

PERANCANGAN SISTEM INFORMASI DATA *MINING* DENGAN ALGORITMA *NAIVE BAYES* UNTUK PENENTUAN PEMINATAN MAHASISWA PROGRAM STUDI TEKNIK INFORMATIKA UNIVERSITAS MUHAMMADIYAH PROF. DR HAMKA (UHAMKA)

Erizal¹⁾, M. Fathul Hanif B.²⁾

^{1,2}Program Studi Teknik Informatika Fakultas Teknik Universitas Muhammadiyah Prof.DR.HAMKA
Jl.Tanah Merdeka No.6, Kp.Rambutan, Ps.Rebo, Jakarta Timur
E-mail: erizal@uhamka.ac.id¹⁾, Mfathulhanif@gmail.com²⁾

Corresponding author : erizal@uhamka.ac.id

Abstrak

Pemilihan peminatan dari program studi adalah suatu masalah bagi para calon mahasiswa yang belum memilih peminatannya dari program studi yang diampu, karena kurangnya informasi yang ia dapatkan membuat calon mahasiswa buta akan jurusan dan peminatan jurusan yang akan dipilih. Para psikologi menggunakan metode menilai karakter dan bakat siswa untuk mengetahui jurusannya serta peminatan jurusannya, ini tidak salah. Tetapi ada cara yang lebih mudah yaitu dengan penilaian atau pengambilan keputusan melalui riwayat nilai KRS mahasiswa yang telah diperoleh. UHAMKA sebagai Universitas be-rakreditasi A, seharusnya memiliki sistem yang dapat membantu merekomendasikan para mahasiswanya untuk memilih peminatan program studi yang ingin dipilih ketika berkuliah di UHAMKA khususnya di Fakultas Teknik Program Studi Informatika, ini akan sangat membantu mahasiswa Program Studi Teknik Informatika SDM yang benar-benar cocok untuk peminatan pada Program Studi Teknik Informatika. Sistem ini akan menggunakan nilai KRS mahasiswa Teknik Informatika dari semester 1 (satu) sampai dengan 4 (empat), lalu data akan diolah hingga mendapatkan rules dengan menggunakan metode classifier/klasifikasi dengan algoritma Naive Bayes. Sehingga akan mendapatkan kesimpulan untuk membantu menyelesaikan permasalahan pemilihan peminatan Program Studi Teknik Informatika.

Kata Kunci : Data Mining, Naive Bayes, pemilihan peminatan kuliah, KRS.

Abstract

The choice of specialization from the study program is a problem for prospective students who have not chosen their specialization from the study program being taught, because the lack of information they get makes prospective students blind to the majors and specializations of the majors to be chosen. Psychologists use the method of assessing the character and talents of students to find out their majors and specializations in their majors, this is not wrong. But there is an easier way, namely by assessing or making decisions through the history of students' KRS scores that have been obtained. UHAMKA as an A-accredited University, should have a system that can help recommend its students to choose the study program they want to choose when studying at UHAMKA, especially at the Faculty of Engineering, Informatics Study Program, this will really help students of the Informatics Engineering Study Program HR who really suitable for specialization in the Informatics Engineering Study Program. This system will use the KRS value of Informatics Engineering students from semester 1 (one) to 4 (four), then the data will be processed to obtain rules using the classifier/classification method with the Naive Bayes algorithm. So that it will get a conclusion to help solve the problem of choosing a specialization in the Informatics Engineering Study Program.

Keyword: Data Mining, Naive Bayes, Concentration Choice of Study Program, KRS

1 PENDAHULUAN

Fakultas Teknik UHAMKA merupakan salah satu Fakultas di UHAMKA yang hingga saat ini masih memiliki akreditasi B, dan Teknik Informatika sebagai salah satu dari Program Studi Fakultas Teknik pun memiliki akreditasi B. Pada Program Studi Teknik Informatika terdapat 3(tiga) buah peminatan, yaitu RPL, Jaringan Komputer dan Multimedia.

Berdasarkan wawancara peneliti dengan dosen pembimbing akademik dan mahasiswa Teknik Informatika minimnya informasi terkait peminatan ini sangat dirasakan dikarenakan tidak adanya acuan dan panduan baku dari Fakultas Teknik yang tidak memberikan rekomendasi mengenai peminatan untuk mahasiswa Teknik Informatika memilih peminatan tersebut, sehingga terjadi kesalahan mahasiswa dalam memilih peminatan yang tidak sesuai dengan dirinya.

Maka dari itu seharusnya untuk tingkat Perguruan Tinggi UHAMKA seharusnya memiliki cara/sistem untuk menentukan peminatan untuk mahasiswa kampus UHAMKA, sehingga Universitas dan fakultas dapat merekomendasikan peminatan yang sesuai dengan mahasiswanya untuk meningkatkan performa belajar mahasiswa yang masuk kedalam instrumen penilaian. (PT, 2019).

Hal ini sangat menguntungkan UHAMKA terutama Fakultas Teknik dan Program Studi Teknik Informatika karena akan mendapatkan mahasiswa-mahasiswa yang cocok dan tepat dengan konsentrasi belajarnya. Data *Mining* dengan metode klasifikasi dapat mengklasifikasi konsentrasi yang ingin sortir dengan menggunakan data, data ini akan diolah sedemikian rupa sehingga mmenjadi rules tertentu yang dapat menentukan konsentrasi dari tujuan pembuatan sistem keputusan tersebut (Saleh, 2017).

Metode penentuan jumlah kelas masih memiliki kekurangan ditandai dengan adanya mahasiswa yang belum bisa mengambil mata kuliah tertentu karena semua kelas yang dibuka untuk mata kuliah tersebut sudah terisi penuh. Ada juga jumlah kelas untuk mata kuliah tertentu yang dirasa berlebihan karena ada salah satu kelas yang dibuka dengan jumlah mahasiswa yang lebih sedikit dibanding dengan kelas lain dengan mata kuliah yang sama.

Sistem perkuliahan yang mengharuskan mengulang mata kuliah yang tidak lulus mempengaruhi proses penentuan jumlah kelas, karena mahasiswa memiliki kemungkinan tidak akan memilih mata kuliah yang telah disediakan sesuai kurikulum per semester dan mengulang terlebih dahulu mata kuliah yang mengulang atau bahkan sebaliknya.

Sistem pembatasan jumlah SKS mata kuliah yang bisa diambil mahasiswa dalam 1 semester juga mempengaruhi proses penentuan jumlah kelas, karena akan terbentuk banyak kemungkinan-kemungkinan dari kombinasi mata kuliah yang akan dipilih oleh mahasiswa dalam 1 semester. Mahasiswa memiliki kemungkinan meninggalkan beberapa mata kuliah, bisa juga mengambil mata kuliah yang disediakan untuk angkatan di atas-nya.

Ketika ada pembatasan pengambilan mata kuliah berdasarkan indeks prestasi, hal ini menimbulkan masalah dalam pengambilan mata kuliah yang memiliki prasyarat telah mengambil mata kuliah tertentu di semester sebelumnya. Ketika mahasiswa belum mengambil mata kuliah prasyarat, maka dia tidak bisa mengambil mata kuliah yang memiliki mata kuliah prasyarat

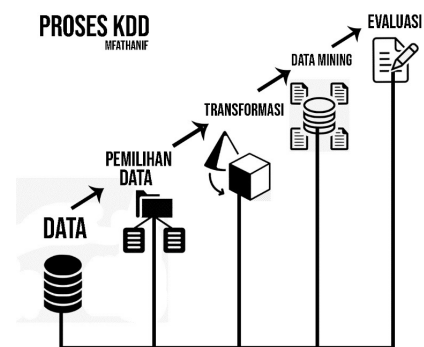
Dengan menggunakan 3 algoritma klasifikasi yang akan dibandingkan dan variabel yang digunakan dalam prediksi ini yang terdiri dari semester mahasiswa, jumlah maksimal SKS, angkatan, status pengambilan mata kuliah mengulang/baru, status nilai mata kuliah mengulang nilai C/mengulang nilai D dan E, jenis mata kuliah, SKS mata kuliah dan nilai IPK. Hasil dari pengujian model klasifikasi dengan k-Fold Cross-Validation adalah algoritma naïve bayes memiliki akurasi 66.69%, SVM 85.25% dan neural network 81.16%, sehingga algoritma yang memiliki akurasi

tertinggi adalah SVM. Sistem prediksi akan menggunakan algoritma SVM untuk memprediksi pengambilan matakuliah dari masing-masing mahasiswa untuk mendapatkan jumlah mahasiswa dalam sebuah kelas perkuliahan.

2 DASAR TEORI

Website atau sistus dapat diartikan sebagai kumpulan halaman-halaman yang digunakan untuk menampilkan informasi teks, gambar diam atau gerak, animasi, suara dan atau gabungan dari semuanya, baik bersifat statis maupun dinamis yang membentuk suatu rangkaian bangunan yang saling terkait, yang masing- masing dihubungkan dengan jaring-jaringan halaman (Nurmi, 2017).

Data Mining merupakan proses untuk mencari sebuah pola atau *code* unik dalam data terpilih yang didasari teknik atau metode tertentu. Pada data mining memiliki berbagai macam variasi teknik-teknik, metode-metode dan algoritma-algoritma yang dimiliki. Oleh karena itu dibutuhkan pemilihan teknik-teknik, metode-metode dan algoritma-algoritma yang tepat untuk dapat dipadukan dalam sebuah data *mining* untuk mendapatkan hasil yang akurat., seperti pada gambar 1 berikut ini.



Gambar 1 Proses KDD

Klasifikasi proses penentuan model pola dengan menggambarkan perbedaan kelas data sehingga bisa digunakan untuk mencapai tujuan prediksi kelas yang belum diketahui (Annur Haditsah, 2018). Teknik pada klasifikasi yang sering digunakan antara lain *Neural*, *Rough sets*, *K-nearest neighbor*, *Bayesian classifier*, *network* dan lain-lain (Saputra & Primadasa, 2019). Model tersebut dibangun dengan menggunakan analisa dari *database tuple*, setiap tuple diberikan asumsi menjadi predefined class yang ditentukan oleh atribut yang disebut *class label attribut*, seperti ada gambar 1 berikut ini.

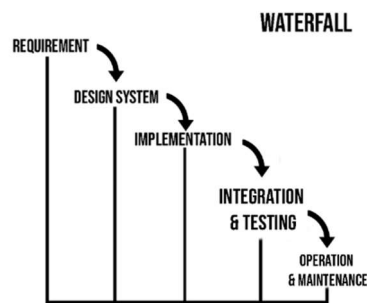


Gambar 2 Klasifikasi Model

Prediksi, pada data *mining* dikenal terdapat juga prediksi yaitu menarik nilai kesimpulan pada variabel-variabel yang ditentukan pada data yang terpilih. Pada sifatnya prediksi dibagi menjadi dua jenis, yaitu kualitatif dan kuantitatif.

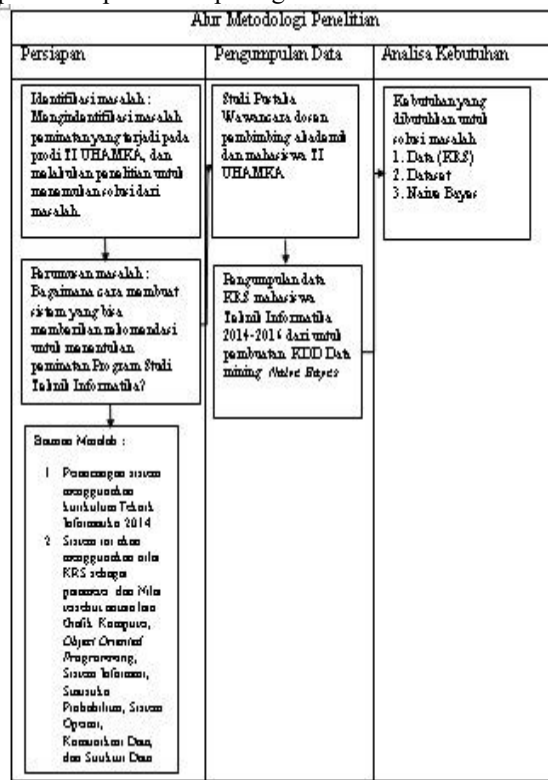
Naive Bayes, merupakan salah-satu algoritma *learning* induktif yang paling efektif dan efisien pada *machine learning* dan data *mining*. *Naive bayes* memiliki performa yang kompetitif pada proses klasifikasi walaupun menggunakan asumsi keidependenan atribut (tidak ada keterkaitan antar atribut). Biasanya asumsi keidependenan atribut ini (tidak ada keterkaitan antar atribut) pada data sebenarnya jarang terjadi, namun walaupun asumsi keidependenan atribut tersebut terjadi performa pengklasifikasian *naive bayes* cukup tinggi, hal ini dibuktikan pada berbagai penelitian empiris (Syarli & Muin, 2018).

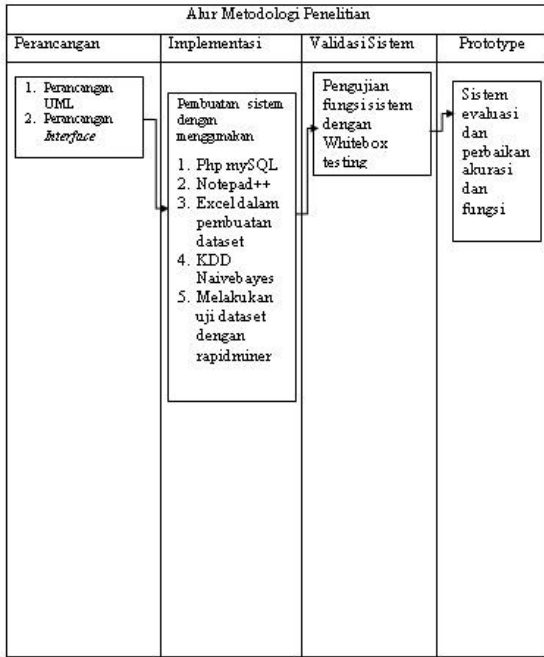
1. Teknik Pengujian
pada teknik pengujian digunakan *confusion matrix* sebagai suatu metode yang biasanya digunakan untuk melakukan perhitungan akurasi pada konsep data mining. *Confusion matrix* bila digambarkan dengan tabel akan menyatakan jumlah data uji yang benar diklasifikasikan dan jumlah data uji yang salah diklasifikasikan (Rahman M. Fadly, M.IIham Darmawidjadja, & Dion Alamsah, 2017).
2. Metode Pengembangan Aplikasi
Peneliti pada pengembangan aplikasi ini menggunakan metode *waterfall*, karena dirasa cukup dan masih relevan dengan aplikasi yang peneliti bangun. Metode *waterfall* merupakan metode yang pertama kali dikemukakan oleh seorang Winston Royce pada tahun 70-an. Dengan model klasik dan aliran sistem linier (Junaedy & Munir, 2017).



3 METODE PENELITIAN

Peneliti menggunakan metode kualitatif dengan observasi dan melakukan wawancara yang digambarkan pada alur penelitian pada gambar berikut ini:





1. Persiapan

Pada tahap ini penelitian dimulai dengan melengkapi seluruh perizinan yang berkaitan dengan penelitian dan dalam persiapan dibagi menjadi beberapa tahap bagian antara lain.

A. Identifikasi Masalah

Pada tahap ini penulis melakukan identifikasi masalah dengan mencatat seluruh masalah yang ada pada peminatan mahasiswa Teknik Informatika di UHAMKA dan masalah-masalah mengenai Data Mining dengan metode klasifikasi algoritma *Naive bayes* yang digunakan penulis untuk menyelesaikan masalah penelitian.

B. Perumusan Masalah

Pada tahap ini penulis mengumpulkan masalah-masalah yang dialami oleh mahasiswa Teknik Informatika UHAMKA disaat melakukan pemilihan peminatan Program Studi Teknik Informatika dengan wawancara.

C. Batasan Masalah

Pada tahap ini penulis membuat batasan agar penelitian terfokus pada tujuan dari masalah yang akan selesaikan.

2. Pengumpulan Data

Pengumpulan data dilakukan peneliti untuk menguatkan landasan dan menemukan solusi dari penelitian yang sedang dilakukan oleh peneliti. Pada pengumpulan data dibagi menjadi beberapa bagian antara lain :

A. Studi Pustaka

Pada tahap ini penulis mengumpulkan jurnal-jurnal yang berkaitan dengan penelitian yang sedang dilakukan oleh penulis.

B. Wawancara

Pada tahap ini penulis melakukan wawancara kepada dosen pembimbing akademik dan mahasiswa teknik informatika UHAMKA untuk menemukan masalah dan solusi peminatan Teknik Informatika UHAMKA

C. Pengumpulan KRS mahasiswa Teknik Informatika UHAMKA

Pada tahap ini penulis mengumpulkan data-data berupa KRS mahasiswa Teknik Informatika UHAMKA.

3. Analisa Kebutuhan

Pada tahap ini peneliti mengumpulkan kebutuhan-kebutuhan yang diperlukan dalam pembangunan sistem penentuan peminatan yang akan digunakan sebagai rekomendasi dari program studi teknik informatika UHAMKA.

A. Data KRS

Data KRS mahasiswa Teknik Informatika berdasarkan kurikulum TI UHAMKA 2014.

B. Dataset

Dataset yang akan digunakan sebagai data training dari sistem yang akan dibuat.

C. Naive Bayes

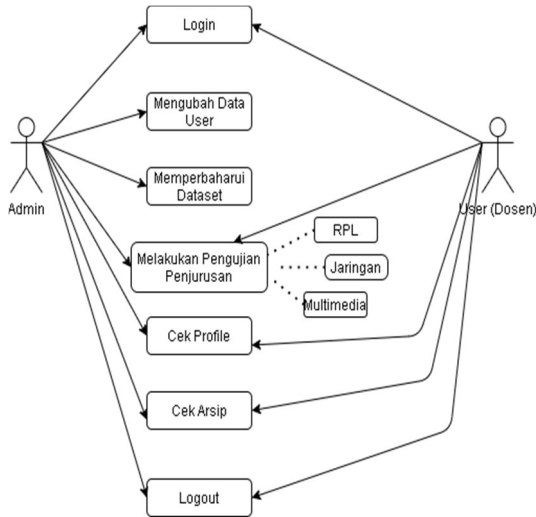
Algoritma Naive Bayes sebagai algoritma yang akan digunakan oleh sistem yang akan dibuat.

4. Hasil dan Pembahasan

Pada tahap ini peneliti melakukan proses klasifikasi mulai dari analisis masalah, analisis kebutuhan sistem, dataset, pembuatan UML dan pengujian akurasi sistem dengan *rapid*.

1. Usecase

Use case diagram yang digunakan untuk menjelaskan interaksi dan kegiatan antara pengguna/aktor dan sistem pada sistem informasi. *Use case diagram* untuk sistem informasi penentuan peminatan program studi teknik informatika memiliki 2 (dua) aktor yaitu user yang merupakan dosen pembimbing akademik dan admin yang merupakan karyawan sekretariat, seperti yang digambarkan oleh gambar 4.



Gambar 3 Usecase

3. Implementasi Antarmuka

Halaman Login

Halaman *login* adalah jendela yang digunakan oleh *user* untuk masuk ke dalam sistem, *user* yang telah memiliki hak akses akan dapat masuk dengan mudah ke dalam sistem dengan menggunakan *username* dan *password* yang telah diberikan oleh admin. Jendela *login* adalah yang utama dalam sebuah sistem karena sebagai salah satu keamanan hak akses dalam sistem dan merupakan halaman pertama ketika mengakses system, seperti yang terlihat pada Gambar 5.



Gambar 4 Halaman Login

Halaman Home

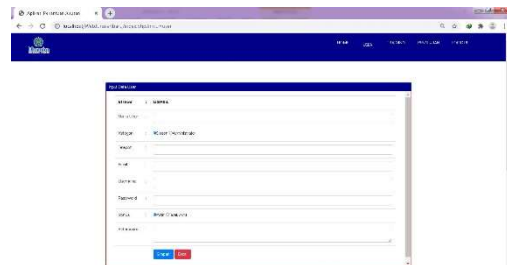
halaman utama yang ditampilkan setelah melalui proses *login*, oleh karena itu halaman *home* memiliki beberapa fungsi antara lain menampilkan latar dari sistem tersebut, karena sistem yang peneliti kembangkan adalah sistem penentuan jurusan peminatan Teknik Informatika di UHAMKA maka peneliti mengisi halaman *Home* dengan latar Teknik UHAMKA khususnya prodi Teknik Informatika, seperti yang terlihat pada Gambar 6.



Gambar 5 Halaman Home

Halaman User

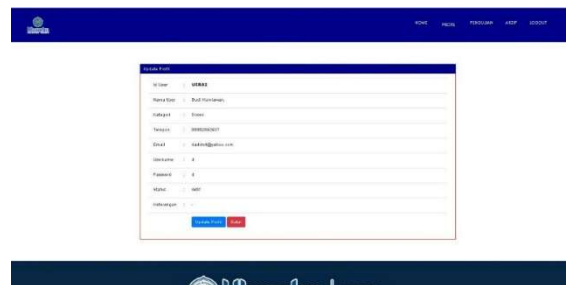
Halaman *User* adalah halaman yang digunakan oleh admin untuk memberikan hak akses kepada *user* atau admin lain, sehingga sistem dapat digunakan oleh banyak user yang telah ditentukan yaitu dosen akademik dari mahasiswa teknik informatika, seperti yang terlihat pada Gambar 7.



Gambar 6 Halaman User

Halaman Profil

Halaman Profil hak akses diberikan kepada user, user dapat mengganti profil yang diisi sebelumnya oleh admin dengan menggunakan tombol *update*, profil yang dapat diganti antara lain, nama, *email*, telepon, *username* dan *password*. Untuk status aktif atau tidak aktif hanya dapat diganti oleh admin pada halaman user seperti yang terlihat pada Gambar 8.

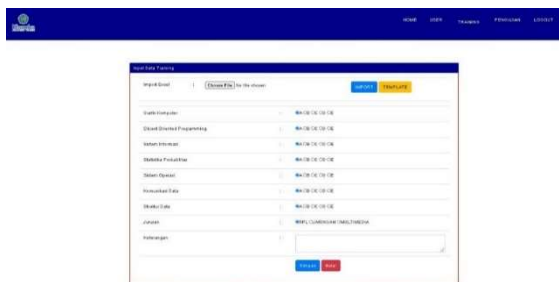


Gambar 7 Halaman Profile

Halaman Training

Halaman *Training* adalah halaman yang berisikan dataset yang telah dimasukan sebelumnya oleh admin sebagai *basic* data dari pengambilan keputusan rekomendasi peminatan

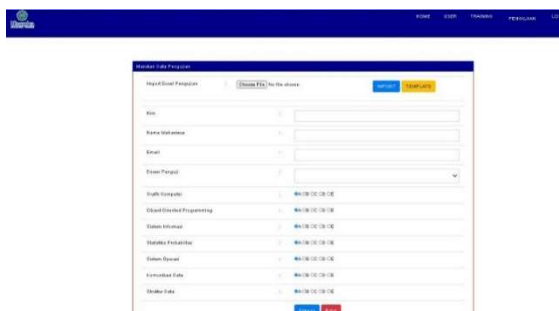
teknik informatika mahasiswa UHAMKA berdasarkan system, seperti yang terlihat pada Gambar 9



Gambar 8 Halaman Training

Halaman Pengujian

Pada halaman pengujian *user* dan admin memiliki hak akses pada halaman ini, *user* yang merupakan dosen pembimbing akademik mahasiswa dapat melakukan pengujian terhadap anak didiknya, melalui pemilihan kolom atau dengan menggunakan *import excel* yang merupakan data berkelompok sehingga *user* tidak perlu mengisi data satu persatu, seperti yang digambarkan oleh Gambar 10.



Gambar 9 Halaman Pengujian

Halaman Arsip

Pada halaman arsip hak akses diberikan kepada *user* untuk melihat data-data mahasiswa yang telah mereka uji, data-data pada halaman ini dibedakan menjadi RPL Jaringan dan Multimedia sehingga terlihat rapi dan tidak berantakan, seperti yang terlihat pada Gambar 11.



Gambar 10 Halaman Arsip

3. Pengujian Akurasi Dataset Naive Bayes

Pada uji akurasi menggunakan *rapidminer* dimana peneliti hanya tinggal memasukan data *excel* dan operator yang tepat untuk digunakan sebagai penilai akurasi dari Dataset yang akan digunakan oleh sistem yang dibuat oleh peneliti. Disini peneliti mendapatkan akurasi sebesar 61.30 % seperti yang terlihat pada gambar 12.

accuracy 61.30%

	true MULTIMEDIA	true JARINGAN	true RPL	class precision
pred. MULTIMEDIA	7	0	1	87.50%
pred. JARINGAN	33	110	31	63.22%
pred. RPL	8	16	24	50.00%
class recall	14.58%	97.30%	42.86%	

Gambar 11 Rapid Akurasi

4. Kesimpulan dan Saran

1. Kesimpulan

Berdasarkan Algoritma *Naive Bayes* dapat digunakan untuk memutuskan rekomendasi peminatan mahasiswa Teknik Informatika UHAMKA dengan nilai akurasi sebesar 61.30%

2. Saran

- A. Gunakan algoritma selain *Naive Bayes* untuk menemukan hasil yang berbeda dari yang telah peneliti buat. Dalam rangka perbandingan efektifitas.
- B. Gunakan *dataset* lebih banyak dari yang peneliti lakukan, karena akan mempengaruhi tingkat akurasi dari sistem tersebut.
- C. Tambahkan data mata kuliah lain, seperti mata kuliah wajib atau tambahan untuk menguji akurasi dari sistem

KEPUSTAKAAN

[1] Alfa Saleh. (2017 (Yosepta Dwi Anggih, 2017))."Klasifikasi Metode Naive Bayes Dalam Data Mining Untuk Menentukan Konsentrasi Siswa (Studi Kasus di MAS PAB 2 Medan)" KeTIK. ISBN : 979-458-766-4.

[2] Annur Haditsah. (2018). Klasifikasi Masyarakat Miskin Menggunakan Metode Naive Bayes. *ILKOM Jurnal Ilmiah*.

[3] Hj Nurmi.(2014). Membangun Website Sistem Informasi Dinas Pariwisata. *Jurnal Edik Informatika*.

[4] Rahman M Fadly., M.Ihham Darmawidjadja .,& Dion Alamsah. Klasifikasi Untuk Diagnosa Diabetes Menggunakan Metode Bayesian Neural Network (RBNN). *JURNAL INFORMATIKA*

[5] Syarli, S., & Muin, A. A. (2018). Metode Naive Bayes Untuk Prediksi Kelulusan (Studi Kasus: Data Mahasiswa Baru Perguruan Tinggi). *Jurnal Ilmu Komputer*, 2(1). Retrieved from <https://fikom-unasman.ac.id/ejournal/index.php/jikom/article/view/26/17>

- [6] BAN-PT. (2019). Naskah Akademik Instrumen Akreditasi Perguruan Tinggi 3.0. Jurnal BAN-PT
- [7] Widayu, H., Nasution, S. D., Silalahi, N., & Mesran, M. (2017). DATA MINING UNTUK MEMPREDIKSI JENIS TRANSAKSI NASABAH PADA KOPERASI SIMPAN PINJAM DENGAN ALGORITMA C4.5. JURNAL
- [8] Mardi Yuli. (2017). Data Mining : Klasifikasi Menggunakan Algoritma C4.5. *Jurnal Edik Informatika*.
- [9] Buntoro, G. A. (2016). Analisis Sentimen Hatespeech Pada Twitter Dengan Metode Naïve Bayes Classifier Dan Support Vector Machine. *Jurnal Dinamika Informatika*, 3(1), 56. <https://doi.org/https://doi.org/10.3929/ethz-b-000238666>