil data mereka

2.1.2. Manfaat *Data Mining*

Menurut (Efori Buulolo, 2017) dilihat dari sudut pandang terdapat dua manfaat yaitu:

1. Sudut Pandang Komersial.

   Dari sudut pandang komersial, *data mining* dapat digunakan untuk menentukan jumlah data. Cara lain untuk bersaing adalah dengan menggunakan metode *data mining* untuk menangkap data dan hasil.

2. Perspektif ilmuwan

   Data mining merupakan bagian yang berfungsi sebagai analisis data yang dapat menyimpan data besar.

2.3. Metode Naïve Bayes

Naïve Bayes merupakan salah satu metode yang terdapat pada *data mining*. Algoritma Naïve Bayes adalah teknik klasifikasi berdasarkan penerapan logika bayes dengan asumsi yang kuat bahwa semua prediktor independen satu sama lain dengan kata lain bahwa adanya kehadiran fitur kelas tidak tergantung pada kehadiran fitur kelas lain dikelas yang sama. Agoritma Naïve Bayes banyak digunakan dalam proses prediksi karena memiliki kemampuan prediksi di bulan yang akan datang (Wijaya & Dwiasnati, 2020)

Algoritma Naïve Bayes adalah cara langsung memprediksikan pola data menggunakan ramalan penjualan bulan depannya. Oleh karena itu, Naïve Bayes adalah algoritma yang baik untuk meramalkan penjualan (Wijaya & Dwiasnati, 2020). Peneliti menggunakan metode algoritma Naïve Bayes merupakan metode statistik yang berdasarkan pengenalan pola, menggunakan *Probability* dan biaya yang dihasilkan dari keputusan ini untuk menyelidiki proses klasifikasi disebut *Teorema Bayes* (Fitri, 2017).

Pengklasifikasi *Probability* dengan mudah menghitung satu set probabilitas dan kombinasi beberapa level atau hasil dari kumpulan data yang diperoleh. Kedekatan dan kecepatan Naive Bayesian dalam database data besar (Saleh, 2015). Berikut rumus 2.1 *Probability Bayes*:

$$P(x|y) = \frac{P(x \cap y)}{P(y)} \qquad 2.1$$

Probability Bayes Probability X pada Y artinya *Probability* interaksi X dan Y dari probabilitas Y. Atau disebut dengan P(X|Y) artinya persentase banyak nya X di dalam Y. Berikut rumus 2.2 *Teorema Bayes*:

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)} \qquad 2.2$$

Penjelasan rumus 2.2 X artinya ciri-ciri, H artinya hipotesis, P(H|X) artinya *Probability* bahwa hipotesis H benar untuk ciri X atau disebut dengan P(H|X) adalah *Probability posterior* H dengan ketentuan X, P(X|H) artinya *Probability* bahwa ciri X benar untuk hipotesis H atau *Probability posterior* X dengan ketentuan H, P(H) adalah *Probability prior* hipotesis H, dan P(X) artinya *Probability prior* bukti X. Berikut Rumus 2.3 *Probability Bayes*:

$$P(Y) = \frac{P(X \cap Y)}{P(Y)} \qquad 2.3$$

Probability X didalam Y adalah *Probability* interseksi X dan Y dari *Probability* Y, atau disebut dengan P(X|Y) artinya tingkatan banyaknya X didalam Y. Berikut rumus 2.4 *Teorema bayes*:

$$P(X) = \frac{P(H)P(H)}{P(X)} \qquad 2.4$$

Penjelasan diatas X artinya ciri-ciri, H artinya hipotesis, P(H|X) artinya *Probability* bahwa hipotesis H benar untuk ciri x atau disebut dengan P(H|X) artinya *Probability posterior* H dengan ketentuan X, P(X|C) artinya *Probability*.

2.4. *Software* Pendukung

Software pendukung merupakan suatu perangkat lunak yang digunakan oleh peneliti untuk mendukung penelitian, RapidMiner merupakan software pendukung yang digunakan oleh peneliti. RapidMiner merupakan salah satu cara yang sering

digunakan dalam menemukan solusi pada *data mining* karena mampu dijalankan pada berbagai sistem operasi. RapidMiner adalah sebuah perangkat lunak yang memiliki sifat *open source.* Sebelum dikenal dengan nama RapidMiner, perangkat ini memiliki nama lain yaitu *YALE (Yet Another Learning Environment)* yang dikembangkan pada tahun 2001 oleh Ralf Klinkenberg.

Menurut Knuggets RapidMiner adalah bagian yang teratas dalam sebagai *software data mining*. RapidMiner menyediakan *GUI (Graphic Interface User)* untuk desain detail *pipeline*. *GUI* ini membuat file *XML* (bahasa penandaan tambahan) yang menentukan metode penyortiran yang harus digunakan pengguna untuk data. File ini dibaca oleh RapidMiner dan diaktifkan langsung dari penganalisis.

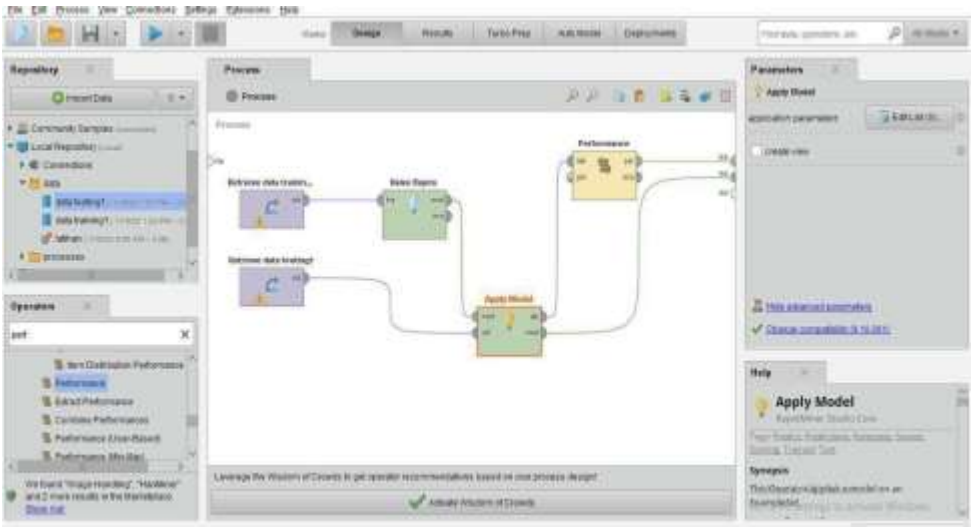
RapidMiner memiliki beberapa sifat sebagai berikut:

a) Java ditulis dalam bahasa pemrograman sehingga dapat berjalan di sistem operasi yang berbeda
b) Cara memperoleh informasi ada dalam model pohon
c) *XML Internal*
d) Bahasa *scripting* memungkinkan untuk eksperimen skala besar
e) Rancangan *multi-layer* untuk menjaga bentuk data yang efisien dan menjaga penangan data.
f) Mempunyai *GUI, command line* mode dan *Java API* yang dapat dipanggil dari program lain.

Beberapa fitur dari RapidMiner, antara lain:

1) Seperti *decision tree* dan *self organizing map* adalah algoritma *data mining*.

2) Tree *chart* dan *3D scatter plots* bentuk grafis yang terbaru.
3) Seperti *text plugin* untuk melakukan analisis teks.
4) Menyuplai prosedur *data mining* dan *machine learning* termasuk: *ETL (extraction, transformation, loading) data preprocessing, visualisasi, modeling dan evaluasi.*
5) *Data mining* memiliki proses yang teratur dideskripsikan dengan *XML* dan dibuat dengan *GUI.*
6) Menggabungkan proyek *data mining* weka dan statistik R.



Gambar 2.3 Aplikasi *Rapidminner* (2025)

2.5 Penelitian Terkait

Penelitian ini dilakukan tidak terlepas dari hasil penelitian-penelitian terkait yang pernah dilakukan sebagai bahan perbandingan dan kajian. Penelitian terkait dipergunakan untuk mempermudah dalam menemukan langkah-langkah untuk penyusunan penelitian dari segi teori maupun konsep. Berikut merupakan beberapa jurnal yang terkait dengan penelitian yang dilakukan:

Tabel 2.1 Penelitian Terkait

No	Nama Peneliti, Judul, Tahun	Hasil Penelitian
1	Nur Ajijah, Dwiza Riana (2019) Algoritma Naïve Bayes, *Decision Tree*, dan SVM untuk Klasifikasi Persetujuan Pembiayaan Nasabah Koperasi Syariah	SVM untuk klasifikasi data pembiayaan nasabah koperasi syariah ke dalam kelas macet dan lancar menghasilkan kinerja terbaik dengan akurasi 89,86% dan AUC 0,939 dibandingkan Naïve Bayes dan *Decision Tree*. Sistem Pendukung Persetujuan Pembiayaan (SPPP) menggunakan SVM menghasilkan akurasi klasifikasi sebesar 94% dari 50 *record* dataset.
2	Herry Derajad Wijaya, Saruni Dwiasnati (2020) Implementasi Data Mining dengan Algoritma Naïve Bayes pada Penjualan Obat	Pengujian pada data rekapitulasi penjualan Obat dengan proses mining Algoritma Naïve Bayes menghasilkan tingkat *accuracy* dengan nilai 88.00%, dimana dalam pengujian model data, keseluruhan data set digunakan sebagai data testing. Penelitian ini dilakukan menggunakan *tools* RapidMiner pada dataset penjualan obat dengan menggunakan algoritma Naïve Bayes.
3	Mujib Ridwan, Hadi Suyono, dan M. Sarosa (2023) Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naïve Bayes *Classifier*	Hasil pengujian menunjukkan bahwa faktor yang paling berpengaruh dalam penentuan klasifikasi kinerja akademik mahasiswa yaitu Indeks Prestasi Kumulatif (IPK), Indeks Prestasi (IP) semester 1, IP semester 4, dan jenis kelamin. Sehingga faktor-faktor tersebut dapat digunakan sebagai bahan evaluasi bagi pihak pengelola perguruan tinggi. Pengujian pada data mahasiswa angkatan 2005-2009, algoritma NBC menghasilkan nilai *precision*, *recall*, dan *accuracy* masing-masing 83%, 50%, dan 70%.

No	Nama Peneliti, Judul, Tahun	Hasil Penelitian
4	Haditsah Annur (2018) Klasifikasi Masyarakat Miskin Menggunakan Metode Naïve Bayes	Sistem klasifikasi masyarakat miskin di wilayah pemerintahan Kecamatan Tibawa Kab. Gorontalo dapat direkayasa dan Berdasarkan hasil pengujian *confusion matrix* dengan teknik split validasi, penggunaan metode klasifikasi Naïve Bayes terhadap dataset yang telah diambil pada objek penelitian diperoleh tingkat akurasi sebesar 73% atau termasuk dalam kategori *Good*. Sementara nilai *Precision* sebesar 92% dan *Recall* sebesar 86%.
5	Aji Ghassa dkk., (2022) Penerapan *Data Mining* Algoritma Naïve Bayes *Classifier* Untuk Mengetahui Minat Beli Pelanggan Terhadap INDIHOME	Sistem klasifikasi data INDIHOME digunakan untuk menampilkan informasi klasifikasi layak atau Tidak layak minat masyarakat pada INDIHOME ini dengan menggunakan algoritma Naïve Bayes. Algoritma Naïve Bayes sangat cocok diterapkan dalam memprediksi peluang di masadepan berdasarkan pengalaman dimasa sebelumnya sehingga memudahkan perusahaan untuk memprediksi permintaan masyarakat terhadap jenis paket INDIHOME yang baru.Dengan mengetahui Layak atau Tidak Layak INDIHOME ini, akan meminimalisir kerugian pada perusahaan
6	Putro dkk., (2020) Penerapan Metode Naïve Bayes Untuk Klasifikasi Pelanggan	Berdasarkan master pelanggan yang dijadikan data latih, telah berhasil mengklasifikasikan 23 data dari 25 data yang diuji. Sehingga berhasil memprediksi pelanggannya dengan nilai *precision* mencapai 100%, nilai *recall* mencapai 91%, nilai *accuracy* mencapai 92%.
7	Yogiek Indra Kurniawan (2018) Perbandingan Algoritma Naïve Bayes Dan C.45 Dalam Klasifikasi Data Mining	Berdasarkan hasil pengujian, semakin banyaknya data *training* yang digunakan, maka nilai *precision*, *recall* dan *accuracy* akan semakin meningkat. Selain itu, hasil

No	Nama Peneliti, Judul, Tahun	Hasil Penelitian
		klasifikasi pada algoritma Naïve Bayes dan C.45 tidak dapat memberikan nilai yang absolut atau mutlak di setiap kasus. Pada kasus penentuan penerimaan Kartu Indonesia Sehat, kedua buah algoritma tersebut sama-sama efektif untuk digunakan. Untuk kasus pengajuan kartu kredit di sebuah bank, C.45 lebih baik daripada Naïve Bayes. Pada kasus penentuan usia kelahiran, Naïve Bayes lebih baik daripada C.45. Sedangkan pada kasus penentuan kelayakan calon anggota kredit di koperasi, Naïve Bayes memberikan nilai yang lebih baik pada *precision*, tapi untuk *recall* dan *accuracy*, C.45 memberikan hasil yang lebih baik. Sehingga untuk menentukan algoritma terbaik yang akan dipakai di sebuah kasus, harus melihat kriteria, *variable* maupun jumlah data di kasus tersebut.
8	K.Vembandasamy dkk., (2015) *Heart Diseases Detection Using* Naïve Bayes *Algorithm*	Hasil penelitian menunjukkan algoritma *naive bayes* mampu digunakan untuk mengklasifikasikan kumpulan data karena *naive bayes* memberikan hasil yang akurat, dengan hasil ini penyakit jantung pada manusia dapat diprediksi. Dengan demikian sistem prediksi penyakit jantung berhasil mendiagnosis data medis dan memprediksi penyakit jantung. Hasil yang diperoleh menunjukkan bahwa algoritma *naive bayes* memberikan akurasi 86.4198% dengan waktu minimum.
9	Sri Wahyuni, Murni Marbun (2020) *Implementation of Data Mining In Predicting the Study Period of Student Using the* Naïve Bayes *Algorithm*	Hasil penelitian menunjukkan bahwa aplikasi *data mining* yang dirancang dapat memprediksi masa studi mahasiswa Universitas Pembangunan Panca Budi, hasil implementasi sistem data mining

No	Nama Peneliti, Judul, Tahun	Hasil Penelitian
		yang dirancang dapat dimanfaatkan untuk membuat strategi kelulusan mahasiswa tepat waktu dan menekankan pada pencegahan mahasiswa *drop out*,
10	Heliyanti Susana dkk., (2022) Penerapan Model Klasifikasi Metode Naïve Bayes Terhadap Penggunaan Akses Internet	Konsep dasar yang digunakan oleh Naïve Bayes adalah *Teorema Bayes*, yaitu melakukan klasifikasi dengan melakukan perhitungan nilai probabilitas hasil dari penelitian ini memiliki akurasi sebesar 89.83% Hasil Prediksi Ya dan ternyata Ya sebanyak 34. Hasil Prediksi Ya dan ternyata Tidak sebanyak 6. Hasil Prediksi tidak dan ternyata Ya sebanyak 0. Hasil Prediksi tidak dan ternyata tidak sebanyak 19. Hasil dari prediksi dengan uji 59 data baru maka mendapatkan hasil ya sebanyak 40 siswa dan tidak ada 19 siswa.

Berdasarkan uraian, terdapat persamaan antara penelitian sebelumnya dengan penelitian yang dilakukan yaitu memprediksi penjualan dan persediaan barang serta ada beberapa penelitian yang sama menggunakan algoritma Naïve Bayes. Adapun keterbaruan yang dilakukan dari penelitian ini membuat model perhitungan dengan algoritma Naïve Bayes dalam menentukan persediaan produk yang berpotensi terjual dan tidak terjual berdasarkan laporan penjualan bulanan perusahaan, metode yang digunakan berupa Observasi dan wawancara untuk memperoleh 31 atribut Naïve Bayes dan Prediksi dilakukan dengan *Tools* RapidMiner.

2.6 Relevansi Penelitian

Penelitian terdekat dijadikan perbandingan dengan penelitian yang akan

dilakukan sehingga dapat diketahui perbedaan apa saja yang ada pada penelitian. Penelitian dibuat oleh Herry Derajad Wijaya, Saruni Dwiasnati (2020) tujuan penelitian ini adalah untuk menganalisis masalah-masalah pada penentuan sebuah produk mana yang dapat di kategorikan LAKU atau TIDAK LAKU. Penelitian ini menggunakan *tools* RapidMiner versi 8 sebagai media untuk menguji data yang akan diolah untuk mendapatkan hasil *accuracy* dan nilai ROC. Nilai *accuracy* tersebut menunjukkan dinilai 88.00%. Kedekatan dari penelitian ini adalah adanya penggunaan salah metode Naïve Bayes dalam memprediksi penjualan suatu produk serta penggunaan *tools* RapidMiner.

Penelitian dibuat oleh Aji Ghassa dkk., (2022) penelitian ini membahas tentang pengujian akurasi merupakan pengujian yang dilakukan untuk memperkirakan seberapa tepat hasil klasifikasi terhadap data yang ada. Hasil akhir yang diperoleh dari penelitian ini adalah berdasarkan perancangan, analisis, implementasi dan pengujian pada Penerapan dengan Algoritma *Naive Bayes Classifier* untuk mengetahui minat beli pelanggan terhadap INDIHOME dengan menggunakan metode klasifikasi, maka dapat ditarik beberapa kesimpulan yaitu sistem klasifikasi data INDIHOME digunakan untuk menampilkan informasi klasifikasi layak atau Tidak layak minat masyarakat pada INDIHOME ini dengan menggunakan algoritma Naïve Bayes. Kedekatan dari penelitian ini adalah adanya penggunaan metode Naïve Bayes dalam memprediksi penjualan suatu produk.