

Penerapan Data Mining untuk Memprediksi Hasil Kelulusan Siswa Menggunakan Metode *Naïve Bayes*

Rizky Amalia

Abstrak— Pertumbuhan yang pesat dari akumulasi data telah menciptakan kondisi kaya akan data tetapi minim informasi. Data mining merupakan cara untuk menemukan informasi dengan mencari pola atau aturan tertentu dari data dalam jumlah besar yang diharapkan dapat mengatasi kondisi tersebut. Dengan memanfaatkan data kelulusan siswa sebagai sumber datanya, diharapkan dapat menghasilkan informasi tentang pola tingkat kelulusan siswa melalui teknik *datamining*. Kategori tingkat kelulusan diukur dari daftar kolektif hasil ujian sekolah berstandar nasional (DKHUSBN). Algoritma yang digunakan adalah algoritma *Naïve Bayes*. Proses pada aplikasi ini ada 2 macam yaitu, proses analisis pola data kelulusan siswa yang telah ada sebelumnya (*Learning Phase*) berdasarkan atribut-atribut yang diujikan dan proses dari analisa pola data baru yang diujikan berdasarkan pola yang telah ada (*Testing Phase*). Informasi yang ditampilkan pada aplikasi tersebut ada 2 macam yaitu, informasi hasil proses *Learning Phase* dan informasi data berupa nilai probabilitas posterior (kemungkinan kemunculan) dari masing-masing kategori tingkat kelulusan.

Kata Kunci: *Data Mining, Naïve Bayes, Learning Phase, Testing Phase, Tingkat Kelulusan.*

Abstract—The rapid growth of data accumulation has created conditions that are rich in data but lack information. Data mining is a way to find information by looking for certain patterns or rules from large amounts of data that are expected to overcome these conditions. By utilizing student graduation data as the source of the data, it is expected to produce information about the graduation rate patterns through data mining techniques. The graduation level category is measured from a collective list of national standard school examination results (DKHUSBN). The algorithm used is the Naïve Bayes algorithm. The process in this application there are 2 kinds, namely, the process of analyzing student graduation data patterns that have been there before (*Learning Phase*) based on the attributes tested and the process of analyzing new data patterns that are tested based on existing patterns (*Testing Phase*). There are 2 kinds of information displayed in the application, namely, information on the results of the *Learning Phase* process and data information in the form of posterior probability values (probability of occurrence) of each graduation level category.

Mahasiswa, Jurusan Sistem Informasi Fakultas Ilmu Komputer Universitas Darwan Ali, Jl. Batu Berlian Kab. Kotawaringin Timur 74322 INDONESIA (telp: 0531-2065891; fax: 0531-2065891; e-mail: info@unda.ac.id)

Keywords: *Data Mining, Naïve Bayes, Learning Phase, Testing Phase, Graduation Level.*

I. PENDAHULUAN

1.1 Latar Belakang

Dengan berbagai kemajuan teknologi informasi dewasa ini, segala kebutuhan demi kebutuhan akan informasi yang akurat tentunya akan sangat dibutuhkan dalam kehidupan, sehingga informasi akan menjadi suatu elemen penting dalam perkembangan masyarakat saat ini hingga waktu mendatang. Namun sejatinya 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. Salah satunya yaitu pada bidang Pendidikan, dengan kemajuan teknologi informasi yang dibutuhkan, Pada Pemahamannya Pendidikan merupakan salah satu faktor kemajuan ataupun sebuah kemandirian bangsa. Semakin majunya pendidikan suatu bangsa, dari hal ini akan semakin maju ataupun sebuah mandiri bangsa tersebut. Melalui pendidikan para generasi penerus bangsa dibentuk kualitas siswa tersebut, Pendidikan nasional berfungsi untuk mengembangkan berbagai kemampuan dan membentuk watak serta peradaban bangsa yang bermartabat dalam rangka mencerdaskan kehidupan bangsa. Dari hal itulah dapat dinyatakan juga Pendidikan juga sangat membutuhkan informasi dari berbagai sistem informasi seperti halnya Penggunaan Teknik *datamining*. Penggunaan teknik *datamining* diharapkan dapat memberikan pengetahuan-pengetahuan yang sebelumnya tersembunyi di dalam gudang data sehingga menjadi informasi yang berharga [1].

Didalam dunia Pendidikan tentunya ada standarisasi dalam mengukur kelulusan siswa ataupun peserta didiknya, standarisasi merupakan salah satu bentuk kesesuaian dan kepadaman mengikuti suatu pedoman, hal ini dapat diartikan bahwa Pendidikan juga mempunya kesesuaian dan standar kelulusan siswa tersebut. Standarisasi Dunia Pendidikan Sekolah dasar hingga Sekolah Menengah adalah Paduan Nilai Ujian Nasional dan Nilai Rapor sekolah, yang menunjukkan kemampuan siswa itu di Sekolah. 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 [2].

Sejatinya Sistem yang akan dibuat ini tentu sejatinya akan membantu pihak sekolah mengetahui pola kelulusan dari siswa-siswinya dengan memanfaatkan data siswa dan data kelulusan, Hal ini dapat diartikan bahwa sistem ini nantinya akan diharapkan diharapkan bisa menganalisa faktor-faktor yang sangat berpengaruh pada tingkat kelulusan. Dalam masalah ini, dilakukan analisis berbagai faktor seperti apa yang akan diharapkan penelitian ini, faktor ini tentunya yang akan 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 dan peranan penting dalam menentukan tingkat kelulusan siswa tersebut. Sebagai penilaianya, siswa akan menggunakan nilai belajarnya selama disekolah sebagai standar ataupun akan dijadikan sebuah acuan untuk tingkat kelulusan siswa tersebut. Dengan demikian, nilai belajar siswa tersebut dapat menentukan mengapa tingkat kelulusan siswa yang satu dengan yang lain dapat berbeda [3].

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.

1.2 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*?

1.3 Tujuan Penelitian

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

1.4 Manfaat Penelitian

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

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

1.5 Batasan Masalah

Batasan masalah dalam penelitian ini adalah :

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

II. LANDASAN TEORI

2.1 Data Mining

Datamining adalah kegiatan menemukan pola yang menarik dari data dalam jumlah besar, data dapat disimpan dalam *database*, *data warehouse*, atau penyimpanan informasi lainnya. *Datamining* berkaitan dengan bidang ilmu-ilmu lain, seperti *databasesystem*, *data warehousing*, *statistik*, *machine learning*, *informationretrieval*, dan komputasi tingkat tinggi. Selain itu, *datamining* didukung oleh ilmu lain seperti *neuralnetwork*, pengenalan pola, *spatialdataanalysis*, *imagedatabase*, *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 [4]. Pada Prinsipnya Teknik data mining dimanfaatkan untuk memeriksa basis data berukuran besar sebagai cara untuk menemukan pola yang baru dan berguna tentunya, Sejatinya tidak semua pekerjaan ataupun berbagai pekerjaan pencarian informasi dinyatakan sebagai data mining. Data mining adalah bagian integral dari *knowledge discovery in databases* (KDD). Keseluruhan proses KDD untuk konversi raw data ke dalam informasi yang berguna[5].

Proses dalam pelaksanaan pada sistem Data Mining ini, yang diawali Data input sejatinya dapat disimpan dalam berbagai format seperti flat file, spreadsheet, ataupun tabel-tabel relasional, hingga dapat menempati tempat penyimpanan data terpusat dan terdistribusi pada banyak tempat. Tujuan dari preprocessing adalah mentransformasikan data input mentah ke dalam format yang sesuai untuk analisis selanjutnya. Selanjutnya Langkah-langkah yang terlibat dalam preprocessing data meliputi mengabungkan data dari berbagai sumber, membersihkan (cleaning) data agar membuang noise dan observasi duplikat, dan menyeleksi record hingga fitur yang relevan untuk pekerjaan data mining. Karena terdapat banyak cara mengumpulkan dan menyimpan data, tahapan preprocessing data merupakan langkah yang

banyak menghabiskan waktu dalam KDD[6]. Kemudian untuk hal itu seperti halnya 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 post processing yang menjamin bahwa hanya hasil yang valid dan berguna yang akan digabungkan dengan DSS, salah satu pekerjaan dan postprocessing adalah visualisasi yang memungkinkan analis untuk mengeksplor data dan hasil data mining dari berbagai sudut pandang. Ukuran-ukuran statistik dan metode pengujian hipotesis dapat digunakan selama post processing untuk membuang hasil data mining yang palsu.

Secara khusus pada pemahamannya 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.

Selain itu, Sejatinya datamining ini memiliki tujuan tujuan dari Penggunaan datamining ini, seperti halnya tujuannya yaitu Explanatory merupakan Untuk menjelaskan beberapa kondisi penelitian, seperti mengapa penjualan truk pick up meningkat di colorado, selanjutnya Confirmatory Untuk mempertegas hipotesis, seperti halnya 2 kali pendapatan keluarga lebih suka di pakai untuk membeli peralatan keluarga, di bandingkan dengan satu kali pendapatan keluarga, tidak dapat dipungkiri juga tujuan dari datamining ini Exploratory Menganalisis data untuk hubungan yang baru yang tidak di harapkan, seperti halnya pola apa yang cocok untuk kasus penggelapan kartu kredit.

Data mining merupakan sebuah proses interatif dan interaktif untuk mendapatkan sebuah pola baru yang menarik. Pola tersebut tentunya akan sangat bermanfaat. Model yang dihasilkan dari proses data mining biasanya sudah sempurna sehingga dapat digeneralisasi untuk kepentingan di masa depan[7].

Karena prosesnya yang cukup panjang dan rumit, maka dari proses awal biasanya akan menghasilkan sesuatu yang baru, yang tidak diketahui sebelumnya. Sesuatu yang baru ini akan menambah pengetahuan para pengguna ataupun peneliti dan tentunya akan sangat bermanfaat karena dapat digunakan untuk melakukan tindakan tertentu.

Penggalian data juga sering dikatakan sebagai proses interaktif dan interatif. Proses interaktif maksudnya yaitu proses yang masih memerlukan interaksi manusia agar

bisa terlaksana. Sedangkan proses interatif, maksudnya adalah proses yang tidak hanya dilakukan sekali, perlu proses yang berulang-ulang untuk mendapatkan data penting yang dimaksud.

Kelebihannya membuat analisa suatu data besar menjadi semakin mudah. Pencarian pola baru atau trend baru bisa dilakukan dengan mudah sehingga bisa membantu mengambil keputusan di masa yang akan datang atau bisa memprediksi data tertentu sehingga bisa menganalisis apa yang harus dilakukan.

2.2 Fungsi Data Mining

Seperti yang 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, tentunya fungsi deskripsi dalam data mining juga merupakan sebuah fungsi untuk memahami lebih jauh dan lebih dalam 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. Prediction atau fungsi prediksi merupakan salah satu fungsi data mining. Maksudnya yaitu dari proses nanti akan menemukan pola tertentu dari suatu data. Pola tersebut dapat diketahui dari variabel-variabel yang ada pada data. Pola yang didapat bisa digunakan untuk memprediksi variabel lain yang belum diketahui nilai ataupun jenisnya, Karena itulah fungsi satu ini dikatakan sebagai fungsi prediksi. Nantinya bisa digunakan untuk memprediksi variabel tertentu yang tidak ada dalam suatu data. Hal ini tentunya memudahkan dan menguntungkan bagi mereka pemilik kepentingan yang memerlukan prediksi akurat untuk membuat hal penting tersebut menjadi lebih baik. 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.

2. Predictive atau Prediksi, 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. 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 memprediksi adanya pengurangan jumlah pelanggan dalam waktu dekat dan prediksi harga saham dalam tiga bulan yang akan datang. 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.
3. Fungsi Description, Fungsi selanjutnya adalah description atau fungsi deskripsi. Maksud dari fungsi deskripsi ini yaitu untuk memahami lebih jauh tentang data yang diamati. Jadi dengan melakukan proses, diharap mampu mengetahui perilaku dari data tersebut yang nantinya bisa digunakan untuk mengetahui karakteristik dari data yang dimaksud, Data mining nantinya bisa menemukan pola tertentu yang tersembunyi dalam sebuah data. Dengan pola yang berulang dan bernilai itulah karakteristik data bisa diketahui. Hal satu ini tentunya memberikan banyak manfaat dan dapat meningkatkan pengetahuan.
4. Fungsi Klasifikasi, Fungsi Fungsi lainnya adalah fungsi klasifikasi atau classification. Maksud dari fungsi klasifikasi yaitu data yang ada akan diproses sehingga akan ditemukan fungsi atau model tertentu yang menggambarkan konsep dari suatu data. Model atau fungsi tersebut nantinya akan memisahkan tiap data menjadi kelompok-kelompok tertentu, Kelompok data tersebut nantinya bisa digunakan untuk meramalkan kecenderungan suatu data di masa depan. Pengelompokan atau pengklasifikasian data juga dapat

memudahkan pemilik data saat mencari data yang dibutuhkan.

5. Fungsi Asosiasi, Fungsi selanjutnya ialah fungsi asosiasi. Maksud dari fungsi asosiasi atau analisis asosiasi yaitu penggunaannya untuk menemukan kombinasi atau aturan assosiatif dari suatu data. Jadi data yang ada nantinya diproses sehingga akan menemukan informasi tentang hubungan variabel satu dengan lainnya. Agar mudah dipahami, contoh permisalannya pada analisis pembelian barang di swalayan. Semisal dari data pembelian diproses dan ternyata memperoleh hasil hubungan antara pembelian mie dan kecap. Jika besar kemungkinan pelanggan membeli mie dan kecap secara bersamaan, maka pihak swalayan bisa memanfaatkan informasi tersebut untuk mengatur penempatan mie dan kecap. Kecap bisa diletakkan di rak yang tidak jauh dari mie. Bisa juga menggunakan kecap sebagai bonus pemesanan mie atau bisa juga menggunakan cara lain, yang jelas menggambarkan bentuk hubungan mie dan kecap, Hal ini tentu saja menguntungkan.

Frequent patterns, association, correlation

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)

Cluster analysis, Membuat data grup untuk membentuk kelas baru. Misalnya, Memaksimalkan kesamaan intrakelas & meminimalkan kesamaan antar kelas, Outlier analysis, Objek data yang tidak sesuai dengan perilaku umum dari data, Berguna dalam deteksi penipuan, analisis peristiwa langka.

Multidimensional concept description, Karakterisasi dan diskriminasi, Atau berfungsi untuk Menggeneralisasikan, meringkas, dan membedakan karakteristik data, dll.

Trend and evolution analysis, Trend dan penyimpangan: misalnya Analisis regresi atau Mining Penambangan pola berurutan: misalnya, Kamera digital, atau Analisis periodik dan Analisis berbasis kesamaan. Other pattern-directed or statistical analyses.

2.3 Tahapan Data Mining

Tahap-tahap *data mining* ada 7, yaitu : [8]

1. Pembersihan data (*datacleaning*), Pembersihan data merupakan proses menghilangkan *noise* dan data yang tidak konsisten atau data tidak relevan. Integrasi data (*dataintegration*). Pada umumnya data yang diperoleh, baik dari database memiliki isian-isian

yang tidak sempurna seperti data yang hilang, data yang tidak valid atau juga hanya sekedar salah ketik, selanjutnya Data-data yang tidak relevan itu juga lebih baik dibuang, hal ini juga diartikan Pembersihan data juga akan mempengaruhi performasi dari teknik data mining karena data yang ditangani akan berkurang jumlah dan kompleksitasnya.

2. Integrasi data merupakan penggabungan data dari berbagai *database* ke dalam satu *database* baru. Integrasi data perlu dilakukan secara cermat karena kesalahan pada integrasi data bisa menghasilkan hasil yang menyimpang dan bahkan menyesatkan pengambilan aksi nantinya. Hal ini dapat di contohkan bila integrasi data berdasarkan jenis produk ternyata menggabungkan produk dari kategori yang berbeda maka akan didapatkan korelasi antar produk yang sebenarnya tidak ada.
3. Seleksi Data (*DataSelection*), Tahapan selanjutnya dalam adalah seleksi. Proses seleksi merupakan proses penyeleksian data. Data yang diseleksi akan ditransformasikan ke format yang sesuai untuk analisis data. Seleksi data menggunakan beberapa kriteria. Data hasil seleksi kemudian akan disimpan di suatu berkas terpisah yang kemudian akan diolah atau dilakukan proses data mining, data yang ada pada *database* sering kali tidak semuanya dipakai, oleh karena itu hanya data yang sesuai untuk dianalisis yang akan diambil dari *database*. sebagai contoh, sebuah kasus yang meneliti faktor kecenderungan orang membeli dalam kasus market basket analysis, tidak perlu mengambil nama pelanggan, cukup dengan id pelanggan saja.
4. Transformasi data (*DataTransformation*), data diubah atau digabung ke dalam format yang sesuai untuk diproses dalam data mining. Proses transformasi atau coding merupakan proses transformasi data ke dalam format tertentu sehingga nantinya data dapat digunakan dan ditelusuri. Sebagai contoh beberapa metode standar seperti analisis asosiasi dan clustering hanya bisa menerima input data kategorikal. Karenanya data berupa angka numerik yang berlanjut perlu dibagi-bagi menjadi beberapa interval. Proses ini sering disebut transformasi data.
5. Proses *mining*, merupakan suatu proses utama saat metode diterapkan untuk menemukan pengetahuan berharga dan tersembunyi dari data.
6. Evaluasi pola (*patternevaluation*), untuk mengidentifikasi pola-pola menarik ke dalam *knowledge based* yang ditemukan. Dalam tahap ini hasilnya berupa pola-pola yang khas maupun model prediksi dievaluasi untuk menilai apakah hipotesis yang ada memang tercapai. Setelah menemukan pola dan data menarik, selanjutnya adalah menampilkan

data tersebut ke dalam bentuk yang mudah dipahami oleh pengguna atau pihak yang berkepentingan. Jadi pola yang ditemukan nanti akan diperiksa dan dicek apakah bertentangan dengan hipotesis sebelumnya ataukah tidak. Intinya data sudah bisa dibaca dan tentunya akan bermanfaat bagi pihak yang berkepentingan.

7. Presentasi pengetahuan (*knowledgepresentation*), merupakan visualisasi dan penyajian pengetahuan mengenai metode yang digunakan untuk memperoleh pengetahuan yang diperoleh pengguna. Tahap terakhir adalah bagaimana memformulasikan keputusan atau aksi dari hasil analisis yang didapat. Karena presentasi dalam bentuk pengetahuan yang bisa dipahami semua orang adalah satu tahapan yang diperlukan. Dalam presentasi ini, visualisasi juga bisa membantu mengkomunikasikan hasil data mining .

2.3 Metode Naïve Bayes Clasifier

Naïve Bayes merupakan metode klasifikasi yang berdasarkan pada Teorema Bayes. Dinamakan Teorema Bayes karena disesuaikan dengan nama penemunya, yaitu Reverend Thomas Bayes, walaupun sebenarnya ada beberapa penelitian yang mengatakan bahwa Teorema Bayes telah ditemukan oleh orang lain sebelum Reverend Thomas Bayes.

Reverend Thomas Bayes merupakan seorang ilmuwan dari Inggris. Reverend Thomas Bayes mempelajari hal-hal mengenai klasifikasi, namun setelah beliau meninggal, temannya yang mengantikannya untuk mempresentasikan penelitiannya. 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. *Simplenaiive Bayesian classifier* merupakan salah satu metode pengklasifikasi berpeluang sederhana yang berdasarkan pada penerapan *Teorema Bayes* dengan asumsi antar variabel penjelas saling bebas (*independen*). 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 *MicrosoftResearch* memperkenalkan metode statistik *Bayesian* ini pada teknologi anti *spamfilter*. Tetapi yang membuat algoritma *Bayesianfiltering* ini popular adalah pendekatan yang dilakukan oleh Paul Graham. Dasar dari *teoremanaiive* digunakan dalam pemrograman adalah rumus berikut ini : [9]

$$P(A|B) = (P(B|A) * P(A)) / P(B)$$

Artinya Peluang kejadian A sebagai B ditentukan dari peluang B saat A, peluang A, dan peluang B. Contoh penggunaan Algoritma *Naïve Bayes* antara lain :

- Untuk klasifikasi dokumen

- Untuk deteksi SPAM atau filtering SPAM
- Dan masalah klasifikasi lainnya

Teorema Bayes:

$$P(C|X) = P(X|C) \cdot P(C) / P(X)$$

Di mana :

- $P(X)$ bernilai konstan untuk semua *class*
- $P(C)$ merupakan frek relatif *sample class* C

Dicari $P(C|X)$ bernilai maksimum, sama halnya dengan $P(X|C) \cdot P(C)$ juga bernilai maksimum.

Untuk perhitungan *mean* dan standar deviasi hanya dapat berlaku untuk setiap *variable* yang memiliki nilai kontinu atau bisa dikatakan nilainya terus menerus saling berhubungan sesuai dengan atributnya masing-masing yaitu seperti : berat badan, usia, kadar hemoglobin, tensi atas dan tensi bawah. Perhitungan *Mean* (μ) dan Standar Deviasi (S) dapat dilihat sebagai berikut :

$$\bar{X} = \frac{x_1 + x_2 + \dots + x_n}{n}$$

Keterangan :

\bar{X} = Rata-rata hitung

x = Nilai sampel

n = Jumlah seluruh sampel

dan persamaan untuk menghitung nilai simpangan baku (standar deviasi) dapat dilihat sebagai berikut :

$$S = \sqrt{\frac{\sum(x - \bar{X})^2}{n - 1}}$$

Keterangan :

S = Standar deviasi

\bar{X} = Rata-rata hitung

x = Nilai sampel

n = Jumlah seluruh sampel

- a. Untuk nilai probabilitas dalam kategori penentuan calon pendonor darah dan probabilitas untuk setiap kategori itu sendiri, maka *teorema Bayes* dirumuskan :

$$P(E) = \frac{x}{n}$$

Keterangan :

P = Probabilitas

E = Event (Kejadian)

x = Nilai sampel

n = Jumlah seluruh sampel

- b. Perhitungan *Naïve Bayes* dari parameter-parameter calon pendonor darah akan di dapat fungsi densitas probabilitas relatif, maka fungsi densitas probabilitas yaitu :

$$f(x) = \frac{1}{\sqrt{2\pi\sigma}} \exp^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Keterangan :

μ = *mean* atau nilai rata-rata (dari nilai kontinu)

σ = Standar deviasi

x = Nilai dari variabel pada inputan tertentu

\exp = 2.71828

Fungsi dari metode ini sejatinya banyak, hal ini alasanya penulis menerapkan metode ini dengan system tertentu, dalam metode ini berfungsi untuk Mengklasifikasikan dokumen teks seperti teks berita ataupun teks akademis, yang sangat bermanfaat jika bagian akademis ataupun teks berita dalam berbagai informasi yang akan disajikan nantinya, Kaitan antara Naïve Bayes dengan klasifikasi, korelasi hipotesis dan bukti klasifikasi adalah bahwa hipotesis dalam teorema Bayes merupakan salah satu label kelas yang menjadi target pemetaan dalam klasifikasi, sedangkan bukti merupakan fitur-fitur yang menjadikan masukkan dalam model klasifikasi. Jika X adalah vektor masukkan yang berisi fitur dan Y adalah label kelas, Naïve Bayes dituliskan dengan $P(X|Y)$. Selain itu metode ini juga digunakan sebagai metode machine learning yang menggunakan probabilitas, baik itu dalam probabilitas berbagai bidang tentunya. Selanjutnya yaitu metode ini juga berfungsi untuk membuat diagnosis medis secara otomatis, kemudian juga digunakan untuk Mendeteksi atau menyaring spam. Selain dari kegunaan yang dimiliki oleh metode ini tentu ada kelebihan dan kekurangan dalam penerapan metode ini, kelebihannya meliputi Bisa dipakai untuk data kuantitatif maupun kualitatif, Tidak memerlukan jumlah data yang banyak, tidak perlu melakukan data training yang banyak, kemudian Jika ada nilai yang hilang, maka bisa diabaikan dalam perhitungan. Perhitungannya cepat dan efisien, Mudah dipahami, Mudah dibuat, tidak hanya itu dalam penggunaan system ini ataupun pada metode ini sejatinya juga terkait dengan pengklasifikasikan, dimana pengklasifikasikan ini dapat menjadi lebih menarik dimana Pengklasifikasian dokumen bisa dipersonalisasi, disesuaikan dengan kebutuhan setiap orang, dan tentunya Jika digunakan dalam bahasa pemrograman, code-nya sederhana. Sedangkan kekurangan dari Metode ini ialah Apabila probabilitas kondisionalnya bernilai nol, maka probabilitas prediksi juga akan bernilai nol, Asumsi bahwa masing-masing variabel independen membuat berkurangnya akurasi, karena biasanya ada korelasi antara variabel yang satu dengan variabel yang lain, Keakuratannya tidak bisa diukur menggunakan satu probabilitas saja, hingga Butuh bukti-bukti lain untuk membuktikannya seperti halnya pada penelitian ini menggunakan nilai ujian siswa sebagai objek dan subjek dalam penelitian ini yaitu Daftar Kolektif Hasil Ujian Sekolah Berstandar Nasional (DKHUSBN) tahun ajaran 2018/2019 provinsi Kalimantan Tengah Kabupaten Kotawaringin Timur kecamatan Baamang dan Seranau.

III. METODOLOGI PENELITIAN

3.1 Rancangan Penelitian

Dalam pengerjaan penelitian ini diperlukan langkah-langkah penelitian yang dapat mendukung dan memaksimalkan dalam penyelesaian penelitian ini. Pada

metode penelitian terdapat langkah-langkah sebagai berikut [10]:

1. Studi Literatur

Studi literatur adalah mencari referensi teori yang relevan dengan kasus atau permasalahan yang ditemukan. Referensi ini dapat dicari dari buku, jurnal, artikel laporan penelitian, dan situs-situs di internet. Output dari studi literatur ini adalah terkoleksinya referensi yang relevan dengan perumusan masalah. Hal ini seirama dengan pemahaman mengenai makna dari Studi Literatur, Studi literatur dimaksudkan untuk mengungkapkan berbagai teori-teori yang tentunya juga akan relevan dengan permasalahan yang sedang diteliti, hal ini dapat dipahami bahwa Teknik ini dilakukan dengan cara membaca, mempelajari dan mengkaji literatur-literatur yang berhubungan dengan materi pada penelitian ini. Sejatinya Pengkajian literatur untuk kepentingan penelitian ini merupakan berupa literatur teknis dan literatur non-teknis, Literatur teknis, seperti laporan tentang kajian penelitian dan karya tulis profesional atau disipliner dalam bentuk makalah teoretik atau filosofis. Sedangkan literatur non-teknis seperti: buku harian, dokumen, naskah, catatan, sehingga dapat dipahami pada penulisan ini Literatur yang didapat dari sumber ilmiah seperti situs internet, artikel dan dokumen yang berhubungan.

2. Pengumpulan Data

Pada tahap ini, data-data yang dibutuhkan untuk *training* dan *testing*, yaitu daftar kelokatif hasil ujian sekolah berstandar nasional dari dinas pendidikan Kabupaten Kotawaringin Timur. Format data memiliki 3 variabel, yaitu nilai bahasa Indonesia, matematika dan IPA pada penelitian ini peneliti menjadikan salah satu datanya yaitu nilai ujian siswa pada mata pelajaran IPA, Daftar Kolektif Hasil Ujian Sekolah Berstandar Nasional (DKHUSBN) tahun ajaran 2018/2019 Provinsi Kalimantan Tengah Kabupaten Kotawaringin Timur kecamatan Baamang dan Seranau.

3. Implementasi

Detail mengenai implementasi program dilakukan sesuai hasil analisis pada tahap sebelumnya. Uji coba dilakukan dengan proses metode *Naïve Bayes*. Pada tahap implementasi menunjukkan hasil dari aplikasi data mining agar menghasilkan pola kelulusan siswa yang mana menunjukkan pengolahan data induk siswa sesuai dengan atributnya. Pada Metode ini akan menunjukkan pengolahan data kelulusan dengan menunjukkan nilai dari masing-masing siswa. Sedangkan hal selanjutnya atau gambar selanjutnya menunjukkan proses pembelajaran *naïve bayes* mulai dari proses *learning* dan *testing*, pada implementasi penelitian ini dilakukan pada data nilai ujian siswa, Daftar Kolektif Hasil Ujian Sekolah Berstandar Nasional (DKHUSBN) tahun ajaran 2018/2019 provinsi Kalimantan Tengah Kabupaten Kotawaringin Timur kecamatan Baamang dan Seranau.

4. Pengujian

Dilakukan dengan menggunakan hasil nilai ujian siswa, yaitu Daftar Kolektif Hasil Ujian Sekolah Berstandar Nasional (DKHUSBN) tahun ajaran 2018/2019 provinsi

Kalimantan Tengah Kabupaten Kotawaringin Timur kecamatan Baamang dan Seranau.

5. Pembuatan Laporan

Dokumentasi penelitian dilakukan dalam bentuk pembuatan laporan yang meliputi semua tahap dari proses-proses penelitian ini, hingga penelitian ini dapat menjadi sebuah Peneliti memanfaatkan sumber-sumber berupa catatan dan dokumen (non human resources) untuk pengembangan analisis kajian. Terdapat pemahaman lain mengenai hal ini, pembuatan laporan ini menjelaskan bahwa catatan dan dokumen ini dapat dimanfaatkan sebagai saksi dari kejadian-kejadian tertentu atau sebagai bentuk pertanggungjawaban.

3.2 Metode Pengumpulan Data

Pengumpulan data dilakukan untuk memperoleh informasi yang dibutuhkan dalam rangka mencapai tujuan penelitian Untuk membuktikan hipotesis secara empiris, seorang peneliti membutuhkan pengumpulan data untuk diteliti secara lebih mendalam, tidak hanya itu sejatinya Metode pengumpulan data adalah teknik atau cara yang dilakukan oleh peneliti untuk mengumpulkan data. Pengumpulan data dilakukan untuk memperoleh informasi yang dibutuhkan dalam rangka mencapai tujuan penelitian, Proses pengumpulan data ditentukan oleh variabel-variabel yang ada dalam hipotesis. Pengumpulan data dilakukan terhadap sampel yang telah ditentukan sebelumnya. Data adalah sesuatu yang belum memiliki arti bagi penerimanya dan masih membutuhkan adanya suatu pengolahan. Data bisa memiliki berbagai wujud, mulai dari gambar, suara, huruf, angka, bahasa, simbol, bahkan keadaan. Semua hal tersebut dapat disebut sebagai data asalkan dapat kita gunakan sebagai bahan untuk melihat lingkungan, obyek, kejadian, ataupun suatu konsep [11].

3.3 Desain Penelitian

Dalam penelitian ini penulis menggunakan model standarisasi data mining yaitu CRISP-DM (Cross Industry Standart Process for Data Mining), dengan langkah-langkah sebagai berikut:

1. Pemahaman Data (Data Understanding)

Data yang digunakan dalam penelitian ini adalah sumber data primer. Data primer adalah sumber data penelitian yang diperoleh secara langsung dari sumber aslinya yang berupa wawancara, jajak pendapat dari individu atau kelompok (orang) maupun hasil observasi dari suatu obyek, kejadian atau hasil pengujian (benda), Dengan kata lain, peneliti membutuhkan pengumpulan data dengan cara menjawab pertanyaan riset (metode survei) atau penelitian benda (metode observasi), Kelebihan dari data primer merupakan sebuah bentuk data lebih mencerminkan kebenaran berdasarkan dengan apa yang dilihat dan didengar langsung oleh peneliti sehingga unsur-unsur kebohongan dari sumber yang fenomenal dapat dihindari, selain itu data primer atau data utama atau data pokok yang digunakan dalam penelitian. Data pokok

dapat dideskripsikan sebagai jenis data yang diperoleh langsung dari tangan pertama subjek penelitian atau responden atau informan. Perkecualian pada riset kuantitatif. Kekurangan dari data primer adalah membutuhkan waktu yang relatif lama serta biaya yang dikeluarkan relatif cukup besar. Data diperoleh dari lembaga pendidikan. Data yang dikumpulkan yaitu data nilai tryout siswa, terhadapa nilai ujian siswa, dimana pada penelitian ini penulis mengambil Daftar Kolektif Hasil Ujian Sekolah Berstandar Nasional (DKHUSBN) tahun ajaran 2018/2019 provinsi Kalimantan Tengah Kabupaten Kotawaringin Timur kecamatan Baamang dan Seranau sebagai bahan untuk penelitian ini.

2. Pengolahan Data (Data Preparation)

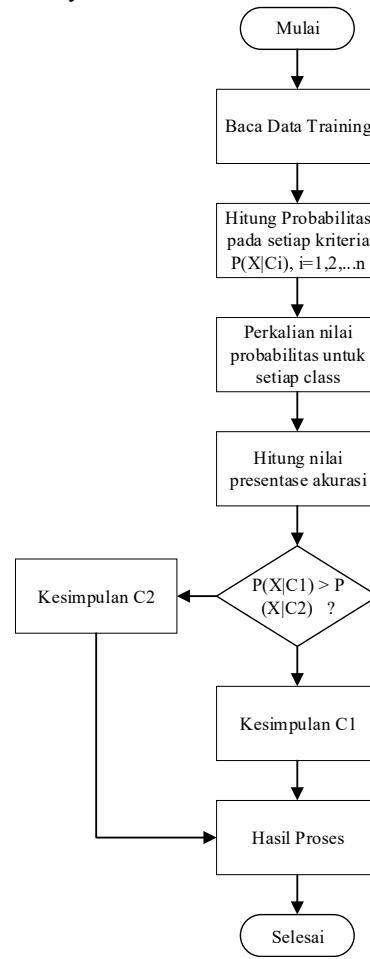
Persiapan data (atau yang lebih dikenal dengan *prapemrosesan data*) merupakan suatu mengambil data yang diidentifikasi pada tahap sebelumnya ataupun sejatinya menyiapkan nya untuk analisa dengan menggunakan metode-metode DM, hal ini diharapkan dapat Dibandingkan dengan tahapan-tahapan lainnya dalam CRISP-DM, *prapemrosesan data* menyita waktu dan usaha paling banyak; banyak orang percaya bahwa tahap ini bertanggungjawab atas sekitar 80 persen dari total waktu yang diluangkan untuk project DM. Penyebab dari usaha yang sedemikian besar itu yang dihabiskan untuk tahap ini adalah karena data riil (di 'real-world') yang ada pada umumnya tidak lengkap (tidakadanya nilai pada atribut-atributnya, tidakadanya atribut tertentu yang menjadi perhatian, atau hanya berisi data yang sudah ringkas dan digabungkan), 'noisy' (berisi data yang error atau data yang tidak diinginkan), dan data yang tidak konsisten (berisi data yang berbeda antara kode-kode dan nama-nama).

- Tahap Pertama, penentuan data yang akan diolah. Dari data yang telah diperoleh, tidak semua data akan diolah karena penelitian yang akan dilakukan memiliki batasan-batasan data yang akan digunakan.
- Tahap Kedua, penanganan data missing value. Missing value adalah data yang tidak lengkap dikarenakan atribut tidak tercatat maupun atribut memang tidak dimiliki dsb. Penanganan missing value dilakukan dengan penghapusan record yang kosong, diharapkan data Daftar Kolektif Hasil Ujian Sekolah Berstandar Nasional (DKHUSBN) tahun ajaran 2018/2019 provinsi Kalimantan Tengah Kabupaten Kotawaringin Timur kecamatan Baamang dan Seranau.
- Tahap Ketiga, menentukan atribut yang akan digunakan dari tahap pertama. Atribut yang akan digunakan adalah nama, nilai tryout yang meliputi : nilai bahasa indonesia, nilai bahasa inggris, nilai matematika, nilai IPA, total nilai, rata-rata, pengujian ini dilakukan Daftar Kolektif Hasil Ujian Sekolah Berstandar Nasional (DKHUSBN) tahun ajaran 2018/2019 provinsi Kalimantan Tengah Kabupaten Kotawaringin Timur kecamatan Baamang dan Seranau.

3. Permodelan (*Modelling*)

Pada tahap ini data yang telah diklasifikasikan akan diproses ke tahap selanjutnya denan beberapa permodelan.

Permodelan ini dilakukan untuk mengoptimalkan hasil yang ingin dicapai. Seletah melakukan semua tahap pemrosesan data, maka akan dihasilkan data training. Data training merupakan data yang digunakan untuk melakukan proses perhitungan menggunakan metode Naïve Bayes. Berikut gambaran proses perhitungan menggunakan metode Naïve Bayes dalam sebuah flowchart.



Gambar 1 Flowchart Naïve Bayes

IV. HASIL DAN PEMBAHASAN

4.1 Hasil Pengujian Model

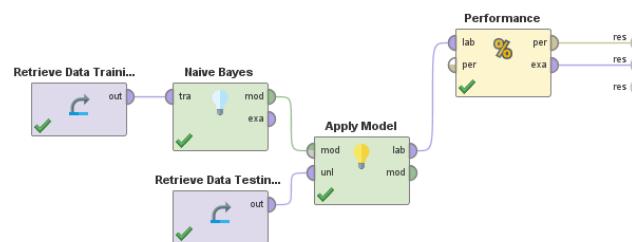
Penelitian ini menggunakan metode *Naïve Bayesian Classification* (NBC). NBC merupakan algoritma klasifikasi yang sangat efektif (mendapatkan hasil yang tepat) dan efisien (proses penalaran dilakukan memanfaatkan input yang ada dengan cara yang relative cepat). Algoritma NBC bertujuan untuk melakukan klasifikasi data pada kelas tertentu. Model statistic merupakan salah satu model yang terpercaya sangat andal sebagai pendukung pengambilan keputusan. Konsep probabilitas merupakan salah satu bentuk model statistik. Salah satu metode yang menggunakan konsep probabilitas adalah *Naïve Bayesian Classification* (NBC). Pada metode ini, semua atribut akan memberikan kontribusinya dalam pengambilan keputusan, dengan bobot atribut yang

sama penting dan setiap atribut saling bebas satu sama lain tidak hanya itu juga menggunakan variable variable tertentu ataupun variable variable data pada proses pengujian penelitian ini.

Naïve Bayes dalam studi kasus yang kami gunakan adalah untuk membuat sebuah prediksi masa depan tentang atribut yang dimiliki siswa yaitu nilai Bahasa Indonesia, Matematika, IPA dan Jumlah Nilai, mengambil Daftar Kolektif Hasil Ujian Sekolah Berstandar Nasional (DKHUSBN) tahun ajaran 2018/2019 provinsi Kalimantan Tengah Kabupaten Kotawaringin Timur kecamatan Baamang dan Seranau sebagai bahan untuk penelitian ini.

Penggunaan Rapid Miner untuk mendapatkan kalkulasi klasifikasi *Naïve Bayes*. Import dataset ke rapid miner dengan Nilai Bahasa Indonesia sebagai *integer*, Matematika sebagai *integer*, IPA sebagai *integer*, nilai Bahasa Indonesia, Matematika, IPA dan Jumlah Nilai, mengambil Daftar Kolektif Hasil Ujian Sekolah Berstandar Nasional (DKHUSBN) tahun ajaran 2018/2019 provinsi Kalimantan Tengah Kabupaten Kotawaringin Timur kecamatan Baamang dan Seranau sebagai bahan untuk penelitian ini.

Jumlah nilai sebagai *real* dan hasil sebagai *binominal*. Dibuat sebuah design pada rapid miner dimana tabel acuan (base) digunakan sebagai acuan untuk prediksi dataset testing dengan *n* *Naïve Bayes* dan performance monitoring.



Gambar 2 Implementasi *Naïve Bayes* pada *RapidMiner*

4.2 Evaluasi Model Confusion Matrix

Tabel 1 adalah perhitungan berdasarkan data *training*, diketahui dari 100 data, 58 diklasifikasikan LULUS sesuai dengan prediksi yang dilakukan dengan metode algoritma *Naïve Bayes*, 2 data diprediksi LULUS tetapi ternyata hasilnya TIDAK LULUS, 16 data diprediksi TIDAK LULUS tetapi ternyata hasilnya LULUS dan 24 *class* TIDAK LULUS diprediksi sesuai.

TABEL I

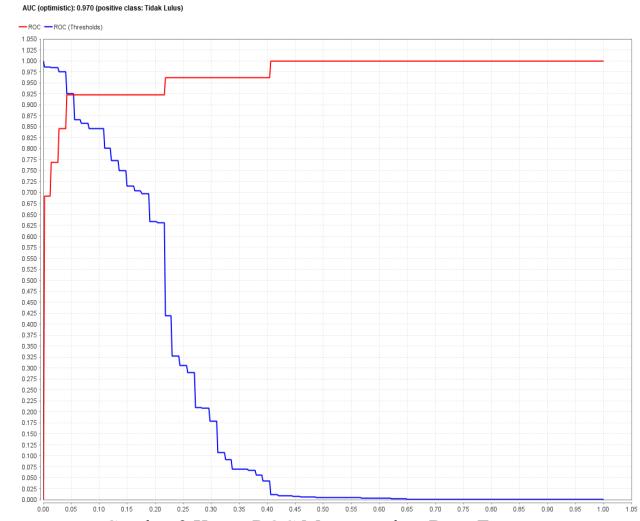
MODEL CONFUSION MATRIX UNTUK METODE ALGORITMA *NAÏVE BAYES*.

Accuracy : 82,00%			
	true Lulus	true Tidak Lulus	class precision
pred. Lulus	58	2	96.67%
pred. Tidak Lulus	16	24	60.00%

class recall	78.38%	92.31%	
--------------	--------	--------	--

4.3 Kurva ROC (Receiver Operating Characteristic)

Hasil perhitungan kurva ROC untuk algoritma *Naïve Bayes* dengan menggunakan data *training* hasilnya sebesar 0,970 yang mengartikan bahwa pada perhitungan ini NBC dalam system ini ataupun penggunaan untuk penentuan masalah ini sangat baik dan penggunaanya sangat tepat dalam menyelesaikan permasalahan ini tentunya dan dapat dilihat dari kurvanya seperti berikut ini.



Gambar 3 Kurva ROC Menggunakan Data *Training*

Gambar 4.1 menunjukkan bahwa hasil algoritma *Naïve Bayes* mendapatkan nilai **0,970**, di mana nilai tersebut menunjukkan nilai AUC klasifikasi sangat baik, dikarenakan memiliki nilai di antara 0,90 – 1,00.

V. KESIMPULAN

Dalam penelitian ini dilakukan pengujian model algoritma *Naïve Bayes*, hasil evaluasi dan validasi, diketahui bahwa *Naïve Bayes* memiliki nilai akurasi cukup tinggi yaitu sebesar 82,00%. Selain itu, berdasarkan data Daftar Kolektif Hasil Ujian Sekolah Berstandar Nasional (DKHUSBN) tahun ajaran 2018/2019 provinsi Kalimantan Tengah Kabupaten Kotawaringin Timur kecamatan Baamang dan Seranau tepat dan berpengaruh pada penelitian ini. Dengan demikian, metode *Naïve Bayes* merupakan metode yang cukup baik dalam menentukan kelulusan siswa secara lebih efektif dan efisien.

DAFTAR PUSTAKA

- [1] Erdogan, S.Z. dan M. Timor. 2005. A Data Mining Application In A StudentDatabase. Journal Of Aeronautics And Space Technologies. Volume 2 Number 2: 53-57.
- [2] Garcia, E.P.I. dan P.M. Mora. 2011. Model Prediction of Academic Performance forFirst Year Students. IEEE Computer Society.
- [3] Hien, N.T.N. dan P. Haddawy. 2007. A Decision Support System for EvaluatingInternational Student Applications. IEEE Frontiers in Education Conference.

- [4] Santosa, B. 2007. Data Mining: Teknik Pemanfaatan Data untuk Keperluan Bisnis. Grahalmu. Yogyakarta.
- [5] Kusrini dan Emha Taufiq Luthfi. 2009 Algoritma Data Mining, Andi Offset.
- [6] Larose, Daniel T. 2005 Discovering Knowledge in Data Mining An Introduction to Data Mining, Wiley Interscience
- [7] Umaedi. 2001. Manajemen peningkatan mutu berbasis sekolah. Jakarta: Departemen Pendidikan Nasional Direktorat Jendral Pendidikan Dasar dan Menengah Direktorat Sekolah Lanjutan Tingkat Pertama
- [8] Ridwan, M., Suyono, H., dan Sarosa, M. 2013. Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier.
- [9] Kusumadewi, Sri. 2009. Klasifikasi Status Gizi menggunakan Naive Bayesian Classification.
- [10] Martono, Nanang. 2011. Metode Penelitian Kuantitatif. Jakarta: PT Raya Grafindo Persada.
- [11] Mahmud. 2011. Metode Penelitian Pendidikan. Bandung: CV Pustaka Setia.