

## **BAB II**

### **LANDASAN TEORI**

#### **2.1 DATA MINING**

*Data mining* adalah proses mengidentifikasi dan mengekstrak informasi yang berarti dari pola tersembunyi dalam data, hubungan antara item data, atau pemodelan untuk prediksi data. Ini menggunakan teknik statistik, matematika, kecerdasan buatan, dan *Machine Learning* [3].

*Data Mining* atau *Knowledge Discovery in Database* (KDD) yang semua prosesnya adalah kegiatan yang meliputi pengumpulan, pemakaian data historis dalam menemukan keteraturan, pola atau hubungan dengan set data yang berukuran besar. Keluaran dari data mining ini bisa dipakai untuk memperbaiki pengambilan keputusan di masa depan [4].

##### **2.1.1 Tahapan Data Mining**

Secara sederhana *data mining* biasa dikatakan sebagai proses penyaring atau menambang pengetahuan dari sejumlah data yang besar. Istilah *data mining* dan *Knowlegde Discovering in Database* (KDD) sering kali digunakan secara bergantian, Karena data mining adalah sebuah untain proses, maka pecah menjadi beberapa tahap. Tahapan tersebut akan bersifat interaktif, pengguna akan terlibat langsung atau dengan perantara KDD [5]:

a. Pembersihan data (*Data cleaning*)

Sebelum proses *data mining* dapat dilaksanakan, perlu dilakukan proses *cleaning* pada data yang menjadi fokus KDD. Proses *cleaning* mencakup antara lain membuang 4 duplikasi data, memeriksa data yang *inkonsisten*, dan memperbaiki kesalahan pada data, seperti kesalahan cetak (*tipografi*). Juga dilakukan proses *enrichment*, yaitu proses “memperkaya” data yang sudah ada dengan data atau informasi lain yang relevan dan diperlukan untuk KDD, seperti data atau informasi *eksternal*.

b. Integrasi data (*Data integration*)

Integrasi data merupakan penggabungan data dari berbagai *database* ke dalam satu *database* baru. Tidak jarang data yang diperlukan untuk *data mining* tidak hanya berasal dari satu *database* tetapi juga berasal dari beberapa *database* atau *file* teks.

c. Seleksi Data (*Data Selection*)

Data yang ada pada *database* sering kali tidak semuanya dipakai, oleh karena itu hanya data yang sesuai untuk dianalisis yang akan diambil dari *database*.

d. Transformasi data (*Data Transformation*)

Data diubah atau digabung ke dalam format yang sesuai untuk diproses dalam *data mining*.

e. Proses *mining*

Sebuah proses yang paling utama pada saat metode diterapkan untuk mencari pengetahuan tersembunyi dan berharga dari data.

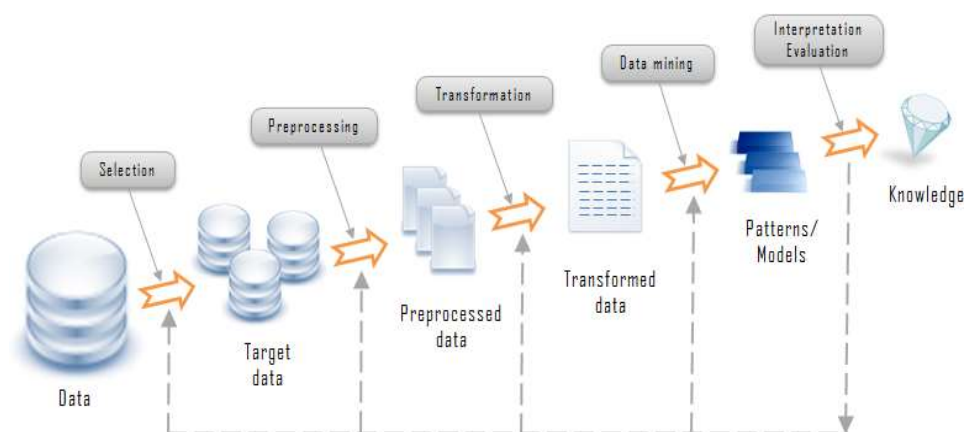
f. Evaluasi pola (*Pattern evaluation*)

Untuk mengidentifikasi pola-pola menarik kedalam *knowledge based* yang ditemukan.

g. Presentasi pengetahuan (*Knowledge presentation*)

Merupakan penyajian dan *visualisasi* pengetahuan mengenai metode yang digunakan untuk memperoleh pengetahuan yang diperoleh pengguna. Tahap terakhir dari proses *data mining* adalah bagaimana memformulasikan keputusan atau aksi dari hasil analisis yang didapat.

Berikut ini adalah gambar 2.1 yang menunjukkan Proses dari *Knowledge Discovering in Database (KDD)*:



**Gambar 2.1 Proses Dalam KDD**

Sehingga dapat disimpulkan bahwa *data mining* merupakan salah satu tahapan dari proses KDD. Sebelum sebuah data dapat diolah menggunakan *data*

*mining* sebagai sebuah informasi, diperlukan proses *preprocessing* data untuk seleksi atribut dari data tersebut.

### 2.1.2 Pengelompokan Data Mining

*Data Mining* dibagi menjadi beberapa kelompok berdasarkan tugas yang dapat dilakukan, yaitu [6]:

a. Deskripsi

Terkadang peneliti dan analisis secara sederhana ingin mencoba mencari cara untuk menggambarkan pola dan kecenderungan yang terdapat dalam data. Sebagai contoh, petugas pengumpulan suara mungkin tidak dapat menemukan keterangan atau fakta bahwa siapa yang tidak cukup profesional akan sedikit didukung dalam pemilihan presiden. Deskripsi dari pola dan kecenderungan sering memberikan kemungkinan penjelasan untuk suatu pola atau kecenderungan.

b. Estimasi

Estimasi hampir sama dengan klasifikasi, kecuali variabel target estimasi lebih ke arah numerik daripada ke arah kategori. Model dibangun menggunakan *record* lengkap yang menyediakan nilai dari variabel target sebagai nilai prediksi. Selanjutnya, pada peninjauan berikutnya estimasi nilai dari variabel target dibuat berdasarkan nilai variabel prediksi.

c. Prediksi

Prediksi hampir sama dengan klasifikasi dan estimasi, kecuali bahwa dalam prediksi nilai dari hasil akan ada di masa mendatang. Beberapa metode

dan teknik yang digunakan dalam klasifikasi dan estimasi dapat pula digunakan (untuk keadaan yang tepat) untuk prediksi.

d. Klasifikasi

Dalam klasifikasi, terdapat target variabel kategori. Sebagai contoh, penggolongan pendapatan dapat dipisahkan dalam tiga kategori, yaitu pendapatan tinggi, pendapatan sedang, dan pendapatan rendah.

e. Pengklasteran

Pengklasteran merupakan pengelompokan *record*, pengamatan, atau memperhatikan dan membentuk kelas objek-objek yang memiliki kemiripan. Pengklasteran berbeda dengan klasifikasi yaitu tidak adanya variabel target dalam pengklasteran. Pengklasteran tidak mencoba untuk melakukan klasifikasi, mengestimasi, atau memprediksi nilai dari variabel target. Akan tetapi, algoritma pengklasteran mencoba untuk melakukan pembagian terhadap keseluruhan data menjadi kelompok-kelompok yang memiliki kemiripan (homogen), yang mana kemiripan *record* dalam satu kelompok akan bernilai maksimal, sedangkan kemiripan dengan *record* dalam kelompok lain akan bernilai minimal.

f. Asosiasi

Tugas asosiasi dalam data mining adalah menemukan atribut yang muncul dalam satu waktu. Dalam dunia bisnis lebih umum disebut *market basket analysis*. Sehingga peranan data mining dalam hal ini adalah mencari aturan yang tidak *tercover* untuk mendapatkan hubungan antara dua

atau lebih atribut. *Association rule* adalah bentuk *if antecendant, then consequent* bersama-sama dengan suatu ukuran *support* dan *confidence*.

## 2.2 PERSIAPAN DATA (*DATA PREPARATION*)

Persiapan data mencakup semua kegiatan untuk membangun dataset siswa/i yang akan diterapkan ke dalam alat pemodelan, dari data mentah awal berupa dataset siswa/i dan selanjutnya akan melakukan proses *data mining*.

Dari dataset siswa/i maka akan dilakukan teknik *Data Preparation* agar kualitas data diperoleh lebih baik dengan cara [7]:

### a. *Data Validation*

*Data Validation* digunakan untuk mengidentifikasi dan menghapus data yang ganjil (*outlier/noise*), data yang tidak konsisten dan data yang tidak lengkap (*missing value*).

### b. *Data Integration and Transformation*

*Data integration* dan *transformation* digunakan untuk meningkatkan akurasi dan efisiensi algoritma. Data yang digunakan dalam penelitian ini bernilai kategorikal untuk model klasifikasi, data ditransformasi ke dalam angka menggunakan *software rapid miner*.

### c. *Data Size Reduction and Discretization*

*Data size reduction* dan *discretization* digunakan untuk memperoleh data set dengan jumlah atribut dan *record* yang lebih sedikit tetapi bersifat *informative*. Di dalam data training yang digunakan dalam penelitian ini, dilakukan seleksi atribut dan penghapusan data duplikasi.

### 2.3 KONSEP ANALISIS

Analisis adalah suatu upaya dalam penyelidikan untuk melihat, mengamati, mengetahui, menemukan, memahami, menelaah, mengklasifikasi, dan mendalami serta menginterpretasikan fenomena yang ada. Analisis pada sebuah estimasi (perkiraan) sangat penting dilakukan pada sebuah penelitian, agar penelitian menjadi lebih tepat dan terarah [8].

Analisis adalah aktivitas yang memuat sejumlah kegiatan seperti mengurai, membedakan, memilah sesuatu untuk digolongkan dan dikelompokkan kembali menurut kriteria tertentu kemudian dicari kaitannya dan ditafsirkan maknanya [9].

Analisis adalah penguraian suatu pokok atas berbagai bagiannya dan penelaahan pada bagian itu sendiri, serta hubungan antar bagian untuk memperoleh pengertian yang benar dan tepat serta pemahaman arti keseluruhan [10].

Berdasarkan uraian diatas, maka dapat disimpulkan bahwa analisis adalah suatu kegiatan dalam rangka mempelajari suatu bentuk benda, kegiatan, fakta maupun fenomena dalam mempelajari hubungan antar bagian yang terkait dari berbagai fungsi yang menjadi satu.

### 2.4 KONSEP KLASIFIKASI

Klasifikasi adalah teknik yang dilakukan untuk memprediksi *class* atau properti dari setiap *instance data*. Model prediksi memungkinkan untuk memprediksi nilai-nilai variabel yang tidak diketahui berdasarkan nilai variabel lainnya. Klasifikasi memetakan data ke dalam kelompok-kelompok kelas yang telah ditetapkan sebelumnya. Klasifikasi disebut juga dengan *supervised learning* karena kelas data telah ditentukan sebelumnya [11].

Konsep dasar dari klasifikasi adalah sejumlah data yang mempunyai struktur data yang hampir sama atau serupa akan menghasilkan klasifikasi yang hampir sama atau serupa juga. Metode klasifikasi ini yang banyak diterapkan di berbagai bidang dalam *data mining* [12].

Jadi dapat disimpulkan bahwa klasifikasi adalah pengelompokan kelas berdasarkan kelas data yang sudah ada untuk memprediksi kelas dari objek yang belum diketahui. Berikut ini adalah beberapa algoritma dalam metode klasifikasi, diantaranya yaitu :

a. Pohon Keputusan (*Decision tree*)

*Decision tree* merupakan sebuah struktur yang dapat digunakan untuk membagi kumpulan data yang besar menjadi himpunan-himpunan *record* yang lebih kecil dengan menerapkan serangkaian aturan keputusan [13].

Data dalam pohon keputusan biasanya dinyatakan dalam bentuk tabel dengan atribut dan *record*. Atribut menyatakan suatu parameter yang dibuat sebagai kriteria dalam pembentukan pohon. Proses pada pohon keputusan adalah mengubah bentuk data dalam tabel menjadi model pohon dan mengubah bentuk pohon menjadi aturan dan menyederhanakan aturan [14].

Jadi dapat diketahui bahwa Sebuah pohon keputusan adalah sebuah struktur yang dapat digunakan untuk membagi kumpulan data yang besar menjadi himpunan-himpunan *record* yang lebih kecil dengan menerapkan serangkaian aturan keputusan.



b. *Iterative Dichotomiser 3 (ID3)*

*Iterative Dichotomiser 3 (ID3)* adalah algoritma *decision tree learning* (algoritma pembelajaran pohon keputusan) yang paling dasar. Algoritma ini melakukan pencarian secara rakus/menyeluruh (*greedy*) pada semua kemungkinan pohon keputusan [14]. Jadi dapat diketahui bahwa Algoritma ID3 merupakan salah satu algoritma yang digunakan untuk melakukan klasifikasi atau segmentasi atau pengelompokan yang bersifat prediktif.

c. *K-nearest neighbor*

*K-nearest neighbor (KNN)* adalah algoritma pembelajaran yang banyak digunakan untuk tugas belajar yang diawasi. Dalam praktiknya, tantangan utama saat menggunakan KNN adalah sensitivitasnya yang tinggi terhadap pengaturan hyperparameternya, Termasuk jumlah tetangga terdekat, fungsi jarak dan fungsi pembobotan [15]. Jadi dapat diketahui bahwa KNN merupakan salah satu algoritma pembelajaran mesin sederhana. Hal ini hanya didasarkan pada gagasan bahwa suatu objek yang 'dekat' satu sama lain juga akan memiliki karakteristik yang mirip.

d. *Naïve Bayes*

*Naïve Bayes Classifier* merupakan sebuah metoda klasifikasi yang berakar pada teorema *Bayes*. Metode pengklasifikasian dengan menggunakan metode probabilitas dan statistik. Ciri utama dari *Naïve Bayes Classifier* ini adalah asumsi yang sangat kuat (naif) akan independensi dari masing-masing kondisi/kejadian [16]. Jadi dapat diketahui bahwa *naïve bayes*

merupakan salah satu algoritma klasifikasi untuk memprediksi probabilitas suatu *class* atau peluang dari data yang sudah ada.

e. *Support Vector Machine*

*Support Vector Machine* (SVM) merupakan bagian dari metode pembelajaran yang digunakan untuk klasifikasi. SVM memetakan vektor input ke sebuah ruang dimensi yang lebih tinggi dimana *hyperplane* pemisah dibangun [17]. Jadi dapat diketahui bahwa SVM (*Support Vector Machine*) adalah metode *supervised* terkait dengan *learning algorithm* untuk analisa pola data yang digunakan untuk klasifikasi dan regresi.

## 2.5 ANALISIS DAN PENERAPAN KLASIFIKASI

Data siswa/i yang telah diolah akan diklasifikasikan ke dalam kelas “Layak” atau “Tidak Layak” pada Penerimaan Beasiswa di SDN 25/IV Kota Jambi. Sehingga proses penyaluran Beasiswa bisa tepat sasaran. Selain itu akan diukur nilai akurasi dari penggunaan metode klasifikasi dengan algoritma *naïve bayes*. Data yang digunakan dalam penelitian ini akan ditransformasi ke dalam angka menggunakan *software Rapid Miner*.

## 2.6 ALGORITMA NAÏVE BAYES

*Bayesian classification* adalah pengklasifikasian statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu *class*. *Bayesian classification* didasarkan pada teorema *Bayes* yang memiliki kemampuan klasifikasi serupa dengan *decision tree* dan *neural network* [18].

Keuntungan penggunaan adalah bahwa metoda ini hanya membutuhkan jumlah data pelatihan (training data) yang kecil untuk menentukan estimasi parameter yang diperlukan dalam proses pengklasifikasian. Karena yang diasumsikan sebagai *variable independent*, maka hanya *varians* dari suatu variable dalam sebuah kelas yang dibutuhkan untuk menentukan klasifikasi, bukan keseluruhan dari *matriks kovarians* [19]. Persamaan dari *Naïve Bayes* sebagai berikut [20]:

$$P(C_i|X) = \frac{P(X|C_i)}{P(X)} \dots\dots\dots(2.1)$$

Penjelasan dari formula tersebut adalah sebagai berikut :

- X : Kriteria suatu kasus berdasarkan masukan.
- C<sub>i</sub> : Kelas solusi pola ke-i, dimana i adalah jumlah label kelas.
- P( C<sub>i</sub> | X ) : Probabilitas kemunculan label kelas C<sub>i</sub> dengan kriteria masukan X.
- P( X | C<sub>i</sub> ) : Probabilitas kriteria masukan X dengan label kelas C<sub>i</sub>.
- P(X) : Probabilitas labek kelas C<sub>i</sub>.

Untuk menjelaskan metode *Naïve Bayes*, perlu diketahui bahwa proses klasifikasi memerlukan sejumlah petunjuk untuk menentukan kelas apa yang cocok bagi sampel yang dianalisis tersebut. Karena itu metode *Naïve Bayes* di atas disesuaikan sebagai berikut [19]:

$$P(C|F_1 \dots F_n) = \frac{P(C)P(F_1 \dots F_n|C)}{P(F_1 \dots F_n)} \dots\dots\dots(2.2)$$

Keterangan :

- C : kelas.
- F<sub>1</sub>...F<sub>n</sub> : variabel karakteristik.

$P(C|F_1 \dots F_n)$  : probabilitas hipotesis C berdasarkan kondisi  $F_1$  sampai dengan  $F_n$

$P(F_1 \dots F_n|C)$  : probabilitas  $F_1$  sampai dengan  $F_n$  berdasarkan kondisi pada hipotesis C.

$P(F_1 \dots F_n)$  : probabilitas X.

Peluang masuknya sampel karakteristik tertentu dalam kelas C (*posterior*) adalah peluang munculnya kelas C (sebelum masuknya sampel tersebut, seringkali disebut *prior*), dikali dengan peluang kemunculan karakteristik - karakteristik sampel pada kelas C (disebut juga *likelihood*), dibagi dengan peluang kemunculan karakteristik-karakteristik sampel secara global (disebut juga *evidence*).

Nilai *evidence* selalu tetap untuk setiap kelas pada satu sampel. Nilai dari *posterior* tersebut nantinya akan dibandingkan dengan nilai-nilai *posterior* kelas lainnya untuk menentukan ke kelas apa suatu sampel akan diklasifikasikan. Penjabaran lebih lanjut rumus *bayes* tersebut dilakukan dengan menjabarkan ( $C|F_1, \dots, F_n$ ) menggunakan aturan perkalian sebagai berikut:

$$\begin{aligned}
 (C|F_1, \dots, F_n) &= (C)(F_1, \dots, F_n|C) \\
 &= (C)P(F_1|C)P(F_2, \dots, F_n|C, F_1) \\
 &= P(C)P(F_1|C)P(F_2|C, F_1)P(F_3, \dots, F_n|C, F_1, F_2) \\
 &= (C)P(F_1|C)P(F_2|C, F_1)P(F_3|C, F_1, F_2)P(F_4, \dots, F_n|C, F_1, F_2, F_3) \\
 &= P(C)P(F_1|C)P(F_2|C, F_1)P(F_3|C, F_1, F_2) \dots \\
 &P(F_n|C, F_1, F_2, F_3, \dots, F_{n-1}) \text{ -----(2.3)}
 \end{aligned}$$

Hasil penjabaran tersebut menyebabkan semakin kompleksnya faktor-faktor syarat yang mempengaruhi nilai probabilitas. Untuk itu digunakan asumsi independensi, bahwa masing-masing petunjuk ( $F_1, F_2, \dots, F_n$ ) saling bebas

(independen) satu sama lain. Dengan asumsi tersebut, maka berlaku suatu kesamaan sebagai berikut [21]:

$$P(F_i|F_j) = \frac{P(F_i \cap F_j)}{P(F_j)} = \frac{P(F_i)P(F_j)}{P(F_j)} = P(F_i) \text{----- (2.4)}$$

Untuk  $i \neq j$ , sehingga :

$$P(F_i / C, F_j) = P(F_i / C) \text{----- (2.5)}$$

## 2.7 BEASISWA

Beasiswa adalah suatu bantuan yang diberikan oleh pemerintah dinas pendidikan dalam bentuk uang yang akan diberi kepada siswa/siswi yang benar-benar layak untuk mendapatkan bantuan yang nantinya dapat digunakan untuk keberlangsungan pendidikan yang sedang ditempuh, sehingga siswa/siswi yang mendapat bantuan tidak perlu lagi khawatir dengan biaya pendidikan.

Dengan demikian dapat disimpulkan bahwa beasiswa merupakan suatu bentuk penghargaan yang berupa pemberian dan dukungan biaya pendidikan yang diberikan kepada mahasiswa yang masih aktif mengikuti perkuliahan di suatu perguruan tinggi atas keunggulan yang dimiliki mahasiswa tersebut baik secara intelektualitas maupun personalitas [22].

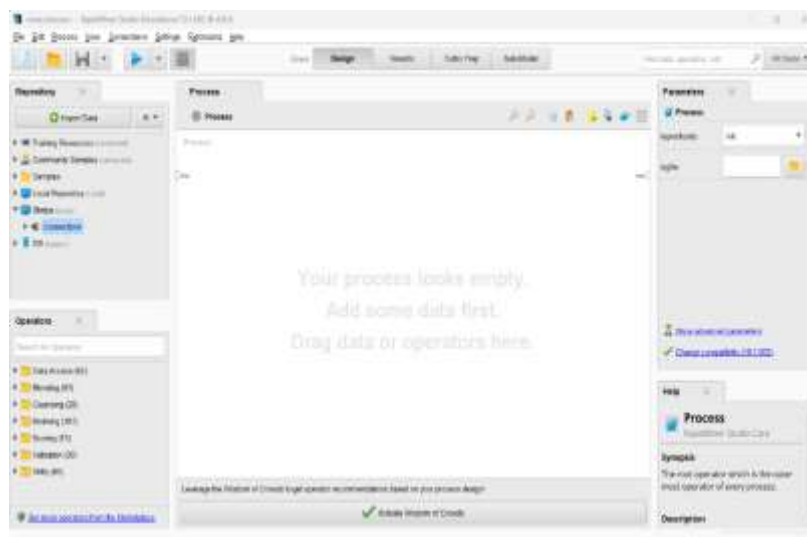
## 2.8 RAPID MINER

*Rapid Miner* adalah *platform* perangkat lunak data ilmu pengetahuan yang dikembangkan oleh perusahaan dengan nama yang sama, yang menyediakan lingkungan terpadu untuk pembelajaran mesin (*machine learning*), pembelajaran mendalam (*deep learning*), penambangan teks (*text mining*), dan analisis prediktif (*predictive analytics*). Aplikasi ini digunakan untuk aplikasi bisnis dan komersial

serta untuk penelitian, pendidikan, pelatihan, pembuatan *prototype* dengan cepat, dan pengembangan aplikasi serta mendukung semua langkah proses pembelajaran mesin termasuk persiapan data, visualisasi hasil, validasi dan pengoptimalan. *Rapid Miner* dikembangkan dengan model *open core* [23].

Penggunaan *Rapid Miner* dalam pembuatan model tidak memerlukan program dan semua yang digunakan dalam pembuatan model sudah tersedia dalam bentuk *operator*. Pembuatan model menggunakan berbagai *operator* yang sesuai dan saling dikaitkan dalam membentuk suatu model [24].

*Rapid Miner* memudahkan penggunaanya dalam melakukan perhitungan data yang sangat banyak dengan menggunakan *operator-operator*. *Operator* ini juga berfungsi untuk memodifikasi data. Data dihubungkan dengan node-node pada operator kemudian kita hanya tinggal menghubungkannya ke *node* hasil untuk melihat hasilnya. Hasil yang diperlihatkan *Rapid Miner* pun dapat ditampilkan secara visual dengan grafik. Menjadikan *Rapid Miner* adalah salah satu *software* dalam melakukan ekstraksi data dengan metode-metode *data mining* [25].



**Gambar 2.2 Halaman Utama Aplikasi *Rapid Miner***

## 2.9 TINJAUAN PENELITIAN SEJENIS

Penulis memasukkan beberapa kajian dari penulis – penulis yang telah melakukan penelitian sejenis untuk menjadi bahan perbandingan atau acuan bagi penulisan Tugas Akhir berjudul “Analisis dan Penerapan Algoritma *Naïve Bayes* Untuk Klasifikasi Kelayakan Penerimaan Beasiswa (Studi Kasus : SDN 25/IV Kota Jambi)”. Berikut ini merupakan kajian penelitian sejenis dan terkait yang peneliti jadikan acuan untuk melakukan penelitian ini :

1. Penelitian yang dilakukan oleh Debby Alita, Indah Sari, Auliya Rahman Isnain, Styawati, pada tahun 2021 dengan judul “Penerapan *Naïve Bayes Classifier* Untuk Pendukung Keputusan Penerima Beasiswa” [26].

Metode *Naive Bayes Classifier* merupakan salah satu metode yang dapat digunakan dalam hal pengambilan keputusan untuk mendapatkan hasil yang lebih baik pada suatu permasalahan klasifikasi. Tujuan dari penelitian ini adalah Untuk membangun sistem pendukung keputusan penerima beasiswa dengan metode *Naïve Bayes Classifier*. Hasil dari penelitian ini Berdasarkan perhitungan persentase skor aktual yang telah dilakukan memberikan hasil sebesar 85.96% dan dibandingkan dengan range kelayakan bahwa nilai 85.96% berada pada range 85-100% yaitu sangat baik. Maka dapat ditarik kesimpulan bahwa sistem pendukung keputusan penerima beasiswa dinyatakan sangat baik dengan menerapkan metode *naïve bayes* pada sistem pendukung keputusan penerima beasiswa dapat membantu pihak sekolah dalam melakukan penentuan penerima beasiswa dengan lebih cepat dan tepat.

2. Penelitian yang dilakukan oleh Tiara Nur Safitri, Pada tahun 2020 dengan judul “Penerapan Algoritma *Naïve Bayes* Untuk Penentuan Calon Penerimaan Beasiswa Pada Sd Negeri 6 Ketapang” [27].

Penelitian ini bertujuan untuk membantu bagian proses seleksi untuk penentuan penerima beasiswa SD Negeri 6 Ketapang. Penentuan penerimaan beasiswa menggunakan beberapa kriteria antara lain: pekerjaan orang tua, penghasilan orang tua dan jumlah tanggungan. Kelayakan calon penerimaan beasiswa ditentukan dengan menerapkan klasifikasi *Data Mining* dengan metode *Naïve Bayes*. Metode ini dipilih karena mampu mempelajari data kasus sebelumnya yang digunakan sebagai data uji. Pada Penelitian ini didapatkan hasil evaluasi dan validasi, diketahui bahwa *Naïve Bayes* memiliki nilai akurasi cukup tinggi yaitu sebesar 76,00%, dengan artian bahwa *Naïve Bayes* dapat dijadikan sebagai penentuan calon penerimaan Beasiswa pada Sekolah Dasar Negeri 6 Ketapang.

3. Penelitian ini dilakukan oleh Haditsah Annur, Pada tahun 2019 dengan judul “Klasifikasi Masyarakat Miskin Menggunakan Metode *Naïve Bayes*” [28].

Metode yang akan digunakan adalah metode *Naïve Bayes Classifier*, yang merupakan salah satu teknik pengklasifikasian dalam *data mining*. Tujuan dari penelitian ini adalah : Klasifikasi masyarakat miskin di Kecamatan Tibawa Menggunakan Algoritma Naive Bayes dan Menerapkan Algoritma Naïve Bayes klasifikasi masyarakat miskin di Kecamatan Tibawa. Hasil yang didapat berdasarkan hasil pengujian *confussion 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%.

4. Penelitian ini dilakukan oleh Ratih Yulia Hayuningtyas, Pada tahun 2019 dengan judul “Penerapan Algoritma *Naïve Bayes* untuk Rekomendasi Pakaian Wanita” [29].

Pada penelitian ini menggunakan metode *Naive Bayes* dalam menentukan rekomendasi pakaian wanita. Didapatkan hasil bahwa *Naïve Bayes* merupakan metode yang dapat membantu dalam menyelesaikan masalah klasifikasi seperti rekomendasi pakaian wanita serta *Naïve Bayes* sangat baik digunakan untuk mendukung penunjang keputusan.

5. Penelitian ini dilakukan oleh Hery Mustofa, Adzhal Arwani Mahfudh, Pada tahun 2019 dengan judul “Klasifikasi Berita *Hoax* Dengan Menggunakan Metode *Naïve bayes*” [30].

Pada penelitian ini menggunakan metode *Naïve Bayes*, Tujuannya adalah untuk mengetahui tingkat akurasi metode *Naïve Bayes* digunakan untuk klasifikasi berita *hoax* berbahasa Indonesia. Pengukuran dilakukan dengan metode *10-fold cross validation*. Dari pengukuran tersebut diperoleh hasil, nilai *fold 6* mempunyai keakuratan tertinggi, yaitu sebesar 85.28 % yang mana dokumen terklasifikasi yang relevan sebanyak 307 dan dokumen tidak relevan sebanyak 53 atau *error rate* sebesar 14.72%. Sedangkan nilai rata-rata berdasarkan dokumen berita *hoax* dan dokumen berita benar nilai *precision* 0,896 dan *recall* 0.853

6. Penelitian ini dilakukan oleh Trya Ayu Pratiwi, Muhammad Irsyad, Rahmad Kurniawan, Surya Agustian, Benny Sukma Negara, Pada tahun 2021 dengan judul “Klasifikasi Kebakaran Hutan Dan Lahan Menggunakan Algoritma *Naïve Bayes* di Kabupaten Pelalawan” [31].

Pada penelitian ini menggunakan metode Algoritma *Naïve Bayes*. Penelitian ini bertujuan untuk mengklasifikasi kebakaran hutan dan lahan di Kabupaten Pelalawan, Riau. Penelitian ini telah menggunakan 792 data dalam lima tahun (2015-2019). Atribut yang digunakan untuk klasifikasi terdiri dari suhu, kelembaban, curah hujan, kecepatan angin, dan kelas dari keempat faktor tersebut didapatkan hasil probabilitas tertinggi yaitu suhu dengan nilai 0.978260870 dan nilai akurasi tertinggi adalah dataset tahun 2017 dengan nilai akurasi adalah 81.03% sehingga nilai akurasi tertinggi tersebut dapat diterapkan pada dataset baru yaitu pada tahun 2019 dengan nilai akurasi nya adalah 82%.

7. Penelitian ini dilakukan oleh Andhini Asri Awaliyah Arifin, Wiwin Handoko, Zulfan Efendi, Pada tahun 2022 dengan judul “Implementasi Metode *Naïve Bayes* Untuk Klasifikasi Penerima Program Keluarga Harapan” [32].

Pada penelitian ini menggunakan metode Algoritma *Naïve Bayes*. Penelitian ini bertujuan untuk menerapkan Algoritma *Naïve Bayes* untuk mengklasifikasi warga tidak mampu berdasarkan pengalaman di masa sebelumnya yang berhak memperoleh bantuan pemerintah dalam bentuk PKH. Hasil pemodelan klasifikasi dengan Algoritma *Naïve Bayes* menghasilkan besaran nilai *precision* untuk kelas positif 100%, untuk kelas negatif 77%, nilai *recall* untuk

kelas positif 80%, untuk kelas negatif 100%, nilai *f1-score* untuk kelas positif 89%, untuk kelas negatif 87%, dan nilai akurasi 88%.

8. Penelitian ini dilakukan oleh Wilianti Aliman, Pada tahun 2022 dengan judul “Implementasi Metode *Naïve bayes* untuk Menentukan Persetujuan Pemberian beasiswa Penuh Pada Penerimaan Mahasiswa Baru Di Institusi Pendidikan X” [33].

Pada penelitian ini menggunakan metode Algoritma *Naïve Bayes*. Penelitian menggunakan metode *Naïve Bayes* ini menghasilkan tingkat akurasi 83,3%, menurut penelitian yang dilakukan *probability* keberhasilan menjadi mahasiswa (calon mahasiswa menuntaskan proses pendaftaran hingga menjadi mahasiswa) lebih tinggi apabila pemberian beasiswa dilakukan melalui ujian atau prestasi dari calon mahasiswa.

9. Penelitian ini dilakukan oleh Rizal Rachman, Rissa Nurfitriana Handayani, Pada tahun 2021 dengan judul “Klasifikasi Algoritma *Naïve Bayes* Dalam Memprediksi Tingkat Kelancaran Pembayaran Sewa Teras UMKM ” [34].

Pada penelitian ini menggunakan metode Algoritma *Naïve Bayes*. Berdasarkan hasil pengolahan data dengan menggunakan aplikasi *Rapidminer* dengan menggunakan algoritma *Naïve Bayes* menghasilkan nilai *Accuracy* 81.81%, *Precision* 66.66%, dan *Recall* 100% dengan hasil *AUC* 0.800. Pengolahan data yang dilakukan menghasilkan informasi jumlah status pembayaran yang *out standing* berdasarkan beberapa kriteria dan lokasi, sehingga pihak manajemen bisa mengambil langkah untuk meminimalisir penyewa yang sering telat dalam pembayaran sewa teras UMKM

10. Penelitian ini dilakukan oleh Lutfi Abdullah, Rosmawati Tamin, A. Akhmad Qashlim, Pada tahun 2021 dengan judul “Klasifikasi Penerimaan Beasiswa Menggunakan Algoritma *Naive Bayes* Di Universitas Al Asyariah Mandar Kabupaten Polewali Mandar” [2].

Pada penelitian ini menggunakan metode Algoritma *Naive Bayes*. Di penelitian ini ada beberapa kriteria yang digunakan untuk penyeleksian beasiswa, yaitu Nilai IPK, Jarak rumah dari kampus, kendaraan yang digunakan saat kekampus, tempat tinggal, jumlah tanggungan, dan pendapatan orang tua. Proses testing memiliki jumlah *data training* dan *data testing* yang berbeda. Dihasilkan akurasi tertinggi terdapat proses ketujuh dengan akurasi sebesar 88%.

Dari beberapa penelitian diatas, dapat disimpulkan bahwa metode *Naive Bayes* memiliki akurasi yang baik dalam pengklasifikasian data. Maka dari itu dalam penelitian ini penulis menggunakan algoritma *Naive Bayes* untuk Klasifikasi Kelayakan calon penerima beasiswa di SDN 25/IV KOTA JAMBI.

Berdasarkan dari beberapa penelitian tersebut, dapat dijelaskan bahwa yang membedakan penelitian ini dengan penelitian sebelumnya adalah penulis akan menambahkan atribut yang digunakan dalam proses *data mining* yang digunakan dalam penelitian, jumlah data dan alat bantu yang digunakan yaitu aplikasi *Rapid Miner*. Sehingga diharapkan nantinya akan mendapatkan hasil yang diperoleh dapat dijadikan acuan oleh pihak sekolah dalam menentukan kelayakan penerima beasiswa.