# KLASIFIKASI PENJURUSAN PROGRAM STUDI SEKOLAH MENENGAH ATAS DENGAN ALGORITMA NAÏVE BAYES CLASSIFIER PADA SMA N 1 SUBAH

Nikmatul Hidayah[1]

[1,3]*Jurusan Teknik Informatika-S1, Fakultas Ilmu Komputer,*
*Universitas Dian Nuswantoro Semarang*
*Jln. Nakula I no 5-17 Semarang 50131 INDONESIA*
[1]111201005243@mhs.dinus.ac.id

**Penjurusan siswa kelas X SMA yang akan naik ke kelas XI bertujuan mengarahkan peserta didik agar lebih fokus mengembangkan kemampuan dan minat yang dimiliki. Strategi ini diharapkan dapat memaksimalkan potensi, bakat atau talenta individu, sehingga juga akan memaksimalkan nilai akademisnya. Penentuan jurusan akan berdampak terhadap kegiatan akademik selanjutnya dan mempengaruhi pemilihan bidang ilmu atau studi bagi siswa-siswi yang ingin melanjutkan ke perguruan tinggi nantinya. Jurusan yang tidak tepat bisa sangat merugikan siswa dan masa depannya.**

**Atas dasar permasalahan tersebut ,maka dilakukan penelitian untuk menerapkan metode data mining yaitu algoritma Naïve Bayes Classifier untuk mengklasifikasikan jurusan program studi. Naïve Bayes adalah suatu metode pengklasifikasian data dengan model statistic yang dapat digunakan untuk memprediksi probabilitas keanggotaan pada suatu kelas dan digunakan ntuk menganalisis dalam membantu tercapainya hasil keputusan terbaik suatu permasalahan dari sejumlah alternatif.**

**Hasil akurasi klasifikasi jurusan siswa SMA N 1 Subah menggunakan naïve bayes memiliki akurasi sebesar 98,00% dan nilai AUC 0,999% . Akurasi yang dihasilkan oleh algoritma naïve bayes merupakan akurasi yang excellent dan dapat diterapkan untuk meningkatkan akurasi klasifikasi penjurusan siswa SMA N 1 Subah.**
*Kata kunci—data mining, naïve bayes classifier , klasifikasi penjurusan siswa SMA N 1 Subah , AUC .*

**Placement of students of class X who will rise to a high school class XI students aim to be more focus on developing skills and interests owned. This strategy is expected to maximize the potential, talent or talents of the individual, so also will maximize academic value. Determination of the majors will have an impact on the activities and influence the selection of the next academic science or study for students who want to go to college later. Programs that are not appropriate can be very detrimental to the students and their future.**

**On the basis of these problems, then do the research to apply data mining methods, namely Naïve Bayes classifier algorithm to classify major courses. Naïve Bayes is a method of classifying the data with statistical models that can be used to predict the probability of membership in a class and used ntuk analyzing the decision help achieve the best result of a problem from a number of alternatives.**
***Keywords****—data mining, naïve Bayes classifier, the classification of SMA N 1 student majors Subah*

## I. PENDAHULUAN

Penjurusan siswa kelas X SMA yang akan naik ke kelas XI bertujuan untuk mengarahkan peserta didik agar dapat lebih fokus mengembangkan kemampuan dan minat yang dimiliki. Strategi ini diharapkan dapat memaksimalkan potensi, bakat atau talenta individu, sehingga juga akan memaksimalkan nilai akademisnya. Penentuan jurusan akan berdampak terhadap kegiatan akademik selanjutnya dan mempengaruhi pemilihan bidang ilmu atau studi bagi siswa-siswi yang ingin melanjutkan ke perguruan tinggi nantinya. Jurusan yang tidak tepat bisa sangat merugikan siswa dan masa depannya. Pengambilan keputusan penjurusan oleh sekolah dipertimbangkan dengan melihat beberapa faktor, antara lain nilai akademis, hasil test IQ, minat siswa, dan lain sebagainya.

Pihak sekolah yang dalam hal ini adalah guru BK dituntut sebijaksana mungkin dalam memutuskan jurusan yang tepat. Menentukan jurusan dengan memperhatikan banyak faktor yang kompleks dan dilakukan secara manual mempunyai banyak kelemahan. Data yang banyak cukup menyita waktu dan menguras tenaga, serta menuntut ketelitian ekstra. Selain itu, cara seperti ini memungkinkan terjadinya kesalahan baik yang manusiawi maupun yang disengaja. Penilaian subjektif dengan memberikan keistimewaan tersendiri kepada pihak tertentu sering kali menimbulkan ketidakadilan.

Perkembangan teknologi yang pesat diiringi dengan kebutuhan akan informasi yang cepat guna meningkatkan efektifitas pelayanan dan keakuratan memungkinkan dibangunnya sebuah soft computing untuk membantu mengklasifikasikan penjurusan dengan menerapkan metode

Bayesian Classification. Pengolahan berbagai data, berbagai informasi dan berbagai metode dengan kemampuan teknologi yang canggih akan sangat membantu meminimalisasi kesalahan, sehingga apat memutuskan jurusan secara cepat, tepat dan adil.

## II. STUDI PUSTAKA

### *2.1. Penelitian Terkait*

Ada beberapa referensi yang diambil penulis sebagai bahan pertimbangan untuk penelitian yang dilakukan, referensi tersebut diambil dari beberapa penulisan yang dilakukan

sebelumnya yang membahas permasalahan yang hampir sama, antara lain :

Yetli Oslan dalam Jurnal EKSIS penelitian vol 6 implementasi metode bayesian , 2013 telah melakukan eksperimen untuk data penjurusan SMA dan menghasilkan tingkat Keakuratan tertinggi dari hasil proses penjurusan dengan menggunakan range yang ditentukan secara manual berada pada range dengan interval 20 dan 25. Pada interval 20, angkatan 2009 mendapatkan sebesar 59 %, 2010 sebesar 66 %, 2011 sebesar 61 %. Sedangkan pada Interval 25, angkatan 2009 mendapatkan sebesar 62 %, 2010 sebesar 58 %, 2011 sebesar 54 %

Hasil proses penjurusan dengan menggunakan range yang dapatkan dari proses Box Plot ratarata memiliki tingkat keakuratan lebih tinggi dibanding dengan range yang ditentukan secara manual. Pada range Box Plot, tingkat keakuratan angkatan 2009 adalah 62%, tingkat keakuratan angkatan 2010 adalah 63% dan angkatan 2011 adalah 66% . Berdasarkan proses yang telah dilakukan, hasil penjurusan dengan cara tersebut rata-rata memiliki keakuratan lebih tinggi karena seluruh siswa mendapatkan saran jurusan tanpa terkecuali.

*2.2. Tinjauan Pustaka*

*A. Pengertian Data Mining*

Secara sederhana data mining adalah penambangan atau penemuan informasi baru dengan mencari pola atau aturan tertentu dari sejumlah data yang sangat besar[10]. Data mining juga disebut sebagai serangkaian proses untuk menggali nilai tambah berupa pengetahuan yang selama ini tidak diketahui secara manual dari suatu kumpulan data[11]. Data mining, sering juga disebut sebagai knowledge discovery in database (KDD). KDD adalah kegiatan yang meliputi pengumpulan, pemakaian data, historis untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran 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[8]. 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[12].

Karakteristik data mining sebagai berikut :

• Data mining berhubungan dengan penemuan sesuatu yang tersembunyi dan pola data tertentu yang tidak diketahui sebelumnya.

• Data mining biasa menggunakan data yang sangat besar. Biasanya data yang besar digunakan untuk membuat hasil lebih dipercaya.

• Data mining berguna untuk membuat keputusan yang kritis, terutama dalam strategi[23].

Berdasarkan beberapa pengertian tersebut dapat ditarik kesimpulan bahwa data mining adalah suatu teknik menggali informasi berharga yang terpendam atau tersembunyi pada suatu koleksi data (database) yang sangat besar sehingga ditemukan suatu pola yang menarik yang sebelumnya tidak diketahui. Kata mining sendiri berarti usaha untuk mendapatkan sedikit barang berharga dari sejumlah besar material dasar. Karena itu data mining sebenarnya memiliki akar yang panjang dari bidang ilmu seperti kecerdasan buatan (artificial intelligent), machine learning, statistik dan database. Beberapa metode yang sering disebut-sebut dalam literatur data mining antara lain clustering, classification, association rules mining, neural network, genetic algorithm dan lain-lain[24].

*B. Pengenalan Pola, Data Mining, dan Machine Learning*

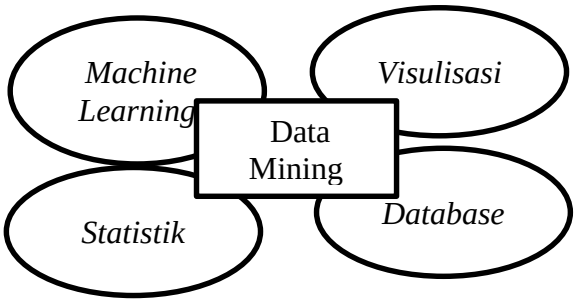
Ada beberapa metode yang dapat digunakan untuk menyelesaikan masalah MADM [4], salah satunya yaitu Simple Additive Weighting (SAW). Metode SAW sering juga dikenal sebagai metode penjumlahan terbobot. Konsep dasar metode SAW adalah mencari penjumlahan terbobot dari rating kinerja pada setiap alternatif pada semua atribut. Metode SAW membutuhkan proses normalisasi matriks keputusan (X) ke suatu skala yang dapat dibandingkan dengan semua rating alternatif yang ada.

Pengenalan pola adalah suatu disiplin ilmu yang mempelajari cara-cara mengklasifikasikan obyek ke beberapa kelas atau kategori dan mengenali kecenderungan data. Tergantung pada aplikasinya, obyek-obyek ini bisa berupa pasien, mahasiswa, pemohon kredit, image atau signal atau pengukuran lain yang perlu diklasifikasikan atau dicari fungsi regresinya[10].

Data mining, sering juga disebut knowledge discovery in database (KDD), adalah kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar. Keluaran dari data mining ini bisa dipakai untuk memperbaiki pengambilan keputusan di masa depan. Sehingga istilah pattern recognition jarang digunakan karena termasuk bagian dari data mining[10].

Machine Learning adalah suatu area dalam artificial intelligence atau kecerdasan buatan yang berhubungan dengan pengembangan teknik-teknik yang bisa diprogramkan dan belajar dari data masa lalu. Pengenalan pola, data mining dan machine learning sering dipakai untuk menyebut sesuatu yang sama. Bidang ini bersinggungan dengan ilmu probabilitas dan statistik kadang juga optimasi.

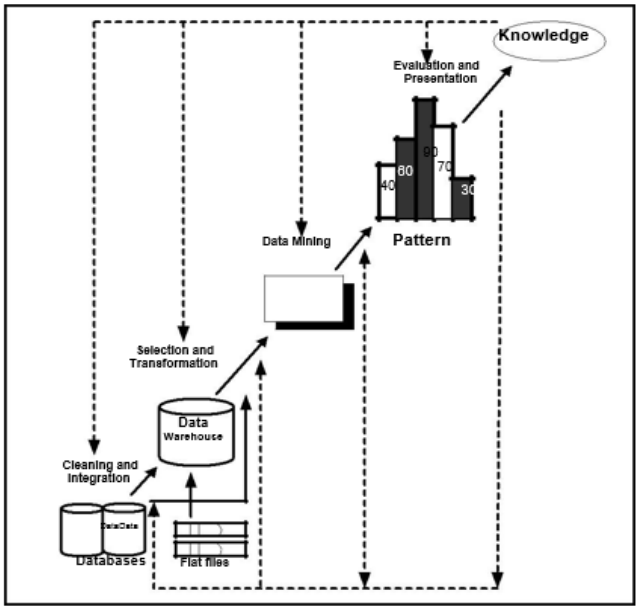
Machine learning menjadi alat analisis dalam data mining. Bagaimana bidang-bidang ini berhubungan bisa dilihat dalam gambar 2.1 [10].




Gambar 2.1. Data mining merupakan irisan dari berbagai disiplin

*C. Tahap tahap Data Mining*

Sebagai suatu rangkaian proses, data mining dapat dibagi menjadi beberapa tahap yang diilustrasikan di Gambar 2.5. Tahap-tahap tersebut bersifat interaktif, pemakai terlibat langsung atau dengan perantaraan knowledge base.




Gambar 2.2. Tahap-Tahap Data Mining[8]

Tahap-tahap data mining ada 6 yaitu :

1.Pembersihan data (data cleaning)

Pembersihan data merupakan proses menghilangkan noise dan data yang tidak konsisten atau data tidak relevan. Pada umumnya data yang diperoleh, baik dari database suatu perusahaan maupun hasil eksperimen, memiliki isian-isian yang tidak sempurna seperti data yang hilang, data yang tidak valid atau juga hanya sekedar salah ketik. Selain itu, ada juga atribut-atribut data yang tidak relevan dengan hipotesa data mining yang dimiliki. Data-data yang tidak relevan itu juga lebih baik dibuang. Pembersihan data juga akan mempengaruhi performasi dari teknik data mining karena data yang ditangani akan berkurang jumlah dan kompleksitasnya.

2. Integrasi data (data integration)

Integrasi data merupakan penggabungan data dari berbagai database ke dalam satu database baru. Tidak jarang data yang diperlukan untuk data mining tidak hanya berasal dari satu database tetapi juga berasal dari beberapa database atau file teks. Integrasi data dilakukan pada atribut-aribut yang mengidentifikasikan entitas-entitas yang unik seperti atribut nama, jenis produk, nomor pelanggan dan lainnya. Integrasi data perlu dilakukan secara cermat karena kesalahan pada integrasi data bisa menghasilkan hasil yang menyimpang dan bahkan menyesatkan 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.

3.Seleksi Data (Data Selection)

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

Data diubah atau digabung ke dalam format yang sesuai untuk diproses dalam data mining. Beberapa metode data mining membutuhkan format data yang khusus sebelum bisa diaplikasikan. 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 (pattern evaluation),

Untuk mengidentifikasi pola-pola menarik kedalam knowledge based yang ditemukan. Dalam tahap ini hasil dari teknik data mining berupa pola-pola yang khas maupun model prediksi dievaluasi untuk menilai apakah hipotesa yang ada memang tercapai. Bila ternyata hasil yang diperoleh tidak sesuai hipotesa ada beberapa alternatif yang dapat diambil seperti menjadikannya umpan balik untuk memperbaiki proses data mining, mencoba metode data mining lain yang lebih sesuai, atau menerima hasil ini sebagai suatu hasil yang di luar dugaan yang mungkin bermanfaat.

7. Presentasi pengetahuan (knowledge presentation),

Merupakan visualisasi dan penyajian pengetahuan mengenai metode yang digunakan untuk memperoleh pengetahuan yang diperoleh pengguna. Tahap terakhir dari proses data mining adalah bagaimana memformulasikan keputusan atau aksi dari hasil analisis yang didapat. Ada kalanya hal ini harus melibatkan orang-orang yang tidak memahami data mining. Karenanya presentasi hasil data mining dalam bentuk pengetahuan yang bisa dipahami semua orang adalah satu tahapan yang diperlukan dalam proses data mining. Dalam presentasi ini, visualisasi juga bisa membantu mengkomunikasikan hasil data mining[8].

*D. CRISP-DM*

CRISP-DM (Cross-Industry Standard Process for Data Mining) telah banyak digunakan dalam industri oleh para ahli saat ini sebagai salah satu proses data mining untuk memecahkan suatu masalah". Metodologi ini terdiri dari enam tahap proses siklus. Metodologi ini membuat data mining yang besar dapat dilakukan dengan lebih cepat, lebih ekonomis, dan mudah untuk diatur. Bahkan, data mining yang berukuran kecil pun dapat memperoleh keuntungan dari CRISP-DM (Olson & Delen, 2008:9). Berikut adalah enam tahap yang disebut sebagai siklus[23]:

1. Business understanding

Business understanding meliputi penentuan tujuan bisnis, menilai situasi saat ini, menetapkan tujuan data mining, dan mengembangkan rencana proyek.

2. Data understanding

Setelah tujuan bisnis dan rencana proyek ditetapkan, Data understanding mempertimbangkan persyaratan data. Langkah ini dapat mencakup pengumpulan data awal, deskripsi data, eksplorasi data, dan verifikasi data yang berkualitas.

3. Data preparation

Setelah sumber data telah tersedia untuk diidentifikasi. Data tersebut perlu untuk dipilih, dibersihkan, dibangun ke dalam model yang diinginkan, dan diformat. Pembersihan data dan transformasi data dalam penyusunan pemodelan data perlu terjadi di tahap ini.

4. Modeling

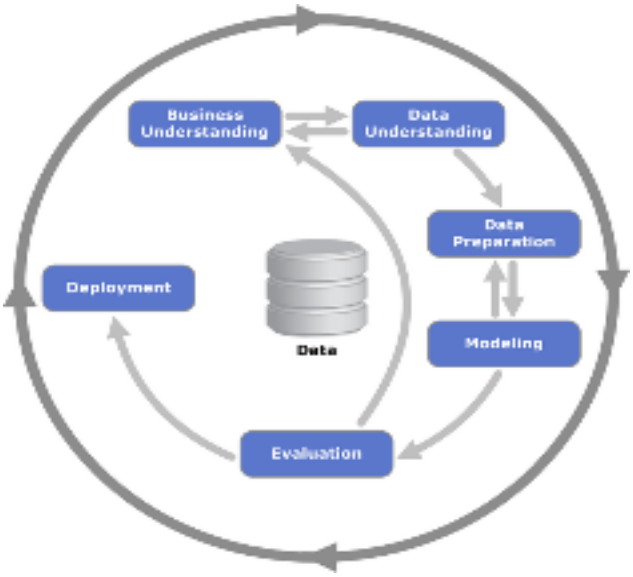
Tujuan dari pemodelan data mining adalah untuk mencari hasil dari berbagai situasi yang ada. Alat perangkat lunak untuk data mining seperti visualisasi (mensplit data dan membangun hubungan) dan analisis kluster (untuk mengidentifikasikan variable berjalan dengan baik secara bersamaan) dapat berguna untuk analisis awal model yang akan digunakan. Pembagian data ke dalam set pelatihan dan pengujian juga diperlukan untuk pemodelan.

5. Evaluation

Hasil model harus dievaluasi sesuai tujuan bisnis pada tahap pertama (pemahaman bisnis). Evaluasi dilakukan dari hasil visualisasi dan perhitungan statistik pengujian berdasarkan pemodelan yang dibuat. Pada akhir dari tahap ini, keputusan penggunaan hasil data mining telah ditentukan.

6. Deployment

Pembuatan dari model bukanlah akhir dari projek data mining. Meskipun tujuan dari pemodelan adalah untuk meningkatkan pengetahuan dari data, pengetahuan data tersebut perlu dibangun dengan terorganisasi dan dibuat pada satu bentuk yang dapat digunakan oleh pengguna.




Gambar 2.3. Proses CRISP-DM[23]

*E. Metode Naive Bayes Classifier*

Bayesian Classification didasarkan pada Teorema Bayes yang memiliki kemampuan hampir serupa dengan Decision Tree dan Neural Network. Teorema Bayes adalah teorema yang digunakan dalam statistika untuk menghitung peluang suatu hipotesis. Teorema Bayes ini dibuat oleh Thomas Bayes yang ditulis pada paper yang berjudul "An Essay toward Solving a Problem in The Doctrine of Chance". Inti dari Teorema Bayes tersebut adalah memprediksi probabilitas dimasa yang akan datang berdasarkan pengalaman dimasa sebelumnya[25].

Bayesian Classification adalah suatu metode pengklasifikasian data dengan model statistic yang dapat digunakan untuk memprediksi probabilitas keanggotaan pada suatu kelas[10]. Metode Bayesian Classification digunakan menganalisis dalam membantu tercapainya pengambilan keputusan terbaik suatu permasalahan dari sejumlah alternatif. Bayesian Classification merupakan salah satu metode yang sederhana yang dapat digunakan untuk data yang tidak konsisten dan data bias. Metode Bayes juga merupakan metode yang baik dalam mesin pembelajaran berdasarkan data training dengan berdasarkan pada probabilitas bersyarat[10].

Formulasi Naive Bayes Classifier adalah sebagai berikut :

$$P(H|X) = (P(X|H)P(H))/(P(X))$$

Dimana untuk formula diatas :

X : Sampel data dengan kelas yang belum diketahui

H : Hipotesis data X merupakan suatu kelas spesifik

P(H|X) :Probabilitas hipotesis H berdasar kondisi X (posterior probability probability)

P(H) : Probabilitas hipotesis H (prior probability)

P(X|H) : Probabilitas X berdasar kondisi pada hipotesis H

P(X) : Probabilitas dari X

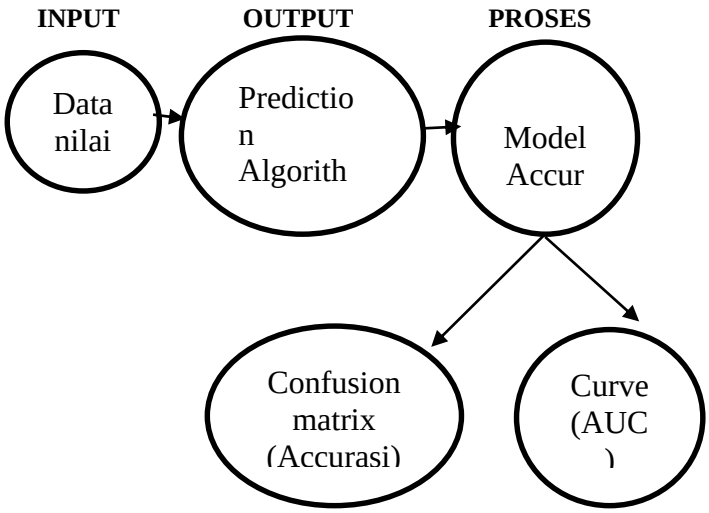
Dalam terminologi sederhana,sebuah NBC mengasumsikan bahwa kehadiran atau ketiadaan fitur tertentu dri suatu kelas tidak berhubungan dengan kehadiran atau ketiadaan fitur lainya.Sebagai contoh,buah mungkin dianggap apel jika

merah,bulat,dan berdiameter sekitar 4 inchi.Bahkan jika fitur ini bergantung pada satu sama lain atau atas keberadaan fitur lain.Sebuah NBC menganggap bahwa seluruh sifat-sifat berkontribusi mandiri untuk probabilitas bahwa buah ini adalah apel.Tergantung pada situasi yag tepat dari model probabilitas,NBC dapat dilatih sangat efisien dalam supervised learning.

Naive Bayes Classifier (NBC) membutuhkan jumlah record data yang besar untuk mendapatkan hasil yang baik. Jika kategori prediktor tidak ada dalam data training, maka naive bayes classifier mengasumsikan bahwa record baru dengan kategori predictor memiliki probabilitas nol.

*F. Kerangka Pemikiran*

Komponen dari model kerangka pemikiran penelitian ini adalah Input , Proses,dan Output.Kerangka pemikiran dimulai dari klasifikasi hasil penjurusan.Maka penulis membuat soft computing dengan menggunakan naïve bayes classifier




Gambar 2.4 Kerangka Pemikiran

Dari gambar 2.4 Terlihat bahwa data yang digunakan adalah data nilai siswa SMA N 1 Subah sebagai inputan awal.Sedangkan pada metode yang digunakan dalam pemrosesan data adalah naïve bayes classifier . Tujuan objektiv pada penelitian ini adalah peningkatan akurasi model.Pengukuran akurasi tersebut menggunakan confusion matrix dan ROC Curve.

.

## III. METODOLOGI PENELITIAN

A. Instrumen Penelitian

Instrumen penelitian dalam penelitian ini menggunakan pegukuran akurasi dengan rapidminer 5 dan menguji proposed method dengan program soft computing. Implementasi dari Naïve Bayes Classifier sendiri untuk klasifikasi penjurusan siswa SMA N 1 Subah menggunakan soft computing yang akan dibangun dengan matlab.

B. Pengumpulan Data

Jenis data yang digunakan dalam penelitian ini adalah jenis data kuantitatif. Data kuantitatif adalah data yang berupa angka. Dalam penelitian ini, terdapat sampel data yang berupa nilai mata pelajaran siswa kelas X semester dua SMA Negeri 1 Subah.Kemudian dari sampel data nilai siswa akan diproses dan digunakan sebagai data uji yang akan melalui pengolahan yang nantinya akan menghasilkan suatu output penjurusan yaitu Ilmu Pengetahuan Alam(IPA) atau Ilmu Pengetahuan Sosial(IPS).

C. Eksperimen

Penelitian ini merupakan penelitian eksperimnen. Untuk melakukan eksperimen dibutuhkan alat bantu berupa spesifikasi software dan hardware.

Berikut spesifikasi software dan hardware yang digunakan

Software meliputi :

- SistemOperasi Windows 8
- Pengolahan data dan angka : Microsoft Office dan Microsof Excel 2013
- Data Mining :Rapidminer
- Implementasi : Matlab

Hardware meliputi :

- CPU Intel Core i3
- RAM Memory 4 GB
- HD 500 GB

Dalam pengujian model eksperiman dengan jumlah 11 atribut prediktor yaitu Kimia , Fisika , Biologi , Ekonomi , Sosiologi , Ekonomi , Geografi , Sosiologi ,Jumlah IPA,Jumlah IPS , IQ dan Minat . Semua atribut tersebut akan dihitung peluangnya oleh naïve bayes sehingga menghasilkan prediksi jurusan program studi.

D. Pengujian dan implementasi

Dari hasil pemodelan melalui algoritma naïve bayes penulis akan mengimplementasikan prediksi penjurusan ke sebuh program matlab. Berdasarkan proses training pada data nilai mata pelajaran siswa SMAN 1 Subah 2012/2013 naïve bayes akan melakukan perhitungan hingga menghasilkan output penjurusan IPA atau IPS

## IV. HASIL PENELITIAN DAN PEMBAHASAN

*A. Hasil Penelitian*

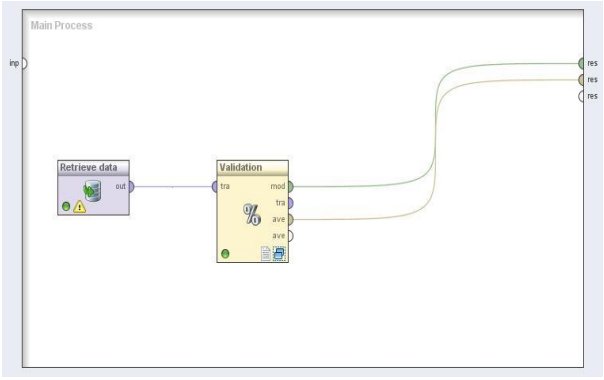
Pada penelitian ini, akan dijelaskan mengenai analisis hasil dan pembahasan selama eksperimen dengan algoritmanaive bayes.

*B.* Hasil Pengolahan Data

Pada Percobaan pertama adalah dengan menerapkan algoritma naïve bayes untuk memprediksi jurusan dengan data set yang terdiri dari 300 examples, 4 special attributes meliputi final,confidence IPA,confidence IPS,Prediction Final,13 regular attributes meliputi Kimia,Fisika,Biologi, Jumlah IPA,Ekonomi,Geografi,Sosiologi,Jumlah IPS,IQ,dan Minat. Dengan total 17 atribut yang akan diproses dan dimodelkan dengan algoritma naïve bayes classifier .Kemudian inputan proses klasifikasi dengan 8 atribut berupa nilai Kimia,nilai Fisika,nilai Biologi,nilai Ekonomi,nilai Geografi,nilai Sosiologi,IQ,dan Minat.

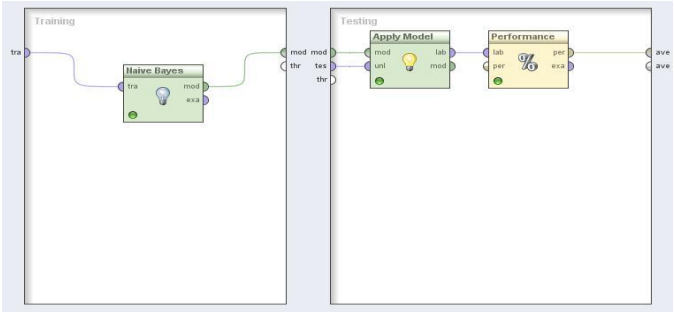
*C.* Proses Pemodelan

a. Pengujian dengan menggunakan cross validation untuk melakukan pengujian model.




Gambar 4.1 Pemodelan naïve bayes dan *cross validation*

Pada gambar di atas dilakukan pengujian model naïve bayes dengan menghubungkan dataset nilai siswa ke operator cross validation yang di dalamnya terdapat proses seperti pada gambar 4.2.




Cross validation yang digunakan dalam penelitian ini adalah 10-fold validation. Dataset yang berisi 300 data dengan 8 atribut akan dipecah menjadi 10 bagian. Dimana setiap bagian akan dibentuk secara random. Prinsip 10-fold validation adalah 1:9, 1 bagian menjadi data testing, data lainnya menjadi data training. Demikian sehingga 10 bagian tersebut bisa menjadi data testing. Setelah proses training dan testing maka dapat diukur akurasinya.

*D.* Hasil Pengujian Metode Naïve Bayes

a. Confusion Matrix

Berdasarkan data training sebanyak 300 exampleset dengan 8 atribut berupa nilai Kimia,nilai Fisika,nilai Biologi,nilai Ekonomi,nilai Geografi,nilai Sosiologi,IQ,dan Minat yang dimodelkan dengan algoritma naïve bayes diperoleh hasil akurasi sebanyak 98.00 % dengan rincian sebagai berikut :

1. Accurasi

accuracy: 98.00%

| | true IPA | true IPS | class precision |
|---|---|---|---|
| pred. IPA | 142 | 3 | 97.93% |
| pred. IPS | 3 | 152 | 98.06% |
| class recall | 97.93% | 98.06% | |

Gambar 4.4 Nilai akurasi model naïve bayes

Jumlah true positif (tp) adalah 142 jumlah siswa yang diklasifikasikan ke dalam kelas IPA dan false negative (fn) sebanyak 3 jumlah siswa yang diklasifikasikan ke dalam kelas IPA tetapi masuk kelas IPS. Jumlah true negative (tn) adalah sebanyak 152 jumlah siswa yang diklasifikasikan ke dalam kelas IPS dan false positif (fp) sebanyak 3 jumlah siswa yang diklasifikasikan IPS tetapi masuk kelas IPA.

Berdasarkan hasil confusion matrix , menunjukkan bahwa hasil akurasi yang didapat dengan menggunakan algoritma naïve bayes classifier adalah sebesar 98.00 % . Perhitungan dari akurasi , sensitifity , ppv , dan npv adalah sebagai berikut :

acc = (tp+tn)/(tp+tn+fp+fn) (142+152)/(142+152+3+3) = 0,980

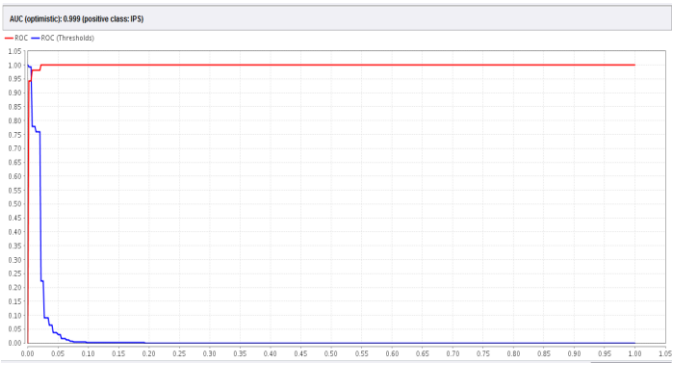
sensitivity = tp/(tp+fn) = 142/(142+3) = 0,979

specifity = tn/(tn+fp) = 152/(152+3) = 0,980

PPV = tp/(tp+fp) = 142/(142+3) = 0,979

NPV = tn/(tn+fn) = 152/(152+3) = 0,980

2. Evaluasi ROC Curve

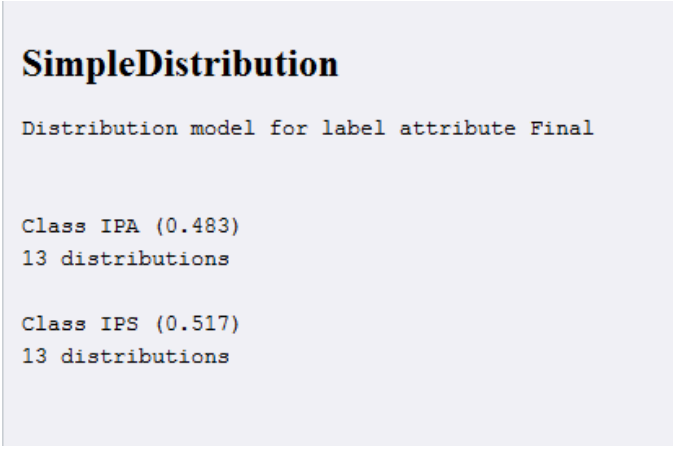



Gambar 4.5 Nilai AUC model naïve bayes

Berdasarkan grafik ROC diatas , telah menunjukkan bahwa nilai AUC (Area Under Curve) sebesar 0,999 dengan tingkat akurasi Excellent Classification.

3. Simple Distribution

Dari hasil pemodelan didapatkan simple distribution model untuk label atribut final dari masing – masing kelas IPA dan IPS.




Gambar 4.6 Simple Distribution

Dari kesimpulan simpledistribution , presentase dari 300 dataset nilai siswa SMA menghasilkan 0,483 = 48,3 % untuk kelas IPA dan 0,517 = 51,7 % kelas IPS**.**

## V. PENUTUP

algoritma naïve bayes classifier dapat digunakan dan diterapkan untuk mengklasifikasikan jurusan siswa SMA . Hasil dari proses data mining ini dapat digunakan sebagai pertimbangan dalam penjurusan lebih lanjut. Klasifikasi menggunakan naïve bayes classifier menghasilkan akurasi yang excellent . Akurasi yang dihasilkan dari klasifikasi jurusan siswa SMA N 1 Subah menggunakan naïve bayes memiliki akurasi sebesar 98,00% dan nilai AUC 0,999% . Penelitian yang telah dilakukan dengan algoritma data mining naïve bayes ini diharapkan dapat membantu proses klasifikasi penjurusan SMA dengan tepat dan mengurangi resiko terjadinya kesalahan perhitungan sehingga dapat memaksimalkan kinerja penjurusan di SMA N 1 Subah.

## REFERENSI


[1] Yetli Oslan,et.al “Implementasi Metode Bayesian Dalam Penjurusan Di SMA Bruderan Purworejo,2013”

[2] Satrianto, Mehmed. 2003. Teorema Bayes.http://Pdf.searchengine/teorema_bayes.pdf.

[3] Saraswati, N.W.S., 2011, Text Mining dengan Metode Naïve Bayes Classifier dan support Vector Machine untuk Sentimen Analysis, Thesis Program Studi Teknik Elektro, Program pasca Sarjana Universitas Udayana, Bali.

[4] Daihani,2001 . Sistem Informasi .

[5] Wibisono, Y. 2005. Klasifikasi Berita Berbahasa Indonesia menggunakan Naïve Bayes Classifier.

[6] Anonym. 2010. Naïve Bayes Classifier. [Online]. Tersedia di: http://en.wikipedia.org/wiki/Naive_Bayes_classifier.

[7] Rainardi,Vincent.2008.”Building a Data Warehouse with Example in SQL Server”.

[8] Han,J.and Kamber , M , 2006 “ Data Mining Concepts and Techniques Second Edition.Morgan Kauffman.

[9] Vercellis, C 2009 .Bussines Intelegence : Data Mining Optimization for Decision Making.John Wiley & Sons,Ltd.

[10] Santoso,Budi . 2007 . Data Mining . Teknik Pemanfaatan Data Untuk Keperluan Bisnis Yogyakarta ; Graha Ilmu,

[11] Kusrini, dan Emha, T.L. 2009. Algoritma Data Mining. Yogyakarta : Andi.

[12] Witten,I.H and Frank,E.2005. Data Mining : Practical Machine Learning Tools and Techniques Second Edition.Morgan Kauffman : San Fransisco

[13] Amir Hamzah.2012."Klasifikasi teks dengan naive bayes classifier untuk pengelompokan teks berita dan abstrak Akademis".

[14] Sri Kusumadewi.vol 3.2009."klasifikasi gizi menggunakan NBC".

[15] Tresna Yudha Prawira,Vol1.2011."Sistem Pendukun Keputusan untuk penjurusan(IPA/IPS/Bahasa)"

[16] Dwi Kurnia Basuki, et.al."DSS untuk rekomendasi Pemilihan Jurusan Pada Perguruan Tinggi Bagi siswa SMU".

[17] Erlan Darmawan,"Implementasi Metode Weighted Product pada Sistem Penduung Keputusan Untuk Menentukan Penjurusan Di Sekolah Menengah Atas”.

[18] Zhang, H., dan Su, J. (2007). Naive bayesian classi_ers for ranking. Retrieve December 2007, from www.cs.unb.ca/profs/hzhang/publications/NBRanking.

[20] Rish, Irina, 2001 , An Empirical Study of the Naïve Bayes Classifier, T.J. Watson