

Data Mining Prestasi Akademik Dengan Naive Bayes Berdasarkan *Attribut Importance* (AI)

Ni Komang Sri Julyantari, I Ketut Dedy Suryawan
STMIK STIKOM Bali

Jl. Raya Puputan No. 68 Renon, Denpasar – Bali, (+62 361) 244445
tari@stikom-bali.ac.id, dedymeng@stikom-bali.ac.id

Abstrak

Institusi pendidikan memegang peranan penting dalam menghasilkan kualitas lulusan maupun prestasi yang baik bagi anak didiknya. Diperlukan suatu usaha yang konsisten dan sistem pengajaran yang terkelola dengan baik sehingga apa yang disampaikan, diterapkan dan diajarkan kepada peserta didik bisa dipahami, dikembangkan bahkan diterapkan sehingga mencapai hasil yang diinginkan. Namun, prestasi akademik pasti selalu ada baik dan buruk yang dipengaruhi banyak faktor selain proses dan tata kelola belajar mengajar pada institusi pendidikan tersebut. Prestasi akademik yang dimaksud disini berupa nilai Indek Prestasi Akademik yang baik dan masa studi yang tepat waktu. Akibat banyaknya peserta didik yang indek prestasi akademiknya buruk mengakibatkan masa studi menjadi lebih lama sehingga terjadi penumpukan dalam hal jumlah siswa yang berakibat ratio dan keseimbangan kurang baik. Saat ini, STIKOM Bali dengan jumlah mahasiswa yang mencapai 6000 orang yang disertai dengan jumlah lulusan yang masih kurang ideal setiap periode kelulusan, maka perlu diterapkan dan dimulai pola penerimaan mahasiswa baru yang efektif bagi pencapaian prestasi dan kelulusan tepat waktu tersebut. Dalam penjurangan mahasiswa baru dan kaitannya dengan potensi prestasi akademik perlu dilihat latar belakang calon mahasiswa sehingga didapat klasifikasi yang tepat mengenai kriteria apakah calon mahasiswa baru diterima atau ditolak untuk menjadi mahasiswa di STIKOM Bali. Berdasarkan data – data siswa sebelumnya dan hasil prestasi akademik yang dicapai tentunya banyak faktor seperti nilai UAN, daerah asal, jenis kelamin, nilai tes masuk, keuangan, pergaulan dan lain - lain yang memberikan pengaruh terhadap prestasi akademik. Pemilihan faktor atau atribut yang paling berpengaruh ditentukan dengan cara menggunakan *AI(Attribut importance)* terhadap indek prestasi mahasiswa. Indek prestasi mahasiswa dikategorikan dalam 3 skala yaitu indek prestasi sedang, rendah dan baik. Setelah pemilihan atribut menggunakan teknik *Attribute Importance (AI)*, data sampling dianalisa menggunakan teknik klasifikasi menggunakan *Algoritma Naïve Bayes* yang kemudian digunakan untuk melakukan scoring terhadap prestasi akademik. Data training yang digunakan yaitu data mahasiswa angkatan 2007-2011 yang telah lulus. Sebagai data test yang akan dicoba kedalam klasifikasi yang dihasilkan menggunakan data mahasiswa baru angkatan 2012-2013 yang belum lulus.

Kata Kunci : *Naive Bayes, Attribut Importance*, Prestasi Akademik

Abstract

Educational institutions play an important role in producing quality graduates as well as a good achievement for the students. It takes a consistent effort and teaching system managed properly so that what is delivered, implemented and taught to students can be understood, developed and even implemented so as to achieve the desired results. However, academic achievement must always be good and bad that is influenced by many factors in addition to the process of governance and learning in the educational institutions. Academic achievement is here in the form Academic Performance Index score is good and timely study period. Due to the number of learners that lead to poor academic performance index study period becomes longer sehingga a buildup in the number of students that result in unfavorable ratio and balance.

Currently, STIKOM Bali the number of students who reach 6000 people, along with the number of graduates are still less than ideal every passing period, it is necessary to apply and begin the pattern of new admissions to the achievement of effective and timely graduation is. In networking new students and their relationship to academic achievement potential to be seen the background of prospective students in order to get the appropriate classification criteria regarding whether new students accepted or rejected for a student in STIKOM Bali. Based on the data - previous student data and achievement outcomes

certainly achieved many factors such as UAN, place of origin, gender, entrance test score, financial, and other relationships - Another influence on academic achievement. Selection factors or attributes are most influential determined by using AI (Attribute importance) on student achievement index. Index of student achievement scale are categorized into three achievement index medium, low and well. After the selection of attributes using the technique Attribute Importance (AI), sampling data were analyzed using the technique of classification using Naïve Bayes algorithm is then used to scoring on achievement. Training data used is data 2007-2011 generation students who have graduated. As the test data to be tested into klasifikasi generated using the 2012-2013 freshmen class of data that has not been passed.

Keywords: Naive Bayes, Attribute Importance, Academic Achievement

1. Pendahuluan

Berkembangnya penggunaan teknologi informasi dan komputer dalam bidang pengelolaan data menyebabkan akumulasi data dalam jumlah sangat besar di berbagai macam perusahaan. Apalagi dengan semakin berkembangnya pengetahuan mengenai penambangan data yang besar tersebut menjadi informasi yang lebih berguna bagi pengelola data untuk lebih dioptimalkan dalam menunjang keputusan – keputusan bisnis yang menguntungkan. Data dalam jumlah besar kadangkala belum dioptimalkan untuk memperoleh informasi yang lebih mendalam untuk mendukung tujuan – tujuan strategis yang diperlukan sehingga seolah – olah dibiarkan begitu saja menjadi kuburan data

Demikian halnya data dalam institusi perguruan tinggi seperti STIKOM Bali yang menyimpan kumpulan data yang banyak. Sejak berdiri tahun 2002 dan sekarang terdapat kurang lebih 6000 mahasiswa sehingga terdapat banyak data yang bisa digali. Peranan sebagai institusi pendidikan diharapkan menyelenggarakan pendidikan yang berkualitas bagi mahasiswa sehingga menghasilkan sumber daya manusia yang berilmu, cakap, kreatif dan berkualitas. Perlu digali informasi yang bisa digunakan untuk lebih meningkatkan daya saing yang salah satunya untuk meningkatkan prestasi akademik mahasiswa. Prestasi akademik sebagai keberhasilan seorang mahasiswa tidak terlepas dari latar belakang mahasiswa itu sendiri disamping sistem dan iklim belajar mengajar yang tercipta di lingkungan pendidikannya. IPK yang baik tentunya membuat target masa studi tercapai dengan kualitas yang bagus. Masa studi yang tepat waktu mendorong berkurangnya penumpukan mahasiswa di semester akhir yang bisa mengakibatkan ratio dan kualitas yang tidak baik. Prestasi Akademik biasanya diukur melalui Indeks Prestasi Akademik (IPK). Keberhasilan dalam memperoleh IPK yang tinggi biasanya dipengaruhi oleh banyak faktor seperti nilai UAN, nilai ujian masuk, jenis kelamin, jalur masuk, daerah asal, keuangan, pekerjaan orang tua, pergaulan dan lain-lain.

Data training yang akan digunakan yaitu data mahasiswa angkatan 2007-2011 yang telah lulus. Sebagai data test yang akan dicoba kedalam klasifikasi yang dihasilkan akan menggunakan data mahasiswa baru angkatan 2012-2013 yang belum lulus Dalam penentuan atribut, tidak semua atribut dipilih. Penentuan atribut yang paling berpengaruh dilakukan dengan menggunakan *AI(Attribute importance)* terhadap indek prestasi mahasiswa. Indek prestasi mahasiswa dikategorikan dalam 3 skala yaitu indek prestasi sedang, rendah dan baik. Setelah pemilihan atribut ditentukan dan dianalisa menggunakan teknik Attribute Importance (AI), data akan dianalisa menggunakan teknik klasifikasi menggunakan *Algoritma Naïve Bayes* yang kemudian akan digunakan untuk melakukan scoring terhadap prestasi akademik

Dengan adanya klasifikasi ini diharapkan adanya klasifikasi tingkat prestasi akademik mahasiswa berdasarkan atribut terpilih dengan teknik *atribut importance(AI)* yang kemudian dianalisa dengan algoritma *Naive Bayes*. Hasil yang diperoleh dapat dijadikan acuan dalam proses penerimaan mahasiswa baru berikutnya, sehingga bisa lebih selektif dalam menganalisa, dan memproses latar belakang calon mahasiswa baru

2. Tinjauan Pustaka/ State of the Art

2.1 Literatur Riview

Penelitian mengenai data mining sudah banyak dilakukan dimana – mana. Setiap bidang pekerjaan mempunyai data dan apabila data sudah dimanfaatkan dengan lebih maksimal akan bisa memberi manfaat lebih kepada perilaku dan strategi kebijakan dalam mengambil keputusan. Terlebih lagi di dunia pendidikan, sudah cukup banyak penelitian mengenai data mining akademik yang telah dilakukan karena ddata mining merupakan teknologi yang tepat untuyk meningkatkan perbaikan kualitas pendidikan dan prestasi akademik mahasiswa. Beberapa penelitian mengenai dunia pendidikan dengan data mining antara lain :

- a. Analisis komparasi algoritma klasifikasi data mining untuk prediksi mahasiswa non aktif (Khafiizh Hastuti, 2012)
 Dalam jurnalnya tersebut dilakukan penelitian untuk komparasi beberapa algoritma klasifikasi seperti *logistic regression*, *decision tree*, *naive bayes* dan *neural network* untuk mengetahui tingkat akurasi dari masing – masing algoritma dalam memprediksi tingkat mahasiswa non aktif. Hasil proses klasifikasi dievaluasi dengan menggunakan *cross validation*, *confusion matrix*, *ROC Curve* dan *T-test*
- b. Sistem pendukung Keputusan berbasis *decision tree* dalam pemberian beasiswa studi kasus : AMIK “BSI Yogyakarta” (Anik Andriani, 2013)
 Pada penelitian ini menggunakan algoritma C4.5 sebagai alat bantu dalam memberi keputusan penerimaan beasiswa. Hasil klasifikasi dievaluasi menggunakan *confusion matrix* dan kurva ROC/AUC(Area under Curve). Hasil evaluasi *confusion matrix* hasil klasifikasi penerima beasiswa menghasilkan tingkat akurasi sebesar 71,43%, nilai presisi sebesar 76,67% dan nilai *recall* sebesar 63,89%. Evaluasi dengan kurva ROC sebesar 0,660
- c. *Improving Academic Performance of Students of Defence University Based on Data Warehousing and Data Mining* (Sreenivasarao & Yohannes, 2012)
 Penelitian ini dilakukan di jurusan Teknik Defence University College untuk mencari faktor-faktor yang mempengaruhi prestasi akademik mahasiswa. Prestasi akademik mahasiswa jurusan Teknik kebanyakan relatif rendah, yang diukur dari GPA setiap mahasiswa. Maka daripada itu diterapkan *data mining k-Means clustering* dan *Decision tree* untuk mengekstrak informasi dan variabel penting yang signifikan berpengaruh terhadap prestasi akademik mahasiswa. Penelitian ini bertujuan membantu para guru mengurangi angka mahasiswa yang di *drop out* dan memperbaiki prestasi akademik para mahasiswa
- d. Sistem informasi klasifikasi tingkat prestasi akademik mahasiswa berdasarkan seleksi ujian masuk perguruan tinggi dengan algoritma *k-means* (Gede Rasben Dentes, Ketut Agustini, 2008)
 Metode yang digunakan pada penelitian ini adalah *k-means* dengan menggunakan atribut : id, jenis ujian, nilai ujian masuk dan IPK. Hasil dari penelitian menyatakan bahwa IPK paling bagus antara 3.03-4.00 dengan ujian masuk jalur SPMB (34.23%), PMJK (34.11%) dan PMJL (31.66%), dan paling buruk (1.10–2.07) dengan ujian masuk jalur SPMB (31.93%), PMJK (33.92%) dan PMJL (34.15%)
- e. Penerapan Data Mining untuk evaluasi kinerja Akademik Mahasiswa menggunakan algoritma *Naive Bayes Classifier*(Mujib Ridwan dkk, 2013)
 Dalam jurnal EECCIS Vol 7, No 1, Juni 2013, dilakukan penelitian untuk mengevaluasi kinerja akademik mahasiswa di tahun ke-2 dan diklasifikasi dalam kategori mahasiswa yang dapat lulus tepat waktu atau tidak. Dengan *Naive Bayes Classifier* data input diproses untuk membentuk tabel probabilitas sebagai dasar klasifikasi kelulusan mahasiswa. Hasil pengujian menghasilkan nilai *precision*, *recall*, dan *accuracy* masing – masing 83%, 50%, dan 70%. Secara umum, nilai hasil pengujian menunjukkan nilai akurasi yang sedang karena faktor penentu kelulusan mahasiswa tepat atau tidak tepat waktu kenyataannya memiliki nilai yang tidak konsisten. Atribut yang digunakan yaitu IPK, IP Semester 1, IP Semester 4 dan jenis kelamin
- f. *An Application of Predicting Student Performance Using Kernel K-Means and Smooth Support Vector Machine* (Sajadin Sembiring, 2012) Pada penelitian ini menghasilkan model peramalan prestasi akademik bagi intitusi perguruan tinggi berdasarkan faktor – faktor psikometri seperti minat, sikap pelajar, kepercayaan, dukungan keluarga dan penggunaan masa. Model dibangun dengan kernel *K-Means Clustering* dan *Smooth Support Vector Machine*. Ketepatan peramalan mempunyai ketepatan terendah 61% dalam membuat indek prestasi “Baik” dan ketepatan tertinggi 93,67% dalam membuat peramalan Indek Prestasi “Lemah

2.2 Data Mining

Data mining adalah sebuah istilah yang digunakan untuk menggambarkan penemuan ilmu pengetahuan dalam bidang database, sebuah bidang analisis informasi yang mencari pola tersembunyi dalam sekelompok data yang dapat digunakan untuk memprediksi perilaku masa depan (Turban et al, 2007:202)

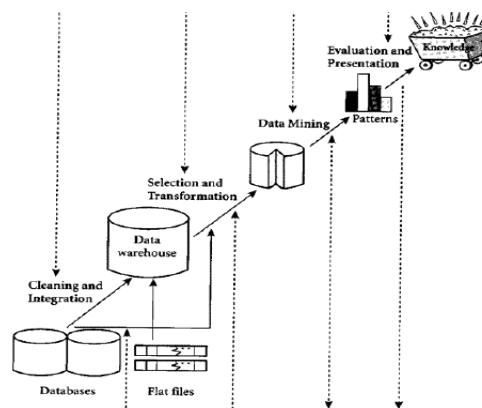
Menurut Gartner Group *data mining* adalah suatu proses menemukan hubungan yang berarti, pola, dan kecenderungan dengan memeriksa dalam sekumpulan besar data yang tersimpan dalam penyimpanan dengan menggunakan teknik pengenalan pola seperti teknik statistik dan matematika (Larose,2005)

Data Mining adalah proses menemukan pengetahuan yang menarik dari sejumlah besar data yang tersimpan dalam *database*, *data warehouse*, atau repositori penyimpanan data dan informasi lainnya (Han dan Kamber, 2006).

Data mining adalah proses menemukan hubungan dalam data yang tidak diketahui oleh pengguna dan menyajikannya dengan cara yang dapat dipahami hingga hubungan tersebut dapat menjadi dasar pengambilan keputusan (McLeod, 2007).

Data mining memegang peranan penting dalam berbagai bidang seperti industri, keuangan, cuaca, medical, teknologi dan lain sebagainya. Data mining berkenaan dengan jumlah pengolahan data dalam skala besar. Data-data yang ada, tidak dapat langsung diolah dengan menggunakan sistem data mining. Data-data tersebut harus dipersiapkan terlebih dahulu agar hasil yang diperoleh dapat lebih maksimal, dan waktu komputasinya lebih minimal. Adapun tahapan-tahapan yang harus dilalui dalam proses data mining antara lain :

- a. Data cleaning, untuk membersihkan data dari noise data dan data yang tidak konsisten.
- b. Data integration, mengkombinasikan atau mengintegrasikan beberapa sumber data.
- c. Data selection, mengambil data-data yang relevan dari database untuk dianalisis.
- d. Data transformation, mentransformasikan data summary ataupun operasi agregasi.
- e. Data mining, merupakan proses yang esensial dimana metode digunakan untuk mengekstrak pola data yang tersembunyi.
- f. Pattern evaluation, untuk mengidentifikasi pola sehingga merepresentasikan pengetahuan berdasarkan nilai-nilai yang menarik
- g. Knowledge presentation, dimana teknik representasi dan visualisasi data digunakan untuk mempresentasikan pengetahuan yang didapat kepada user



Gambar 1. Tahapan Data Mining

2.3. Atribut Importance (AI)

Attribute Importance (AI) memberi peringkat atribut dengan menghilangkan atribut yang berulang, tidak relevan, atau tidak informatif dan mengidentifikasi atribut yang mungkin memiliki pengaruh yang paling tinggi dalam membuat prediksi. AI menggunakan algoritma *Minimum Description Length* (MDL). Algoritma MDL mempertimbangkan setiap atribut sebagai model prediktif sederhana dari kelas target. Teknik AI digunakan untuk mengoptimalkan analisa model *classification* dengan mengurangi atribut yang digunakan dan akan meningkatkan kecepatan dan akurasi saat membangun model. AI memberikan solusi otomatis untuk meningkatkan kecepatan dan akurasi model klasifikasi yang dibangun di atas tabel data dengan sejumlah besar atribut

2.4. Prestasi Akademik

Prestasi akademik adalah hasil pelajaran yang diperoleh dari kegiatan belajar di sekolah atau perguruan tinggi yang bersifat kognitif dan biasanya ditentukan melalui pengukuran dan penilaian. Prestasi akademik pada penelitian ini dinilai berdasarkan IPK (Indeks Prestasi Kumulatif). Indeks Prestasi (IP) adalah penilaian keberhasilan studi semester yang dilakukan pada tiap akhir semester. Penilaian ini meliputi semua mata kuliah yang direncanakan mahasiswa dalam Kartu Rencana Studi (KRS).

$$IP = \frac{\sum_{i=1}^n N_i k_i}{\sum_{i=1}^n k_i} \dots\dots\dots(\text{persamaan 1})$$

Dimana :

- IP = Indeks Prestasi
- N_i = Nilai mutu mata kuliah i
- K = Bobot sks mata kuliah i
- N = Jumlah mata kuliah i

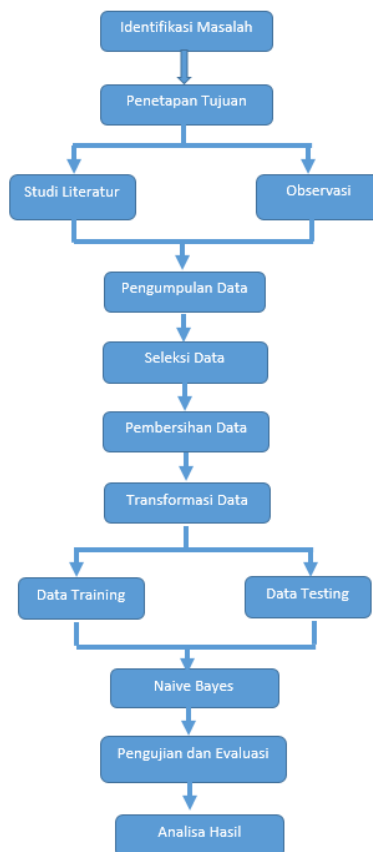
3. Metode Penelitian

3.1 Tempat dan Waktu Penelitian

Penelitian dilakukan di STMIK STIKOM Bali, Jl. Raya Puputan No. 86 Renon, Denpasar-Bali

3.2 Alur Penelitian

Untuk alur penelitian yang digunakan dimulai dari mengidentifikasi masalah yang ada di STIKOM Bali, lalu dilanjutkan dengan menetapkan tujuan dengan melakukan observasi dan studi literatur setelah itu melakukan pengumpulan data, seleksi data, pembersihan data, transformasi data yang kan menghasilkan data training dan data set lalu dilanjutkan dengan metode naïve bayes dan setelah melakukan perhitungan tersebut , maka dilakukan pengujian dan evaluasi dari dat yang telah dihitung, lalu hasil akhirnya akan di analisa. Untuk gambaran dari alur penelitian dapat dilihat dari gambar 2.



Gambar 2. Alur Penelitian

3.1. Teknik Pengumpulan Data

- a. Studi Literatur
Studi literatur merupakan metode pengumpulan data dan informasi yang diperoleh melalui sumber-sumber seperti buku, karya tulis, dan sumber lain yang berhubungan dengan objek penelitian terutama mengenai *data mining*, *Naive Bayes* sehingga dapat membantu dalam proses perancangan sistem yang dibuat.
- b. Observasi
Observasi merupakan metode pengumpulan data dan informasi yang diperoleh dengan cara melakukan pengamatan langsung terhadap objek penelitian seperti pengumpulan data kunjungan, data nilai, data transactional baik yang didapat secara *hardcopy* maupun *softcopy* serta database akademik yang ada
- c. Wawancara
Wawancara dilakukan terhadap Unit Akademik dan Pemasaran yang menangani Penerimaan Mahasiswa Baru STIKOM Bali untuk mencari data mahasiswa baru dan nilai prestasi akademik.

3.2. Analisis dan Perancangan Sistem

- a. Analisis Data dan Sistem
Digunakan untuk menganalisa kebutuhan data, proses dan jaringan. Berdasarkan kebutuhan yang telah ditentukan sebelumnya, maka dibutuhkan data – data yang diperlukan untuk data sample dari data akademik yang tersedia, diseleksi, dibersihkan dan ditransformasi sehingga menjadi data siap pakai sebagai data training. Hasil dari analisis data dan sistem ini kemudian akan digunakan untuk melakukan perancangan sistem.
- b. Perancangan Sistem
Berdasarkan hasil dari analisis sistem, maka tahap selanjutnya adalah perancangan sistem yang meliputi beberapa bagian, yaitu : *Use Case Diagram*, *Class Diagram*, *Activity Diagram*, *Sequence Diagram*, *User Interface Design*

4. Hasil dan Pembahasan

Dalam membangun rancangan sistem Data mining prestasi akademik dengan *Naive Bayes*, maka diperlukan analisa kebutuhan yang berkaitan dengan kebutuhan sistem secara keseluruhan. Analisa yang tidak bagus akan berakibat buruk pada sistem secara keseluruhan, misalnya data-data apa saja yang dibutuhkan pada sistem yang akan dibangun, proses-proses yang terjadi pada sistem serta konfigurasi jaringan yang dibutuhkan dalam mendesain dan merancang sistem tersebut

Untuk menganalisa kebutuhan dalam penelitian ini maka digunakan pemodelan menggunakan UML yaitu *Use Case Diagram*, *Class Diagram*, *Activity Diagram* dan *Sequence Diagram*.

4.1 Analisa Kebutuhan Data

Proses analisa ini dimulai dari identifikasi data yang dibutuhkan untuk keperluan informasi yang akan diberikan kepada pengguna. Kebutuhan yang dimaksud adalah kebutuhan dari sisi pengguna untuk keperluan klasifikasi data prestasi akademik dengan *Naive Bayes*

Tabel 1. Analisa Kebutuhan

No	Nama Data	Atribut Data	Deskripsi Data
1	User	Username Password Kode Karyawan Level	Daftar user yang berhak mengakses aplikasi
2	Mahasiswa Baru	Kode Cama Nama Cama Jurusan SMA/SMK Kelas	Data ini digunakan untuk memperoleh data mahasiswa baru STIKOM Bali

		Angkatan Pekerjaan Ortu Nem	
3	Ujian Calon Mahasiswa	Kode Cama Nilai Ujian PMB Status Lulus Jurusan Kelas	Data ini digunakan untuk memperoleh nilai dan status kelulusan ujian mahasiswa baru
4	Mahasiswa	NIM Kode Cama Nama Jurusan Kelas Angkatan	Data mahasiswa
5	Data Nilai	NIM IPS IPK Masa Studi Lama Skripsi/TA	Data nilai beserta lama studi mahasiswa

4.2 Analisa Kebutuhan Proses

Pada tahap ini akan menjelaskan proses – proses yang terjadi pada desain dan rancangan Data Mining Prestasi Akademik dengan *Naive Bayes* berdasarkan *Atribut Importance*. Prosesnya seperti terlihat pada table berikut ini:

Tabel 2. Analisa Kebutuhan Proses

No	Nama Proses	Deskripsi Proses	Data Input	Aktor/user
1	Login	Proses validasi user id dan password dilakukan untuk mengakses halaman admin atau user lain.	Username Password	Admin/User
2	Seleksi Data	Filter data dari database mentah yang didapat ke data warehouse	Data database server asal, Data database server tujuan, Data tabel	Admin
3	Pembersihan Data	Data yang terpilih dibersihkan dari isian yang tidak valid	Data Tabel	Admin
4	Transformasi Data	Penyimpana data ke dalam bentuk data warehouse yang siap digunakan sebagai data training untuk proses data baru	Data Tabel	Admin
5	Klasifikasi Data	Proses klasifikasi data prestasi akademik	Jurusan SMA/SMK Kelas Angkatan Pekerjaan Ortu Nem	User

4.3 Use Case Diagram

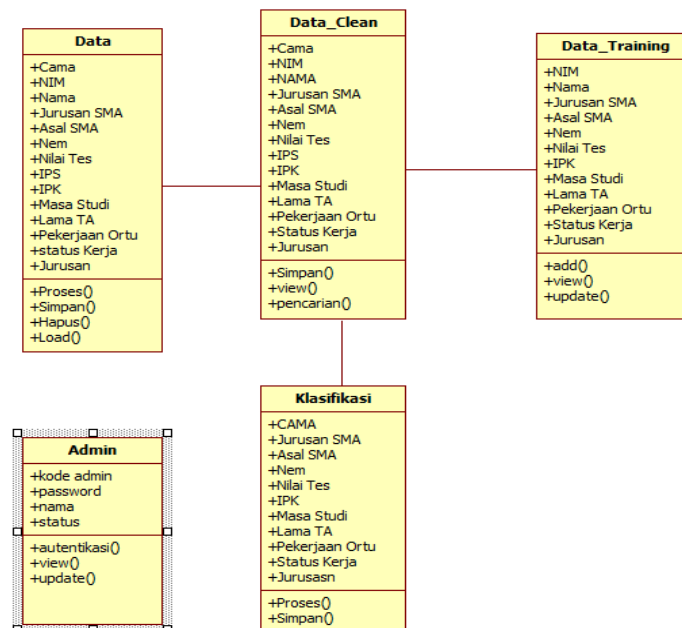
Use Case diagram menggambarkan fungsionalitas yang diharapkan dari sebuah sistem. Yang ditekankan adalah “apa” yang diperbuat sistem, dan bukan “bagaimana”. Sebuah *use case* menggambarkan suatu urutan interaksi antara satu atau lebih aktor dengan sistem. Berbeda dengan *class* diagram yang lebih cocok dibaca oleh disainer/analisis, *use case* diagram sangat cocok untuk dipahami oleh pengguna sistem. Selain itu *use case* diagram hanya menggambarkan apa yang dilakukan oleh sistem dan tidak menggambarkan bagaimana sistem melakukannya. Aktor adalah pemain, sedangkan *use case* adalah apa yang dimainkan/dilakukannya dengan relation sebagai penunjuknya. Berdasarkan kebutuhan sistem yang akan dirancang terdapat dua aktor, yaitu Admin dan KuA.



Gambar 3. Use Case Diagram

4.4. Class Diagram

Setelah menganalisa *use case diagram* diatas maka dapat diketahui kebutuhan sistem dan menentukan *class* apa saja yang diperlukan dalam aplikasi ini. Untuk memodelkan *class-class* yang ada maka digunakan *class diagram* yang akan menggambarkan *class-class* yang ada beserta relasi antar *class*. Berikut ini adalah *class diagram* dari sistem yang akan dirancang



Gambar 4. Class Diagram

4.5 Implementasi Sistem

A. Form Login

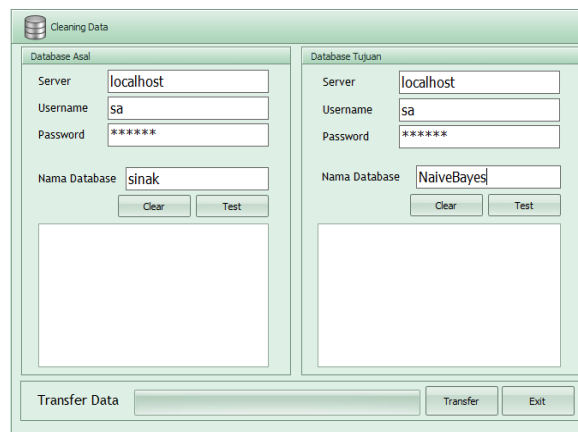
Pada form login ini, user memasukkan account berupa *username* dan *password* untuk dapat menggunakan sistem sesuai hak akses yang diberikan. *User* ada 2 yaitu user *admin* dan *KuA*



Gambar 5 Form Login

B. Form Transfer Data

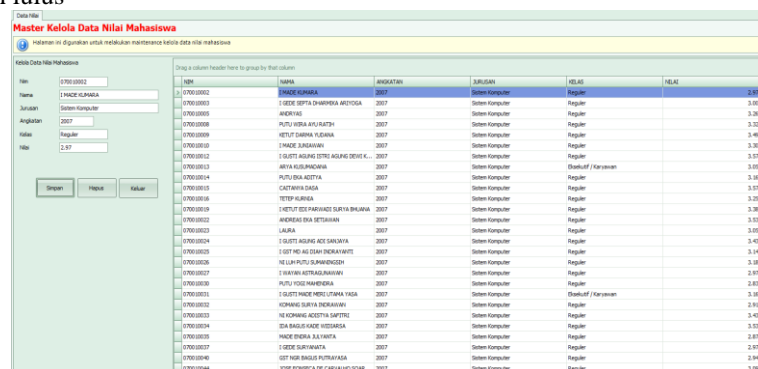
Gambar dibawah merupakan form untuk transfer data dari database asal ke database warehouse yang nantinya digunakan dalam penelitian ini



Gambar 6. Form Transfer Data

C. Form Kelola Data Nilai IPK

Form ini merupakan form untuk mengelola data nilai IPK (Indek Prestasi Akademik) mahasiswa yang telah lulus



Gambar 7. Form Kelola Data Nilai IPK

D. Form Transaksi Cleaning Data

Form ini merupakan form untuk membersihkan data – data dari isian yang tidak valid untuk nantinya dipakai sebagai *datawarehouse*

Gambar 8. Form Transaksi Cleaning Data

E. Form Klasifikasi Data Training

Form ini merupakan form uji coba klasifikasi data training dengan *Naive Bayes*. Terdapat filter jurusan, angkatan dan kelas dari mahasiswa yang dipakai sebagai data training. Terdapat dua hasil yaitu status kelulusan meliputi “Tepat Waktu” dan “Tidak Tepat Waktu” serta status prestasi meliputi “Kurang Bagus”, “Bagus” dan “Sangat Bagus”

Gambar 9. Form Data Training

F. Form Klasifikasi Naive Bayes Data Uji

Form ini merupakan form klasifikasi Naive Bayes Data Uji dengan mencoba data yang akan diuji berdasarkan klasifikasi data training.

Gambar 10 Form Data Uji

5. Simpulan

- a. Hasil analisa kebutuhan data yang mencakup data mahasiswa baru, data nilai akademik.
- b. Hasil analisa kebutuhan proses mencakup proses login, seleksi, cleaning dan transformasi data, proses data mining dan Pendukung keputusan.
- c. Konfigurasi jaringan (*network configuration*) yang digunakan adalah jaringan Local Area Network, sehingga sistem aplikasi ini dapat diakses melalui jaringan lokal dan berbasis client server.
- d. *Business Model* telah didefinisikan tata laksana sistem yang digambarkan dalam *use case diagram*. *Use case diagram* tersebut terdiri dari 5 (lima) *use case*, yaitu: *use case* autentikasi, Seleksi data, *cleaning* data, Transformasi Data, kelola data admin, Klasifikasi Data. Selain itu, pada *use case diagram* juga terdapat 2 (dua) *actor* yaitu *Admin* dan *KuA*.
- e. Model-model logikal yang dihasilkan telah digambarkan dalam *class diagram*. *Class diagram* yang terbentuk sejumlah 5 (lima) *class*, yaitu *class* data, data_clean, data_training, admin dan klasifikasi
- f. Aplikasi yang dihasilkan membantu dalam mendukung keputusan hasil prestasi akademik dan status kelulusan berdasarkan atribut yang dimasukkan

Daftar Pustaka

- [1] Han, Jiawei, dan Micheline Kamber. (2006). *Data Mining Concept and Techniques*, second edition. Elsevier, San Francisco.
- [2] Kusriani, Emha Taufiq Luthfi. 2009. *Algoritma Data Mining*, Penerbit Andi, Yogyakarta
- [3] Larose, D.T. (2005). *Discovering Knowledge in Data*. New Jersey : John Wiley & Sons Inc.
- [4] McLeod, Jr.R. dan G.P. Schell. 2007. *Management Information System*. 10th ed. Pearson Education, Inc.
- [5] Santoso, Budi. (2007). *Data Mining, Teknik pemanfaatan Data untuk keperluan Bisnis*, Graha Ilmu, Yogyakarta
- [6] Tang Zaohui, MacLennan Jamie. 2005. *Data Mining with SQL Server 2005*. Indianapolis : Wiley
- [7] Turban, E., Aronson, J. E., Liang, T. P., & Sharda, R. (2007). *Decision support and business intelligencesystems (Eighth ed.)*. Pearson Education
- [8] Turban, E., et al. (2011). *Business Intelligence: A Managerial Approach*, 2nd ed. Upper Saddle River, NJ: Prentice Hall