

# PENERAPAN ALGORITMA *NAÏVE BAYES* UNTUK PENENTUAN CALON PENERIMAAN BEASISWA PADA SD NEGERI 6 KETAPANG

Tiara Nur Safitri<sup>1</sup>

**Abstrak**— Penelitian ini dilatar belakangi karena dibutuhkan suatu algoritma penentuan calon penerima beasiswa yang tepat dan efektif, hal ini disebabkan kurang lebih 300 pendaftar beasiswa SD Negeri 6 Ketapang setiap tahun, menyebabkan proses seleksinya berjalan lambat dan berpotensi tidak konsisten. Tidak konsistennya pada sistem penentuan penerima menyebabkan tujuan penyelenggaraan beasiswa menjadi samar, tidak transparan dan tidak tepat sasaran. 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.

**Kata Kunci** : *Data Mining, Naïve Bayes, Kriteria, Beasiswa.*

**Abstract**— This research is motivated because it takes an algorithm to determine the right and effective scholarship recipients, this is because About 300 applicants for the 6 Ketapang Public Elementary School scholarship each year, causing the selection process to run slowly and potentially inconsistent. The inconsistency in the system of determining recipients causes the purpose of organizing scholarships to be vague, not transparent and not on target. This study aims to assist part of the selection process for determining the recipients of 6 Ketapang Public Elementary Scholarships. Determination of scholarship acceptance uses several criteria including: parents' occupation, parents' income and the number of dependents. The eligibility of prospective scholarship recipients is determined by applying the Data Mining classification to the *Naïve Bayes* method. This method was chosen because it is able to study previous case data that is used as test data. In this study, the results of evaluation and validation show that *Naïve Bayes* has a high curative value of 76.00%, with the meaning that *Naïve Bayes* can be used as a determination for candidates for

scholarship acceptance at Ketapang 6 Public Elementary Schools.

**Keywords** : *DataMining, NaïveBayes, Criteria, Scholarship.*

## I. PENDAHULUAN

### 1.1 Latar Belakang

Pendidikan di Indonesia diwajibkan 12 tahun, tetapi pada nyatanya masih banyak masyarakat yang tidak bisa sekolah karena adanya kendala biaya, Hal ini dapat didasari bahwa Pendidikan sejatinya merupakan salah satu kegiatan yang dilakukan untuk mengembangkan potensi diri. Institusi yang bertanggung jawab atas penyebaran ilmu pengetahuan adalah institusi pendidikan. Setiap Institusi pendidikan bertanggungjawab untuk mendidik, mengajar dan memberikan hak serta kewajiban pada seluruh siswanya, dengan tujuan untuk mencerdaskan kehidupan bangsa, mengembangkan karakter dan menguatkan karakter siswa itu sendiri, dengan meningkatkan ilmu pengetahuannya seiring menguatkan iman dan ketaqwaan pada tuhan yang maha kuasa. Salah satu Institusi Pendidikan yaitu Sekolah Dasar, yang mengajar, mendidik dan memberikan hak serta kewajiban anak mengenai Pendidikan dasar serta menguatkan karakter, Setiap warga negara Indonesia berhak mendapatkan pendidikan sebagaimana tertuang dalam UUD 1945 pasal 31 ayat 1-2. Namun, tidak semua warga negara Indonesia mendapatkan kesempatan untuk memperoleh pendidikan yang layak dan sesuai yang berkelanjutan [1]. Oleh karena itu bagi peserta didik yang kurang mampu berhak untuk memperoleh biaya pendidikan dan bagi peserta didik yang berprestasi berhak mendapatkan beasiswa [2]. namun tidak dapat dipungkiri bahwa masih banyak anak yang tidak dapat merasakan dan menikmati dunia pendidikan tentunya, salah satu sebabnya karena kurang biaya, dari hal itu lah diperlukan berbagai solusi, salah satunya adalah program beasiswa. Beasiswa merupakan pembiayaan yang tidak bersumber dari pendanaan sendiri ataupun orang tua yang bertujuan untuk meringankan beban biaya pendidikan dan juga dapat menjadi motivasi belajar bagi siswa yang menerimanya. Namun seringkali pemberian beasiswa tidak tepat sasaran, karena kesulitan dalam menyeleksi siswa. Di setiap lembaga pendidikan khususnya sekolah banyak sekali beasiswa yang ditujukan kepada siswa. Beasiswa ditujukan untuk membantu meringankan beban

<sup>1</sup>Status penulis, Mahasiswa Jurusan Sistem Informasi Fakultas Ilmu Komputer Universitas Darwan Ali, Jl. Batu Berlian No.10 Kab. Kotawaringin Timur 74322 INDONESIA (telp:053-121527; fax: 053-121527; e-mail: info@unda.ac.id)

biaya bagi siswa yang mendapatkannya dan dengan beasiswa bisa nantinya diharapkan untuk meningkatkan partisipasi pendidikan hingga mencerdaskan kehidupan setiap anak bangsa, yang memiliki kesempatan sama untuk merasakan Pendidikan. Beasiswa tersebut nantinya dimanfaatkan sebaik mungkin, hingga dari bantuan tersebut, seorang siswa dapat memenuhi kebutuhan pokok selama masa studi. Untuk mendapatkan beasiswa tersebut maka harus sesuai dengan aturan yang diterapkan. Kriteria yang diterapkan antara lain pekerjaan orang tua, penghasilan orang tua dan jumlah tanggungan [3].

Banyaknya jumlah siswa yang mengajukan permohonan untuk menerima beasiswa dan kriteria penilaian yang banyak pula maka tidak semua siswa yang mengajukan permohonan beasiswa dapat dikabulkan. Permasalahan yang biasanya dihadapi oleh sekolah sehubungan dengan penentuan beasiswa adalah tidak adanya sistem yang dapat membantu untuk melakukan penyeleksian atau penentuan penerima beasiswa secara otomatis sesuai dengan kriteria-kriteria yang telah ditentukan [4].

SD Negeri 6 Ketapang memiliki program pemberian beasiswa untuk siswa yang berprestasi namun memiliki keterbatasan finansial. Kriteria penilaian yang diterapkan di antaranya penghasilan orang tua, pekerjaan orang tua dan jumlah tanggungan. Dalam hal penentuan penerimaan beasiswa, SD Negeri 6 Ketapang belum efektif dan efisien dalam penentuan penerima beasiswa.

Penelitian ini akan menerapkan algoritma *Naïve Bayes* dalam penentuan calon penerima beasiswa. *Naïve Bayes* merupakan sebuah pengklasifikasian probabilistik sederhana yang menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari *dataset* yang diberikan. Algoritma *Naïve Bayes* adalah salah satu metodologi *data mining* termasuk ke dalam sepuluh klasifikasi *data mining* yang paling populer di antara algoritma-algoritma lainnya. Metode *Naïve Bayes* juga dinilai berpotensi baik dalam mengklasifikasi dokumen dibandingkan metode pengklasifikasian yang lain dalam hal akurasi dan efisiensi dalam penentuan sebuah keputusan dalam sebuah perhitungan dengan metode ini tentunya.

## 1.2 Identifikasi Masalah

Berdasarkan latar belakang yang penulis tuliskan, maka dari itu penulis akan mengidentifikasi Masalah yang akan diteliti, sebagai berikut :

1. Banyaknya jumlah siswa yang mengajukan permohonan untuk mendapatkan beasiswa. Sehingga penentuan penerima beasiswa kurang efektif dan efisien.
2. Mengetahui akurasi klasifikasi penerimaan beasiswa pada algoritma *Naïve Bayes*.
3. Menerapkan algoritma *Naïve Bayes* dalam menentukan calon penerima beasiswa.

Kebutuhan informasi yang tinggi kadang tidak diimbangi dengan penyajian informasi yang memadai,

sering kali informasi tersebut masih harus digali ulang dari data yang jumlahnya sangat besar, dapat dipahami bahwa Kemajuan teknologi informasi sudah semakin berkembang pesat disegala bidang kehidupan. Banyak sekali data yang dihasilkan oleh teknologi informasi yang canggih, mulai dari bidang industri, ekonomi, ilmu dan teknologi serta berbagai bidang kehidupan lainnya. Penerapan teknologi informasi dalam dunia pendidikan juga dapat menghasilkan data yang berlimpah mengenai siswa dan proses pembelajaran yang dihasilkan, dari itulah diperlukan Penggunaan teknik *data mining* diharapkan dapat memberikan pengetahuan-pengetahuan yang sebelumnya tersembunyi di dalam gudang data sehingga menjadi informasi yang berharga [5].

Siswa mempunyai kewajiban untuk belajar, dalam hal ini harus lebih giat lagi dalam belajar karena akan menghadapi Ujian Nasional (UN). Sebagai salah satu syarat untuk bisa melanjutkan jenjang pendidikan ke tingkat selanjutnya dan seperti yang kita ketahui, bahwa akhir-akhir ini standar kelulusan Ujian Nasional (UN) di Indonesia semakin tinggi. Oleh karena itu, sekolah seharusnya mengetahui apa yang menjadi faktor-faktor yang menentukan tingkat kelulusan siswanya [6].

Dalam masalah ini, dilakukan analisis faktor yang memprediksi tingkat kelulusan siswa berdasarkan nilai belajar siswa selama belajar di jenjang tingkat pendidikan SD. Nilai belajar siswa selama belajar di sekolah diyakini mempunyai peranan yang besar dalam menentukan tingkat kelulusan siswa tersebut. Sebagai penilaiannya, siswa akan menggunakan nilai belajarnya selama disekolah sebagai standar atau acuan untuk tingkat kelulusannya. Dengan demikian, nilai belajar siswa dapat menentukan mengapa tingkat kelulusan siswa yang satu dengan yang lain dapat berbeda [7].

Oleh karena itu, dapat diuraikan faktor apa saja yang membedakan nilai belajar siswa dalam memprediksi tingkat kelulusan siswa SD. Faktor-faktor tersebut adalah nilai ujian bahasa Indonesia, matematika dan IPA. Responden yang digunakan dalam penelitian ini adalah data siswa-siswi kelas 6 dari Dinas Pendidikan Kabupaten Kotawaringin Timur. Data yang akan digunakan adalah nilai-nilai siswa. Metode *Naïve Bayes* adalah metode klasifikasi statistik yang dapat memprediksi kelas suatu anggota probabilitas, algoritma ini memanfaatkan teori probabilitas yang dikemukakan oleh ilmuwan Inggris yaitu memprediksi probabilitas di masa depan berdasarkan pengalaman di masa sekarang. Dalam penelitian ini menggunakan metode *Naïve Bayes* karena tujuan penelitian ini adalah memprediksi hasil kelulusan di masa depan berdasarkan pengalaman di masa sekarang.

## 1.3 Rumusan Masalah

Sesuai dengan latar belakang yang telah diuraikan, maka dapat ditentukan rumusan masalahnya adalah bagaimana cara memprediksi hasil kelulusan siswa dengan metode *Naïve Bayes*?

### 1.4 Tujuan Penelitian

Penelitian ini bertujuan untuk memprediksi hasil kelulusan siswa dengan metode *NaïveBayes* yang akan menghasilkan lulus atau tidak lulus.

### 1.5 Manfaat Penelitian

Sesuai dengan permasalahan dan tujuan penelitian di atas, maka manfaat penelitian dapat diuraikan sebagai berikut :

- Mempermudah analisis data kelulusan yang jumlahnya besar agar dapat diketahui faktor-faktor yang sangat berpengaruh pada tingkat kelulusan.
- Membuat sistem pendukung keputusan untuk membantu meningkatkan kualitas kelulusan siswa.

### 1.6 Batasan Masalah

Batasan masalah dalam penelitian ini adalah :

- Semua proses perhitungan yang disediakan oleh sistem menggunakan teknik *data mining* dengan metode *NaïveBayes*.
- Informasi yang ditampilkan berupa laporan analisis pola *data mining* tingkat kelulusan dan nilai kalkulasi probabilitas posterior pada data siswa.

## II. LANDASAN TEORI

### 2.1 Beasiswa

Pengertian beasiswa menurut Sihotang adalah pembiayaan yang tidak bersumber dari pendanaan sendiri atau orang tua, akan tetapi diberikan oleh pemerintah, perusahaan swasta, kedutaan, universitas, serta lembaga pendidik atau peneliti, atau juga dari kantor tempat seseorang bekerja untuk meningkatkan kapasitas sumber daya manusia melalui pendidikan. Sedangkan menurut Putra dan Hardiyanti beasiswa adalah pemberian berupa bantuan keuangan yang diberikan kepada perorangan yang bertujuan untuk digunakan demi keberlangsungan pendidikan yang ditempuh [8].

Beasiswa adalah pemberian berupa bantuan keuangan yang diberikan kepada perorangan yang bertujuan untuk digunakan demi keberlangsungan pendidikan yang ditempuh. Beasiswa dapat diberikan oleh Lembaga pemerintah, perusahaan ataupun yayasan. Pemberian beasiswa dapat dikategorikan pada pemberian cuma-cuma ataupun pemberian dengan ikatan kerja (biasa disebut ikatan dinas) setelah selesainya pendidikan. Lama ikatan dinas ini berbeda-beda, tergantung pada lembaga yang memberikan beasiswa tersebut.

Sejatinya banyak tujuan dari Beasiswa ini, terlebih bagi siswa Sekolah Dasar yaitu Untuk membantu para pelajar atau mahasiswa agar dapat mencari ilmu sesuai dengan bidang yang ingin dikuasai. Terutama bagi yang memiliki masalah dalam pembiayaan, Membuat pemerataan ilmu pengetahuan dan pendidikan kepada masing-masing individu yang membutuhkan, Mencetak generasi baru yang lebih pintar dan cerdas. Karena adanya

bantuan beasiswa ini, dapat memberikan kesempatan kepada para penerus bangsa untuk mendapatkan pendidikan di jenjang yang lebih tinggi, hingga Meningkatkan kesejahteraan. Setelah mendapatkan sumber daya manusia yang lebih cerdas dan pintar, mereka tentu diharapkan bisa saling memberi bantuan berupa ilmu pengetahuan yang sudah didapat pada saat sedang menimba ilmu.

Persyaratan untuk mendapatkan beasiswa pada SDN 6 Ketapang yakni ialah Berasal dari Keluarga Kurang Mampu dan nilai rapor tinggi.

### 2.2 Data Mining

*Data Mining* adalah analisis terhadap data (biasanya data yang berukuran besar) untuk menemukan hubungan yang jelas serta menyimpulkan yang belum diketahui sebelumnya dengan cara terkini dipahami dan berguna bagi pemilik data tersebut [8]. *Data Mining* adalah suatu proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan pembelajaran mesin untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai big data. Istilah *data mining* memiliki hakikat sebagai suatu disiplin ilmu yang tujuannya yaitu untuk menemukan, menggali, atau mendapat pengetahuan dari data atau informasi yang telah dimiliki [9]. *Data Mining* merupakan suatu proses menemukan hubungan yang berarti kecenderungan dengan memeriksa sekumpulan besar data yang tersimpan dalam penyimpanan dengan teknik pengenalan pola seperti teknik statistik dan matematika. Ketersediaan data yang melimpah, kebutuhan akan informasi atau pengetahuan sebagai pendukung pengambilan keputusan untuk membuat solusi bisnis, dan dukungan infrastruktur di bidang teknologi informasi merupakan pendukung dari pemrosesan *data mining* ini. Berbagai penerapan *data mining* yakni Analisa pasar dan manajemen. Analisa perusahaan dan manajemen resiko, Telekomunikasi, Keuangan, Asuransi, Olahraga dan Astronomi. Sebagai suatu teknologi yang dapat menghasilkan knowledge, *data mining* melakukan pemrosesan dalam beberapa tahap. *Data mining* juga memiliki sebutan lain yang disebut *Knowledge Discovery in Database (KDD)*. *Knowledge Discovery in Database (KDD)* merupakan kegiatan yang meliputi pengumpulan data, pemakaian data, historis untuk menemukan keteraturan, pola atau hubungan dalam set data yang memiliki ukuran besar [10]. *Data mining* adalah kegiatan menemukan pola yang menarik dari data dalam jumlah besar, data dapat disimpan dalam database, data warehouse, atau penyimpanan informasi lainnya. *Data mining* berkaitan dengan bidang ilmu – ilmu lain, seperti database system, data warehousing, statistik, machine learning, information retrieval, dan komputasi tingkat tinggi. Selain itu, *data mining* didukung oleh ilmu lain seperti neural network, pengenalan pola, spatial data analysis, image database, signal processing. *Data mining* didefinisikan sebagai proses menemukan pola-pola

dalam data. Proses ini otomatis atau seringnya semiotomatis. Pola yang ditemukan harus penuh arti dan pola tersebut memberikan keuntungan, biasanya keuntungan secara ekonomi. Data yang dibutuhkan dalam jumlah besar. Tentunya tidak terlepas dari pembahasan Data Mining erat kaitan dengan klasifikasi. Klasifikasi bertujuan untuk mengklasifikasikan item data menjadi satu dari beberapa kelas standar. Sebagai contoh, suatu program email dapat mengklasifikasikan email yang sah dengan email spam. Beberapa algoritma klasifikasi antara lain pohon keputusan, nearest neighbor, naive bayes, neural networks dan support vector machines. Hasil dari data mining sering kali diintegrasikan dengan decision support system (DSS). Sebagai contoh, dalam aplikasi bisnis informasi yang dihasilkan oleh data mining dapat diintegrasikan dengan tool manajemen kampanye produk sehingga promosi pemasaran yang efektif yang dilaksanakan dan dapat diuji. Integrasi demikian memerlukan langkah postprocessing yang menjamin bahwa hanya hasil yang valid dan berguna yang akan digabungkan dengan DSS. Salah satu pekerjaan dan postprocessing adalah visualisasi yang memungkinkan analyst untuk mengeksplor data dan hasil data mining, dari berbagai sudut pandang. Ukuran-ukuran statistik dan metode pengujian hipotesis dapat digunakan selama postprocessing untuk membuang hasil data mining yang palsu. Secara khusus, data mining menggunakan ide-ide seperti (1) pengambilan contoh, estimasi, dan pengujian hipotesis, dari statistika dan (2) algoritme pencarian, teknik pemodelan, dan teori pembelajaran dari kecerdasan buatan, pengenalan pola, dan machine learning. Data mining juga telah mengadopsi ide-ide dari area lain meliputi optimisasi, evolutionary computing, teori informasi, pemrosesan sinyal, visualisasi dan information retrieval. Sejumlah area lain juga memberikan peran pendukung dalam data mining, seperti sistem basis data yang dibutuhkan untuk menyediakan tempat penyimpanan yang efisien, indexing dan pemrosesan kueri.

### 2.3 Fungsi Data Mining

Fungsi *datamining* dapat dipahami dan diketahui bahwa Data mining memiliki banyak sekali fungsi. Untuk fungsi utamanya sendiri yaitu ada dua; Yaitu fungsi descriptive dan fungsi predictive. Untuk fungsi lainnya akan dibahas di bawah :

1. Descriptive, Tujuan dari tugas prediktif adalah untuk memprediksi nilai dari atribut tertentu berdasarkan pada nilai dari atribut-atribut lain. Atribut yang diprediksi umumnya dikenal sebagai target atau variabel tak bebas, sedangkan atribut-atribut yang digunakan untuk membuat prediksi dikenal sebagai explanatory atau variabel bebas. Pada Prinsipnya Fungsi Descriptive dari *datamining* ini menggambarkan, Dengan menggunakan Fungsi descriptive Data mining, Maka nantinya bisa menemukan pola tertentu yang tersembunyi dalam sebuah data. Dengan kata lain jika pola yang berulang

dan bernilai itulah karakteristik sebuah data bisa diketahui. Selain itu, terdapat pemahaman juga mengenai Fungsi Deskripsi pada *datamining* ini tentunya, Deskripsi bertujuan untuk mengidentifikasi pola yang muncul secara berulang pada suatu data dan mengubah pola tersebut menjadi aturan dan kriteria yang dapat mudah dimengerti oleh para ahli pada domain aplikasinya. Aturan yang dihasilkan harus mudah dimengerti agar dapat dengan efektif meningkatkan tingkat pengetahuan (knowledge) pada sistem. Tugas deskriptif merupakan tugas data mining yang sering dibutuhkan pada teknik postprocessing untuk melakukan validasi dan menjelaskan hasil dari proses data mining, hal ini tentunya Postprocessing merupakan proses yang digunakan untuk memastikan hanya hasil yang valid dan berguna yang dapat digunakan oleh pihak yang berkepentingan, tentunya fungsi deskripsi dalam data mining juga merupakan sebuah fungsi untuk memahami lebih jauh dan lebih dalam tentunya tentang data yang diamati, terlebih pada penggunaan informasi yang dalam satuan pendidikan seperti yang akan dilakukan. Hal ini tentunya Fungsi *datamining* ini dapat diharapkan menggambarkan lebih dalam informasi pada suatu hal yang diteliti. Dengan melakukan sebuah proses diharap bisa mengetahui perilaku dari sebuah data tersebut. Data tersebut itulah yang nantinya dapat digunakan untuk mengetahui karakteristik dari data yang dimaksud.

2. Predictive atau Prediksi, Fungsi ini juga bisa digunakan untuk memprediksi sebuah variabel tertentu yang tidak ada dalam suatu data. Sehingga fungsi ini memudahkan dan menguntungkan bagi siapapun yang memerlukan prediksi yang akurat untuk membuat hal penting tersebut menjadi lebih baik. Fungsi prediksi pada *datamining* merupakan sebuah fungsi bagaimana sebuah proses nantinya akan menemukan pola tertentu dari suatu data. Pola-pola tersebut dapat diketahui dari berbagai variabel-variabel yang ada pada data. Ketika sudah menemukan pola, Maka pola yang didapat tersebut bisa digunakan untuk memprediksi variabel lain yang belum diketahui nilai ataupun jenisnya. Karena itulah fungsi satu ini dikatakan sebagai fungsi prediksi sama halnya dengan melakukan predictive analisis, tentunya berbagai Prediksi memiliki kemiripan dengan klasifikasi, akan tetapi data diklasifikasikan berdasarkan perilaku atau nilai yang diperkirakan pada masa yang akan datang, dapat dipahami Contoh dari tugas prediksi misalnya untuk memprediksikan adanya pengurangan jumlah pelanggan dalam waktu dekat dan prediksi harga saham dalam tiga bulan yang akan datang. Tujuan dari tugas deskriptif adalah untuk menurunkan pola-pola (korelasi, trend, cluster, trayektori, dan anomali) yang meringkas hubungan yang pokok dalam data. Tugas data mining deskriptif sering merupakan penyelidikan dan seringkali memerlukan teknik post processing untuk validasi dan penjelasan hasil.

Fungsi Data mining yang lainnya yaitu :

1. Characterization, discrimination, association, classification, clustering, outlier and trend analysis, dll.
2. Multidimensional concept description, Karakterisasi dan diskriminasi, Atau berfungsi untuk Menggeneralisasikan, meringkas, dan membedakan karakteristik data, dll.
3. Frequent patterns, association, correlation.
4. Classification and prediction, Membangun model (fungsi) yang menggambarkan dan membedakan kelas atau konsep untuk prediksi masa depan. Misalnya, Mengklasifikasikan negara berdasarkan (iklim), atau mengklasifikasikan mobil berdasarkan (jarak tempuh gas).
5. Cluster analysis, Membuat data grup untuk membentuk kelas baru. Misalnya, Memaksimalkan kesamaan intra-kelas & meminimalkan kesamaan antar kelas, Outlier analysis, Objek data yang tidak sesuai dengan perilaku umum dari data, Berguna dalam deteksi penipuan, analisis peristiwa langka.
6. Trend and evolution analysis, Trend dan penyimpangan: misalnya Analisis regresi atau Mining Penambangan pola berurutan: misalnya, Kamera digital, atau Analisis periodisitas dan Analisis berbasis kesamaan. Other pattern-directed or statistical analyses

### 2.3 Naïve Bayes

*Naïve Bayes* merupakan sebuah pengklasifikasian probabilistik sederhana yang menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari *dataset* yang diberikan. Menurut Bustami dalam Saleh *Naïve Bayes* merupakan pengklasifikasian dengan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya. Metode *Naïve Bayes* juga dinilai berpotensi baik dalam pengklasifikasian lain dalam hal akurasi dan efisiensi [11]. Algoritma *Naïve Bayes* kurang lebih ditemukan pada pertengahan abad ke-18. Pada saat itu, algoritma ini dikenal dengan banyak nama. Meskipun begitu, algoritma ini populer dikenal sebagai metode pengelompokan teks dan pengkategorian menggunakan frekuensi kata-kata. *Simplenaïve Bayesian classifier* merupakan salah satu metode pengklasifikasi berpeluang sederhana yang berdasarkan pada penerapan *Teorema Bayes* dengan asumsi antar variabel penjelas saling bebas (independen). Algoritma menggunakan teorema Bayes dan mengasumsikan semua atribut independen atau tidak saling ketergantungan yang diberikan oleh nilai pada variabel kelas. Definisi lain dari *Naïve Bayes* adalah pengklasifikasian dengan menggunakan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu cara memprediksi peluang di masa depan berdasarkan hasil pengalaman di masa sebelumnya. *Naïve Bayes* didasarkan pada asumsi penyederhanaan bahwa nilai atribut secara kondisional saling bebas jika diberikan nilai

output. maka dari itu, diberikan nilai output, probabilitas mengamati secara bersama merupakan produk dari probabilitas individu. Keuntungan menggunakan metode *Naïve Bayes* adalah hanya membutuhkan jumlah data pelatihan (Training Data) yang kecil untuk menentukan estimasi parameter yang diperlukan dalam proses pengklasifikasian. *Naïve Bayes* dapat bekerja jauh lebih baik dalam suatu situasi dunia nyata yang kompleks daripada yang diharapkan. Algoritma *Naïve Bayes* adalah salah satu algoritma yang terdapat pada teknik klasifikasi. *Naïve Bayes* dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya sehingga disebut sebagai Teorema Bayes. Teorema Bayes dikombinasikan dengan *Naïve* dimana diasumsikan dengan kondisi antar atribut yang saling bebas. Klasifikasi *Naïve Bayes* diasumsikan bahwa ada atau tidak ciri tertentu dari sebuah kelas tidak berhubungan dengan ciri dari kelas lain, Klasifikasi adalah proses menemukan model (fungsi) yang menjelaskan dan membedakan kelas-kelas atau konsep, dengan tujuan agar model yang diperoleh dapat digunakan untuk memprediksi kelas atau objek yang memiliki label kelas tidak diketahui. Model yang turunkan didasarkan pada analisis dari training data (yaitu objek data yang memiliki label kelas yang diketahui). Model yang diturunkan dapat direpresentasikan dalam berbagai bentuk seperti aturan IF-THEN klasifikasi, pohon keputusan, Dalam banyak kasus, pengguna ingin memprediksi nilai-nilai data yang tidak tersedia atau hilang (bukan label dari kelas). Dalam kasus ini biasanya nilai data yang akan diprediksi merupakan data numeric. Kasus ini seringkali dirujuk sebagai prediksi. Di samping itu, prediksi lebih menekankan pada identifikasi trend dari distribusi berdasarkan pada data yang tersedia, tidak hanya itu tentunya akan formula matematika atau jaringan syaraf tiruan, Algoritma ini memanfaatkan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi probabilitas di masa depan berdasarkan pengalaman di masa sebelumnya. Dua kelompok peneliti, satu oleh Pantel dan Lin, dan yang lain oleh Microsoft Research memperkenalkan metode statistik Bayesian ini pada teknologi anti spamfilter.

Menurut Bustami dalam Saleh persamaan dari teorema Bayes adalah : [12]

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

Di mana :

$X$  : Data dengan *class* yang belum diketahui

$H$  : Hipotesis data merupakan suatu *class* yang spesifik

$P(H|X)$  : Probabilistik hipotesis  $H$  berdasarkan kondisi  $X$  (posteriori probabilistik)

$P(H)$  : Probabilistik hipotesis  $H$  (prior probabilitas)

$P(X|H)$  : Probabilistik hipotesis  $X$  berdasar kondisi pada hipotesis  $H$

$P(X)$  : Probabilitas  $X$

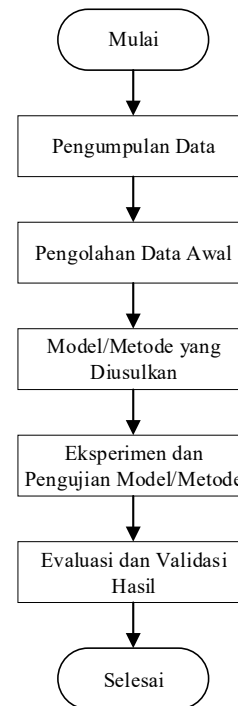
### III. METODOLOGI PENELITIAN

#### 3.1 Jenis Penelitian

Di dalam penelitian ini digunakan penelitian dengan metode eksperimen, Metode eksperimen adalah cara penyajian bahan pelajaran dimana siswa melakukan percobaan dengan mengalami untuk membuktikan sendiri sesuatu pertanyaan atau hipotesis yang dipelajari., Metode Eksperimen (Menguji coba), adalah penelitian untuk menguji apakah variabel-variabel eksperimen efektif atau tidak. Untuk menguji efektif tidaknya harus digunakan variabel kontrol. Penelitian eksperimen adalah untuk menguji hipotesis yang dirumuskan secara ketat. Penelitian eksperimen biasanya dilakukan untuk bidang yang bersifat eksak. Sedangkan untuk bidang sosial biasanya digunakan metode survey eksplanatory, metode deskriptif, dan historis. metode penelitian ini sejatinya juga dengan menggunakan data siswa yang mengajukan beasiswa di “SD Negeri 6 Ketapang” sejumlah 1000 orang, yang terdiri dari 460 siswa yang dinyatakan lulus seleksi dan 640 di antaranya yang dinyatakan tidak lulus seleksi.

#### 3.2 Metode Pengumpulan Data

Dalam penelitian ini metode pengumpulan data yang digunakan adalah metode pengumpulan data sekunder. Metode pengumpulan data sekunder sering disebut metode penggunaan bahan dokumen, karena dalam hal ini peneliti tidak secara langsung mengambil data sendiri tetapi meneliti dan memanfaatkan data atau dokumen yang dihasilkan oleh pihak-pihak lain. Dalam hal ini, data sekunder merupakan data primer yang diperoleh oleh pihak lain atau data primer yang telah diolah lebih lanjut dan disajikan baik oleh pengumpul data primer atau oleh pihak lain yang pada umumnya disajikan dalam bentuk tabel-tabel atau diagram-diagram, Data sekunder pada umumnya digunakan oleh peneliti untuk memberikan gambaran tambahan, gambaran pelengkap ataupun untuk diproses lebih lanjut [13]. Sebagai contoh, banyak informasi tentang manajer potensial yang diperoleh oleh suatu perusahaan dari terbitan yang dikeluarkan oleh badan riset yang dikelola oleh swasta. Data utama diperoleh dari data siswa yang lulus seleksi beasiswa dan yang tidak lulus seleksi beasiswa, sedangkan data pendukung lainnya didapat dari buku, jurnal dan publikasi lainnya. Berikut merupakan langkah-langkah yang digunakan dalam penelitian ini.



Gambar 1. Langkah-Langkah Penelitian

#### 3.3 Instrumen Penelitian

Pada penelitian ini peneliti membuat sebuah instrumen penelitian, dimana instrumen penelitian dapat diartikan bahwa instrumen penelitian adalah sebuah alat yang digunakan untuk mengumpulkan data atau informasi yang bermanfaat untuk menjawab permasalahan penelitian. Alat ini harus dipilih sesuai dengan jenis data yang diinginkan dalam sebuah penelitian. Instrumen penelitian yaitu sebagai berikut :

1. Penelitian ini menggunakan data sekunder berupa data siswa yang lulus seleksi dan tidak lulus seleksi yang akan digunakan sebagai instrumen guna memperoleh data dalam proses menentukan calon penerima beasiswa.
2. Perangkat lunak yang digunakan untuk menganalisis adalah *Rapidminer*.

#### 3.4 Metode Analisis dan Pengujian Data

Teknik analisis data menggunakan data kuantitatif berupa matematika terhadap angka atau numerik dan nominal. Data kuantitatif merupakan salah satu jenis data yang dapat diukur (measurable) ataupun dihitung secara langsung sebagai variabel angka dan bilangan. Variabel dalam ilmu statistika adalah atribut, karakteristik, atau pengukuran yang mendeskripsikan suatu kasus atau objek penelitian. Nilai data bisa berubah-ubah atau bersifat variatif. Proses pengumpulan data kuantitatif tidak membutuhkan banyak waktu dan sangat mudah dilakukan. Pada penelitian ini, analisis data dilakukan melalui data siswa SD dengan nilai rata-rata siswa lulus seleksi dan tidak lulus seleksi, data diolah dan diuji dalam pengujian pada algoritma *Naïve Bayes*. Kemudian pengujian *rule*

yang diperoleh algoritma *NaïveBayes* tersebut selanjutnya diuji dengan *Confusion Matrix* dan Kurva *ReceiverOperatingCharacteristic* (ROC) untuk mengukur tingkat akurasi yang akan dihasilkan.

Dengan pengujian di atas dapat diperoleh nilai akurasi dari *rule* algoritma *NaïveBayes*. Sehingga dapat diterapkan pada *GraphicalUserInterface* (GUI) dengan baik. Selanjutnya aplikasi yang dibuat akan dievaluasi untuk menghasilkan pengetahuan (*knowledge*) baru.

Dalam penelitian ini pengujian data menggunakan metode eksperimen dengan model *Cross Industry Standard Processfor Data Mining* (CRISP-DM) yang terdiri dari enam tahap, yaitu :

#### 1. Tahap Pemahaman Bisnis (*Business Understanding*)

Berdasarkan dari data penerimaan beasiswa tahun 2016 – 2018 siswa dinyatakan lulus menjadi penerima beasiswa berjumlah 460 orang, dengan memanfaatkan teknik klasifikasi pada *data mining* model *NaïveBayes* maka proses penyeleksian beasiswa dapat menjadi lebih efisien dan efektif, sejatinya pemahaman mengenai tahapan ini, Penentuan tujuan proyek dan kebutuhan secara detail dalam lingkup bisnis atau unit penelitian secara keseluruhan, kemudian fase pemahaman bisnis ini, Menerjemahkan tujuan dan batasan menjadi formula dari permasalahan data mining, selanjutnya Menyiapkan strategi awal untuk mencapai tujuan, Pengimplementasian data dikarenakan Seperti yang sudah dijelaskan bahwa pengimplementasian data mining pada aplikasi ini adalah berhubungan dengan data penentuan beasiswa SD Negeri 6 Ketapang dengan berbagai factor nantinya. Faktor-faktor apa saja yang mempengaruhi beasiswa SD Negeri 6 Ketapang sesuai data yang sudah ada. Memeriksa keterkaitan antara tiap atribut dan factor lainnya, factor tersebut dapat berupa penghasilan orang tua, jumlah saudara, dan keadaan rumahnya siswa SD Negeri 6 Ketapang. Tujuan dari pengimplementasian data mining pada aplikasi ini adalah agar didapatkan suatu pengetahuan tambahan mengenai produksi ikan sehingga dapat menjad tolak ukur mengenai potensi produksi ikan pada data yang ada sebagai review bagi pengguna sistem dengan melihat faktor-faktor yang mempengaruhi produksi tersebut.

#### 2. Tahap Pemahaman Data (*Data Understanding*)

Data yang digunakan adalah data sekunder, yang didapat dari data siswa yang mengajukan permohonan beasiswa, di dalam data tersebut dapat diketahui status siswa yang lulus seleksi dan tidak lulus seleksi yang terdiri dari 3 atribut *preditor* dan 1 atribut hasil. Tahap ini memberikan fondasi analitik untuk sebuah proyek dengan membuat ringkasan (*summary*) dan mengidentifikasi potensi masalah dalam data. Tahap ini juga harus dilakukan secara cermat dan tidak terburu - buru, seperti pada visualisasi data, yang terkadang insightnya sangat sulit didapat jika dihubungkan dengan *summary* datanya. Jika ada masalah pada tahap ini yang belum terjawab, maka akan mengganggu pada tahap modeling. Ringkasan atau *summary* dari data dapat berguna untuk mengkonfirmasi apakah data terdistribusi seperti yang

diharapkan, atau mengungkapkan penyimpangan tak terduga yang perlu ditangani pada tahap selanjutnya, yaitu Data Preperation, Masalah dalam data biasanya seperti nilai-nilai yang hilang, outlier, berdistribusi spike, berdistribusi bimodal harus diidentifikasi dan diukur sehingga dapat diperbaiki dalam Data Preperation.

Atribut-atribut yang menjadi parameter terlihat pada Tabel 1.

TABEL I  
ATRIBUT DAN NILAI KATEGORI SELEKSI BEASISWA

Atribut	Kategori
NILAI	Nilai => 80 Nilai <80
PEKERJAAN_ORTU	Pedagang Kecil Buruh Wiraswasta Karyawan Swasta Petani Sudah Meninggal Tidak bekerja Lainnya PNS/TNI/Polri Pedagang Besar Peternak Wirasaha Nelayan
PENGHASILAN ORTU	0 – 4.000.000
JUMLAH TANGGUGAN	0 – 4

#### 3. Tahap Data Persiapan (*Data Preparation*)

Tahap ini jelas membutuhkan pemikiran yang cukup matang dan usaha yang cukup tinggi untuk memastikan data tepat untuk algoritma yang digunakan. Bukan berarti saat Data Preperation pertama kali dimana masalah-masalah pada data sudah diselesaikan, data sudah dapat digunakan hingga tahap terakhir. Tahap ini merupakan tahap yang sering ditinjau kembali saat menemukan masalah pada saat pembangunan model. Sehingga dilakukan iterasi sampai menemukan hal yang cocok dengan data. Tahap sampling dapat dilakukan disini dan data secara umum dibagi menjadi dua, data training dan data testing. Jumlah data yang diperoleh pada penelitian ini sebanyak 1000 *record*, baik siswa yang lulus seleksi maupun yang tidak lulus seleksi, akan tetapi data tersebut masih mengandung duplikasi dan anomali atau inkonsisten data. Untuk mendapatkan data yang berkualitas, ada beberapa teknik *preprocessing* yang digunakan yaitu :

a. *Data Validation*, untuk mengidentifikasi dan menghapus anomali dan inkonsistensi data. Anomali dan inkonsistensi data dapat terjadi dikarenakan adanya *missingvalues* dan juga karena data yang dipakai mengandung nilai-nilai yang salah atau disebut *noise* atau *outlier*. Pada penelitian ini digunakan *data validation excel* yang tentunya akan mengefektifkan dan mengefisien kan penulis dalam meneliti hal ini tentunya, Data validation Excel adalah salah satu fitur excel yang digunakan untuk membatasi input data yang bisa dimasukkan ke dalam sebuah sel atau range, sehingga data yang diinputkan ke dalam sebuah sel

atau range valid atau memenuhi kriteria tertentu yang ditetapkan., *Data Validation* Manfaat dari data ini yakni ialah Membatasi isi sel excel dengan nilai teks tertentu, Membatasi isi sel dengan rentang nilai angka, tanggal atau waktu tertentu, Membuat daftar atau list data yang bisa dimasukkan ke dalam sel, Menampilkan pesan tentang jenis data yang bisa Anda masukkan ke dalam sel excel, Menampilkan pesan tertentu jika data yang Anda masukkan ke dalam sel tidak sesuai.

- b. *Data Integration and Transformation*, untuk meningkatkan akurasi dan efisiensi algoritma. Data yang digunakan dalam penulisan ini kategorial, Data Integration atau integrasi data merupakan proses menggabungkan atau menyatukan dua atau lebih sebuah data dari berbagai sumber database yang berbeda ke dalam sebuah penyimpanan seperti gudang data (data warehouse), Integrasi data dilakukan pada atribut-atribut yang mengidentifikasi entitas-entitas yang unik seperti atribut nama, jenis produk, nomor pelanggan dsb. Integrasi data perlu dilakukan secara cermat karena kesalahan pada integrasi data bisa menghasilkan hasil yang menyimpang dan bahkan menyakutkan pengambilan aksi nantinya. Sebagai contoh bila integrasi data berdasarkan jenis produk ternyata menggabungkan produk dari kategori yang berbeda maka akan didapatkan korelasi antar produk yang sebenarnya tidak ada. Dalam integrasi data ini juga perlu dilakukan transformasi dan pembersihan data karena seringkali data dari dua database berbeda tidak sama cara penulisannya atau bahkan data yang ada di satu database ternyata tidak ada di database lainnya, Transformasi data adalah merubah skala data kedalam bentuk lain sehingga data memiliki distribusi yang diharapkan, Setiap data dilakukan operasi matematika yang sama pada data aslinya, Beberapa teknik data mining membutuhkan format data yang khusus sebelum bisa diaplikasikan. Sebagai contoh beberapa teknik standar seperti analisis asosiasi dan klustering hanya bisa menerima input data kategorial. Karenanya data berupa angka numerik yang berlanjut perlu dibagi-bagi menjadi beberapa interval. Proses ini sering disebut binning. Disini juga dilakukan pemilihan data yang diperlukan oleh teknik data mining yang dipakai. Transformasi dan pemilihan data ini juga menentukan kualitas dari hasil data mining nantinya karena ada beberapa karakteristik dari teknik-teknik data mining tertentu yang tergantung pada tahapan ini.
- c. *Data Size Reduction and Discretization*, untuk memperoleh *dataset* dengan jumlah atribut dan *record* yang lebih sedikit tetapi bersifat *informative*. Di dalam data *training* yang digunakan dalam penelitian ini, dilakukan penghapusan data duplikasi angka menggunakan *tools Ms. Excel 2013*.

Setelah dilakukan *preprocessing* data yang didapat dari data internal sekolah sebanyak 1000 *record*, 581 *record* lulus seleksi dan 419 tidak lulus seleksi.

#### 4. Tahap Pemodelan (*Modelling*)

Tahap ini juga disebut tahap *learning* karena pada tahap ini data diklasifikasikan oleh model dan kemudian menghasilkan sejumlah aturan. Pada penelitian ini, pembuatan model menggunakan algoritma *Naive Bayes*. Secara garis besar untuk membuat model prediktif atau deskriptif, Pada tahap ini akhirnya kita dapat menggunakan Statistika dan Machine Learning untuk mendapatkan insight yang berguna dari data untuk mencapai tujuan proyek. Pada penelitian yang berjudul 'Data Clustering menggunakan metodologi CRISP-DM untuk pengenalan pola proporsi pelaksanaan tridharma, didefinisikan 'tahapan pemodelan merupakan tahapan yang secara langsung melibatkan data mining. Pemilihan teknik datamining, algoritma dan menentukan parameter dengan nilai yang optimal'. Pada tahapan pemodelan, ada beberapa hal yang dilakukan antara lain, memilih teknik pemodelan, membangun model, dan menilai model. Tahapan yang dilakukan mengacu pada penelitian tersebut. Beberapa modeling yang biasa dilakukan adalah classification, scoring, ranking, clustering, finding relation, characterization.

#### 5. Tahap Evaluasi (*Evaluation*)

Pada tahap ini dilakukan pengujian terhadap model-model untuk mendapatkan informasi model yang akurat. Evaluasi dan validasi menggunakan metode *Confusion Matrix* dan Kurva ROC. Secara garis besar untuk menilai model agar dapat melaporkan efek yang diharapkan dari model, Setelah mempunyai model, kita harus menentukan apakah sesuai dengan tujuan kita. Beberapa pertanyaan di bawah ini bisa membantu apakah model kita sudah sesuai dengan tujuan atau belum.

#### 6. Tahap *Deployment*

Pada tahap ini diterapkan algoritma *Naive Bayes* pada GUI yang akan diterapkan pada instansi terkait. Secara garis besar untuk rencana penggunaan model, Tahap ini adalah tahap yang paling dihargai dari proses CRISP-DM. Perencanaan untuk Deployment dimulai selama Business Understanding dan harus menggabungkan tidak hanya bagaimana untuk menghasilkan nilai model, tetapi juga bagaimana mengkonversi skor keputusan, dan bagaimana untuk menggabungkan keputusan dalam sistem operasional. Pada akhirnya, rencana sistem Deployment mengakui bahwa tidak ada model yang statis. Model tersebut dibangun dari data yang diwakili data pada waktu tertentu, sehingga perubahan waktu dapat menyebabkan berubahnya karakteristik data. Modelpun harus dipantau dan mungkin diganti dengan model yang sudah diperbaiki.

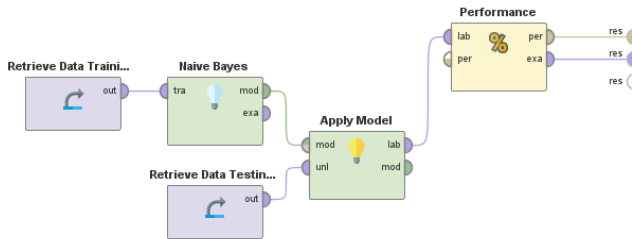
### IV. HASIL DAN PEMBAHASAN

#### 4.1 Hasil Pengujian Model

Setelah tahap *preprocessing* dilakukan kemudian langkah selanjutnya adalah tahapan uji coba (eksperimental). Pada tahapan ini uji coba pertama yang dilakukan adalah mengimplementasi model *Naive Bayes* dengan mengabaikan atribut yang berisikan jenis data



*numerical*. Pemahaman mengenai Jenis Data Numerical ini ialah Data numerik atau dalam bahasa Inggris dikenal dengan sebutan numerical adalah data di mana ia direpresentasikan dalam sebuah angka. Karakteristik dari angka adalah ia merepresentasikan sifat kuantitatif dari data itu sendiri. Maksud dari kuantitatif adalah jumlah, kadar, kapasitas yang bisa diukur dengan jelas. Gambar 2 merupakan hasil dari uji coba *supervised learning NaïveBayes*.



Gambar 2 Implementasi *Naïve Bayes* pada *RapidMiner*

Dalam membuat model *NaïveBayes* terlebih dahulu kita mencari probabilitas hipotesis untuk masing-masing kelas  $P(H)$  tujuannya untuk menetapkan suatu dasar sehingga dapat mengumpulkan bukti yang berupa data-data dalam menentukan keputusan apakah menolak atau menerima kebenaran dari pernyataan atau asumsi yang telah dibuat. Sedangkan Probabilitas Hipotesis yang digunakan pada penelitian ini Membantu peneliti dalam pengambilan keputusan yang lebih tepat. Pengambilan keputusan yang lebih tepat dimaksudkan tidak ada keputusan yang sudah pasti karena kehidupan mendatang tidak ada yang pasti kita ketahui dari sekarang, karena informasi yang didapat tidaklah sempurna. Dengan teori probabilitas kita dapat menarik kesimpulan secara tepat atas hipotesis yang terkait tentang karakteristik populasi. Menarik kesimpulan secara tepat atas hipotesis (perkiraan sementara yang belum teruji kebenarannya) yang terkait tentang karakteristik populasi pada situasi ini kita hanya mengambil atau menarik kesimpulan dari hipotesis bukan berarti kejadian yang akan datang kita sudah ketahui apa yang akan terjadi. Hipotesis yang ada yaitu siswa yang lulus seleksi beasiswa dan siswa yang tidak lulus seleksi beasiswa. Data yang digunakan adalah data utama yang diperoleh dari data beasiswa SD 6 Ketapang, dengan total data 1000 data. 581 siswa yang lulus seleksi dan 491 siswa yang tidak lulus seleksi, perhitungan probabilitas yaitu seperti di bawah ini :

TABEL II  
PROBABILITAS PRIOR

Atribut		Jml Kasus	Lulus	Tidak Lulus	P(X Ci)	
					Lulus	Tidak Lulus
Total		1000	581	419		
NILAI	Nilai $\Rightarrow$ 80	485	272	213	0.468	0.508
	Nilai $<$ 80	515	308	207	0.530	0.494
PEKERJAAN_ORTU	Buruh	136	92	44	0.158	0.105
	Karyawan Swasta	303	178	125	0.306	0.298
	Lainnya	46	30	16	0.052	0.038
	Nelayan	2	2	0	0.003	0.000
	Pedagang	6	2	4	0.003	0.010

Atribut	Jml Kasus	Lulus	Tidak Lulus	P(X Ci)	
				Lulus	Tidak Lulus
Besar					
Pedagang Kecil	58	36	22	0.062	0.053
Petani	35	16	19	0.028	0.045
Peternak	10	6	4	0.010	0.010
PNS/TNI/Polri	37	24	13	0.041	0.031
Wiraswasta	292	160	132	0.275	0.315
Wirausaha	9	2	7	0.003	0.017
Sudah Meninggal	8	6	2	0.010	0.005
Tidak Bekerja	58	29	29	0.050	0.069
PENGHASILAN_ORTU	0	8	6	0.010	0.005
	500000	69	31	0.053	0.091
	700000	43	34	0.059	0.045
	800000	116	77	0.133	0.093
	1000000	110	56	0.096	0.129
	1300000	2	2	0.003	0.000
	1500000	111	67	0.115	0.105
	1700000	57	43	0.074	0.033
	1950000	57	32	0.055	0.060
	2000000	52	28	0.048	0.057
	2175000	57	30	0.052	0.064
	2400000	57	26	0.045	0.074
	2500000	67	49	0.084	0.043
	2950000	57	28	0.048	0.069
	3500000	57	29	0.050	0.067
	3864000	52	28	0.048	0.057
	4000000	18	14	0.024	0.010
JUMLAH TANGGUNGAN	0	8	6	0.010	0.005
	1	192	112	0.193	0.191
	2	441	259	0.446	0.434
	3	293	167	0.287	0.301
	4	60	32	0.055	0.067
	5	6	4	0.007	0.005

$$P(\text{Lolos}) = 581:1000 = 0,581$$

$$P(\text{Tidak Lolos}) = 419:1000 = 0,419$$

Setelah probabilitas untuk tiap hipotesis diketahui, langkah selanjutnya adalah menghitung probabilitas kondisi tertentu (probabilitas X) berdasarkan probabilitas tiap hipotesis (probabilitas H) atau dinamakan probabilitas prior. Hasil perhitungan probabilitas prior dengan menggunakan *Naïve Bayes* dapat dilihat pada tabel dibawah ini:

Nilai *accuracy* dapat diartikan bahwa Data Accuracy dapat diartikan bahwa kebenaran data, karena data harus mewakili apa yang dimaksudkan dan didefinisikan oleh sumber aslinya. Nilai *Accuracy* dari data *training* dapat dihitung dengan menggunakan *RapidMiner*. Setelah diuji coba dengan metode *cross validation*, didapat hasil pengukuran terhadap data *training* yaitu *accuracy*=76,00%. *RapidMiner* adalah platform perangkat lunak ilmu data yang dikembangkan oleh perusahaan bernama sama dengan yang menyediakan lingkungan terintegrasi untuk persiapan data, pembelajaran mesin, pembelajaran dalam, penambangan teks, dan analisis prediktif. Hal ini digunakan untuk bisnis dan komersial, juga untuk penelitian, pendidikan, pelatihan, rapid prototyping, dan pengembangan aplikasi

serta mendukung semua langkah dalam proses pembelajaran mesin termasuk persiapan data, hasil visualisasi, validasi model, dan optimasi.

#### 4.2 Evaluasi Model Confusion Matrix

Tabel 3 adalah perhitungan berdasarkan data *training*, Data training digunakan untuk melatih algoritma, sedangkan data testing dipakai untuk mengetahui performa algoritma yang sudah dilatih sebelumnya ketika menemukan data baru yang belum pernah dilihat sebelumnya. Ini biasanya disebut dengan generalisasi. Hasil dari pelatihan tersebut bisa disebut dengan model. diketahui dari 100 data, 76 diklasifikasikan LOLOS sesuai dengan prediksi yang dilakukan dengan metode algoritma *NaïveBayes*, 21 data diprediksi LOLOS tetapi ternyata hasilnya TIDAK LOLOS, 3 data *class* TIDAK LOLOS diprediksi tidak sesuai.

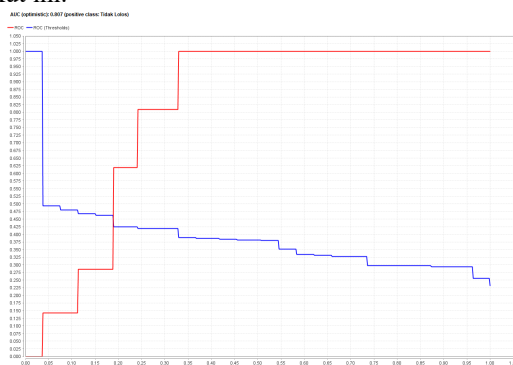
**TABEL III**

MODEL CONFUSION MATRIX UNTUK METODE ALGORITMA *NAÏVEBAYES*.

Accuracy : 76,00%			
	true Lolos	true Tidak Lolos	class precision
pred. Lolos	76	21	78.35%
pred. Tidak Lolos	3	0	0.00%
class recall	96.20%	0.00%	

#### 4.3 Kurva ROC (ReceiverOperatingCharacteristic)

Hasil perhitungan kurva ROC untuk algoritma *NaïveBayes* dengan menggunakan data *training* hasilnya sebesar 0,807 dan dapat dilihat dari kurvanya seperti berikut ini.



Gambar 3 Kurva ROC Menggunakan Data *Training*

Gambar 3 menunjukkan bahwa hasil algoritma *Naïve Bayes* mendapatkan nilai **0,807**, di mana nilai tersebut menunjukkan nilai AUC dengan klasifikasi baik, dikarenakan memiliki nilai di atas 0,80.

#### V. KESIMPULAN

Dalam penelitian ini dilakukan pengujian model algoritma *NaïveBayes*, hasil evaluasi dan validasi,

diketahui bahwa *NaïveBayes* memiliki nilai akurasi cukup tinggi yaitu sebesar 76,00%. Dengan demikian, metode *NaïveBayes* merupakan metode yang cukup baik dalam menentukan calon penerima beasiswa secara lebih efektif dan efisien, terebih penentuan beasiswa pada SD Negeri 6 Ketapang menjadi Solusi terbaik dalam penentuannya.

#### DAFTAR PUSTAKA

- [1] F. Wiranata, "Perancangan Sistem Pendukung Keputusan Penerimaan Beasiswa SMA Muhammadiyah 1 Gisting Dengan Metode SAW (Simple Additive Weighting)," *Jur. Sist. Inf. STMIK*, vol. 4, no. 1, 2016.
- [2] Gunawan, R. P. Kesuma, and R. R. Wigati, "Pengembangan Sistem Penunjang Keputusan Penentuan Pemberian Beasiswa Tingkat Sekolah," *JSM STMIK Mikroskil*, 2013.
- [3] A. Saleh, "Implementasi Metode Klasifikasi Naïve Bayes Dalam Memprediksi Besarnya Penggunaan Listrik RumahTangga," *Citec J.*, p. 209, 2015.
- [4] R. A. Saputra, "Penerapan Algoritma Naive Bayes Untuk Prediksi Penyakit Tuberculosis," *Swabumi*, vol. 19, 2014.
- [5] H. Wasiati and D. Wijayanti, "Sistem Pendukung Keputusan Penentuan Kelayakan Calon Tenaga Kerja Indonesia Menggunakan Naive Bayes," *IJNS*, vol. 45, 2014.
- [6] N. Hijriana and M. Rasyidan, "Penerapan Metode Decision Tree Algoritma C4.5 Untuk Seleksi Calon Penerima Beasiswa Tingkat Universitas," *Al Ulum Sains Dan Teknol.*, vol. 3, no. 1, Nov. 2017.
- [7] A. Putra and D. Y. Hardiyanti, "Penentuan Penerima Beasiswa Dengan Menggunakan Fuzzy Multiple Atribute Decision Making," *JSI*, 2011.
- [8] J. Augusto and D. Guterres, "Kelayakan Algoritma C4.5 Sebagai Pendukung Keputusan Dalam Pengajuan Penerima Beasiswa," pp. 142–147, 2015.
- [9] S. A. Pattekari and A. Parveen, "Prediction System for Heart Disease Using Naive Bayes," *Int. J. Adv. Comput. Math. Sci.*, vol. 3, no. 3, pp. 290–294, 2012.
- [10] S. Riswanto, "Penerapan Data Mining Untuk Memprediksi Fluktuasi Harga Saham Menggunakan Metode Classification Dengan Teknik Decision Tree," Universitas Komputer Indonesia, 2011.
- [11] B. Setyaji and Pujiono, "Sistem Pendukung Keputusan Penentuan Kelayakan Calon Tenaga Kerja Menggunakan Metode Naive Bayes (Studi Kasus CV. Lingkar Aksi)," pp. 1–9, 2015.
- [12] A. Saleh, "Penerapan Data Mining Dengan Metode Klasifikasi Naïve Bayes Untuk Memprediksi Kelulusan Mahasiswa Dalam Mengikuti English Proficiency Test," *ResearchGate*, Feb. 2015.
- [13] N. Martono, *Metode Penelitian Kuantitatif: Analisis Isi dan Analisis Data Sekunder*, 2nd ed. Jakarta: PT. RajaGrafindo Persada, 2010.