

**PENERAPAN *TEXT MINING* DALAM KLASIFIKASI JUDUL SKRIPSI  
MENGUNAKAN METODE *NAÏVE BAYES CLASSIFIER***

**SKRIPSI**

untuk memenuhi sebagian persyaratan  
mencapai gelar Sarjana  
pada Program Studi Informatika



disusun oleh  
**Cakra Amiyantoro**  
15.11.9319

**PROGRAM SARJANA  
PROGRAM STUDI INFORMATIKA  
FAKULTAS ILMU KOMPUTER  
UNIVERSITAS AMIKOM YOGYAKARTA  
YOGYAKARTA  
2018**

**PENERAPAN *TEXT MINING* DALAM KLASIFIKASI JUDUL SKRIPSI  
MENGUNAKAN METODE *NAÏVE BAYES CLASSIFIER***

**SKRIPSI**

untuk memenuhi sebagian persyaratan  
mencapai gelar Sarjana  
pada Program Studi Informatika



**disusun oleh**  
**Cakra Amiyantoro**  
**15.11.9319**

**PROGRAM SARJANA**  
**PROGRAM STUDI INFORMATIKA**  
**FAKULTAS ILMU KOMPUTER**  
**UNIVERSITAS AMIKOM YOGYAKARTA**  
**YOGYAKARTA**  
**2018**



**PERSETUJUAN**

**SKRIPSI**

**PENERAPAN *TEXT MINING* DALAM KLASIFIKASI JUDUL SKRIPSI  
MENGUNAKAN METODE *NAÏVE BAYES CLASSIFIER***

yang dipersiapkan dan disusun oleh

**Cakra Amiyantoro**

**15.11.9319**

telah disetujui oleh Dosen Pembimbing Skripsi  
pada tanggal 6 Oktober 2018

**Dosen Pembimbing,**

  
**Andi Sunyoto, M.Kom.**

**NIK. 190302052**

## PENGESAHAN

### SKRIPSI

#### PENERAPAN *TEXT MINING* DALAM KLASIFIKASI JUDUL SKRIPSI MENGUNAKAN METODE *NAÏVE BAYES CLASSIFIER*

yang dipersiapkan dan disusun oleh

**Cakra Amiyantoro**

**15.11.9319**

telah dipertahankan di depan Dewan Penguji  
pada tanggal 17 November 2018

#### Susunan Dewan Penguji

**Nama Penguji**

**Tanda Tangan**

**Windha Mega P.D., M.Kom.**  
NIK. 190302185



**Yuli Astuti, M.Kom.**  
NIK. 190302146



**Andi Sunyoto, M.Kom.**  
NIK. 190302052



Skripsi ini telah diterima sebagai salah satu persyaratan  
untuk memperoleh gelar Sarjana Komputer  
Tanggal 26 November 2018



**DEKAN FAKULTAS ILMU KOMPUTER**

**Krisnawati, S.Si., M.T.**  
NIK. 190302038

## PERNYATAAN

Saya yang bertandatangan di bawah ini menyatakan bahwa, skripsi ini merupakan karya saya sendiri (ASLI), dan isi dalam skripsi ini tidak terdapat karya yang pernah diajukan oleh orang lain untuk memperoleh gelar akademis di suatu institusi pendidikan tinggi manapun, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis dan/atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Segala sesuatu yang terkait dengan naskah dan karya yang telah dibuat adalah menjadi tanggungjawab saya pribadi.

Yogyakarta, 23 November 2018



Cakra Amiyantoro

NIM. 15.11.9319

## MOTTO

“Kehidupan layaknya drama tanpa naskah. Jangan sedih, jangan takut, jangan galau, jangan baperan, percaya kita bisa, semuanya enggak mustahil kalau diperjuangkan. Aku percaya mimpi itu bisa jadi nyata, selucu apapun mimpinya. Maksimalin waktu produktif sebaik yang kita bisa dan jangan lupa berdoa.”

(Penulis)



## PERSEMBAHAN

Skripsi berjudul “Penerapan *Text Mining* dalam Klasifikasi Judul Skripsi Menggunakan Metode *Naïve Bayes Classifier*” ini dipersembahkan kepada:

1. Allah SWT yang telah memberikan nikmat luar biasa kepada saya.
2. Kedua orang tua saya yang telah memberikan segala yang terbaik kepada anak-anaknya.
3. Universitas AMIKOM Yogyakarta sebagai perguruan tinggi tempat saya bernaung.
4. Bapak Andi Sunyoto, M.Kom. yang telah membimbing saya dalam penyusunan skripsi ini dengan penuh kesabaran.
5. Teman-teman IF12 angkatan 2015 yang telah menemani saya selama perkuliahan.
6. Para sahabat saya yang tidak bisa saya sebutkan satu per satu.

## KATA PENGANTAR

Segala puji bagi Allah SWT tuhan semesta alam yang telah memberikan kenikmatan luar biasa yang tak ada henti-hentinya sehingga skripsi ini dapat terselesaikan dengan sebaik-baiknya. Shalawat serta salam semoga selalu tercurah kepada nabi agung Nabi Muhammad SAW beserta keluarga, para sahabat dan umatnya hingga akhir zaman.

Dengan terselesaikannya skripsi ini yang berjudul “Penerapan *Text Mining* dalam Klasifikasi Judul Skripsi Menggunakan Metode *Naïve Bayes Classifier*” penulis mengucapkan terima kasih yang sebanyak-banyaknya kepada:

1. Kedua orang tua penulis bernama Arif Priyantoro dan Siti Amisah, serta kedua kakak penulis bernama Ari Yunita dan Betha Amitasari yang telah mendukung penuh seluruh kegiatan penulis, sehingga mustahil penulis untuk menghitung kebaikannya.
2. Bapak Prof. Dr. M. Suyanto selaku Rektor Universitas AMIKOM Yogyakarta.
3. Bapak Sudarmawan, M.Kom. selaku Dekan Fakultas Sains dan Teknologi, dan Ketua Program Studi S1 Informatika.
4. Bapak Andi Sunyoto, M.Kom. selaku pembimbing skripsi ini yang selalu sabar dalam memberikan arahan-arahan serta kesempatan waktu bimbingan yang banyak.



5. Seluruh dosen dan staff Universitas AMIKOM Yogyakarta yang pernah berkontribusi terhadap segala aktivitas penulis selama menjalani perkuliahan.
6. Teman-teman IF12 angkatan 2015 yang telah menemani penulis selama perkuliahan.
7. Erlin Idris yang menjadi *support system* penuh penulis selama menjalani perkuliahan hingga skripsi ini terselesaikan.
8. Asvarizal Filcha, Subraga Islammada Sunjiwo, dan Utsman Fajar yang berkontribusi banyak kepada penulis selama menjalani perkuliahan.
9. Puguh Dwiyanto dan Danang Adi Wicaksono sebagai sahabat yang selalu mengingatkan penulis akan pentingnya kebahagiaan.
10. Muthi'atun Nasyithoh dan Sabila Rafiqah Zahra sebagai sahabat penulis yang selalu menawarkan bantuan, memberikan motivasi, serta mengingatkan berdoa.
11. Rian Rahmawan Putra yang mengajari penulis terhadap hebatnya kekuatan optimisme dalam tiap sudut pandang kehidupan.

Penulis menyaradi skripsi ini masih banyak kekurangan, namun penulis berharap skripsi ini dapat memberikan manfaat bagi para pembacanya. Dan semoga Allah SWT membalas kebaikan semua orang yang telah memberikan dukungan dalam bentuk apapun kepada penulis.

Yogyakarta, 23 November 2018

Cakra Amiyantoro

## DAFTAR ISI

JUDUL .....	i
PERSETUJUAN .....	ii
PENGESAHAN .....	iii
PERNYATAAN.....	iv
MOTTO .....	v
PERSEMBAHAN .....	vi
KATA PENGANTAR .....	vii
DAFTAR ISI.....	ix
DAFTAR TABEL.....	xii
DAFTAR GAMBAR .....	xv
INTISARI.....	xxi
<i>ABSTRACT</i> .....	xxii
<b>BAB I</b> Pendahuluan .....	1
1.1 Latar Belakang Masalah .....	1
1.2 Rumusan Masalah .....	3
1.3 Batasan Masalah.....	3
1.4 Maksud dan Tujuan Penelitian .....	4
1.5 Manfaat Penelitian.....	4
1.6 Metodologi Penelitian .....	5
1.6.1 Metode Pengumpulan Data .....	5
1.6.2 Metode Analisis Data .....	5
1.6.3 Metode Perancangan Sistem .....	6
1.6.4 Metode Pengembangan Sistem .....	6

1.6.5	Metode Pengujian.....	6
1.7	Sistematika Penulisan.....	6
<b>BAB II LANDASAN TEORI .....</b>		<b>8</b>
2.1	Tinjauan Pustaka .....	8
2.2	Landasan Teori .....	10
2.2.1	Pengertian Klasifikasi .....	10
2.2.2	Manfaat Klasifikasi .....	11
2.2.3	<i>DDC (Dewey Decimal Classification)</i> .....	11
2.2.4	<i>Text Mining</i> .....	12
2.2.5	<i>Text Preprocessing</i> .....	12
2.2.6	<i>Tf-Idf (Term Frequency Inverse Document Frequency)</i> .....	16
2.2.7	<i>Naïve Bayes Classifier</i> .....	17
2.2.8	Metode Evaluasi.....	20
2.3	Model Pengembangan <i>Waterfall</i> .....	22
2.3.1	Pengertian Model <i>Waterfall</i> .....	22
2.3.2	Tahapan Pengembangan <i>Waterfall</i> .....	22
2.4	Bahasa Pemrograman <i>Python</i> .....	24
2.5	Analisis Kebutuhan Sistem .....	25
2.6	Perancangan Sistem.....	26
2.6.1	<i>Flowchart</i> .....	26
2.6.2	<i>ERD (Entity Relationship Diagram)</i> .....	27
2.6.3	<i>DFD (Data Flow Diagram)</i> .....	30
<b>BAB III ANALISIS DAN PERANCANGAN .....</b>		<b>32</b>
3.1	Analisis Masalah .....	32
3.2	Analisis Kebutuhan Penelitian .....	34

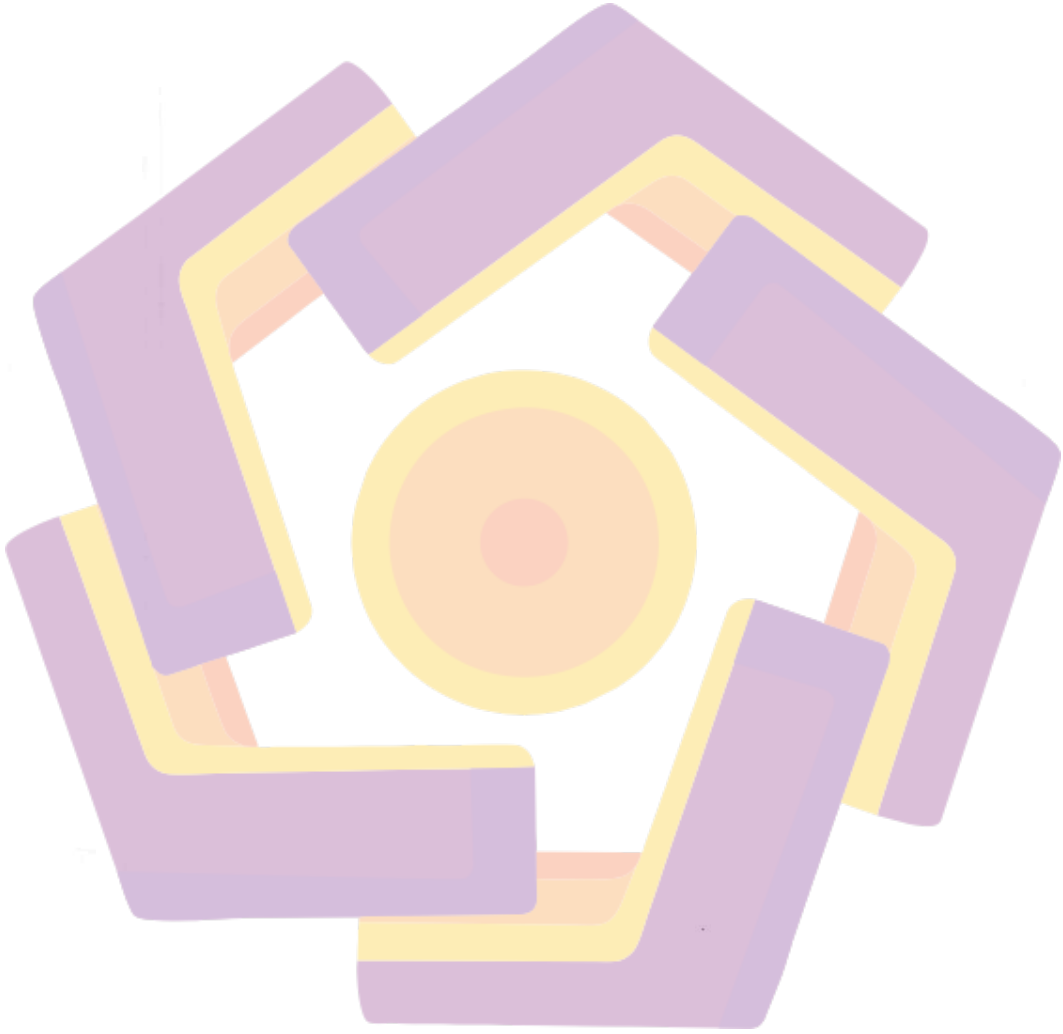
3.2.1	Analisis Kebutuhan Data.....	34
3.2.2	Analisis Kebutuhan Fungsional .....	37
3.2.3	Analisis Kebutuhan Non-Fungsional .....	39
3.3	Gambaran Umum Sistem .....	40
3.4	Perancangan Sistem.....	51
3.4.1	<i>ERD (Entity Relationship Diagram)</i> .....	52
3.4.2	Rancangan Struktur Tabel.....	55
3.4.3	<i>Data Flow Diagram (DFD)</i> .....	60
3.4.4	Flowchart Sistem.....	66
3.4.5	Perancangan Antarmuka Pengguna.....	78
<b>BAB IV IMPLEMENTASI DAN PEMBAHASAN .....</b>		<b>93</b>
4.1	Implementasi .....	93
4.1.1	Implementasi Pengumpulan Data .....	93
4.1.2	Implementasi Program .....	96
4.2	Pengujian Sistem .....	132
4.3	Evaluasi Model.....	133
<b>BAB V PENUTUP.....</b>		<b>152</b>
5.1	Kesimpulan.....	152
5.2	Saran .....	153
<b>DAFTAR PUSTAKA .....</b>		<b>154</b>

## DAFTAR TABEL

Table 2.1	Contoh pemecahan <i>n-gram</i> .....	16
Table 2.2	Tabel <i>Confusion Matrix</i> .....	21
Table 2.3	Simbol <i>Flowchart</i> Sistem.....	26
Table 2.4	Simbol <i>Entity Relationship Diagram</i> .....	28
Table 2.5	Simbol-simbol <i>DFD</i> .....	31
Tabel 3.1	Tabel kelas <i>DDC</i> yang didapat .....	36
Tabel 3.2	Contoh <i>case folding</i> .....	43
Tabel 3.3	Contoh <i>filtering</i> .....	44
Tabel 3.4	Contoh <i>tokenization</i> menggunakan <i>N-Gram</i> .....	44
Tabel 3.5	Contoh <i>stemming</i> menggunakan <i>Sastrawi Stemmer</i> .....	45
Tabel 3.6	Contoh perhitungan <i>Tf-Idf</i> .....	47
Tabel 3.7	Tabel variasi skenario <i>preprocessing</i> .....	49
Tabel 3.8	Contoh Tabel <i>Confusion Matrix</i> .....	50
Tabel 3.9	Rancangan tabel <i>training</i> .....	56
Tabel 3.10	Rancangan tabel <i>comparation</i> .....	56
Tabel 3.11	Rancangan tabel <i>testing</i> .....	57
Tabel 3.12	Rancangan tabel <i>result</i> .....	57
Tabel 3.13	Rancangan tabel <i>ddc</i> .....	58
Tabel 3.14	Rancangan tabel <i>prior_probability</i> .....	58
Tabel 3.15	Rancangan tabel <i>evidence</i> .....	59
Tabel 3.16	Rancangan tabel <i>user</i> .....	59
Tabel 3.17	Rancangan tabel <i>settings</i> .....	60
Tabel 3.18	Rancangan tabel <i>log</i> .....	60

Tabel 4.1	Hasil pengujian sistem .....	132
Tabel 4.2	Hasil tabel <i>Confusion Matrix</i> pada sekenario <i>preprocessing</i> 1.....	134
Tabel 4.3	Hasil kinerja model pada sekenario <i>preprocessing</i> 1.....	135
Tabel 4.4	Hasil tabel <i>Confusion Matrix</i> pada sekenario <i>preprocessing</i> 2.....	135
Tabel 4.5	Hasil kinerja model pada sekenario <i>preprocessing</i> 2.....	136
Tabel 4.6	Hasil tabel <i>Confusion Matrix</i> pada sekenario <i>preprocessing</i> 3.....	136
Tabel 4.7	Hasil kinerja model pada sekenario <i>preprocessing</i> 3.....	137
Tabel 4.8	Hasil tabel <i>Confusion Matrix</i> pada sekenario <i>preprocessing</i> 4.....	137
Tabel 4.9	Hasil kinerja model pada sekenario <i>preprocessing</i> 4.....	138
Tabel 4.10	Hasil tabel <i>Confusion Matrix</i> pada sekenario <i>preprocessing</i> 5.....	138
Tabel 4.11	Hasil kinerja model pada sekenario <i>preprocessing</i> 5.....	139
Tabel 4.12	Hasil tabel <i>Confusion Matrix</i> pada sekenario <i>preprocessing</i> 6.....	139
Tabel 4.13	Hasil kinerja model pada sekenario <i>preprocessing</i> 6.....	140
Tabel 4.14	Hasil tabel <i>Confusion Matrix</i> pada sekenario <i>preprocessing</i> 7.....	140
Tabel 4.15	Hasil kinerja model pada sekenario <i>preprocessing</i> 7.....	141
Tabel 4.16	Hasil tabel <i>Confusion Matrix</i> pada sekenario <i>preprocessing</i> 8.....	141
Tabel 4.17	Hasil kinerja model pada sekenario <i>preprocessing</i> 8.....	142
Tabel 4.18	Hasil tabel <i>Confusion Matrix</i> pada sekenario <i>preprocessing</i> 9.....	142
Tabel 4.19	Hasil kinerja model pada sekenario <i>preprocessing</i> 9.....	143
Tabel 4.20	Hasil tabel <i>Confusion Matrix</i> pada sekenario <i>preprocessing</i> 10...	143
Tabel 4.21	Hasil kinerja model pada sekenario <i>preprocessing</i> 10.....	144
Tabel 4.22	Hasil tabel <i>Confusion Matrix</i> pada sekenario <i>preprocessing</i> 11...	144
Tabel 4.23	Hasil kinerja model pada sekenario <i>preprocessing</i> 11.....	145
Tabel 4.24	Hasil tabel <i>Confusion Matrix</i> pada sekenario <i>preprocessing</i> 12...	145
Tabel 4.25	Hasil kinerja model pada sekenario <i>preprocessing</i> 12.....	146

Tabel 4.26 Hasil tabel *Confusion Matrix* pada sekenario *preprocessing* 13... 146  
Tabel 4.27 Hasil kinerja model pada sekenario *preprocessing* 13..... 147  
Tabel 4.28 Hasil tabel *Confusion Matrix* pada sekenario *preprocessing* 14... 147  
Tabel 4.29 Hasil kinerja model pada sekenario *preprocessing* 14..... 148



## DAFTAR GAMBAR

Gambar 2.1	Contoh proses penghapusan <i>stopword</i> .....	14
Gambar 2.2	Contoh proses <i>stemming</i> kata dengan <i>Sastrawi stemmer</i> .....	15
Gambar 2.3	Tahap klasifikasi <i>Naïve Bayes Classifier</i> .....	18
Gambar 2.4	Tahap pengembangan <i>waterfall</i> .....	24
Gambar 3.1	Contoh data <i>Simpus UII</i> .....	34
Gambar 3.2	Contoh data judul dengan kesalahan kata .....	35
Gambar 3.3	Contoh data judul dengan kelas yang tidak sesuai .....	35
Gambar 3.4	Contoh data dengan kelas yang tidak terdaftar pada <i>DDC</i> .....	35
Gambar 3.5	Alur penelitian perancangan sistem .....	40
Gambar 3.6	Alur <i>web scrapping</i> penelitian .....	41
Gambar 3.7	Proses <i>preprocessing</i> .....	43
Gambar 3.8	Proses klasifikasi <i>Naïve Bayes Classifier</i> .....	48
Gambar 3.9	<i>ERD</i> sistem .....	52
Gambar 3.10	Diagram Konteks Sistem.....	61
Gambar 3.11	<i>Data flow diagram</i> level 0.....	62
Gambar 3.12	<i>Data flow diagram</i> level 1 proses manipulasi data .....	63
Gambar 3.13	<i>Data flow diagram</i> level 1 proses <i>Learning/Pelatihan</i> .....	64
Gambar 3.14	<i>Data flow diagram</i> level 1 proses <i>Classification/Klasifikasi</i> .....	65
Gambar 3.15	<i>Data flow diagram</i> level 1 proses <i>Compare/Komparasi</i> .....	66
Gambar 3.16	<i>Flowchart</i> proses <i>login</i> .....	67
Gambar 3.17	<i>Flowchart</i> edit data <i>user</i> .....	67
Gambar 3.18	<i>Flowchart</i> edit data <i>settings</i> .....	68
Gambar 3.19	<i>Flowchart</i> tampil data, <i>add data</i> , dan <i>delete data DDC</i> .....	69



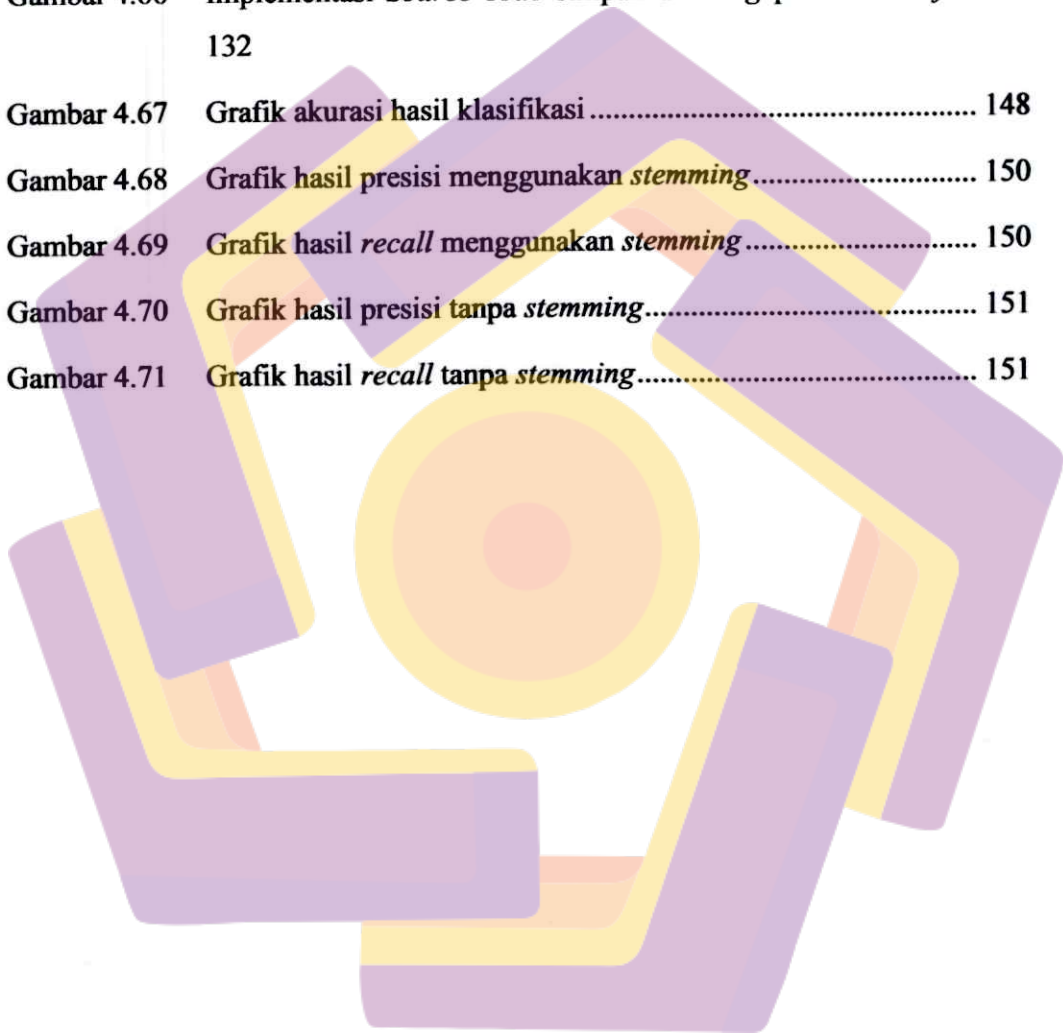
Gambar 3.20	<i>Flowchart edit data, import data, dan clear all data DDC .....</i>	69
Gambar 3.21	<i>Flowchart tampil data, add data, dan delete data training .....</i>	70
Gambar 3.22	<i>Flowchart edit data, import data, dan clear all data training.....</i>	71
Gambar 3.23	<i>Flowchart tampil data, add data, dan delete data testing .....</i>	72
Gambar 3.24	<i>Flowchart edit data, import data, dan clear all data testing .....</i>	72
Gambar 3.25	<i>Flowchart tampil data, add data, dan delete data comparison ..</i>	73
Gambar 3.26	<i>Flowchart edit data, import data, dan clear all data comparison</i> 74	
Gambar 3.27	<i>Flowchart proses preprocessing.....</i>	75
Gambar 3.28	<i>Flowchart proses learning.....</i>	76
Gambar 3.29	<i>Flowchart proses classification/klasifikasi .....</i>	76
Gambar 3.30	<i>Flowchart komparasi data .....</i>	77
Gambar 3.31	<i>Flowchart edit data log.....</i>	77
Gambar 3.32	Rancangan halaman <i>login</i> .....	78
Gambar 3.33	Rancangan navigasi sistem.....	79
Gambar 3.34	Rancangan halaman <i>Home</i> .....	79
Gambar 3.35	Rancangan halaman <i>Classification</i> .....	80
Gambar 3.36	Halaman menu <i>Learning</i> .....	81
Gambar 3.37	Halaman menu <i>Settings</i> .....	81
Gambar 3.38	Rancangan halaman <i>Classification Process</i> .....	82
Gambar 3.39	Rancangan halaman <i>Comparison Result</i> .....	83
Gambar 3.40	Rancangan halaman <i>Testing Data/Data Testing</i> .....	83
Gambar 3.41	Halaman <i>Comparison Data</i> .....	84
Gambar 3.42	Halaman <i>Learning Process</i> .....	85
Gambar 3.43	Rancangan halaman <i>Training Data/Data Training</i> .....	85

Gambar 3.44	Rancangan halaman <i>DDC Data</i> .....	86
Gambar 3.45	Rancangan halaman <i>Preprocessing Settings</i> .....	87
Gambar 3.46	Rancangan halaman <i>Authentication Settings</i> .....	87
Gambar 3.47	Rancangan antarmuka <i>Add Data Training dan Comparison</i> .....	88
Gambar 3.48	Rancangan antarmuka <i>Add Data DDC</i> .....	89
Gambar 3.49	Rancangan antarmuka <i>Add data Testing</i> .....	89
Gambar 3.50	Rancangan antarmuka <i>Edit Data Training dan Comparison</i> .....	90
Gambar 3.51	Rancangan antarmuka <i>Edit Data DDC</i> .....	91
Gambar 3.52	Rancangan antarmuka <i>Edit Data Testing</i> .....	91
Gambar 3.53	Rancangan antarmuka <i>Import Data Training, DDC, Testing, dan Comparison</i> .....	92
Gambar 4.1	Kode program pengambilan dokumen <i>html</i> .....	93
Gambar 4.2	Kode program penyaringan dokumen <i>html</i> .....	94
Gambar 4.3	Kode program penyaringan dokumen <i>html</i> .....	95
Gambar 4.4	Kode program penyimpanan file <i>CSV</i> .....	95
Gambar 4.5	Data dari hasil file <i>CSV</i> .....	96
Gambar 4.6	Hasil implementasi pembuatan <i>database db_classification</i> .....	97
Gambar 4.7	Hasil implementasi relasi tabel <i>database db_classification</i> .....	97
Gambar 4.8	Hasil implementasi tabel <i>training</i> .....	98
Gambar 4.9	Hasil implementasi tabel <i>comparation</i> .....	99
Gambar 4.10	Hasil implementasi tabel <i>testing</i> .....	99
Gambar 4.11	Hasil implementasi tabel <i>result</i> .....	100
Gambar 4.12	Hasil implementasi tabel <i>ddc</i> .....	101
Gambar 4.13	Hasil implementasi tabel <i>prior_probability</i> .....	101
Gambar 4.14	Hasil implementasi tabel <i>evidence</i> .....	102

Gambar 4.15	Hasil implementasi tabel <i>user</i> .....	103
Gambar 4.16	Hasil implementasi tabel <i>settings</i> .....	103
Gambar 4.17	Hasil implementasi tabel <i>log</i> .....	104
Gambar 4.18	Hasil implementasi halaman <i>Login</i> .....	105
Gambar 4.19	Hasil implementasi navigasi sistem .....	105
Gambar 4.20	Hasil implementasi halaman <i>Home</i> .....	106
Gambar 4.21	Hasil implementasi halaman <i>Classification</i> .....	106
Gambar 4.22	Hasil implementasi menu <i>Learning</i> .....	107
Gambar 4.23	Hasil implementasi menu <i>Settings</i> .....	107
Gambar 4.24	Hasil implementasi halaman <i>Classification Process</i> .....	108
Gambar 4.25	Hasil implementasi halaman <i>Comparison Result</i> .....	109
Gambar 4.26	Hasil implementasi halaman <i>Testing Data/Data Testing</i> .....	109
Gambar 4.27	Hasil implementasi <i>Comparison Data</i> .....	110
Gambar 4.28	Hasil implementasi halaman <i>Learning Process</i> .....	111
Gambar 4.29	Hasil implementasi halaman <i>Training Data/Data Training</i> .....	111
Gambar 4.30	Hasil implementasi halaman <i>DDC Data</i> .....	112
Gambar 4.31	Hasil implementasi halaman <i>Preprocessing Settings</i> .....	113
Gambar 4.32	Hasil implementasi halaman <i>Authentication Data</i> .....	113
Gambar 4.33	Hasil implementasi antarmuka <i>Add Data Training</i> .....	114
Gambar 4.34	Hasil implementasi antarmuka <i>Add Data Comparison</i> .....	114
Gambar 4.35	Hasil implementasi antarmuka <i>Add Data DDC</i> .....	115
Gambar 4.36	Hasil implementasi antarmuka <i>Add data Classification</i> .....	116
Gambar 4.37	Hasil implementasi antarmuka <i>Edit Data Training</i> .....	116
Gambar 4.38	Hasil implementasi antarmuka <i>Edit Data Comparison</i> .....	117
Gambar 4.39	Hasil implementasi antarmuka <i>Edit Data DDC</i> .....	117

Gambar 4.40	Hasil implementasi antarmuka Edit Data <i>Testing</i> .....	118
Gambar 4.41	Hasil implementasi antarmuka <i>Import Data Training</i> .....	119
Gambar 4.42	Hasil implementasi antarmuka <i>Import Data DDC</i> .....	119
Gambar 4.43	Hasil implementasi antarmuka <i>Import Data Testing</i> .....	119
Gambar 4.44	Hasil implementasi antarmuka <i>Import Data Comparison</i> .....	120
Gambar 4.45	Potongan <i>source code</i> pada file model.....	121
Gambar 4.46	<i>Source code</i> untuk <i>query database</i> yang rumit .....	121
Gambar 4.47	Implementasi <i>source code</i> inialisasi <i>library</i> yang digunakan.	122
Gambar 4.48	Hasil modifikasi <i>stopword</i> .....	122
Gambar 4.49	Implementasi <i>source code</i> inialisasi pada fungsi <i>preprocessing</i>	123
Gambar 4.50	Implementasi <i>Source code</i> proses case folding.....	123
Gambar 4.51	Implementasi <i>Source code</i> proses <i>filtering</i> .....	123
Gambar 4.52	Implementasi <i>Source code</i> proses <i>tokenizing</i> .....	124
Gambar 4.53	Implementasi <i>Source code</i> fungsi penghapusan tanda baca.....	124
Gambar 4.54	Implementasi <i>Source code</i> fungsi pemecahan kata .....	125
Gambar 4.55	Implementasi <i>Source code</i> inialisasi proses <i>learning</i> .....	125
Gambar 4.56	Implementasi <i>Source code</i> hitung nilai probabilitas kelas .....	126
Gambar 4.57	Implementasi <i>Source code</i> <i>preprocessing</i> data <i>training</i> .....	126
Gambar 4.58	Implementasi <i>Source code</i> proses pertama <i>term weighting</i> .....	127
Gambar 4.59	Implementasi <i>Source code</i> perhitungan nilai <i>tf-idf</i> .....	128
Gambar 4.60	Implementasi <i>Source code</i> hitung probabilitas <i>evidence</i> terhadap kelas hipotesis	128
Gambar 4.61	Implementasi <i>Source code</i> simpan data <i>log</i> proses <i>learning</i> .....	129
Gambar 4.62	Implementasi <i>Source code</i> inialisasi proses <i>classification</i> .....	129

Gambar 4.63	Implementasi <i>Source code preprocessing data testing</i> .....	130
Gambar 4.64	Implementasi <i>Source code penyaringan kata proses classification</i> 130	
Gambar 4.65	Implementasi <i>Source code proses klasifikasi kelas dokumen ...</i>	131
Gambar 4.66	Implementasi <i>Source code simpan data log proses classification</i> 132	
Gambar 4.67	Grafik akurasi hasil klasifikasi .....	148
Gambar 4.68	Grafik hasil presisi menggunakan <i>stemming</i> .....	150
Gambar 4.69	Grafik hasil <i>recall</i> menggunakan <i>stemming</i> .....	150
Gambar 4.70	Grafik hasil presisi tanpa <i>stemming</i> .....	151
Gambar 4.71	Grafik hasil <i>recall</i> tanpa <i>stemming</i> .....	151



## INTISARI

Di perpustakaan perguruan tinggi, dokumen skripsi diklasifikasikan. Umumnya pengklasifikasian dokumen tersebut menggunakan standar *DDC (Dewey Decimal Classification)*. Dengan menggunakan konsep *text mining*, proses klasifikasi dapat dilakukan dengan komputerisasi, menghemat waktu, usaha dan pemikiran.

Penelitian ini menerapkan *term weighting tf-idf* dan menggunakan metode klasifikasi *Naive Bayes Classifier*. Penelitian ini menggunakan data sampel yang diambil dari <https://simpus.uii.ac.id>. Ada 4 kelas dengan masing-masing 300 data. Evaluasi penelitian menggunakan metode *Confusion Matrix* dengan 14 skenario preprocessing. Skenario ini dibedakan berdasarkan bentuk pemecahan *unigram*, *bigram*, *trigram*, beserta kombinasinya, dan ada atau tidaknya proses *stemming*.

Penelitian ini menghasilkan akurasi terbaik dalam menyelesaikan kata menggunakan *unigram*, *bigram*, *trigram* secara bersamaan dan tanpa *stemming*, dengan akurasi 97,92%. Penggunaan kombinasi pemecahan kata meningkatkan akurasi model Klasifikasi *Naive Bayes* yang diimplementasikan dalam penelitian ini. *Stemming* akan bekerja lebih baik dalam skenario yang tidak melibatkan pemecahan *unigram*. Akurasi akan meningkat dengan menerapkan *stemming* dalam skenario pemecahan *bigram* dan *trigram*.

**Kata Kunci:** *text mining, classification, n-gram, tf-idf, naïve bayes classifier, confusion matrix.*

## **ABSTRACT**

*In the college library, the thesis document is classified. Generally the classification of documents uses DDC (Dewey Decimal Classification) standards. By using the concept of text mining, the classification process can be done with computerization, saving time, effort and thought.*

*This study applies the tf-idf weighting term and uses the Naive Bayes Classifier classification method. This study uses sample data taken from <https://simpus.uui.ac.id>. There are 4 classes with 300 data each. Research evaluation using Confusion Matrix method with 14 preprocessing scenarios. This scenario is distinguished based on the form of solving unigram, bigram, trigram, along with its combination, and whether or not there is a stemming process.*

*This study produced the best accuracy in completing words using unigram, bigram, trigram simultaneously and without stemming, with an accuracy of 97.92%. The use of word solving combinations increases the accuracy of the Naive Bayes Classification model implemented in this study. Stemming will work better in scenarios that do not involve solving unigram. Accuracy will increase by applying stemming in bigram and trigram solving scenarios.*

**Keyword:** *text mining, classification, n-gram, tf-idf, naïve bayes classifier, confusion matrix.*