

## Penerapan Naïve Bayes Classifier dengan Gaussian Function Untuk Menentukan Kelompok UKT

Reza Ade Putra<sup>1)</sup>

<sup>1)</sup>Program Studi Sistem Informasi, UIN Raden Fatah Palembang  
Jl. Prof. K.H. Zainal Abidin Fikry KM 3.5 Palembang 30126  
Email : [rezaadepatra\\_uin@radenfatah.ac.id](mailto:rezaadepatra_uin@radenfatah.ac.id)<sup>1)</sup>

### Abstract

Uang Kuliah Tunggal hereinafter abbreviated as UKT is part of a single tuition fees incurred by each student in each department or study program for diploma and degree courses. UKT is the amount of fees to be paid by the student in each semester. Basically, the purpose of UKT is to charge tuition fees according to income and family circumstances students. However, there is a problem regarding the classification UKT improperly. It is caused by several factors, including determining UKT groups still use manual method, as well as the substance of subjectivity in the determination of a new student UKT groups. Based on these problems, we need a decision support system that can help in determining the UKT group of new students. In applying UKT, Cot Kala IAIN Zawiyah Langsa split into 3 (three) categories UKT group. Naïve Bayes classifier methods is used to classify data into three UKT groups. Research results show that the results of validation testing of NBC classification model with a 3-fold cross validation generates an average accuracy of 86.67%. so that it can be concluded that the level of effectiveness of the UKT classification model with the NBC method is included in the fairly good category.

**Keywords :** UKT Groups, Naive Bayes Classifier, K-fold cross validation

### Abstrak

Uang Kuliah Tunggal yang selanjutnya disingkat UKT merupakan sebagian dari biaya kuliah tunggal yang ditanggung oleh setiap mahasiswa pada setiap jurusan atau program studi untuk program diploma dan program sarjana. UKT merupakan besaran biaya yang harus dibayarkan oleh mahasiswa pada setiap semester. Pada dasarnya, tujuan diberlakukannya UKT yaitu untuk membebankan biaya kuliah sesuai dengan penghasilan dan kondisi keluarga mahasiswa yang bersangkutan. Akan tetapi, terjadi permasalahan tentang penggolongan UKT yang tidak tepat. Ini disebabkan oleh beberapa faktor, diantaranya dalam menentukan kelompok UKT masih menggunakan cara manual, serta adanya unsur subjektivitas dalam penentuan kelompok UKT mahasiswa baru. Berdasarkan permasalahan tersebut, dibutuhkan suatu sistem pendukung keputusan yang dapat membantu dalam menentukan kelompok UKT mahasiswa baru. Dalam menerapkan Uang Kuliah Tunggal, IAIN Zawiyah Cot Kala Langsa membagi kedalam 3 (Tiga) kategori kelompok UKT. Metode Naïve Bayes Classifier digunakan untuk mengklasifikasikan data menjadi tiga kelompok UKT. Hasil penelitian menunjukkan bahwa hasil validasi pengujian model klasifikasi NBC dengan 3- fold cross validation menghasilkan rata-rata akurasi sebesar 86.67%, sehingga dapat disimpulkan bahwa tingkat efektivitas model klasifikasi UKT dengan metode NBC termasuk pada kategori cukup baik..

**Kata kunci :** Kelompok UKT, Naive Bayes Classifier, K-fold cross validation

## 1. Pendahuluan

Uang Kuliah Tunggal yang selanjutnya disingkat UKT merupakan sebagian dari biaya kuliah tunggal yang ditanggung oleh setiap mahasiswa pada setiap jurusan atau program studi untuk program diploma dan program sarjana. UKT merupakan besaran biaya yang harus dibayarkan oleh mahasiswa pada setiap semester. Dalam Peraturan Menteri Agama No. 30 tahun 2014 dan Keputusan Menteri Agama No. 124 Tahun 2015, UKT ditentukan berdasarkan kelompok ekonomi masyarakat yang dibagi dalam 3 (tiga) kelompok dari yang terendah hingga yang tertinggi yaitu UKT 1, UKT 2, dan UKT 3. Dalam Pasal 2 Ayat (2) poin (a) Peraturan Menteri Agama No. 30 Tahun 2014 dan Diktum Ketiga poin (a) Keputusan Menteri Agama No. 124 Tahun 2015 bahwa kelompok UKT 1 diperuntukan bagi mahasiswa miskin di luar penerima beasiswa pendidikan mahasiswa miskin dan berprestasi (Bidikmisi) dan paling sedikit diberikan sebanyak 5 (lima) % dari jumlah mahasiswa yang diterima.

Dalam menerapkan Uang Kuliah Tunggal, IAIN Zawiyah Cot Kala Langsa membagi kedalam 3 (Tiga) kategori kelompok UKT. Dalam menentukan kelompok UKT mahasiswa baru, pihak IAIN Zawiyah Cot Kala Langsa melakukan wawancara kepada setiap mahasiswa baru dengan melihat beberapa kriteria yang menggambarkan kemampuan ekonomi orang tua. Pada penerapannya, proses penentuan kelompok UKT mahasiswa baru di IAIN Zawiyah Cot Kala sangat bergantung pada penilaian masing-masing pewawancara. Hal ini menyebabkan acuan penilaian wawancara dalam menentukan kelompok UKT mahasiswa baru menjadi sangat beragam dan tidak konsisten. Oleh sebab itu, pihak IAIN Zawiyah Cot Kala Langsa membutuhkan suatu cara yang efektif sebagai rekomendasi untuk membantu menganalisa data yang diperoleh dari kondisi sosial ekonomi orang tua setiap mahasiswa dalam menentukan kelompok UKT mahasiswa baru.

Dalam penelitian ini digunakan algoritma Naïve Bayes Classifier (NBC) dengan *Gaussian Function* dalam melakukan klasifikasi data. Dalam proses klasifikasi ini, kriteria yang digunakan untuk mengelompokkan Uang Kuliah Tunggal dengan metode NBC adalah penghasilan orang tua, rasio penghasilan terhadap jumlah tanggungan orang tua, dan tagihan rekening listrik. Ketiga kriteria tersebut merupakan parameter utama yang dapat mengidentifikasi faktor keuangan yang memiliki pengaruh besar terhadap tingkat kemampuan ekonomi orang tua.

Berdasarkan latar belakang tersebut, diharapkan dengan menerapkan metode NBC dapat digunakan untuk merancang sebuah sistem pendukung keputusan penentuan UKT mahasiswa baru yang dapat memudahkan proses pengelompokan UKT dan menjadi rekomendasi bagi pihak perguruan tinggi dalam menentukan kelompok UKT mahasiswa baru.

### A. Klasifikasi

Klasifikasi adalah proses untuk menemukan model atau fungsi yang menjelaskan atau membedakan konsep atau kelas data dengan tujuan untuk memperkirakan kelas yang tidak diketahui dari suatu objek. Dengan kata lain, klasifikasi merupakan penempatan objek-objek ke salah satu dari beberapa kategori yang telah ditetapkan sebelumnya. Ada dua langkah dalam proses klasifikasi, yaitu (Han dan Kamber, 2006) :

- Membangun suatu model dengan menganalisis data *training*.
- Klasifikasi, dimana model yang telah dihasilkan digunakan untuk melakukan klasifikasi terhadap data yang belum diketahui label kelasnya.

Tahap pertama disebut juga sebagai tahap pembelajaran (*learning*). Sebuah algoritma klasifikasi akan membangun sebuah model klasifikasi dengan cara menganalisis data *training*.

### B. Teorema Bayes

Pendekatan naive bayes merupakan pendekatan yang mengacu pada teorema bayes yang merupakan prinsip peluang statistika untuk mengkombinasikan pengetahuan sebelumnya dengan pengetahuan baru. Prinsip ini kemudian digunakan untuk memecahkan masalah klasifikasi (Tan dkk, 2006). Prediksi Bayes didasarkan pada teorema bayes dengan bentuk umum seperti Persamaan (1).

$$P(H|E) = \frac{P(E|H) \times P(H)}{P(E)} \quad \dots (1)$$

Hubungan hipotesis dan bukti dengan klasifikasi adalah, hipotesis dalam teorema bayes merupakan label kelas yang menjadi atribut target, sedangkan bukti merupakan himpunan atribut yang menjadi masukan dalam model klasifikasi.

### C. Naïve Bayes Classifier

Metode klasifikasi yang digunakan pada penelitian ini adalah Naive Bayes Classifier (NBC), dimana NBC menggunakan teori probabilitas sebagai dasar teori. Salah satu kelebihan Naive Bayes Classifier adalah sederhana tetapi memiliki akurasi yang tinggi (Rish, 2001).

Pada penelitian ini, atribut yang digunakan bertipe numerik, dimana untuk atribut dengan tipe numerik ada perlakuan khusus sebelum dimasukkan dalam Naïve Bayes. Jika atribut bernilai numerik (kontinyu), maka perlu dilakukan sedikit lebih banyak pekerjaan, tapi perhitungannya cukup sederhana. Sebuah atribut bernilai numerik (kontinyu) biasanya diasumsikan memiliki distribusi gaussian dengan dua parameter, yaitu Mean ( $\mu$ ), dan Standar Deviasi ( $\sigma$ ) (Han dan Kamber, 2006). Untuk perhitungan atribut dengan data numerik, pertama dihitung rata-rata dari data *training* dengan rumus yang dinotasikan dengan Persamaan (2).

$$\mu = \frac{1}{n} \sum_{i=1}^n x_i \quad \dots (2)$$

Setelah mendapatkan rata-rata dari masing masing atribut pada *data training*, selanjutnya dilakukan perhitungan standar deviasi dari masing-masing atribut dengan rumus yang dinotasikan dengan Persamaan (3).

$$\sigma = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \mu)^2} \quad \dots(3)$$

Jika diasumsikan suatu kelas dengan  $C_i$ , maka dapat diketahui Probabilitas *Class Prior*  $P(C_i)$  dengan  $C_i$  adalah label kelas ke- $i$  dan  $i = 1, 2, \dots, m$ , sehingga menjadi seperti Persamaan (4) (Murphy,2006).

$$P(C_i) = \frac{N_i}{N} \quad \dots(4)$$

Selanjutnya pada proses pengujian data, nilai  $\mu$  dan  $\sigma$  dari masing-masing atribut tersebut akan dimasukkan ke rumus Distribusi Gaussian yang dinotasikan dengan Persamaan (5)(Han dan Kamber, 2006).

$$P(x_k|C_i) = g(x_k, \mu_{C_i}, \sigma_{C_i}) = \frac{1}{\sqrt{2\pi}\sigma_{C_i}} e^{-\frac{(x_k - \mu_{C_i})^2}{2\sigma_{C_i}^2}} \quad (5)$$

dengan  $k = 1, 2, \dots, m$ ; dan  $j = 1, 2, \dots, n$

Pada metode NBC, setiap objek direpresentasikan dalam pasangan atribut  $(x_1, x_2, \dots, x_n)$  di mana  $x_1$  adalah atribut pertama,  $x_2$  atribut kedua dan seterusnya. Sedangkan  $C$  adalah himpunan kategori kelas, dimana dalam penelitian ini terdiri dari UKT 1, UKT 2, dan UKT 3. Pada klasifikasi, pendekatan NBC akan menghasilkan label kategori yang paling tinggi probabilitasnya ( $C_{MAP}$ ) dengan masukan atribut  $(x_1, x_2, \dots, x_n)$ , seperti pada Persamaan (6).

$$C_{MAP} = \operatorname{argmax}_{C_i \in C} P(C_i | x_1, x_2, \dots, x_n) \quad \dots(6)$$

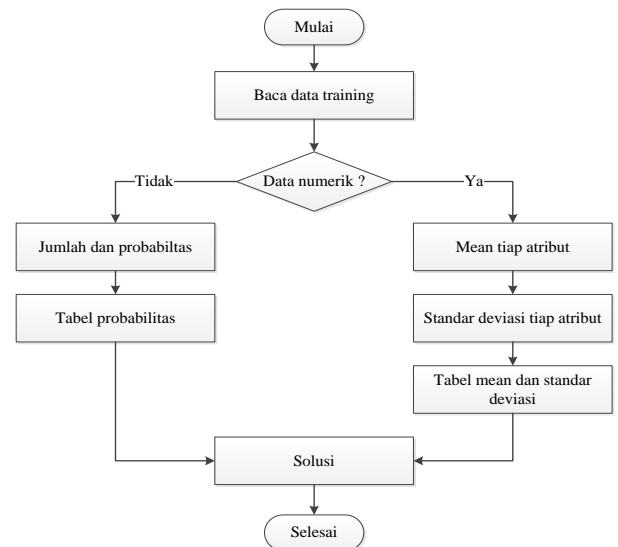
Dari Persamaan (6) jika disubstitusikan ke Persamaan (1) maka menjadi Persamaan (7).

$$C_{MAP} = \operatorname{argmax}_{C_i \in C} \frac{P(x_1, x_2, \dots, x_n | C_i) \times P(C_i)}{P(x_1, x_2, \dots, x_n)} \quad \dots(7)$$

$P(x_1, x_2, \dots, x_n)$  bernilai konstan untuk semua kelas ( $C_i$ ), hanya  $P(x_1, x_2, \dots, x_n | C_i) \times P(C_i)$  yang perlu dimaksimalkan, sehingga persamaan (7) disederhanakan menjadi Persamaan (8).

$$C_{MAP} = \operatorname{argmax}_{C_i \in C} P(x_1, x_2, \dots, x_n | C_i) \times P(C_i) \quad \dots(8)$$

Secara umum diagram alir pada proses klasifikasi NBC ditunjukkan pada Gambar 1.



Gambar 1. Flowchart NBC

#### D. Uang Kuliah Tunggal

Tujuan diberlakukannya UKT adalah untuk meringankan beban mahasiswa terhadap pembiayaan pendidikan, sehingga pemerintah melalui Menteri Agama (Menag) pada tanggal 18 September 2014 telah mengeluarkan ketetapan mengenai besarnya Biaya Kuliah Tunggal (BKT) dan Uang Kuliah Tunggal (UKT) pada Perguruan Tinggi Agama Negeri (PTAN) di lingkungan Kementerian Agama (Kemenag). Ketentuan itu tertuang dalam Keputusan Menteri Agama No. 30 Tahun 2014 yang selanjutnya mengalami pembaharuan berdasarkan Keputusan Menteri Agama No. 124 Tahun 2015. Dalam Keputusan Menteri Agama tersebut disebutkan bahwa :

- Biaya Kuliah Tunggal yang selanjutnya disingkat BKT adalah keseluruhan biaya operasional per mahasiswa per semester pada jurusan/ program studi tertentu di perguruan tinggi keagamaan negeri untuk program diploma dan program sarjana.
- Uang kuliah Tunggal yang selanjutnya disingkat UKT adalah sebagian biaya kuliah tunggal yang ditanggung oleh setiap mahasiswa pada setiap jurusan/ program studi untuk program diploma dan program sarjana.
- UKT terdiri dari 3 (Tiga) kelompok dari yang terendah hingga tertinggi, yang ditentukan berdasarkan kelompok kemampuan ekonomi masyarakat di lingkungan perguruan tinggi agama negeri berada.
- Uang kuliah tunggal kelompok I diperuntukkan bagi mahasiswa miskin di luar penerima beasiswa pendidikan mahasiswa miskin dan berprestasi (Bidikmisi), dan paling sedikit diberikan sebanyak 5 (Lima) persen dari jumlah mahasiswa yang diterima.
- Uang kuliah tunggal kelompok II diperuntukkan bagi mahasiswa yang memiliki kemampuan ekonomi menengah.
- Uang kuliah tunggal kelompok III diperuntukkan bagi mahasiswa yang memiliki kemampuan ekonomi tinggi.

- g. Penetapan mahasiswa berdasarkan kelompok UKT sebagaimana dimaksud dalam Pasal 2 Keputusan Menteri Agama No. 30 Tahun 2014 dilakukan oleh Rektor atau Ketua Perguruan Tinggi Keagamaan Negeri.
- h. Perguruan Tinggi Keagamaan Negeri tidak boleh memungut uang pangkal dan pungutan lain selain UKT dari mahasiswa baru program Sarjana (S1) dan program diploma mulai Tahun Akademik 2014-2015.

Berdasarkan penjelasan tersebut di atas, maka dapat disimpulkan bahwa :

- a. Kriteria-kriteria yang digunakan untuk melakukan penilaian dalam penentuan kelompok UKT ditetapkan oleh masing-masing pimpinan Perguruan Tinggi Keagamaan Negeri (PTAN).
- b. UKT secara umum merupakan suatu kebijakan untuk menepis anggapan mahal biaya kuliah di Perguruan Tinggi Keagamaan Negeri karena dengan adanya UKT dapat meringankan beban mahasiswa terhadap pembiayaan selama menempuh studi.
- c. UKT juga memberikan kepastian kepada mahasiswa, orang tua atau para pihak yang membiayai mahasiswa untuk memperkirakan seberapa besar dana yang harus disiapkan sampai lulus kuliah.

**E. Holdout**

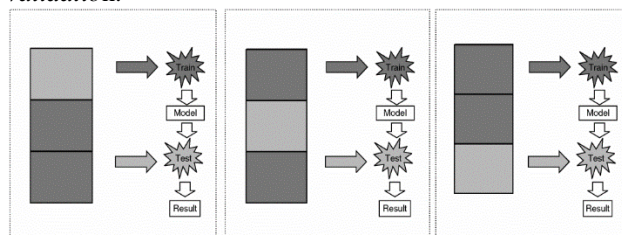
Penelitian ini menggunakan metode *Holdout* dan *K-Fold Cross Validation* dalam proses evaluasi. Metode *Holdout* biasanya menggunakan sebanyak setengah atau dua pertiga dari data keseluruhan untuk keperluan proses *training*, sedangkan sisanya digunakan untuk keperluan *testing* (Witten dkk, 2011). Pembagian data yang berbeda terhadap suatu data tentunya akan memberikan perkiraan hasil yang berbeda pula. Sebuah pengulangan proses dengan kumpulan data *training* dan data uji yang berbeda dipilih secara acak, integrasi hasil error akan digunakan sebagai parameter standar yang akan meningkatkan estimasi dari model yang dibangun (Kantardzic, 2003).

**F. K-fold cross validation**

*Cross validation* adalah metode statistik untuk mengevaluasi algoritma pembelajaran dengan membagi data menjadi dua set, dataset pertama digunakan untuk belajar atau melatih model dan dataset yang lain digunakan untuk memvalidasi model (Madhukar dkk, 2012).

Untuk mengestimasi tingkat kesalahan dari model yang telah dikembangkan dapat menggunakan *K-fold cross validation* yang merupakan salah satu teknik metode *cross validation*. Penggunaan *K-fold cross validation* bertujuan untuk menghilangkan bias pada data. Pada metode ini, data dibagi menjadi k subset atau 'fold', yaitu  $S_1, S_2, \dots, S_k$  dengan ukuran setiap subset kira-kira sama. Pelatihan dan pengujian dilakukan sebanyak k kali. Pada percobaan pertama, subset  $S_1$  diperlakukan sebagai data uji dan subset lainnya diperlakukan sebagai data *training*. Pada percobaan

kedua subset  $S_1, S_3, \dots, S_k$  menjadi data *training* dan  $S_2$  menjadi data uji, dan seterusnya (Bramer, 2007). Gambar 2 menunjukkan contoh prosedur *3-fold cross validation*.



**Gambar 2.** Prosedur 3-fold cross validation

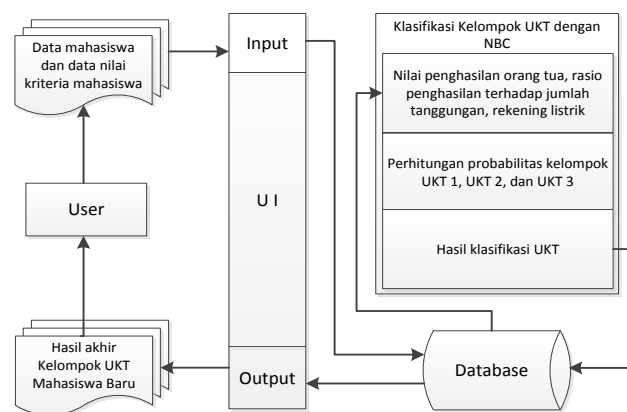
Untuk pengukuran kinerja *classifier* dengan cara menggunakan model klasifikasi untuk memprediksi data uji. Jika data uji mengandung N data dan C adalah data yang benar diklasifikasikan, maka akurasi prediksi P dari *classifier* untuk data uji ditunjukkan pada Persamaan (9).  $P = C / N \dots(9)$

dimana P adalah nilai akurasi, N menunjukkan banyak data uji, dan C menunjukkan banyak data benar hasil klasifikasi.

**2. Pembahasan**

**A. Arsitektur Sistem**

Metode NBC digunakan untuk melakukan proses klasifikasi data mahasiswa menjadi tiga kelompok UKT. Atribut yang digunakan dalam proses klasifikasi meliputi penghasilan orang tua (Ph), Rasio penghasilan terhadap jumlah tanggungan orang tua (Rs), dan Tagihan Rekening Listrik (RL). Alasan pemilihan atribut penghasilan orang tua, rasio penghasilan terhadap jumlah tanggungan orang tua, maupun tagihan rekening listrik didasarkan bahwa atribut ini merupakan parameter utama yang dapat mengidentifikasi faktor keuangan yang memiliki pengaruh besar terhadap tingkat kemampuan ekonomi orang tua. Output yang dihasilkan pada model NBC ini adalah nilai probabilitas kelas tertinggi dan hasil klasifikasi kelompok UKT. Gambaran sistem yang dibangun pada penelitian ini ditunjukkan pada Gambar 3.



**Gambar 3.** Arsitektur Sistem Klasifikasi UKT



**B. Model Klasifikasi UKT dengan NBC**

Proses ini bertujuan untuk membangun model klasifikasi berdasarkan data *training* yang telah diinput ke dalam sistem. Selanjutnya admin memasukkan sebanyak 165 data *training* dengan rincian 49 sampel data mahasiswa UKT 1, 59 sampel data mahasiswa UKT 2, dan 57 sampel data UKT 3. Berdasarkan data *training*, sistem melakukan proses perhitungan nilai *class prior*, mean dan standar deviasi masing-masing atribut klasifikasi pada setiap kelas. Selanjutnya nilai yang dihasilkan tersebut digunakan sebagai model dalam proses perhitungan klasifikasi UKT mahasiswa baru. Perhitungan nilai mean masing-masing atribut independen terhadap kelas dilakukan dengan menggunakan Persamaan (2). Berdasarkan data *training* dan atribut klasifikasi, didapatkan nilai rata-rata yang tersaji pada Tabel 1.

**Tabel 1.** Nilai mean atribut independen terhadap kelas

Kelompok	Nilai rata-rata ( $\mu$ )		
	Ph	Rs	RL
UKT 1	604081.633	184932	48492.204
UKT 2	1342448.305	451732.119	63983.864
UKT 3	2590857.895	873644.491	121251.018

Dari nilai mean yang didapatkan, selanjutnya dihitung nilai standar deviasi masing-masing atribut independen terhadap kelas menggunakan Persamaan (3), sehingga didapatkan nilai standar deviasi yang ditunjukkan pada Tabel 2.

**Tabel 2.** Nilai standar deviasi atribut independen terhadap kelas

Kelompok	Nilai Standar Deviasi ( $\sigma$ )		
	Ph	R	RL
UKT 1	236254.792	126640.095	29718.276
UKT 2	350550.679	237819.504	32967.841
UKT 3	496694.779	544503.587	83451.729

Perhitungan nilai *class prior* menggunakan Persamaan (4). Berdasarkan jumlah data *training* pada setiap kelas dan jumlah data *training* keseluruhan, didapatkan nilai *class prior* yang ditunjukkan pada Tabel 3.

**Tabel 3.** Nilai *class prior*

Kelompok	Class Prior
UKT 1	0.297
UKT 2	0.3576
UKT 3	0.3455

**C. Pengujian Sistem Klasifikasi UKT dengan NBC**

Pengujian sistem menggunakan 51 data uji, dengan komposisi 17 data untuk masing-masing kelompok UKT. Atribut klasifikasi yang digunakan yaitu penghasilan orang tua, rasio penghasilan terhadap jumlah tanggungan, dan tagihan rekening listrik. Pengujian dilakukan menggunakan model klasifikasi yang terbentuk dari data *training*.

Proses perhitungan dimulai dari memasukkan nilai atribut klasifikasi yang diambil dari tabel

nilai\_kriteria\_mhs. Berdasarkan nilai atribut klasifikasi tersebut, selanjutnya dihitung nilai probabilitas masing-masing atribut independen terhadap kelas. Keluaran dari proses klasifikasi kelompok UKT adalah pengelompokan klasifikasi kelompok UKT mahasiswa yang terdiri dari UKT 1, UKT 2, dan UKT 3. Hasil klasifikasi ditunjukkan pada Gambar 4.

Jumlah mahasiswa UKT 1 16 Orang (31.37 %)		Jumlah mahasiswa UKT 2 16 Orang (31.37 %)		Jumlah mahasiswa UKT 3 19 Orang (37.25 %)		Total mahasiswa 51 Orang (100 %)	
Filter: Masukkan Nama							
ID Peserta	Nama Mahasiswa	Nilai Probabilitas Tertinggi	Klasifikasi	Perhitungan NBC			
151062003	ABDUL AZIZ	9.35433592925E-19	2	<a href="#">Lihat Perhitungan</a>			
151062007	ACHMAD IRHAM TAMBI	2.017171602500E-18	2	<a href="#">Lihat Perhitungan</a>			
151062011	AGITA PUTRI AULYA	1.568672647772E-18	2	<a href="#">Lihat Perhitungan</a>			
151062019	ALDY HIMATULLAH AR	9.0673524188209E-19	2	<a href="#">Lihat Perhitungan</a>			
151062021	AMALIA PUSPA DEWI	5.218171042579E-18	2	<a href="#">Lihat Perhitungan</a>			
151062022	AMINAH	2.845793828802E-18	1	<a href="#">Lihat Perhitungan</a>			
151062023	ANANDA TANJUNG	2.227684235988E-19	3	<a href="#">Lihat Perhitungan</a>			
151062029	ANISA SEPTIANA SIREGAR	1.291543016903E-18	2	<a href="#">Lihat Perhitungan</a>			
151062030	ANITA SARI	5.011272374475E-18	2	<a href="#">Lihat Perhitungan</a>			
151062031	ANNISA BELA	6.808379689286E-18	1	<a href="#">Lihat Perhitungan</a>			
151062033	ARJUN WAHYU	1.268588274880E-18	2	<a href="#">Lihat Perhitungan</a>			

**Gambar 4.** Hasil Klasifikasi UKT dengan model NBC

Dari hasil klasifikasi UKT mahasiswa dengan menggunakan model NBC yang terbentuk, diperoleh mahasiswa kelompok UKT 1 sebanyak 16 orang (31.37%), mahasiswa kelompok UKT 2 sebanyak 16 orang (31.37%), dan mahasiswa kelompok UKT 3 sebanyak 19 orang (37.25%).

**D. Pengujian Akurasi model NBC menggunakan K-fold cross validation**

Pengujian *K-Fold cross validation* menggunakan tiga atribut dengan *3-fold cross validation*. Percobaan dilakukan dengan menggunakan 165 data menjadi satu dataset, kemudian membagi dataset tersebut menjadi tiga buah partisi (subset) secara random yaitu  $S_1$ ,  $S_2$ , dan  $S_3$ . Setiap subset berisi 55 data. Eksperimen dikerjakan sebanyak tiga kali untuk mengukur keakuratan data, masing-masing eksperimen menggunakan data subset ke-K sebagai data uji dan memanfaatkan subset lainnya sebagai data *training*. Pengukuran kinerja *classifier* dilakukan dengan menggunakan Persamaan (9).

**E. Hasil pengujian subset  $S_1$**

Tabel 4 menunjukkan hasil pengujian menggunakan subset  $S_1$  sebagai data uji, sedangkan subset  $S_2$  dan  $S_3$  sebagai data *training*. Jumlah data *training* yang diperoleh dari subset  $S_2$  dan  $S_3$  yaitu sebanyak 110 data, dengan rincian 32 data *training* UKT 1, 40 data *training* UKT 2, dan 38 data *training* UKT 3. Pada eksperimen ini, tingkat akurasi hasil klasifikasi tertinggi berada pada UKT 3 sebesar 94.74% dan akurasi terendah 70.59% untuk UKT 1. Akurasi keseluruhan untuk proses identifikasi pada eksperimen ke-1 ini yaitu 81.82%.

**Tabel 4.** Hasil pengujian subset  $S_1$

Kelompok	Prediksi Benar	Prediksi Salah	Jumlah data	Akurasi Klasifikasi (%)
UKT 1	12	5	17	70.59%
UKT 2	15	4	19	78.95%

UKT 3	18	1	19	94.74%
Jumlah	45	10	55	

F. Hasil pengujian subset  $S_2$

Tabel 5 menunjukkan hasil pengujian menggunakan subset  $S_2$  sebagai data uji, sedangkan subset  $S_1$  dan  $S_3$  sebagai data *training*. Jumlah data *training* yang diperoleh dari subset  $S_1$  dan  $S_3$  yaitu sebanyak 110 data, dengan rincian 33 data *training* UKT 1, 39 data *training* UKT 2, dan 38 data *training* UKT 3. Pada eksperimen ini, tingkat akurasi hasil klasifikasi tertinggi berada pada UKT 2 sebesar 90% dan akurasi terendah 87.50% untuk UKT 1. Akurasi keseluruhan untuk proses identifikasi pada eksperimen ke-2 ini yaitu 87.27%.

Tabel 5. Hasil pengujian subset  $S_2$

Kelompok	Prediksi Benar	Prediksi Salah	Jumlah data	Akurasi Klasifikasi (%)
UKT 1	14	2	16	87.50%
UKT 2	18	2	20	90.00%
UKT 3	17	2	19	89.47%
Jumlah	48	7	55	

G. Hasil pengujian subset  $S_3$

Tabel 6 menunjukkan hasil pengujian menggunakan subset  $S_3$  sebagai data uji, sedangkan subset  $S_1$  dan  $S_2$  sebagai data *training*. Jumlah data *training* yang diperoleh dari subset  $S_1$  dan  $S_2$  yaitu sebanyak 110 data, dengan rincian 33 data *training* UKT 1, 39 data *training* UKT 2, dan 38 data *training* UKT 3. Pada eksperimen ini, tingkat akurasi hasil klasifikasi tertinggi berada pada UKT 1 sebesar 100% dan akurasi terendah 80% untuk UKT 2. Akurasi keseluruhan untuk proses identifikasi pada eksperimen ke-3 ini yaitu 90.91%.

Tabel 6. Hasil pengujian subset  $S_3$

Kelompok	Prediksi Benar	Prediksi Salah	Jumlah data	Akurasi Klasifikasi (%)
UKT 1	16	0	16	100%
UKT 2	16	4	20	80.00%
UKT 3	18	1	19	94.74%
Jumlah	50	5	55	

Pengujian dengan *3-fold cross validation* digunakan untuk uji coba dengan memperhitungkan tingkat variasi dan keakuratan data, sehingga eksperimen dikerjakan sebanyak 3 kali yang ditunjukkan pada Tabel 7. Hasil pengujian memperlihatkan bahwa tingkat akurasi klasifikasi tertinggi yaitu 90.91% terjadi pada eksperimen ke-3 yang menggunakan subset  $S_3$  untuk data uji dan subset  $S_1$  dan  $S_2$  sebagai data *training*, sedangkan eksperimen ke-1 menghasilkan akurasi klasifikasi terendah dengan tingkat akurasi 81.82%.

Tabel 7. Hasil pengujian 3-fold cross validation

Eksperimen ke-	Data <i>training</i>	Data uji	Akurasi klasifikasi (%)
1	$S_2, S_3$	$S_1$	81.82%
2	$S_1, S_3$	$S_2$	87.27%
3	$S_1, S_2$	$S_3$	90.91%
Rata-rata akurasi			86.67%

Hasil rata-rata akurasi untuk *3-fold cross validation* dihitung dengan menjumlahkan akurasi klasifikasi untuk setiap eksperimen dibagi dengan jumlah eksperimen yang dikerjakan. Rata-rata akurasi hasil pengujian dengan *3-fold cross validation* yaitu 86.67%.

3. Kesimpulan

Dari proses perancangan hingga pengujian yang telah dilakukan pada sistem pendukung keputusan dalam menentukan kelompok UKT mahasiswa ini, maka dapat diambil kesimpulan bahwa hasil validasi pengujian model klasifikasi NBC dengan *K-fold cross validation* menghasilkan rata-rata akurasi sebesar 86.67%, sehingga tingkat efektivitas model klasifikasi NBC dapat dikatakan cukup baik, sehingga sistem ini dapat digunakan sebagai rekomendasi dalam pengambilan keputusan mengenai penentuan kelompok UKT mahasiswa baru yang diterima di IAIN Zawiyah Cot Kala Langsa.

Daftar Pustaka

Bramer, M., 2007, *Principles of Data Mining*, Springer-Verlag, London.

Han, J., and Kamber, M., 2006, *Data Mining : Concepts and Techniques, Second Edition*, Morgan Kauffmann, San Fransisco.

Kantardzic, M., 2003, *Data Mining : Concept, Models, Methods, and Algorithms*, John Wiley and Sans.

Keputusan Menteri Agama Republik Indonesia Nomor 124 Tahun 2015. *Uang Kuliah Tunggal pada Perguruan Tinggi Keagamaan Negeri Di Kementerian Agama Tahun Akademik 2015-2016*. Kementerian Agama Republik Indonesia. 1 Juni 2015. Jakarta

Murphy, K.P., 2006, *Naïve Bayes Classifier*, Department of Computer Science, University of British Columbia.

Rish, I., 2001, *An Empirical Study Of The Naive Bayes Classifier*, IBM Research Report, Thomas J. Watson Research Center, Yorktown Heights, New York.

Tan, P.N., Stelnbach, M., and Kumar, V., 2006, *Introduction to Data Mining*, Pearson Education, Boston

Witten, I.H., Frank, E., and Hall, M.A., 2011, *Data Mining : Practical Machine Learning Tools and Techniques*, Elsevier.