

Implementasi Algoritma *Naïve Bayes Classifier* Dalam Menentukan Topik Tugas Akhir Mahasiswa Berbasis Web

T. Zakia Maulani¹, Zulfan Khairil Simbolon², Amirullah³

^{1,2,3} *Jurusan Teknologi Informasi dan Komputer Politeknik Negeri Lhokseumawe*

¹tengkuzakiamaulani97@gmail.com, ²zulfan69@gmail.com, ³amir@pnl.ac.id

Abstrak — Tugas akhir merupakan karya tulis ilmiah yang memuat hasil pengamatan dari suatu penelitian terhadap beberapa masalah yang terjadi dengan menggunakan kaidah/aturan maupun metode yang berkaitan dengan bidang ilmu tersebut. Mahasiswa tingkat akhir yang akan menghadapi Tugas Akhir (TA) harus memiliki topik yang akan diajukan dalam Seminar Proposal, di mana topik tersebut akan menjadi pokok bahasan dalam Tugas Akhir. Tidak sedikit mahasiswa yang mengalami kesulitan dalam memilih topik Tugas Akhir. Banyaknya mata kuliah yang telah ditempuh justru membuat mahasiswa semakin sulit menentukan pilihan pada satu bidang yang akan difokuskan sebagai Tugas Akhir. Untuk itu, penelitian ini mengusulkan dibangunnya sebuah sistem penentuan topik tugas akhir dengan melihat nilai probabilitas tertinggi pada suatu *class* dengan menggunakan metode *Naïve Bayes Classifier* yang bertujuan untuk memudahkan mahasiswa dalam menentukan kategori topik tugas akhir yang akan dipilih. Sistem yang dibuat ini menghasilkan sistem informasi berbasis web dengan menggunakan beberapa parameter yaitu nilai mata kuliah inti yang diambil dari semester 1 sampai 6 pada program studi Teknik Informatika Politeknik Negeri Lhokseumawe, data dosen pembimbing tahun ajaran 2018/2019 dan minat kategori atau *class* topik tugas akhir yang dikelompokkan menjadi 8 *class* yaitu Sistem Informasi, Sistem Pendukung Keputusan, Sistem Informasi Geografis, Sistem Pakar, Sistem Cerdas, E-Learning, Citra, dan Jaringan Komputer. Sistem ini menggunakan data mahasiswa Teknik Informatika tahun 2014 dan 2015 sebagai data latih dan data uji yang dibagi menjadi 76 data latih dan 20 data uji. Pada penelitian ini, berhasil dibangun sistem penentuan topik tugas akhir mahasiswa yang menghasilkan satu kategori *class* sebagai hasil rekomendasi yang dilihat berdasarkan nilai probabilitas tertinggi diantara 8 *class* yang ada. Selain itu, akan ditampilkan referensi jurnal-jurnal terkait kategori atau *class* yang telah direkomendasikan sebelumnya. Dari hasil pengujian sistem, salah satu mahasiswa memperoleh kategori Sistem Pendukung Keputusan di mana *class* tersebut memiliki nilai probabilitas tertinggi yaitu 0,982661230600192 di antara *class-class* lainnya. Penelitian yang telah dilakukan dengan metode *naïve bayes* ini diharapkan dapat membantu mahasiswa dalam menentukan kategori topik tugas akhir, serta dengan adanya referensi jurnal-jurnal yang diberikan oleh sistem dapat mempermudah mahasiswa dalam memilih judul untuk diajukan dalam proposal tugas akhir.

Kata kunci — *web, data latih, data uji, Naïve Bayes Classifier, topik tugas akhir*

Abstract — *The final project is a scientific paper that contains the results of observations from a study of several problems that occur using rules / rules and methods related to the field of science. Final level students who will face the Final Project (TA) must have a topic that will be submitted in the Proposal Seminar, where the topic will be the subject of the Final Project. Not a few students have difficulty choosing the Final Project topic. The number of courses taken has made it more difficult for students to make choices in one field that will be focused as a Final Project. For this reason, this research proposes to build a system for determining the final assignment topic by looking at the highest probability value in a class by using the Naïve Bayes Classifier method which aims to facilitate students in determining the topic category of the final assignment to be chosen. This system produces a web-based information system using several parameters, namely the value of core courses taken from semester 1 to 6 in the Lhokseumawe State Polytechnic Informatics Engineering study program, 2018/2019 academic year supervisor data and final task category or class interest which are grouped into 8 classes namely Information Systems, Decision Support Systems, Geographic Information Systems, Expert Systems, Intelligent Systems, E-Learning, Imagery, and Computer Networks. This system uses data from Informatics Engineering students in 2014 and 2015 as training data and test data which is divided into 76 training data and 20 test data. In this study, a system for determining the final assignment topic of students was successful which resulted in one class category as a recommendation based on the highest probability value among the 8 existing classes. In addition, references will be made to journals related to the categories or classes that have been previously recommended. From the results of system testing, one of the students obtained the Decision Support System category where the class has the highest probability value of 0.982661230600192 among other classes. The research that has been done with the naïve Bayes method is expected to help students in determining the topic of the final assignment category, and by reference to the journals provided by the system it can make it easier for students to choose the title to be submitted in the final project proposal.*

Keywords — *web, training data, test data, Naïve Bayes Classifier, final assignment topic*

I. PENDAHULUAN

Salah satu syarat kelulusan mahasiswa untuk mendapatkan gelar sesuai dengan program studinya adalah dengan mengerjakan Tugas Akhir (TA). Tugas Akhir (TA) adalah sebuah mata kuliah yang harus ditempuh oleh seorang mahasiswa menjelang akhir studinya. Mata kuliah ini berbentuk proyek mandiri yang dilakukan oleh mahasiswa di bawah bimbingan dosen pembimbing.[1]

Mahasiswa tingkat akhir yang akan menghadapi Tugas Akhir (TA) harus memiliki topik yang akan diajukan dalam Seminar Proposal, di mana topik tersebut akan menjadi pokok

bahasan dalam Tugas Akhir. Tidak sedikit mahasiswa yang mengalami kesulitan dalam memilih topik Tugas Akhir. Banyaknya mata kuliah yang telah ditempuh justru membuat mahasiswa semakin sulit menentukan pilihan pada satu bidang yang akan difokuskan sebagai Tugas Akhir. Pemilihan topik dapat dilakukan berdasarkan pada kategori topik tugas akhir yang diminati. Selain itu pemilihan topik juga dapat dilakukan dengan cara melihat nilai mata kuliah pada bidang tertentu yang sesuai dengan topik dan melihat ketersediaan referensi yang mudah didapatkan.

Sistem bimbingan tugas akhir di Prodi Teknik Informatika Politeknik Negeri Lhokseumawe yang berlaku selama ini

adalah mahasiswa tidak diwajibkan mengajukan topik Tugas Akhir pada bidang mata kuliah tertentu, melainkan topik tersebut ditentukan secara *random* oleh Mahasiswa sendiri.

Telah ada penelitian terdahulu yang dilakukan oleh Ratih Kumalasari Niswatin dan Ardi Sanjaya pada tahun 2017 dengan judul Skripsi Sistem Informasi Berbasis Web Untuk Klasifikasi Kategori Judul Skripsi.[2] Penelitian ini membuat sistem klasifikasi pemilihan kategori judul skripsi mahasiswa menggunakan metode k-nearest neighbor yang menghasilkan sebuah sistem yang dapat mengklasifikasikan kategori judul skripsi mahasiswa berdasarkan pada peminatan dan nilai-nilai mata kuliah perencanaan dan infrastruktur teknologi informasi, rekayasa perangkat lunak, jaringan, *data mining*, pengolahan citra, algoritma pemrograman dan basis data. Akan tetapi sistem tersebut hanya akan menentukan kategori judul skripsi berdasarkan bidang mata kuliah tertentu yang cocok dijadikan tugas akhir mahasiswa. Belum ada penelitian sebelumnya yang menghasilkan *output* yang lebih spesifik mengenai pemilihan Topik Tugas Akhir dengan menggunakan metode *Naive Bayes Classifier*.

Oleh karena itu, Penulis ingin mengembangkan penelitian agar memperoleh *output* yang lebih spesifik mengenai topik Tugas Akhir dengan metode *Naive Bayes Classifier*. Perbedaan penelitian ini dengan penelitian sebelumnya yang berjudul Sistem Informasi Berbasis Web Untuk Klasifikasi Kategori Judul Skripsi yang ditulis oleh Niswatin dkk yaitu untuk *output* yang dihasilkan berupa topik tugas akhir, misalnya seperti Sistem Pendukung Keputusan. Kemudian akan ditampilkan 5 jurnal yang berkaitan dengan topik tersebut. Selain itu, *user* yang melakukan pemilihan minat adalah Mahasiswa, berbeda dari Sistem Informasi Berbasis Web Untuk Klasifikasi Kategori Judul Skripsi yang hanya dapat dilakukan oleh Admin.

Dari permasalahan di atas, Penulis akan membuat sebuah sistem yang menerapkan metode *Naive Bayes Classifier*. Pemilihan metode ini dikarenakan relatif mudah digunakan karena tidak ada perkalian matrik atau optimasi numerik, lebih efisien apabila digunakan untuk memprediksi dalam jumlah yang sangat besar, dan memiliki tingkat keakurasian yang relatif tinggi dalam hasil prediksi.[3] Ada beberapa parameter yang akan digunakan misalnya seperti dosen pembimbing I dan II yang telah ditentukan oleh mahasiswa, minat bidang mata kuliah dan nilai-nilai mahasiswa dari beberapa mata kuliah inti di program studi Teknik Informatika Politeknik Negeri Lhokseumawe.

Sistem ini dirancang berbasis web agar setiap *user* atau pengguna yang menggunakan sistem ini akan lebih mudah dalam mengakses semua informasi yang disediakan oleh sistem dan dapat diakses di berbagai perangkat.

II. METODOLOGI PENELITIAN

A. Metode *Naive Bayes Classifier*

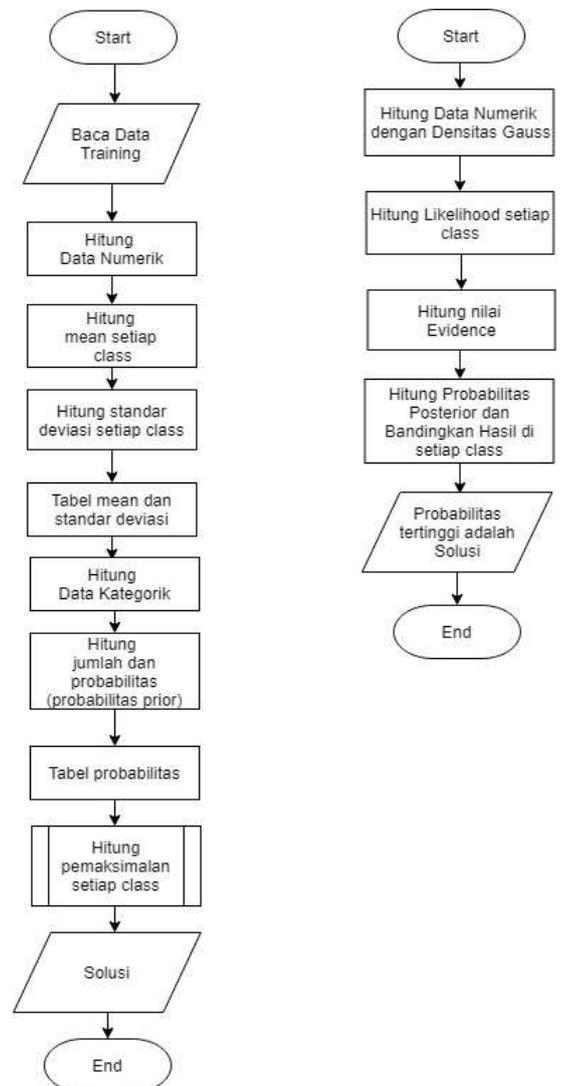
Naive Bayes Classifier merupakan sebuah pengklasifikasian probabilistik sederhana yang menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset yang diberikan. Algoritma menggunakan teorema *Bayes* dan mengasumsikan semua atribut independen atau tidak saling ketergantungan yang diberikan oleh nilai pada variabel *class*. [4]

Definisi lain mengatakan *Naive Bayes* merupakan pengklasifikasian dengan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya.[5]

Naive Bayes didasarkan pada asumsi penyederhanaan bahwa nilai atribut secara kondisional saling bebas jika diberikan nilai *output*. Dengan kata lain, diberikan nilai *output*, probabilitas mengamati secara bersama adalah produk dari probabilitas individu.[6]

Keuntungan penggunaan *Naive Bayes* adalah bahwa metode ini hanya membutuhkan jumlah data pelatihan (*Training Data*) yang kecil untuk menentukan estimasi parameter yang diperlukan dalam proses pengklasifikasian. *Naive Bayes* sering bekerja jauh lebih baik dalam kebanyakan situasi dunia nyata yang kompleks dari pada yang diharapkan.[7]

Berikut ini adalah flowchart algoritma *Naive Bayes Classifier* yang diterapkan pada sistem ditunjukkan pada gambar 1.



Gambar 1. Flowchart *Naive Bayes Classifier*

Adapun keterangan dari Gambar 2.1 sebagai berikut :

1. Baca data training
2. Hitung data numerik, dengan cara :
 - Cari nilai *mean* dari masing-masing atribut yang merupakan data numerik. Adapun rumus yang digunakan untuk menghitung nilai *mean* dapat dilihat pada persamaan berikut:

$$\mu = \frac{\sum_{i=1}^N X_i}{N} \quad (1)$$

di mana :

- μ : rata-rata (*mean*)
- x_i : nilai data ke -i
- N : banyaknya data

- Cari nilai standar deviasi dari masing-masing atribut yang merupakan data numerik. Adapun rumus yang digunakan untuk menghitung nilai standar deviasi dapat dilihat pada persamaan berikut:

$$\sigma = \sqrt{\frac{\sum (X - \mu)^2}{N}} \quad (2)$$

di mana :

- σ : standar deviasi
- x_i : nilai data ke -i
- μ : nilai rata-rata (*mean*)
- N : jumlah data

3. Nilai dari data numerik akan disimpan pada tabel *mean* dan standar deviasi.
4. Untuk data kategorik atau non numerik, maka hitung jumlah data dan probabilitas (*prior probability*), kemudian nilai akan disimpan pada tabel probabilitas.
5. Nilai dari data numerik akan disimpan pada tabel *mean* dan standar deviasi.
6. Nilai dari data numerik dan data kategorik akan dihitung untuk mendapatkan nilai maksimal dari tiap kelas. Hitung pemaksimalan setiap *class*, dengan cara:

- Cari nilai *gaussian* dari tiap atribut menggunakan rumus *Gaussian Distribution* sebagai berikut:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2(\sigma)^2}} \quad (3)$$

di mana :

- σ : standar deviasi
- π : nilai pi (3,146)
- e : 2,7183
- x : nilai data
- μ : nilai rata-rata (*mean*)

- Hitung *likelihood* dari setiap *class*.
 $P(x|c)$: Probabilitas berdasarkan kondisi pada hipotesis
- Hitung nilai *Evidence*.
Nilai *Evidence* selalu tetap untuk setiap kelas pada satu sampel. Nilai dari posterior tersebut nantinya akan dibandingkan dengan nilai nilai posterior kelas lainnya untuk menentukan ke kelas apa suatu sampel akan diklasifikasikan.

Adapun persamaan yang digunakan untuk menghitung nilai *evidence* dapat dilihat sebagai berikut :

$$\begin{aligned} & (\text{Prior}(C_1) \times \text{Likelihood}(C_1)) + \\ & (\text{Prior}(C_2) \times \text{Likelihood}(C_2)) + \\ & \dots (\text{Prior}(C_n) \times \text{Likelihood}(C_n)) \end{aligned} \quad (4)$$

7. Hitung dan bandingkan probabilitas setiap *class* (*prior probability*). Adapun persamaan yang digunakan untuk mencari nilai posterior sebagai berikut:

$$\text{Probabilitas Posterior} = \frac{\text{Prior} \times \text{Likelihood}}{\text{Evidence}} \quad (5)$$

8. Nilai probabilitas tertinggi adalah solusi yang dihasilkan.

B. Parameter

Dalam sistem ini terdapat beberapa parameter untuk menghasilkan sebuah solusi, diantaranya adalah :

1. Dosen Pembimbing, data diambil berdasarkan daftar pembimbing dan komisi sidang tugas akhir mahasiswa tahun ajaran 2018/2019.
2. Minat Mahasiswa yang dikelompokkan ke dalam 8 *class* yaitu Sistem Informasi, Sistem Pendukung Keputusan, Sistem Informasi Geografis, Sistem Pakar, Sistem Cerdas, E-Learning, Citra, dan Jaringan Komputer.
3. Nilai mata kuliah inti yang dikelompokkan ke dalam 5 bidang mata kuliah yaitu Basis Data dan Sistem Informasi, Rekayasa Perangkat Lunak, Citra, Sistem Cerdas, dan Jaringan Komputer dengan total 31 mata kuliah inti.

Berikut adalah tabel dosen pembimbing dapat dilihat pada tabel I.

TABEL I
DOSEN PEMBIMBING

No	Kode	Nama
1	AKM	AKMALUL FATA SST
2	ARH	MUHAMMAD ARHAMI S.Si., M.Kom
3	AZH	AZHAR ST., MT
4	HDW	HENDRAWATY ST., MT
5	HUS	HUSAINI ST., M.Kom
6	HUZ	HUZAENI,SST., M.IT
7	MUL	MULYADI ST., M.Eng
8	RIZ	MUHAMMAD RIZKA, SST., M.Kom
9	SAL	SALAHUDDIN, ST., M.Cs
10	ZUL	ZULFAN KHAIRIL S.,ST., M.Eng
11	AMR	Amirullah SST., M.Kom
12	MAH	Mahdi, ST.M.Cs
13	HAR	Hari Toha Hidayat, S,Si, M.Cs

Berikut adalah tabel pengelompokkan mata kuliah dapat dilihat pada tabel II.

TABEL II
PENGELOMPOKKAN MATA KULIAH

Basis Data & SI	RPL	Citra	Sistem Cerdas	Jarkom	77.71	75.95	79.93	79.55	78.28	SI	HUS	ARH	SI
					75.81	76.62	81.88	76.73	77.76	SI	MAH	SAL	SI
Basis Data 1	Konsep Pemrograman	Komputer Grafik	Kecerdasan Buatan	Konsep Jarkom	78.33	75.76	80.50	73.08	76.92	C	MUL	ZUL	SPK
					77.78	77.87	86.40	86.08	82.03	SI	ARH	HUS	SPK
P. Basis Data 1	P. Konsep Pemrograman	P. Komputer Grafik	Pengenalan Pola	P. Konsep Jarkom	78.90	80.28	84.25	77.25	80.17	SPK	ARH	SAL	SPK
					79.64	82.22	81.60	81.93	81.35	SPK	ARH	HUS	SPK
Basis Data 2	PBO	PCD		Pemrograman Jarkom	79.56	82.33	84.50	76.00	80.60	SI	SAL	AZH	SIG
					78.78	77.86	81.75	71.50	77.47	SIG	SAL	HAR	SIG
P. Basis Data 2	P. PBO	P.PCD		Adm. Jarkom	76.50	81.72	81.10	84.00	80.83	SIG	MAH	MUL	SIG
					77.17	78.92	83.58	76.00	78.92	SIG	AZH	AKM	SIG
Desain Web	Pemrograman Lanjut			P. Adm. Jarkom	73.20	84.75	84.75	83.50	81.55	SP	MAH	RIZ	SP
					75.23	82.77	78.75	78.50	78.81	CC	MAH	RIZ	SP
P. Desain Web	P. Pemrograman Lanjut			Adm. Jarkom	79.44	84.50	81.00	85.50	82.61	SP	ZUL	AMR	SP
					78.83	83.14	82.90	83.50	82.09	SP	MUL	AKM	SP
Administrasi Basis Data	Rancangan Analisa Algoritma				82.34	83.34	83.29	88.00	84.24	SC	RIZ	AMR	SC
					84.39	83.85	87.15	86.00	85.35	SI	AZH	AKM	SC
Struktur Data Algoritma	P. Rancangan Analisa Algoritma				73.06	77.53	69.55	81.50	75.41	SC	MUL	AKM	SC
					74.72	75.55	68.30	82.50	75.27	SC	RIZ	MAH	SC
P. Struktur Data Algoritma	RPL 1				80.97	81.66	81.63	83.00	81.82	SI	HDW	AZH	EL
					74.82	77.20	85.00	84.50	80.38	EL	MUL	ARH	EL
RPL 2	P. RPL				80.25	77.15	87.08	81.50	81.49	EL	ARH	SAL	EL
					79.58	81.94	80.50	80.00	80.51	EL	HUZ	AMR	EL
P. RPL					77.20	81.26	79.00	83.00	80.12	C	MUL	RIZ	C
					77.07	86.59	79.25	91.00	83.48	C	ZUL	MAH	C
					81.22	80.61	88.50	83.50	83.46	C	RIZ	SAL	C
					80.97	81.66	81.63	83.00	81.85	JK	HDW	AZH	JK
					74.82	77.20	85.00	84.50	86.66	JK	ARH	AKM	JK
					80.25	77.15	87.08	81.50	87.80	JK	SAL	AZH	JK
					74.72	75.55	68.30	82.50	84.66	JK	HDW	AZH	JK
				

Berikut adalah tabel pengelompokan minat/kategori class dapat dilihat pada tabel III.

TABEL III
PENGELOMPOKKAN KATEGORI CLASS

No	Kategori
1	Sistem Informasi
2	Sistem Pendukung Keputusan
3	Sistem Informasi Geografis
4	Sistem Pakar
5	Sistem Cerdas
6	E-Learning
7	Citra
8	Jaringan Komputer

Berikut adalah tabel pengelompokan minat/kategori class dapat dilihat pada tabel IV.

TABEL IV
DATA LATIH

Rata-Rata Bidang Mata Kuliah					Minat	Dospem		Class
BDSI	RPL	SC	C	JK		I	II	
77.67	71.20	79.60	74.20	75.67	SC	ARH	HUS	SI
78.11	75.44	81.13	79.38	78.51	SI	HUS	ARH	SI

C. Perhitungan Metode Naïve Bayes Classifier

Metode *Naïve Bayes Classifier* adalah metode pembelajaran mesin yang memanfaatkan perhitungan probabilitas dengan memprediksi probabilitas di masa depan berdasarkan pengalaman di masa sebelumnya.

Perhitungan Probabilitas Class

Berikut adalah jumlah dan nilai probabilitas class yang ditunjukkan pada Tabel V.

TABEL V
JUMLAH DAN PROBABILITAS CLASS

No	Kategori	Jumlah	Probabilitas
1	Sistem Informasi	34	0.4474
2	Sistem Pendukung Keputusan	10	0.1316
3	Sistem Informasi Geografis	8	0.1053
4	Sistem Pakar	5	0.0658
5	Sistem Cerdas	4	0.0526
6	E-Learning	7	0.0921
7	Citra	3	0.0395
8	Jaringan Komputer	5	0.0658
Total		76	

- Perhitungan Probabilitas Minat
Berikut adalah nilai probabilitas minat yang ditunjukkan pada Tabel VI.

TABEL VI
PROBABILITAS MINAT

No	Minat Class	Probabilitas
1	P(minat = Sistem Informasi Sistem Informasi)	0.7353
2	P(minat = Sistem Pendukung Keputusan Sistem Informasi)	0.0294
3	P(minat = Sistem Informasi Geografis Sistem Informasi)	0.0294
4	P(minat = Sistem Pakar Sistem Informasi)	0.0000
5	P(minat = Sistem Cerdas Sistem Informasi)	0.0588
6	P(minat = E-Learning Sistem Informasi)	0.0294
7	P(minat = Citra Sistem Informasi)	0.0294
8	P(minat = Jaringan Komputer Sistem Informasi)	0.0882
...
57	P(minat = Sistem Informasi Jaringan Komputer)	0.0000
58	P(minat = Sistem Pendukung Keputusan Jaringan Komputer)	0.0000
59	P(minat = Sistem Informasi Geografis Jaringan Komputer)	0.0000
60	P(minat = Sistem Pakar Jaringan Komputer)	0.0000
61	P(minat = Sistem Cerdas Jaringan Komputer)	0.0000
62	P(minat = E-Learning Jaringan Komputer)	0.0000
63	P(minat = Citra Jaringan Komputer)	0.0000
64	P(minat = Jaringan Komputer Jaringan Komputer)	1.0000

- Perhitungan Probabilitas Dosen Pembimbing
Berikut adalah nilai probabilitas dosen pembimbing yang ditunjukkan pada Tabel VII.

TABEL VII
PROBABILITAS DOSEN PEMBIMBING

Dosen Pembimbing	Probabilitas	
	PB I	PB II
P(pb = AKM Sistem Informasi)	0.0000	0.1471
P(pb = ARH Sistem Informasi)	0.0882	0.0882
P(pb = AZH Sistem Informasi)	0.2059	0.0882
P(pb = HDW Sistem Informasi)	0.1471	0.0000
P(pb = HUS Sistem Informasi)	0.0588	0.0882
P(pb = HUZ Sistem Informasi)	0.1471	0.0882
P(pb = MUL Sistem Informasi)	0.0588	0.0294
P(pb = RIZ Sistem Informasi)	0.0882	0.0588

P(pb = SAL Sistem Informasi)	0.0000	0.0588
P(pb = ZUL Sistem Informasi)	0.0294	0.0588
P(pb = AMR Sistem Informasi)	0.0000	0.1471
P(pb = MAH Sistem Informasi)	0.1765	0.1471
P(pb = HAR Sistem Informasi)	0.0000	0.0000
...
P(pb = AKM Jaringan Komputer)	0.0000	0.2000
P(pb = ARH Jaringan Komputer)	0.2000	0.0000
P(pb = AZH Jaringan Komputer)	0.0000	0.6000
P(pb = HDW Jaringan Komputer)	0.4000	0.0000
P(pb = HUS Jaringan Komputer)	0.0000	0.0000
P(pb = HUZ Jaringan Komputer)	0.0000	0.0000
P(pb = MUL Jaringan Komputer)	0.0000	0.0000
P(pb = RIZ Jaringan Komputer)	0.0000	0.2000
P(pb = SAL Jaringan Komputer)	0.2000	0.0000
P(pb = ZUL Jaringan Komputer)	0.2000	0.0000
P(pb = AMR Jaringan Komputer)	0.0000	0.0000
P(pb = MAH Jaringan Komputer)	0.0000	0.0000
P(pb = HAR Jaringan Komputer)	0.0000	0.0000

- Perhitungan Data Numerik
Dari tabel-tabel diatas sudah ditentukan nilai probabilitas untuk setiap atribut. Untuk perhitungan data numerik seperti nilai mata kuliah digunakan Distribusi *Gaussian* untuk mencari nilai *mean*, varian, dan standar deviasi yang diklasifikasikan menurut *classnya*. Berikut adalah nilai data numerik yang ditunjukkan pada Tabel VIII.

TABEL VIII
NILAI DATA NUMERIK

Class	Bidang Mata Kuliah	Mean	Standar Deviasi
SI	Basis data dan sistem informasi	77.83	5.0527
	Rekayasa Perangkat Lunak	78.25	4.6284
	Citra	80.01	6.2676
	Sistem Cerdas	78.65	7.3159
	Jaringan Komputer	79.75	6.2217
SPK	Basis data dan sistem informasi	79.43	3.2821
	Rekayasa Perangkat Lunak	81.25	5.7437
	Citra	84.03	5.1071
	Sistem Cerdas	82.28	8.6150
	Jaringan Komputer	80.87	6.1206
SIG	Basis data dan sistem informasi	75.05	5.2019
	Rekayasa Perangkat Lunak	77.87	3.5983
	Citra	79.74	7.8780
	Sistem Cerdas	74.13	6.6603
	Jaringan Komputer	79.42	2.3459
SP	Basis data dan sistem informasi	75.66	3.0639
	Rekayasa Perangkat Lunak	82.87	1.9829
	Citra	81.66	2.0282
	Sistem Cerdas	82.70	2.3152

	Jaringan Komputer	83.61	2.9007
	Basis data dan sistem informasi	78.63	4.8296
	Rekayasa Perangkat Lunak	80.07	3.5992
SC	Citra	77.07	8.2723
	Sistem Cerdas	84.50	2.6220
	Jaringan Komputer	80.26	3.7834
	Basis data dan sistem informasi	78.60	3.5245
	Rekayasa Perangkat Lunak	79.17	3.5151
EL	Citra	83.21	5.4223
	Sistem Cerdas	81.43	4.6092
	Jaringan Komputer	78.19	4.5149
	Basis data dan sistem informasi	78.50	1.9275
	Rekayasa Perangkat Lunak	82.82	2.6791
C	Citra	82.25	4.4206
	Sistem Cerdas	85.83	3.6591
	Jaringan Komputer	81.54	3.9542
	Basis data dan sistem informasi	78.45	3.0313
	Rekayasa Perangkat Lunak	78.46	2.3324
JK	Citra	81.22	6.6905
	Sistem Cerdas	81.70	2.5417
	Jaringan Komputer	84.59	2.4000

D. Perancangan Sistem

Perancangan sistem merupakan tahap awal dari perancangan perangkat lunak yang akan di buat. Perancangan sistem bertujuan untuk memberikan gambaran kebutuhan *user* mengenai alur proses dari perangkat lunak yang akan di buat. Perancangan sistem penentuan topik tugas akhir mahasiswa, yaitu : *Context Diagram* dan *Entity Relationship Diagram (ERD)*

▪ Context Diagram

Diagram konteks digunakan untuk menjelaskan dan menggambarkan cara kerja sistem secara umum, yaitu hubungan antara sistem ke luar, sistem ke dalam, maupun dengan external entity yang terkait.



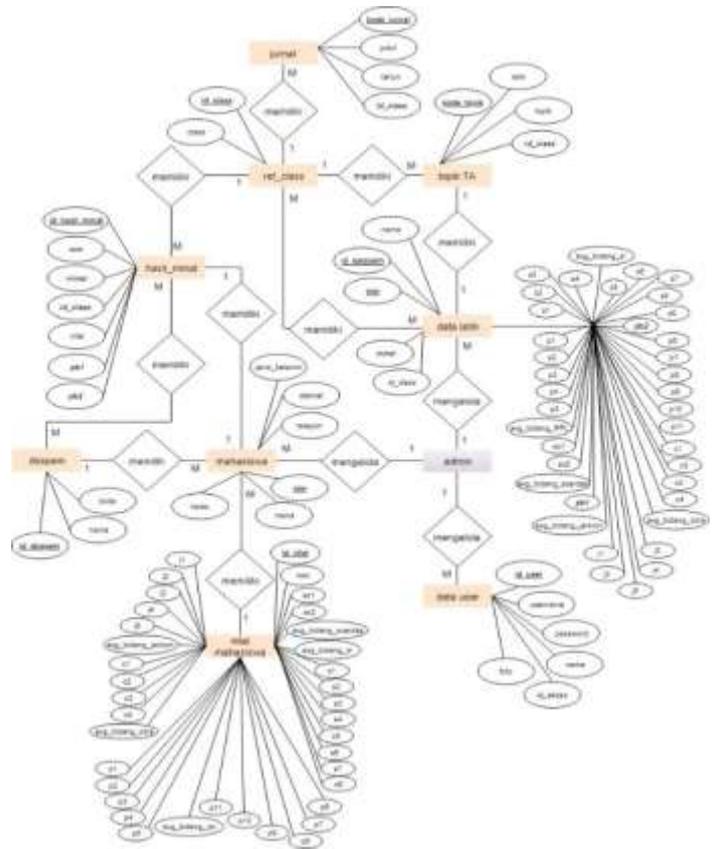
Gambar 2. Context Diagram

Sistem penentuan topik tugas akhir mahasiswa terdiri dari 3 entitas yaitu entitas Admin, Kaprodi, dan Mahasiswa. Dapat dilihat pada Gambar 2. Mahasiswa *login* dan melihat nilai-nilai mata kuliah, kemudian mahasiswa memilih dosen

pembimbing, dan salah satu minat dari beberapa *class* yang disediakan untuk mendapatkan rekomendasi topik tugas akhir sekaligus referensi jurnal-jurnal yang berkaitan dengan topik tersebut. Data-data mahasiswa dan nilai mahasiswa dapat *diinputkan* oleh Kaprodi dan Admin, selain itu admin dapat mengolala beberapa data diantaranya data *user*, data mahasiswa, data nilai mahasiswa, data jurnal, data judul TGA, data dosen pembimbing, data latihan, dan data *class*.

▪ Entity Relationship Diagram

Pada tahap desain akan menjelaskan mengenai desain database, dan desain user interface dari aplikasi atau sistem yang dibuat. Database atau basis data adalah kumpulan data yang saling berelasi. Data sendiri merupakan fakta mengenai objek, orang, dan lain-lain.



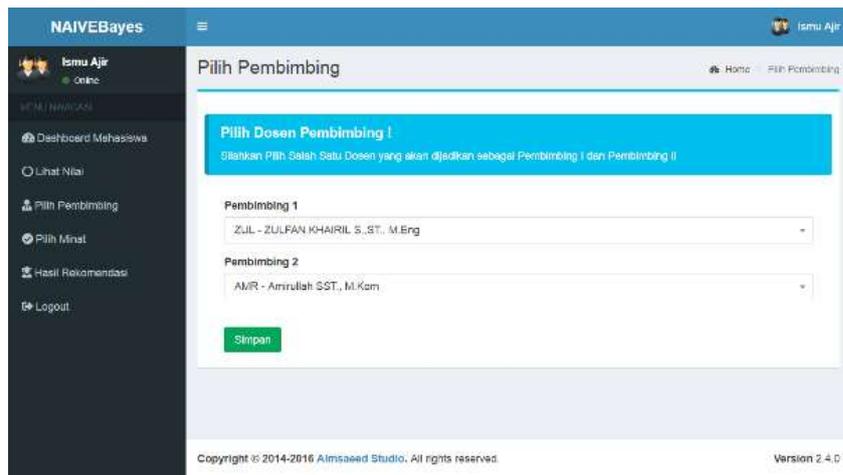
III. HASIL DAN PEMBAHASAN

A. Implementasi Sistem

Implementasi sistem merupakan tahapan dalam memenuhi kebutuhan user, dalam berinteraksi dengan komputer. Fasilitas antar muka yang baik sangat membantu pemakai dalam memahami proses yang sedang dilakukan oleh sistem tersebut dan dapat meningkatkan kinerja sistem.

1. Halaman Pilih Pembimbing

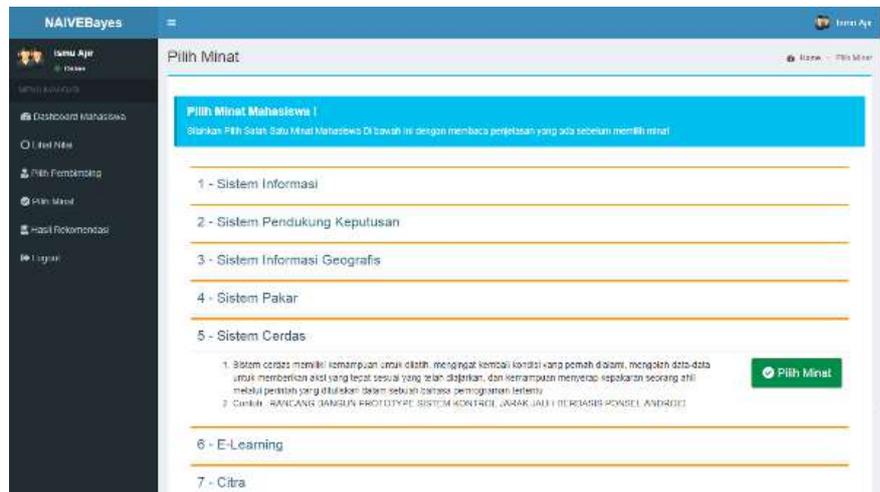
Halaman pilih pembimbing merupakan halaman yang menampilkan *dropdown menu* data dosen pembimbing, di mana mahasiswa harus memilih masing-masing satu dosen untuk pembimbing I dan pembimbing II.



Gambar 4. Tampilan Menu Pilih Pembimbing

2. Halaman Pilih Minat

Halaman pilih minat merupakan halaman yang menampilkan *accordion menu* data kategori topik tugas akhir di mana mahasiswa harus memilih satu kategori yang diminati.



Gambar 5. Tampilan Menu Pilih Minat

3. Halaman Hasil Rekomendasi

Halaman hasil rekomendasi merupakan halaman yang akan menampilkan rekomendasi class untuk mahasiswa, beserta jurnal-jurnal terkait class tersebut.



Gambar 6. Tampilan Menu Hasil Rekomendasi

Pada gambar diatas adalah hasil rekomendasi topik tugas akhir untuk mahasiswa bernama Ismu Ajir. Pada kasus ini, mahasiswa memilih Bapak “Zulfan Khairil S.,ST., M.Eng” sebagai Pembimbing I dan Bapak “Amirullah SST., M.Kom” sebagai Pembimbing II. Setelah itu, mahasiswa memilih “Sistem Cerdas” sebagai kategori topik yang diminati. Dengan perhitungan Naïve Bayes Classifier, maka pada halaman hasil rekomendasi akan tampil Nilai Gauss, Nilai Probabilitas Akhir (*Likelihood*), Hasil Prediksi (*Posterior Probability*), beserta Nama Pembimbing I dan II yang telah di pilih sebelumnya, dan jurnal terkait class yang direkomendasikan.

A. Perhitungan Data Uji

Perhitungan dari metode Naïve Bayesian Classifier akan di cantumkan pada contoh untuk mahasiswa bernama Ismu Ajir. Berikut adalah tabel data uji.

TABEL IX
DATA UJI

NIM	Bidang Mata Kuliah	Nilai	Minat	Pb I	Pb II	Class
1557301108	Basis Data dan Sistem Informasi	81.23	Sistem Cerdas	ZUL	AMR	?
	Rekayasa Perangkat Lunak	84.50				
	Citra	79,00				
	Sistem Cerdas	88,00				
	Jaringan Komputer	86,77				

- Perhitungan Nilai Gaussian untuk Data Numerik
Berikut adalah perhitungan nilai gaussian untuk data numerik.

$$f(SI | Basis Data dan SI) = \frac{1}{5.0527 \sqrt{2 \pi}} e^{-\frac{1(81,23-77,83)^2}{2(5.0527)^2}} = 0.062954$$

$$f(SI | RPL) = \frac{1}{4.6284 \sqrt{2 \pi}} e^{-\frac{1(84,50-78,25)^2}{2(4.6284)^2}} = 0.034610$$

$$f(SI | Citra) = \frac{1}{6.2676 \sqrt{2 \pi}} e^{-\frac{1(79,00-80,01)^2}{2(6.2676)^2}} = 0.062823$$

$$f(SI | Sistem Cerdas) = \frac{1}{7.3159 \sqrt{2 \pi}} e^{-\frac{1(88,00-78,65)^2}{2(7.3159)^2}} = 0.024078$$

$$f(SI | Jarkom) = \frac{1}{6.2217 \sqrt{2 \pi}} e^{-\frac{1(86,77-79,75)^2}{2(6.2217)^2}} = 0.033947$$

$$f(SPK | Basis Data dan SI) = \frac{1}{3.2821 \sqrt{2 \pi}} e^{-\frac{1(81,23-79,43)^2}{2(3.2821)^2}} = 0.104623$$

$$f(SPK | RPL) = \frac{1}{5.7437 \sqrt{2 \pi}} e^{-\frac{1(84,50-81,25)^2}{2(5.7437)^2}} = 0.059170$$

$$f(SPK | Citra) = \frac{1}{5.1071 \sqrt{2 \pi}} e^{-\frac{1(79,00-84,03)^2}{2(5.1071)^2}} = 0.048118$$

$$f(SPK | Sistem Cerdas) = \frac{1}{8.6150 \sqrt{2 \pi}} e^{-\frac{1(88,00-82,28)^2}{2(8.6150)^2}} = 0.037133$$

$$f(SPK | Jarkom) = \frac{1}{6.1206 \sqrt{2 \pi}} e^{-\frac{1(86,77-80,87)^2}{2(6.1206)^2}} = 0.040961$$

...

$$f(Citra | Basis Data dan SI) = \frac{1}{1.9275 \sqrt{2 \pi}} e^{-\frac{1(81,23-78,50)^2}{2(1.9275)^2}} = 0.075770$$

$$f(Jarkom | Basis Data dan SI) = \frac{1}{3.0313 \sqrt{2 \pi}} e^{-\frac{1(81,23-78,45)^2}{2(3.0313)^2}} = 0.086525$$

$$f(Jarkom | RPL) = \frac{1}{2.3324 \sqrt{2 \pi}} e^{-\frac{1(84,50-78,46)^2}{2(2.3324)^2}} = 0.005971$$

$$f(Jarkom | Citra) = \frac{1}{6.6905 \sqrt{2 \pi}} e^{-\frac{1(79,00-81,22)^2}{2(6.6905)^2}} = 0.056449$$

$$f(Jarkom | Sistem Cerdas) = \frac{1}{2.5417 \sqrt{2 \pi}} e^{-\frac{1(88,00-81,70)^2}{2(2.5417)^2}} = 0.007272$$

$$f(Jarkom | Jarkom) = \frac{1}{2.4000 \sqrt{2 \pi}} e^{-\frac{1(86,77-84,59)^2}{2(2.4000)^2}} = 0.110170$$

- Perhitungan Probabilitas Akhir Setiap Class (*Likelihood*)
Berikut adalah tabel nilai likelihood.

TABEL X
NILAI LIKELIHOOD

Nilai Likelihood	
P(X SI)	0.00000000028466
P(X SPK)	0.00000005436745
P(X SIG)	0,00000000000000
P(X Sistem Pakar)	0,00000000000000
P(X Sistem Cerdas)	0,00000000000000
P(X E-Learning)	0,00000000000000
P(X Citra)	0,00000000000000
P(X Jarkom)	0,00000000000000

- Perhitungan Nilai Evidence
Nilai Evidence selalu tetap untuk setiap kelas pada satu sampel. Nilai dari posterior tersebut nantinya akan dibandingkan dengan nilai nilai posterior kelas lainnya untuk menentukan ke kelas apa suatu sampel akan diklasifikasikan. Berikut adalah nilai evidence.

$$Evidence = 0.0000000007281$$

- Hasil Prediksi
Berikut adalah hasil prediksi atau nilai probabilitas posterior.

TABEL XI
NILAI PROBABILITAS POSTERIOR

Nilai Probabilitas Posterior	
P(SI X)	0.017491674769949
P(SPK X)	0.982661230600192
P(SIG X)	0,0000000000000000
P(Sistem Pakar X)	0,0000000000000000
P(Sistem Cerdas X)	0,0000000000000000
P(E-Learning X)	0,0000000000000000
P(Citra X)	0,0000000000000000
P(Jarkom X)	0,0000000000000000

Dari hasil diatas, terlihat bahwa nilai probabilitas akhir terbesar berada pada *class Sistem Pendukung Keputusan*, maka dapat diambil kesimpulan bahwa mahasiswa yang bernama *Ismu Ajir* dapat direkomendasikan untuk mengambil topik "*Sistem Pendukung Keputusan*".

II. KESIMPULAN

1. Aplikasi penentuan topik tugas akhir ini sebagai penerapan dari metode naïve bayesian classifier untuk klasifikasi topik sebagai suatu hasil rekomendasi atau usulan topik tugas akhir untuk mahasiswa teknik informatika.
2. Dengan adanya sistem penentuan topik tugas akhir ini, pengambilan keputusan dilakukan dengan cepat dan mudah karena sistem perhitungan sudah otomatis.
3. Parameter-parameter yang digunakan pada sistem ini adalah data nilai mata kuliah mahasiswa, data dosen pembimbing tahun ajaran 2018/2019, data minat kategori topik tugas akhir mahasiswa.
4. Sistem ini memiliki hasil akhir atau output berupa nilai probabilitas tertinggi yang menjadi acuan dalam penentuan class yang direkomendasikan, serta mahasiswa dapat melihat referensi judul jurnal-jurnal yang berkaitan dengan class yang telah direkomendasikan.

REFERENSI

- [1] Winarno Surakhmad, Pengantar Penelitian Ilmiah, (Bandung: Tarsito, 1994), hal. 131
- [2] Niswatin, Ratih Kumalasari, A. S. (2017). Sistem Informasi Berbasis Web Untuk Klasifikasi Kategori Judul Skripsi. *Seminar Nasional Teknologi Informasi, Komunikasi Dan Aplikasinya*, 04(November), 50–55.
- [3] A. F. Mauriza, "Implementasi Data Mining Untuk Memprediksi Kelulusan Mahasiswa Fakultas Komunikasi dan Informatika UMS Menggunakan Metode Naive Bayes," p. 18, 2014.
- [4] Patil, T. R., Sherekar, M. S., 2013, Performance Analysis of *Naive Bayes* and *J48 Classification Algorithm* for Data

Classification, International Journal of Computer Science and Applications, Vol. 6, No. 2, Hal 256-261.

- [5] Bustami., 2013, Penerapan Algoritma *Naive Bayes* Untuk Mengklasifikasi Data Nasabah Asuransi, *TECHSI : Jurnal Penelitian Teknik Informatika*, Vol. 3, No.2, Hal. 127-146.
- [6] Ridwan, M., Suyono, H., Sarosa, M., (2013), Penerapan Data Mining untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier, *Jurnal EECCIS*, Vol 1, No. 7, Hal. 59-64
- [7] Firman, A. H. (2016). Sistem Informasi Perpustakaan Berbasis Web Application. *E-journal Teknik Elektro dan Komputer*, 29-36.
- [8] Palit, Randi V, dkk. 2015. "Rancangan Sistem Informasi Keuangan Berbasis Web Di Jemaat GMIM Bukit Moria Malalayang." *E-Journal Teknik Elektro Dan Komputer Vol 4(7):1-7*.
- [9] Umam MH. Analisis Perbandingan Algoritma C4.5 dan Algoritma Naïve Bayes untuk Prediksi Kelulusan Mahasiswa. 2017:3.
- [10] Ningsih, Yulia "Perancangan Sistem Informasi Penjualan Baju Batik Berbasis Web (Online Shopping) Pada CV. Selaras Batik" 22 Oktober 2014.