

SKRIPSI
KLASIFIKASI EMOSI PADA LIRIK LAGU BERBAHASA INDONESIA
MENGGUNAKAN ALGORITMA *NAÏVE BAYES CLASSIFIER*



Disusun Oleh :

SLIS AYU WULAN SARI

DBC 116 027

JURUSAN TEKNIK INFORMATIKA
FAKULTAS TEKNIK
UNIVERSITAS PALANGKA RAYA
2020

SKRIPSI

**KLASIFIKASI EMOSI PADA LIRIK LAGU BERBAHASA INDONESIA
MENGUNAKAN ALGORITMA *NAIVE BAYES CLASSIFIER***

Sebagai salah satu syarat menyelesaikan Program Strata - I
pada Jurusan Teknik Informatika Fakultas Teknik Universitas Palangka Raya

OLEH :


SLIS AYU WULAN SARI

NIM. DBC 116 027

Disetujui untuk diajukan dalam Seminar Akhir Skripsi,

Palangka Raya, November 2020

Pembimbing I



ARRESTA LESTARI, S.Kom., M.Cs., Ph.D
NIP. 19800322 200501 2 004

Pembimbing II



SHERLY CHRISTINA, S.Kom., M.Kom
NIP. 19810929 200604 2 001

**JURUSAN TEKNIK INFORMATIKA
FAKULTAS TEKNIK
UNIVERSITAS PALANGKA RAYA**

2020

**KLASIFIKASI EMOSI PADA LIRIK LAGU BERBAHASA INDONESIA
MENGUNAKAN ALGORITMA *NAÏVE BAYES CLASSIFIER***

SKRIPSI

Sebagai salah satu syarat untuk menyelesaikan Program Strata-1 pada Jurusan Teknik
Informatika Fakultas Teknik Universitas Palangka Raya

Oleh

Slis Ayu Wulan Sari

DBC 116 027

Telah dipertahankan didepan tim penguji, pada :

Hari/Tanggal : Kamis, 12 November 2020

Waktu : 11.00-12.30 WIB

- | | |
|---|--|
| 1. Drs. JADIAMAN PARHUSIP, M.Kom
NIP. 19630423 198502 1 001 | :  (Ketun) |
| 2. ARIESTA LESTARI, S.Kom., M.Cs., Ph.D
NIP. 19800322 200501 2 004 | :  (Anggota) |
| 3. SHERLY CHRISTINA, S.Kom., M.Kom
NIP. 19810929 200604 2 001 | :  (Anggota) |
| 4. ADE CHANDRA SAPUTRA, S.Kom., M.Cs.
NIP. 19870203 201404 1 001 | :  (Anggota) |
| 5. EFRANS CHRISTIAN, S.T., M.T
NIP. 19910630 201903 1 013 | :  (Anggota) |

Mengetahui :

Fakultas Teknik
Universitas Palangka Raya
Dekan.



Ir. WALUKO NUSWANTORO, M.T.
NIP. 19651119 199302 1 001

Jurusan / Program Studi Teknik Informatika
Fakultas Teknik Universitas Palangka Raya
Ketua Jurusan,

ABERTUN SAGIT SAHAY, S.T., M.Eng
NIP. 19751212 200312 1 002

PERNYATAAN

Dengan ini saya menyatakan dengan sebenar - benarnya bahwa dalam Skripsi ini tidak terdapat karya ilmiah yang pernah diajukan untuk memperoleh gelar kesarjanaan disuatu Perguruan Tinggi, serta tidak terdapat karya ilmiah atau pendapat yang pernah ditulis atau diterbitkan orang lain, kecuali secara tertulis diacu dalam Skripsi ini dan disebutkan dalam Tinjauan Pustaka.

Palangka Raya, November 2020



Slis Ayu Wulan Sari
DBC 116 027

RIWAYAT PENYUSUN

Data Diri

Nama : Slis Ayu Wulan Sari
NIM : DBC 116 027
Fakultas : Teknik
Jurusan/Program Studi : Teknik Informatika
Jenjang : Strata 1 (S-1)
Jenis Kelamin : Perempuan
Tempat, Tanggal Lahir : Mulyasari, 16 Februari 1999
Agama : Islam
Status dalam Keluarga : Anak Kandung
Anak ke - : I (Satu)
Alamat : Jl. Yogyakarta IV
No. Telpon/HP : +6282247171408



Data Orang Tua

Nama Ayah : Rahmadi
Pekerjaan Ayah : Swasta
Nama Ibu : Sarinem
Pekerjaan Ibu : Ibu Rumah Tangga
Alamat Orang Tua : Jl. Yogyakarta IV
No. Telpon/HP : +6282250004521

Riwayat Pendidikan *)

SD : SDN Pangkoh Sari 4 (Tahun Lulus 2010)
SMP : SMPN 6 Palangka Raya (Tahun Lulus 2013)
SMA : SMAN 4 Palangka Raya (Tahun Lulus 2016)

Palangka Raya, November 2020

Slis Ayu Wulan Sari

DBC 116 027

Keterangan:

*) Nama, Tempat, Tahun Lulus

HALAMAN PERSEMBAHAN

Tuhan Yang Maha Esa

Puji dan syukur selalu saya panjatkan kepada Allah S.W.T, karena berkat rahmat dan hidayah-Nya lah saya dapat menyelesaikan Skripsi, baik dari awal semester perkuliahan hingga saya mendapatkan gelar Sarjana saya selama 4 tahun lebih ini. Skripsi ini saya persembahkan untuk orang-orang yang telah mendukung saya hingga saat ini.

Orang Tua & Adik

Terima kasih kepada Papah, Mamah, dan adik saya, Adhi, yang saya cintai dan sayangi, yang selalu mendoakan dan percaya dengan apa yang saya pilih dari awal hingga saat ini. Terima kasih karena telah mendukung dan memberikan kasih sayang serta perhatian setiap harinya. Terima kasih sudah selalu merasa bangga terhadap hal apapun yang saya capai selama ini.

Dosen Pembimbing & Dosen Penguji

Terima kasih kepada Ibu Ariesta Lestari, S.Kom.,M.Cs.,Ph.D sebagai Dosen Pembimbing I dan Ibu Sherly Christina, S.Kom.,M.Kom sebagai Dosen Pembimbing II yang telah meluangkan waktunya untuk membimbing, memberikan masukan, serta memberikan arahan kepada saya selama saya mengerjakan Skripsi ini dan akhirnya dapat saya selesaikan. Kepada Pak Ade Chandra Saputra, S.Kom., M.Cs, Pak Efrans Christian, S.T.,M.T, dan Pak Drs. Jadianan Parhusip, M. Kom, sebagai Dosen Penguji, saya mengucapkan banyak terima kasih untuk masukan-masukan yang telah diberikan untuk saya.

Sahabat & GSC

Terima kasih kepada Petra Wara Nandayani R., Zakharia, Muhammad Riski Saputra, dan Siti Nor Azizah untuk waktu 24/7 nya selama masa kuliah. Terima kasih untuk GSC, sobat-sobat receh yang suka membual dan ngajak nongkrong

cuma kadang-kadang. Terima kasih sudah menjadi teman-teman yang menghibur, memotivasi, dan teman-teman yang bisa diandalkan. Buat Naomi Floriyentina yang sudah seperti Ibu selama KKN dan selalu memberi semangat sampai saat ini, terima kasih banyak. Terima kasih sudah menjadi salah satu *part* terbaik yang saya miliki.

Teman-Teman TI Angkatan 2016

Terima kasih untuk teman-teman Teknik Informatika Angkatan 2016 atas 4 tahun 4 bulannya selama ada di Jurusan Teknik Informatika semoga selalu semangat dan sukses selalu di masa sekarang atau di masa yang akan datang. Tidak lupa saya ucapkan terima kasih untuk pihak-pihak yang selalu bertanya ‘kapan lulus?’, sehingga memotivasi dan membakar semangat saya dalam mengerjakan skripsi.

Dosen Teknik Informatika & Staf Akademik

Terima kasih untuk semua dosen Teknik Informatika atas ilmu dan waktunya untuk mengajar dan mendidik kami selama ini sebagai bekal kami kedepannya. Terima kasih untuk Staf Akademik yang telah membantu selama disini.

Spesial

Terima kasih untuk Seventeen yang menjadi salah satu teman dan menyalurkan semangat dan memberikan hiburan disela-sela masa perkuliahan. Terima kasih untuk diri saya sendiri dan semuanya, yang tidak pernah menyerah walau banyak kesulitan yang dihadapi. Kita sudah bekerja keras!

*try your best...but maybe not sometimes
it's okay to take a break on our journey of youth
there's a long journey ahead of you*

;

KATA PENGANTAR

Puji syukur kepada Allah S.W.T, Tuhan Yang Maha Esa, atas segala berkat dan rahmat-Nya sehingga penulis dapat menyelesaikan Laporan Skripsi yang berjudul “Klasifikasi Emosi Pada Lirik Lagu Berbahasa Indonesia Menggunakan Algoritma *Naïve Bayes Classifier*”.

Penulis menyadari sepenuhnya, bahwa dalam penyusunan laporan ini masih jauh dari sempurna dalam penyelesaiannya tidak lepas dari bimbingan, arahan, dan bantuan dari berbagai pihak, dalam kesempatan ini perkenankan penulis untuk mengucapkan terima kasih kepada yang terhormat Ibu Ariesta Lestari, S.Kom., M.Cs., Ph.D., sebagai Dosen Pembimbing I dan Ibu Sherly Christuba, S.Kom., M.Kom., sebagai Dosen Pembimbing II, yang telah meluangkan waktunya untuk membimbing dan memberikan arahan serta masukan kepada penulis. Serta seluruh rekan-rekan yang telah membantu dalam pembuatan laporan ini.

Penulis menyadari bahwa laporan ini masih jauh dari sempurna dari penulisan maupun pemahaman. Oleh karena itu, penulis mengharapkan pendapat, kritik, dan saran yang membangun untuk mengembangkan dan menyempurnakan laporan ini. Semoga laporan ini dapat dipahami pembaca, serta dapat memberikan manfaat dan masukan bagi pembaca.

Palangka Raya, November 2020

Penulis

“KLASIFIKASI EMOSI PADA LIRIK LAGU BERBAHASA INDONESIA MENGUNAKAN ALGORITMA *NAIVE BAYES CLASSIFIER*”

Slis Ayu Wulan Sari DBC 116 027

Jurusan Teknik Informatika Fakultas Teknik Universitas Palangka Raya

Kampus Tanjung Nyaho Jl. Yos Sudarso Palangka Raya 73112

E-mail : ayuslis16@gmail.com

Abstrak

Menentukan emosi berdasarkan lirik lagu biasanya dilakukan manual oleh manusia. Hingga saat ini terdapat beberapa situs *website online* yang menyediakan kumpulan lirik lagu tetapi pada situs tersebut tidak menyediakan klasifikasi lirik lagu berdasarkan emosi dari lagu tersebut. Hal ini menyebabkan orang kesulitan dalam mencari lirik lagu berdasarkan emosinya. Oleh karena itu dibutuhkan suatu sistem yang dapat mengklasifikasikan emosi pada lirik lagu secara otomatis agar seseorang dapat memilih lagu berdasarkan emosi yang diinginkan, selain itu terdapat lagu dan video yang dapat diputar oleh pengguna.

Klasifikasi lirik lagu berbahasa Indonesia pada penelitian ini menggunakan metode pengembangan perangkat lunak *Waterfall*. Tahapan yang digunakan, yaitu *communication, planning, modeling, construction, dan deployment*. Analisis sistem dibuat menggunakan *flowchart* dan desain sistem menggunakan *Unified Modeling Language (UML)*. Tahapan *text mining* yang digunakan, yaitu *case folding, tokenizing, filtering*, pembobotan dengan *TF-IDF* dan klasifikasi menggunakan *Naive Bayes Classifier*. Metode *testing* yang digunakan dalam penelitian ini adalah metode akurasi dan *black box*.

Hasil dari pengujian *black box* dari proses analisis hingga operasi diperoleh bahwa sistem dapat digunakan untuk melakukan klasifikasi emosi pada lirik lagu berbahasa Indonesia. Hasil akurasi yang didapat dari empat kali skenario pengujian, didapat komposisi data yang optimal pada skenario pengujian keempat dengan komposisi data latih 90% dan data uji 10% dengan hasil akurasi 85%. Sistem klasifikasi emosi pada lirik lagu berbahasa Indonesia ini dapat ditemukan di pencarian dengan alamat *colourlyrics.com*.

Kata Kunci : *Text Mining, Lirik Lagu Berbahasa Indonesia, Naive Bayes Classifier, Klasifikasi*

“CLASSIFICATION OF EMOTIONS IN INDONESIAN SONG LYRICS USING NAÏVE BAYES CLASSIFIER ALGORITHM”

Slis Ayu Wulan Sari DBC 116 027

Jurusan Teknik Informatika Fakultas Teknik Universitas Palangka Raya

Kampus Tanjung Nyaho Jl. Yos Sudarso Palangka Raya 73112

E-mail : ayuslis16@gmail.com

Abstract

Determining emotions based on song lyrics is usually done manually by humans. Until now, several online websites provide a collection of song lyrics but these sites do not provide a classification of song lyrics based on the emotion of the song. This makes it difficult for people to find song lyrics based on their emotions. Therefore we need a system that can classify emotions in song lyrics automatically so that someone can choose songs based on the desired emotion, besides that some songs and videos can be played by users.

The classification of song lyrics in Indonesian in this study uses the Waterfall software development method. The stages used are communication, planning, modeling, construction, and deployment. System analysis is made using flowcharts and system design using Unified Modeling Language (*UML*). The text mining stages used are case folding, tokenizing, filtering, weighting with *TF-IDF*, and classification using the Naive Bayes Classifier. The testing method used in this research is the accuracy and black box method.

The results of the black box test from the analysis process to the operation show that the system can be used to classify emotions in Indonesian song lyrics. The accuracy results obtained from four test scenarios obtained the optimal data composition in the fourth test scenario with the composition of 90% training data and 10% test data with 85% accuracy results. The emotion classification system for song lyrics in Indonesian can be found in a search with the address *colourlyrics.com*.

Keywords : Text Mining, Indonesian Language Song Lyrics, Naïve Bayes Classifier, Classification

DAFTAR ISI

	<i>Halaman</i>
HALAMAN JUDUL	i
HALAMAN PERSETUJUAN	ii
HALAMAN PENGESAHAN.....	iii
HALAMAN PERNYATAAN.....	iv
HALAMAN RIWAYAT PENYUSUN.....	v
HALAMAN PERSEMBAHAN	vi
KATA PENGANTAR.....	viii
ABSTRAK	ix
ABSTRACT	x
DAFTAR ISI.....	xi
DAFTAR GAMBAR.....	xiv
DAFTAR TABEL	xvii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Batasan Masalah.....	3
1.4 Tujuan.....	6
1.5 Manfaat.....	6
1.6 Sistematika Penulisan.....	7
1.7 Jadwal.....	9
BAB II LANDASAN TEORI	10

2.1	Kajian Pustaka	10
2.2	Emosi.....	12
2.2.1	Macam-Macam Emosi	13
2.3	Lirik Lagu.....	15
2.4	<i>Text Mining</i>	16
2.4.1	Klasifikasi	16
2.5	Tahapan <i>Text Mining</i>	17
2.5.1	<i>Pre-Processing</i>	17
2.5.2	<i>Text Transformation</i>	18
2.5.3	<i>Feature Selection</i>	19
2.5.4	<i>Pattern Discovery</i>	20
2.6	<i>Term Frequency – Inverse Document Frequency</i>	20
2.7	Teorema <i>Bayes</i>	22
2.7.1	<i>Naïve Bayes</i> Untuk Klasifikasi.....	23
2.8	Metode Pengembangan Perangkat Lunak <i>Waterfall</i>	27
2.9	Pengujian	29
2.8.1	Akurasi	29
2.8.2	Black Box Testing.....	31
BAB III METODOLOGI PENELITIAN		32
3.1	Metode Pengumpulan Data	32
3.2	Metode Klasifikasi	33
3.2.1	Ekstraksi Data	34
3.2.2	Pemrosesan Teks	42

3.2.3	Pengujian Model Klasifikasi	65
3.3	Skenario Pengujian.....	75
3.3.1	Black Box Testing.....	76
3.3.2	Akurasi	77
3.4	Metode Pengembangan Perangkat Lunak	78
3.4.1	<i>Communication (Project Initiation & Requirements Gathering)</i> ...	79
3.4.2	<i>Planning (Estimating, Scheduling, & Tracking)</i>	79
3.4.3	<i>Modeling (Analysis & Design)</i>	82
3.4.4	<i>Construction (Code & Test)</i>	108
3.4.5	<i>Deployment (Delivery, Support, & Feedback)</i>	109
BAB IV HASIL DAN PEMBAHASAN		111
4.1	<i>Construction (Code & Test)</i>	110
4.1.1	Implementasi Halaman Admin	110
4.1.2	Implementasi Halaman Pendengar.....	116
4.1.3	Pengujian Sistem.....	120
4.2	<i>Deployment (Delivery, Support, & Feedback)</i>	142
BAB V PENUTUP.....		137
5.1	Kesimpulan.....	137
5.2	Saran	138
DAFTAR PUSTAKA		137

DAFTAR GAMBAR

	<i>Halaman</i>
Gambar 2.1. <i>Waterfall Model</i> Pressman 2015	29
Gambar 3.1. Diagram Alur Pengerjaan.....	33
Gambar 3.2. <i>Flowchart</i> Tahap Klasifikasi.....	35
Gambar 3.3. Tampilan <i>Google Form</i> Data Diri.....	36
Gambar 3.4. Tampilan <i>Google Form</i> Bagian Isi	37
Gambar 3.5. Tampilan <i>Google Form</i> Bagian Hasil Responden	38
Gambar 3.6. <i>Flowchart</i> Perhitungan TF-IDF	55
Gambar 3.7. <i>Flowchart</i> Perhitungan <i>Naïve Bayes Classifier</i>	67
Gambar 3.8. <i>Waterfall Model</i> Pressman, 2015	79
Gambar 3.9. <i>Flowchart</i> Sistem	83
Gambar 3.10. <i>Use Case Diagram</i>	88
Gambar 3.11. <i>Activity Diagram</i> Pendengar	89
Gambar 3.12. <i>Activity Diagram</i> Admin	90
Gambar 3.13. <i>Class Diagram</i>	92
Gambar 3.14. Halaman <i>Login</i>	97
Gambar 3.15. Halaman Utama Admin.....	98
Gambar 3.16. Halaman Kelola Lirik Lagu Baru.....	99
Gambar 3.17. Halaman Kelola <i>Stopword</i>	100
Gambar 3.18. Halaman Kelola Tanda Baca.....	101
Gambar 3.19. Halaman Kelola <i>Dataset</i>	102
Gambar 3.20. Halaman Pelatihan	103

Gambar 3.21. Halaman Hasil Pelatihan	104
Gambar 3.22. Halaman Proses Pengujian	105
Gambar 3.23. Halaman Utama Pendengar	106
Gambar 3.24. Halaman Daftar Lirik Lagu	107
Gambar 3.25. Halaman Isi Lirik Lagu	108
Gambar 3.26. Halaman Tambah Lirik Lagu	109
Gambar 4.1. Halaman <i>Login</i>	111
Gambar 4.2. Halaman Utama Admin.....	112
Gambar 4.3. Halaman Kelola Lirik Lagu Baru.....	112
Gambar 4.4. Halaman Kelola <i>Stopword</i>	113
Gambar 4.5. Halaman Kelola Tanda Baca.....	114
Gambar 4.6. Halaman Kelola <i>Dataset</i>	114
Gambar 4.7. Halaman Pelatihan.....	115
Gambar 4.8. Halaman Hasil Pelatihan	116
Gambar 4.9. Halaman Pengujian	117
Gambar 4.10. Halaman Utama Pendengar.....	117
Gambar 4.11. Halaman Daftar Lirik Lagu	118
Gambar 4.12. Halaman Lirik Lagu	119
Gambar 4.13. Halaman Tambah Lirik Lagu	119
Gambar 4.14. <i>Section</i> Informasi.....	120
Gambar 4.15. Grafik Hasil Akurasi Klasifikasi	125
Gambar 4.16. <i>Word Cloud Senang</i>	126
Gambar 4.17. <i>Word Cloud Sedih</i>	127

Gambar 4.18. <i>Word Cloud</i> Takut.....	128
Gambar 4.19. <i>Word Cloud</i> Marah.....	129

DAFTAR TABEL

	<i>Halaman</i>
Tabel 1.1. Jadwal Penyusunan Tugas Akhir	9
Tabel 2.1. Emosi Dasar dan Jenis-Jenisnya (Setiawan, 2019).....	15
Tabel 2.2. Matriks Confusion Untuk Klasifikasi 2 Kelas	31
Tabel 3.1. Daftar <i>Google Form</i>	39
Tabel 3.1. Daftar <i>Google Form</i> (Lanjutan).....	40
Tabel 3.2. Jumlah Responden <i>Google Form</i>	41
Tabel 3.2 Jumlah Responden <i>Google Form</i> (Lanjutan).....	42
Tabel 3.3. Data Lirik Lagu Hasil <i>Majority Voting</i>	42
Tabel 3.4. Data Lirik Lagu Hasil	43
Tabel 3.5. <i>Case Folding</i>	44
Tabel 3.5. <i>Case Folding</i> (Lanjutan)	45
Tabel 3.5. <i>Case Folding</i> (Lanjutan)	46
Tabel 3.5. <i>Case Folding</i> (Lanjutan)	47
Tabel 3.6. <i>Tokenizing</i>	48
Tabel 3.6. <i>Tokenizing</i> (Lanjutan)	49
Tabel 3.6. <i>Tokenizing</i> (Lanjutan)	50
Tabel 3.6. <i>Tokenizing</i> (Lanjutan)	51
Tabel 3.6. <i>Tokenizing</i> (Lanjutan)	52
Tabel 3.7. <i>Filtering</i>	53
Tabel 3.7. <i>Filtering</i> (Lanjutan)	54
Tabel 3.8. <i>TF-IDF</i>	57

Tabel 3.8. <i>TF-IDF</i> (Lanjutan)	58
Tabel 3.8. <i>TF-IDF</i> (Lanjutan)	59
Tabel 3.8. <i>TF-IDF</i> (Lanjutan)	60
Tabel 3.8. <i>TF-IDF</i> (Lanjutan)	61
Tabel 3.8. <i>TF-IDF</i> (Lanjutan)	62
Tabel 3.8. <i>TF-IDF</i> (Lanjutan)	63
Tabel 3.8. <i>TF-IDF</i> (Lanjutan)	64
Tabel 3.9. Hasil untuk <i>Data Training</i>	68
Tabel 3.10. Prior Probabilitas	69
Tabel 3.11. Matriks Confusion	78
Tabel 3.12. Definisi Aktor	84
Tabel 3.13. Definisi <i>Case</i>	85
Tabel 3.13. Definisi <i>Case</i> (Lanjutan)	86
Tabel 3.13. Definisi <i>Case</i> (Lanjutan)	87
Tabel 3.14. Tabel Admin	93
Tabel 3.15. Tabel Lirik	93
Tabel 3.16. Tabel Kelas	94
Tabel 3.17. Tabel <i>Dataset</i>	94
Tabel 3.18. Tabel Lirik Uji	95
Tabel 3.19. Tabel Lirik Baru	95
Tabel 3.20. Tabel <i>Stopword</i>	96
Tabel 3.21. Tabel Tanda Baca	96
Tabel 3.22. Tabel Kamus	96

Tabel 4.1. Pembagian Data	121
Tabel 4.2. Matriks Confusion Pengujian 60%-40%	122
Tabel 4.3. Matriks Confusion Pengujian 70%-30%	123
Tabel 4.4. Matriks Confusion Pengujian 80%-20%	123
Tabel 4.5. Matriks Confusion Pengujian 90%-10%	124
Tabel 4.6. Pengujian <i>Form Login</i> Admin	130
Tabel 4.6. Pengujian <i>Form Login</i> Admin (Lanjutan)	131
Tabel 4.7. Pengujian Halaman Kelola Lirik Lagu Baru.....	131
Tabel 4.7. Pengujian Halaman Kelola Lirik Lagu Baru (Lanjutan).....	132
Tabel 4.7. Pengujian Halaman Kelola Lirik Lagu Baru (Lanjutan).....	133
Tabel 4.8. Pengujian Halaman Kelola <i>Stopword</i>	133
Tabel 4.8. Pengujian Halaman Kelola <i>Stopword</i> (Lanjutan)	134
Tabel 4.8. Pengujian Halaman Kelola <i>Stopword</i> (Lanjutan)	135
Tabel 4.9. Pengujian Halaman Kelola Tanda Baca.....	135
Tabel 4.9. Pengujian Halaman Kelola Tanda Baca (Lanjutan).....	136
Tabel 4.9. Pengujian Halaman Kelola Tanda Baca (Lanjutan).....	137
Tabel 4.10. Pengujian Halaman Kelola <i>Dataset</i>	137
Tabel 4.10. Pengujian Halaman Kelola <i>Dataset</i> (Lanjutan)	138
Tabel 4.11. Pengujian Halaman Pelatihan	138
Tabel 4.12. Pengujian Halaman Hasil Pelatihan.....	139
Tabel 4.13. Pengujian Halaman Pengujian	139
Tabel 4.13. Pengujian Halaman Pengujian (Lanjutan)	140
Tabel 4.13. Pengujian Halaman Pengujian (Lanjutan)	141

Tabel 4.14. Pengujian Halaman Daftar Lirik Lagu.....	141
Tabel 4.14. Pengujian Halaman Daftar Lirik Lagu (Lanjutan).....	142
Tabel 4.15. Pengujian Halaman Tambah Lirik	143

BAB I

PENDAHULUAN

1.1 Latar Belakang

Lirik lagu merupakan ekspresi seseorang tentang suatu hal yang sudah dilihat, didengar, maupun dialaminya. Dalam mengekspresikan pengalamannya, pencipta lagu melakukan permainan kata-kata untuk menciptakan daya tarik dan kekhasan terhadap lirik atau syair untuk pendengarnya (Kristiyanti, 2012). Pada setiap lirik dari sebuah lagu, terdapat emosi yang tersirat dari pencipta lagu tersebut. Lirik lagu dapat mengekspresikan perasaan atau emosi seseorang terhadap apa yang dialaminya. Ada berbagai macam emosi yang didapatkan seseorang dari berbagai lirik lagu seperti marah, sedih, senang, khawatir, malu, kaget, dan lain sebagainya. Menurut Setiawan(2019), emosi-emosi tersebut digolongkan menjadi empat golongan, yaitu emosi senang, sedih, takut, dan marah.

Menentukan emosi berdasarkan lirik lagu biasanya dilakukan manual oleh manusia, misalnya seseorang yang sedang mendengarkan lagu dengan tempo ceria menyimpulkan bahwa emosi yang ada dalam lagu tersebut adalah emosi bahagia, akan tetapi pengarang lagu tersebut bisa saja ingin menyampaikan bahwa perasaan yang ingin disampaikan adalah malu. Salah satu contohnya adalah lagu “Ya Sudahlah – Bondan Prakoso & Fade2Black”. Lagu ini memiliki suasana musik yang ceria dan terkadang orang menganggap ini adalah musik bahagia, padahal lirik dari lagu tersebut menceritakan tentang seseorang yang patah hati.

Semakin banyak lagu yang tercipta maka semakin banyak pula emosi yang muncul dari lagu tersebut. Hingga saat ini terdapat beberapa situs *website online* yang menyediakan kumpulan lirik lagu seperti *lirik.kapanlagi.com*, *www.elyrics.net*, *www.metrolyrics.com*, *www.azlyrics.com*, dan masih banyak lagi. Pada situs-situs tersebut pengguna dapat mencari lirik lagu berdasarkan penyanyi, genre, atau negara lagu tersebut tetapi tidak menyediakan klasifikasi lirik lagu berdasarkan emosi dari lagu tersebut. Hal ini menyebabkan orang kesulitan dalam mencari lirik lagu berdasarkan emosinya. Oleh karena itu dibutuhkan suatu sistem yang dapat mengklasifikasikan emosi pada lirik lagu secara otomatis agar seseorang dapat memilih lagu berdasarkan emosi yang diinginkan, selain itu terdapat lagu dan video yang dapat diputar oleh pengguna. Jadi, tidak hanya dapat melihat lirik lagu berdasarkan emosinya tetapi juga dapat memainkan musiknya.

Klasifikasi dapat dilakukan dengan metode *data mining*. Jenis data yang dapat digunakan dalam *data mining* ada dua yaitu data terstruktur yang biasanya berbentuk tabel yang terdiri dari kolom dan baris yang tersimpan dalam *database* atau data *warehouse*. Data tidak terstruktur yang tidak memiliki model atau format data secara khusus dimana data tidak terstruktur terdiri dari data teks, data gambar, data video, dan data suara (Akbar, Habibullah, 2017). Pada penelitian ini data yang digunakan adalah data teks yang menjadi bagian dari data tidak terstruktur. Data tersebut dapat di olah menggunakan *text mining*. Sehingga pengklasifikasian teks dilakukan menggunakan *text mining*. Menurut Kurniawan, dkk (2012) *text mining* merupakan variasi dari *data mining* yang berusaha menemukan pola yang menarik dari sekumpulan data tekstual yang berjumlah besar.

Salah satu algoritma yang dapat digunakan untuk klasifikasi teks yaitu *Naïve Bayes Classifier*. Dimana algoritma ini memiliki hasil akurasi yang baik dari beberapa penelitian yang telah dilakukan seperti pada penelitian yang dilakukan oleh Maulana Candra Kusuma (2017) yang digunakan untuk mengklasifikasikan lirik lagu berdasarkan emosi menggunakan metode *naïve bayes classifier* menghasilkan akurasi 88.00%. Pada penelitian lainnya yang dilakukan oleh Afif Hijrah Ferdinan, dkk (2018) untuk mengklasifikasikan lirik lagu menggunakan metode *K-Nearest Neighbour* memiliki rata-rata akurasi 75% dan pada penelitian Irfan Dhiya Abirawa, dkk (2018) pada pengklasifikasi emosi pada lirik lagu menggunakan *Support Vector Machine* menghasilkan akurasi rata-rata 62.50%. Berdasarkan latar belakang yang telah diuraikan, maka judul penelitian yang diangkat adalah “Klasifikasi Emosi Pada Lirik Lagu Berbahasa Indonesia Menggunakan Algoritma *Naïve Bayes Classifier*.”

1.2 Rumusan Masalah

Berdasarkan uraian pada latar belakang, maka rumusan masalah dari penelitian ini adalah bagaimana mengimplementasikan *text mining* dalam sistem pengimplementasian emosi pada lirik lagu berbahasa Indonesia menggunakan algoritma *Naïve Bayes*?

1.3 Batasan Masalah

1. Lirik lagu yang digunakan adalah lirik lagu berbahasa Indonesia.
2. Kelas emosi yang digunakan pada penelitian ini hanya empat, yaitu senang, sedih, takut, dan marah.

3. *Dataset* yang digunakan dalam penelitian ini diambil dari www.kapanlagi.com.
4. Pelabelan emosi pada lirik lagu dilakukan secara manual dimana responden akan diminta untuk menentukan emosi yang tersirat pada lagu.
5. Teknik *majority voting* akan diterapkan untuk menentukan emosi sebuah lagu jika ternyata untuk satu lirik lagu ada beberapa emosi yang terpilih.
6. Tidak menggunakan *stemming* untuk mendapatkan makna yang sebenarnya dari sebuah kata.
7. Fitur Pendengar :
 - i. Daftar Lirik Lagu
Sistem akan menampilkan daftar lirik lagu yang telah diklasifikasikan berdasarkan emosinya.
 - ii. Menampilkan Lirik Lagu, Mp3, dan Video
Sistem akan menampilkan lirik lagu berdasarkan lagu yang dipilih pada halaman daftar lirik lagu. Selain itu, terdapat lagu yang bisa diputar dan link yang menuju ke youtube untuk melihat video dari lirik lagu yang dipilih.
 - iii. Mengunggah Lirik Baru
Pengguna dapat mengunggah lirik lagu ke dalam sistem dan sistem akan mengklasifikasikan lirik lagu tersebut ke dalam empat kelas emosi yang ada.

8. Fitur Admin :

i. Kelola Lirik Lagu Baru

Sistem akan menampilkan data lirik lagu baru yang telah ditambahkan pendengar dan admin dapat mengelola lirik tersebut seperti menambahkan mp3, *link* video, atau menghapus data lirik tersebut.

ii. Kelola *Stopword*

Halaman yang berisikan *stopword* yang digunakan di dalam sistem, dimana admin dapat menambahkan, mengedit, atau menghapus kata yang ada.

iii. Kelola Tanda Baca

Halaman yang menampilkan daftar tanda baca yang digunakan, dimana admin dapat menambahkan, mengedit, dan menghapus tanda baca yang ada di dalam sistem.

iv. Kelola *Dataset*

Admin menambahkan *dataset* atau data lirik lagu yang akan digunakan untuk proses pelatihan atau pembuatan model. Admin juga akan menambahkan label untuk setiap data berdasarkan dengan label data yang telah ditentukan.

v. Pelatihan

Sistem akan melakukan pelatihan terhadap *dataset* yang telah ditambahkan oleh admin pada halaman kelola *dataset* untuk

membentuk model yang akan digunakan dalam pengujian model klasifikasi.

vi. Hasil Pelatihan

Admin dapat melihat hasil pelatihan yang berisi jumlah kata dan informasi dari setiap kata yang telah diolah saat proses pelatihan.

vii. Pengujian

Admin dapat melakukan pengujian model klasifikasi terhadap data uji yang telah dilabelkan secara manual dan melihat hasil klasifikasi yang telah dilakukan oleh sistem.

1.4 Tujuan

Tujuan dari penelitian ini adalah merancang sistem yang dapat mengklasifikasikan emosi berdasarkan lirik lagu berbahasa Indonesia ke dalam empat kelas emosi yaitu senang, sedih, takut, dan marah menggunakan algoritma *Naïve Bayes Classifier*.

1.5 Manfaat

Manfaat penelitian ini yaitu sebuah sistem yang dapat mengklasifikasikan emosi pada lirik lagu berbahasa Indonesia secara otomatis. Selain itu, dapat digunakan untuk membantu pendengar musik memilih lagu sesuai dengan suasana hati pendengar sendiri.

1.6 Sistematika Penulisan

Sistematika penulisan laporan dalam tugas akhir ini antara lain :

BAB I PENDAHULUAN

Bab ini diuraikan tentang latar belakang masalah, rumusan masalah, batasan masalah, tujuan dan manfaat, metode pengumpulan data, sistematika penulisan dan jadwal kegiatan.

BAB II LANDASAN TEORI

Bab ini menjelaskan berbagai konsep dasar dan teori-teori yang berkaitan dengan topik penelitian. Pada bab ini juga mengemukakan teori-teori yang relevan untuk memecahkan masalah.

BAB III METODOLOGI PENELITIAN

Pada bab ini menjelaskan proses perancangan program yang akan dibuat dalam menyelesaikan proyek.

BAB IV HASIL DAN PEMBAHASAN

Bab ini berisi hasil dan pembahasan dari penelitian yang telah dilakukan.

BAB V KESIMPULAN DAN SARAN

Bab ini terdiri dari kesimpulan dan saran. Kesimpulan memuat kesimpulan-kesimpulan yang merupakan rangkuman dari hasil analisis kinerja pada proyek

yang dilakukan. Saran merupakan perbaikan yang diperlukan saat ini atau masa yang akan datang.



1.7 Jadwal

Tabel 1.1 Jadwal Penyusunan Skripsi

No.	Rencana Kegiatan	Bulan dan Minggu																			
		Maret				April				Mei - September				Oktober				November			
		I	II	III	IV	I	II	III	IV	I	II	III	IV	I	II	III	IV	I	II	III	IV
1	Penyusunan Proposal dan Seminar Proposal	■																			
2	Analisis dan Desain Sistem																				
3	Pembuatan Program dan Pengujian																				
4	Pembuatan Laporan	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	
5	Seminar Hasil													■	■						
6	Seminar Tugas Akhir																	■	■		

BAB II

LANDASAN TEORI

2.1 Kajian Pustaka

Pada penelitian sebelumnya, terdapat beberapa penelitian yang memiliki keterkaitan dengan topik penelitian yang akan dilakukan.

Penelitian pertama, penelitian yang dilakukan oleh Rifqi Abdul Aziz, Mohamad Syahrul Mubarak, dan Adiwijaya (2016) yang berjudul “Klasifikasi Topik pada Lirik Lagu dengan Metode *Multinomial Naïve Bayes*”. Tujuan dari penelitian ini untuk membuat sistem yang mampu mengkategorikan lagu berdasarkan topik lagu dengan menggunakan lirik lagu sebagai objek penelitian. Kelas topik lagu yang digunakan yaitu cinta, persahabatan, nasionalisme, keluarga, religi, dan konten negatif. Penelitian ini diawali dengan mengumpulkan lirik lagu dari berbagai *website* penyedia lirik lagu sebagai *dataset* penelitian, kemudian dilakukan *preprocessing* data yang terdiri dari *case folding*, *tokenization*, *stopword removal* dan *stemming*. Tahap selanjutnya yaitu melakukan *feature selection* dengan menggunakan metode *chi-square*. Selanjutnya hasil *feature selection* tersebut dilakukan *feature extraction* dengan menggunakan metode *bag of words*. Selanjutnya dilakukan pembangunan *classifier* yang nantinya digunakan pada saat proses klasifikasi. Tahap selanjutnya yaitu melakukan klasifikasi lirik lagu pada data *testing* kedalam sebuah kelas topik tertentu dengan menggunakan metode *multinomial naïve bayes*. Tahap pengujian pada penelitian ini dilakukan dengan menghitung *accuracy*, *recall*, *precision*, dan *f1-measuer*. Hasil dari penggunaan

metode *Multinomial Naïve Bayes* pada penelitian ini mampu menghasilkan performansi yang baik dengan nilai *f1-measure* 88.91% dan nilai akurasi 96.00%. Persentase pembagian data training dan data *testing* terbaik pada penelitian ini yaitu pada persentase pembagian data 85%-15%. Penggunaan *feature selection chi-square* pada penelitian ini belum mampu meningkatkan performansi sistem namun mampu mereduksi jumlah *feature* yang banyak dan mampu mempercepat waktu pembentukan *classifier* 66%-80% lebih cepat dibandingkan tanpa menggunakan *feature selection*.

Penelitian kedua, penelitian yang dilakukan oleh Afif Hijra Ferdinan, Andrew Briand Osmond, dan Casi Setianingsih (2018) dengan judul penelitian “Klasifikasi Emosi Pada Lirik Lagu Menggunakan Metode K-Nearest Neighbor”. Tujuan dari penelitian untuk mengklasifikasikan emosi berdasarkan lirik lagu secara otomatis agar pengguna dapat memilih lagu yang sesuai dengan suasana hatinya. Tahap pertama yang dilakukan pada penelitian ini adalah pengumpulan lirik lagu dari *website lirik.kapanlagi.com*. Tahap kedua adalah melakukan *preprocessing* untuk mendapatkan data yang bersih. Tahap ketiga adalah *POS Tagging* untuk melabelkan kelas kata secara otomatis pada suatu kata dalam kalimat. Kemudian pembobotan TF-IDF dan tahap terakhir adalah klasifikasi menggunakan algoritma k-Nearest Neighbor. Proses klasifikasi data lirik lagu hanya diberikan empat jenis kelas label emosi yaitu cinta, marah, senang, dan sedih. Pengujian menggunakan akurasi berdasarkan nilai K pada K-NN. Hasilnya adalah sistem berhasil mengklasifikasikan emosi berdasarkan lirik lagu secara otomatis.

Nilai k terbaik adalah $k = 6$ dengan akurasi sebesar 87,5% dan rata-rata pada 10 kali pengujian memiliki nilai akurasi sebesar 75%.

Penelitian ketiga, penelitian yang dilakukan oleh Maftukhah Carolin Noviardini, Andrew Briand Osmond, dan Casi Setianingsih (2018) dengan judul “Klasifikasi Emosi Pada Lirik Lagu Menggunakan Metode Naïve Bayes Classifier”. Tujuan dari penelitian ini adalah mengklasifikasikan teks pada lirik lagu berbahasa Indonesia kedalam empat kelas emosi yaitu marah, sedih, cinta, dan gembira dengan menggunakan algoritma *Naïve Bayes Classifier*. Tahap pertama yang dilakukan adalah pengambilan data lirik lagu. Tahap kedua, *preprocessing* yang terdiri dari *case folding*, *stemming*, *stopword removal*, dan *tokenizing*. Tahap selanjutnya adalah *Part of Speech Tagging (POS Tagging)*. Tahap selanjutnya adalah pembobotan menggunakan TF-IDF. Tahap terakhir adalah pengklasifikasian menggunakan *Naïve Bayes Classifier*. Pengujian dilakukan menggunakan akurasi. Hasil pengujian dan analisa yang dilakukan, maka didapat bahwa sistem yang dirancang berhasil mengklasifikasikan emosi yang dibagi menjadi empat kelas emosi yaitu cinta, marah, sedih, senang berdasarkan lirik lagu secara otomatis menggunakan algoritma *Naïve Bayes Classifier*, dan dari hasil pengujian ditemukan nilai akurasi sebesar 72,5%.

2.2 Emosi

Emosi berasal dari bahasa Latin, yaitu *emovere*, yang berarti bergerak menjauh. Arti kata ini menyiratkan bahwa kecenderungan bertindak merupakan hal mutlak dalam emosi. Goleman (dalam Hudaya, 2002: 411) mengatakan bahwa

emosi merujuk pada suatu perasaan dan pikiran yang khas, suatu keadaan biologis dan psikologis dan serangkaian kecenderungan untuk bertindak. Emosi merupakan reaksi terhadap rangsangan dari luar dan dalam diri individu, sebagai contoh emosi gembira mendorong perubahan suasana hati seseorang, sehingga secara fisiologi terlihat tertawa, emosi sedih mendorong seseorang berperilaku menangis.

Emosi adalah suatu perasaan dan pikiran yang khas, suatu keadaan biologis dan psikologis dan serangkaian kecenderungan untuk bertindak terhadap rangsangan dari luar dan dalam diri individu mencakup perubahan-perubahan yang disadari, yang mendalam sifatnya, dan perubahan perilaku pada umumnya disertai adanya ekspresi kejasmanian (Hudaya, 2015).

2.2.1 Macam-Macam Emosi

Secara garis besarnya emosi digolongkan menjadi dua golongan yaitu emosi positif dan emosi negatif. Emosi positif seperti bahagia, senang, gembira, dan cinta. Sedangkan emosi negatif seperti takut, marah, sedih, dan cemas.

Menurut Ekman (dalam Setiawan, 2019), emosi dapat digolongkan menjadi enam, yaitu marah, muak, takut, bahagia, sedih dan kaget. Dalam perkembangan selanjutnya Ekman (dalam Setiawan, 2019) menggolongkan emosi menjadi tujuh belas macam, yaitu: girang, marah, jijik, suka, muak, memalukan, senang, takut, merasa bersalah, bahagia, bangga, lega, sedih, puas, senang, malu, dan kaget.

Menurut Heider (dalam Setiawan, 2019), emosi sedih, marah, gembira, dan kaget mendekati kesamaan universal, tetapi emosi cinta, takut, jijik, dan muak lebih bersifat khusus dan tergantung budaya.

Jadi, menurut Setiawan (2019), semua emosi tersebut digolongkan menjadi empat golongan, yakni:

1. Emosi Senang

Emosi senang adalah gambaran rasa senang yang dialami seseorang. Emosi senang ini terdiri dari bermacam-macam bentuk, misalnya bahagia, riang, gembira, dan cinta.

2. Emosi Sedih

Emosi sedih adalah gambaran rasa tidak senang yang dialami seseorang. Emosi ini juga banyak macamnya seperti duka, kecewa, hampa, dan malu.

3. Emosi Takut

Emosi takut artinya gambaran rasa tidak senang yang dialami oleh seseorang, baik terhadap objek dari luar diri maupun dalam diri orang tersebut. Objek dari luar diri misalnya takut pada pencuri, takut pada harimau, dan perampok. Rasa takut yang objeknya dalam diri orang tersebut misalnya takut tidak lulus dan takut berbuat salah.

4. Emosi Marah.

Emosi marah merupakan gambaran perasaan terhadap suatu objek seperti peristiwa, perilaku orang, hubungan sosial, dan keadaan lingkungan. Masing-masing emosi dasar terdiri dari berbagai emosi yang sejenis. Masing-masing emosi tidak ada yang bersifat universal, tetapi ada pula yang bersifat khusus, artinya kata sifat emosi tersebut hanya ada pada golongan atau suku tertentu saja sesuai dengan budaya golongan tersebut.

Secara rinci emosi dasar dan jenis-jenisnya dicantumkan pada tabel berikut ini:

Tabel 2.1. Emosi Dasar dan Jenis-Jenisnya (Setiawan, 2019)

Senang	Sedih	Takut	Marah
Gembira	Pilu	-	Jengkel
Bahagia	Duka	Cemas	Jijik
Cinta	Lara	Rindu	Dendam
Suka	Kecewa	Ngeri	Dongkol
Riang	Hampa	Malu	Geram
Sayang	Merana	Ragu-Ragu	Kesal
Takjub	Putus Asa	Khawatir	Sebal
Kagum	Galau	Merinding	Benci
Damai	Frustasi	Gelisah	Muak
-	Rindu	-	-

2.3 Lirik Lagu

Menurut Ferdinan, dkk (dalam S.Satria, 2018) Lirik lagu merupakan bentuk ekspresi seseorang terhadap lingkungan dan pengalamannya yang dituliskan dalam kata-kata. Dalam mengekspresikan perasaannya, penyair atau pencipta lagu perlu melakukan permainan kata-kata dan bahasa untuk menciptakan daya tarik dan kekhasan terhadap lirik atau syairnya. Permainan ini dapat berupa permainan vokal, gaya bahasa maupun penyimpangan makna kata dan diperkuat dengan penggunaan

melodi dan notasi yang disesuaikan dengan lirik lagunya sehingga pendengar semakin terbawa dengan apa yang dipikirkan pengarangnya.

2.4 Text Mining

Text mining, yang juga disebut sebagai Teks Data Mining (TDM) atau *Knowledge Discovery in Text* (KDT), secara umum mengacu pada proses ekstraksi informasi dari dokumen-dokumen teks tak terstruktur (*unstructured*). *Text mining* dapat didefinisikan sebagai penemuan informasi baru dan tidak diketahui sebelumnya oleh komputer, yang secara otomatis mengekstrak informasi dari sumber-sumber teks tak terstruktur yang berbeda. Kunci dari proses ini adalah menggabungkan informasi yang berhasil diekstraksi dari berbagai sumber.

2.4.1 Klasifikasi

Teknik Klasifikasi yaitu cara mempelajari sekumpulan data sehingga dihasilkan aturan yang bisa mengklasifikasikan atau mengenali data-data baru yang belum pernah dipelajari. Klasifikasi dapat didefinisikan sebagai proses untuk menyatakan suatu objek data sebagai salah satu kategori (kelas) yang telah didefinisikan sebelumnya (Zaki et al. dalam Suyanto, 2017). Klasifikasi digunakan dalam berbagai aplikasi, diantaranya adalah deteksi kecurangan (*fraud detection*), pengelolaan pelanggan, diagnosis medis, prediksi penjualan, dan sebagainya.

Model klasifikasi dapat dibangun berdasarkan pengetahuan seorang pakar (ahli). Namun, mengingat himpunan data yang sangat besar, model klasifikasi lebih

sering dibangun menggunakan teknik pembelajaran dalam bidang *machine learning*.

Terdapat banyak teknik klasifikasi yang telah diusulkan para ahli, yang dapat dikelompokkan dalam dua kategori, yaitu teknik klasifikasi global (menghitung semua data latih) dan teknik klasifikasi lokal (hanya memperhitungkan sebagian data latih).

Klasifikasi teks merupakan suatu proses pengklasifikasian terhadap suatu data teks dengan membagi menjadi beberapa bagian kelas yang berbeda. Proses klasifikasi tersebut akan melalui beberapa tahapan *reprocessing* untuk mendapatkan data yang dibutuhkan oleh pengguna sehingga proses klasifikasi memiliki tingkat akurasi yang tinggi. Penggunaan klasifikasi pada teks biasanya dapat digunakan pada beberapa data (Octaviana, 2015).

2.5 Tahapan *Text Mining*

Menurut Sulhaerati, dkk (2017) pada *text mining* terdapat tahapan-tahapan proses untuk mendapatkan pengetahuan, ada empat tahapan yang dilakukan, yaitu *pre-processing*, *text transformation*, *feature selection*, dan *pattern discovery*.

2.5.1 *Pre-Processing*

Pada tahap *pre-processing* dilakukan proses pengolahan teks menjadi data yang akan digunakan dalam proses selanjutnya. Tujuan dilakukannya tahap ini untuk melakukan seleksi informasi yang tidak diperlukan dalam proses selanjutnya seperti menghilangkan tanda baca, *tag-tag HTML*, mengubah huruf besar menjadi huruf kecil, dan menghilangkan kata imbuhan atau kata sisipan. Berikut ini adalah

proses pengolahan teks pada tahap *pre-processing*. Operasi yang dilakukan pada tahap ini adalah sebagai berikut:

1. *Case Folding*

Case folding adalah tahapan dalam mengubah semua huruf dalam dokumen menjadi huruf kecil. Hanya huruf “a” sampai dengan “z” yang diterima.

2. *Tokenizing*

Tokenizing adalah tahapan pemotongan string input berdasarkan tiap kata yang menyusunnya.

2.5.2 *Text Transformation*

Transformasi teks mengacu pada proses untuk mendapatkan representasi dokumen yang diharapkan. Transformasi teks sekaligus juga melakukan perubahan kata-kata ke bentuk dasarnya dan pengurangan dimensi kata di dalam dokumen. Tindakan ini diwujudkan dengan menerapkan *stemming* dan menghapus *stop words*.

1. *Filtering*

Filtering adalah tahapan mengambil kata-kata penting dari hasil token. Bisa menggunakan algoritma *stoplist* (membuang kata yang kurang penting) atau *wordlist* (menyimpan kata penting). *Stoplist/stopword* adalah kata-kata yang tidak deskriptif yang dapat dibuang dalam pendekatan *bag-of-words*.

2. *Stemming*

Stemming adalah tahapan dalam mencari *root* dari kata hasil *filtering*. Pada tahap ini dilakukan proses pengembalian berbagai bentukan kata ke dalam suatu representasi yang sama. Tahap ini kebanyakan dipakai untuk teks berbahasa Inggris dan lebih sulit diterapkan pada teks berbahasa Indonesia. Hal ini dikarenakan bahasa Indonesia tidak memiliki rumus bentuk baku yang permanen.

Proses *stemming* pada teks berbahasa Indonesia berbeda dengan *stemming* pada teks berbahasa Inggris. Pada teks berbahasa Inggris, proses yang diperlukan hanya proses menghilangkan sufiks. Sedangkan pada teks berbahasa Indonesia, selain sufiks, prefiks, dan konfiks juga dihilangkan.

2.5.3 *Feature Selection*

Pemilihan fitur (kata) merupakan tahap lanjut dari pengurangan dimensi pada proses transformasi teks. Walaupun tahap sebelumnya sudah melakukan penghapusan kata-kata yang tidak deskriptif (*stopwords*), namun tidak semua kata-kata di dalam dokumen memiliki arti penting. Oleh karena itu, untuk mengurangi dimensi, pemilihan hanya dilakukan terhadap kata-kata yang relevan yang benar-benar merepresentasikan isi dari suatu dokumen. Ide dasar dari pemilihan fitur adalah menghapus kata-kata yang kemunculannya di suatu dokumen terlalu sedikit atau terlalu banyak.

2.5.4 *Pattern Discovery*

Pattern Discovery merupakan tahap penting untuk menemukan pola atau pengetahuan dari keseluruhan teks. Tindakan yang lazim dilakukan pada tahap ini adalah operasi *text mining*, dan biasanya menggunakan teknik-teknik *data mining*. Dalam penemuan pola ini, proses *text mining* dikombinasikan dengan proses-proses *data mining*. Masukan awal dari proses *text mining* adalah suatu data teks dan menghasilkan keluaran berupa pola sebagai hasil interpretasi atau evaluasi. Apabila hasil keluaran dari penemuan pola belum sesuai untuk aplikasi, dilanjutkan evaluasi dengan melakukan iterasi ke satu atau beberapa tahap sebelumnya. Sebaliknya, hasil interpretasi merupakan tahap akhir dari proses *text mining* dan akan disajikan ke pengguna dalam bentuk visual.

2.6 *Term Frequency – Inverse Document Frequency*

Algoritma *Term Frequency Inverse - Document Frequency* merupakan suatu algoritma yang menggabungkan antara *Term Frequency* dengan *Inverse Document Frequency*. *Term frequency* yaitu jumlah kemunculan sebuah *term* pada sebuah dokumen. *Inverse Document Frequency* yaitu pengurangan dominasi *term* yang sering muncul diberbagai dokumen, dengan memperhitungkan kebalikan frekuensi dokumen yang mengandung suatu kata. (Nurjannah, dkk. 2013)

Metode TF-IDF merupakan suatu cara untuk memberikan bobot hubungan suatu kata (*term*) terhadap dokumen. Metode ini menggabungkan dua konsep untuk perhitungan bobot, yaitu frekuensi kemunculan sebuah kata

di dalam sebuah dokumen tertentu dan inverse frekuensi dokumen yang mengandung kata tersebut. Frekuensi kemunculan kata di dalam dokumen yang diberikan menunjukkan seberapa penting kata itu di dalam dokumen tersebut. Frekuensi dokumen yang mengandung kata tersebut menunjukkan seberapa umum kata tersebut. Sehingga bobot hubungan antara sebuah kata dan sebuah dokumen akan tinggi apabila frekuensi kata tersebut tinggi di dalam dokumen dan frekuensi keseluruhan dokumen yang mengandung kata tersebut yang rendah pada kumpulan dokumen. Rumus untuk TF-IDF:

$$tf = 0,5 + 0,5 \times \frac{tf}{\max(tf)} \quad (2.1)$$

$$idf_t = \log \frac{D}{df_t} \quad (2.2)$$

$$W_{d,t} = tf_{d,t} \times IDF_{d,t} \quad (2.3)$$

Keterangan :

tf = banyaknya kata yang dicari pada sebuah dokumen.

$\max(tf)$ = jumlah kemunculan terbanyak *term* pada dokumen yang sama.

Nilai D = total dokumen.

df_t = jumlah dokumen yang mengandung *term* t .

IDF = *Inversed Document Frequency* ($\log^2 \left(\frac{D}{df} \right)$)

d = dokumen ke- d .

t = kata ke- t dari kata kunci.

W = bobot dokumen ke- d terhadap kata ke- t .

Konsep *term frequency* memandang suatu dokumen sebagai *bag of words* (kantong kata) di mana urutan dari kemunculan suatu kata diabaikan dan hanya jumlah kemunculan dari kata itu saja yang penting. Konsep TF memiliki kelemahan yaitu semua kata dianggap setara. Hal ini mengakibatkan relevansi suatu kata menjadi sangat tinggi jika kata itu sering muncul dalam kumpulan dokumen. Padahal tingginya frekuensi kemunculan suatu kata tidak selalu menyatakan bahwa kata tersebut penting.

Konsep *inverse document frequency (IDF)* dibuat untuk mengurangi efek dari kata yang frekuensinya terlalu tinggi dalam kumpulan dokumen. Ide dasarnya adalah untuk menurunkan bobot dari kata dengan frekuensi kolektif (frekuensi total kemunculan kata disemua dokumen) yang tinggi. Dengan kata lain, semakin banyak dokumen kata tersebut pada suatu kumpulan dokumen, maka semakin rendah bobonya. (Ryansyah, dkk. 2017)

2.7 Teorema Bayes

Bayes merupakan teknik prediksi berbasis probabilitas sederhana yang berdasar pada penerapan teorema *bayes* dengan asumsi independensi (ketidaktergantungan) yaitu kuat (naif). Dengan kata lain, dalam *Naïve Bayes*, model yang digunakan adalah “model fitur independen”. Prediksi *Bayes* didasarkan pada Teorema *Bayes* dengan formula umum sebagai berikut :

$$P(H|E) = \frac{P(E|H) \times P(H)}{P(E)} \quad (2.4)$$

Keterangan :

- $P(H|E)$ = Probabilitas akhir bersyarat (*conditional probability*) suatu hipotesa H terjadi jika diberikan bukti (*evidence*) E terjadi.
- $P(E|H)$ = Probabilitas sebuah bukti E terjadi akan memengaruhi hipotesa H .
- $P(H)$ = Probabilitas awal (priori) hipotesis H terjadi tanpa memandang bukti apapun.
- $P(E)$ = Probabilitas awal (priori) bukti E terjadi tanpa memandang hipotesis atau bukti yang lain.

Ide dasar dari aturan *Bayes* adalah bahwa hasil dari hipotesis atau peristiwa (H) dapat diperkirakan berdasarkan beberapa bukti E yang diamati. Ada beberapa hal penting dari aturan *Bayes* tersebut, yaitu:

1. Sebuah probabilitas awal/priori H atau $P(H)$ adalah probabilitas dari suatu hipotesis sebelum bukti diamati.
2. Sebuah probabilitas akhir H atau $P(H|E)$ adalah probabilitas dari suatu hipotesis setelah bukti diamati.

2.7.1 *Naïve Bayes* Untuk Klasifikasi

Dalam *data mining*, X adalah sebuah *tuple* atau objek data, H adalah hipotesis atau dugaan bahwa *tuple* X adalah kelas C . Secara spesifik, dalam masalah klasifikasi, $P(H|X)$ dapat dihitung sebagai probabilitas bahwa hipotesis H benar untuk *tuple* X atau dengan kata lain $P(H|X)$ adalah probabilitas bahwa *tuple* X

berada dalam kelas C . Sementara itu, $P(H)$ adalah probabilitas *prior* bahwa hipotesis H benar untuk setiap *tuple* tidak peduli nilai-nilai atributnya sedangkan $P(X)$ adalah probabilitas *prior* dari *tuple* X .

Untuk menggunakan Teorema *Bayes* dalam klasifikasi data, atau lebih spesifik lagi, untuk melakukan estimasi $P(H/X)$ untuk menentukan probabilitas bahwa *tuple* X berada dalam kelas C , menggunakan metode *Naïve Bayes Classifier*, yang bekerja dengan lima langkah berikut :

1. Misalnya D adalah himpunan data latih (*training set*) yang berisi sejumlah *tuple* beserta label kelasnya. Setiap *tuple* adalah berdimensi n yang dinyatakan sebagai $X = (x_1, x_2, \dots, x_n)$ yang didapat dari n atribut A_1, A_2, \dots, A_n .
2. Misalkan terdapat m kelas, yaitu $C_1, C_2, \dots, \dots, C_m$. Untuk sebuah *tuple* masukan X , *Naïve Bayes Classifier* memprediksi bahwa *tuple* X termasuk ke dalam kelas C_i jika dan hanya jika

$$P(C_i|X) > P(C_j|X) \text{ untuk } \leq j \leq m, j \neq i \quad (2.5)$$

Dengan kata lain, *Naïve Bayes Classifier* bekerja dengan cara memaksimalkan (mencari nilai maksimal) $P(C_i|X)$. Kelas C_i yang membuat $P(C_i|X)$ bernilai maksimal disebut *maximum posterior hypothesis*. Dengan teorema *Bayes*, $P(C_i|X)$ diestimasi menggunakan formula.

$$P(C_i|X) = \frac{P(X|C_i) \times P(C_i)}{P(X)} \quad (2.6)$$

3. Mengingat $P(X)$ bernilai sama untuk semua kelas (artinya *tuple* X memiliki probabilitas yang sama untuk masuk ke dalam kelas manapun), maka $P(X/C_i) P(C_i)$ yang perlu dimaksimalkan. Jika probabilitas *prior* untuk setiap kelas tidak diketahui, maka probabilitas setiap kelas biasanya diasumsikan sama, $P(C_1) = P(C_2) = P(C_m)$. dengan demikian, *Naïve Bayes Classifier* hanya memaksimalkan $P(X/C_i) P(C_i)$.
4. Jika dihadapkan pada himpunan data yang memiliki banyak atribut, kompleksitas perhitungan $P(X/C_i)$ dapat direduksi dengan asumsi naif tentang independensi bersyarat kelas, yaitu nilai-nilai atribut saling independen, yang berarti antar atribut saling bebas, tidak ada ketergantungan sama sekali. Dengan demikian *Naïve Bayes Classifier* memaksimalkan

$$P(C_i|X) = \prod_{k=1}^n P(x_k|C_i) = P(x_1|C_i) \times \dots \times P(x_n|C_i) \quad (2.7)$$

Berdasarkan tuple-tuple pada himpunan data latih $P(x_1/C_i)$, $P(x_2/C_i)$, ..., $P(x_n/C_i)$ berdasarkan tipe atributnya.

Untuk atribut yang bernilai kategorial, $P(x_k/C_i)$ didefinisikan sebagai jumlah *tuple* di kelas C_i dalam himpunan data D yang memiliki nilai x_k pada atribut A_k dibagi dengan jumlah semua *tuple* di kelas C_i dalam D yang disimbolkan $|C_{i,D}|$.

Sementara itu, untuk atribut yang bernilai kontinu, yang umumnya diasumsikan memiliki distribusi Gaussian, $P(x_k/C_i)$ didefinisikan sebagai

$$P(x_i|C_i) = \frac{1}{\sigma_{C_i} \sqrt{2\pi}} e^{-\frac{(x-\mu_{C_i})^2}{2\sigma_{C_i}^2}} \quad (2.8)$$

Dimana μ_{C_i} dan σ_{C_i} adalah rata-rata dan deviasi standar dari nilai-nilai pada atribut A_k untuk kelas C_i .

5. Untuk memprediksi label kelas dari *tuple* X , harus menghitung probabilitas $P(X/C_i)P(C_i)$ untuk setiap kelas C_i . Selanjutnya hanya perlu memaksimalkan probabilitas tersebut, yaitu mencari kelas C_i yang menghasilkan probabilitas $P(X/C_i)P(C_i)$ maksimum sebagai kelas keputusan. Secara matematis, *tuple* X diberi label kelas C_i jika dan hanya jika

$$P(X|C_i)P(C_i) > P(X|C_j)P(C_j) \text{ untuk } 1 \leq j \leq m, j \neq i \quad (2.9)$$

Jika perhitungan probabilitas menghasilkan nilai 0 karena nilai atribut yang tidak pernah muncul dalam himpunan data latih, cara yang dapat digunakan adalah menambahkan satu *tuple* pura-pura dengan kelas yang belum pernah muncul untuk setiap atribut yang ada (Suyanto, 2018). Hal itu dilakukan untuk menghindari probabilitas bernilai 0.

Naïve Bayes Classifier untuk klasifikasi teks memperhitungkan frekuensi setiap kata yang muncul pada dokumen. Misalnya terdapat dokumen d dengan himpunan kelas c . Untuk memperhitungkan kelas dari dokumen d , maka dapat dihitung dengan rumus :

$$P(c|\text{dokumen } d) = P(c) \times P(x_1|c) \times P(x_2|c) \dots \times P(x_n|c) \quad (2.10)$$

Keterangan :

$P(c)$ = Probabilitas prior dari kelas c .

x_n = Kata dokumen d ke- n .

$P(c / \text{dokumen } d)$ = Probabilitas suatu dokumen termasuk kelas c .

$P(x_n / c)$ = Probabilitas kata ke- n dengan diketahui kelas c .

Probabilitas *prior* kelas c ditentukan dengan rumus :

$$P(c) = \frac{N_c}{N} \quad (2.11)$$

Keterangan :

N_c = Jumlah kelas c pada seluruh dokumen

N = Jumlah seluruh dokumen

Sedangkan rumus yang digunakan dalam *Naïve Bayes Classifier* dengan menggunakan pembobotan kata dari *TF-IDF* adalah sebagai berikut.

$$P(x_n | c) = \frac{W_{ct} + 1}{(\sum W' \in V W'_{ct}) + B'} \quad (2.12)$$

Keterangan :

W_{ct} = Nilai pembobotan *TFIDF* atau W dari *term* t di kelas c .

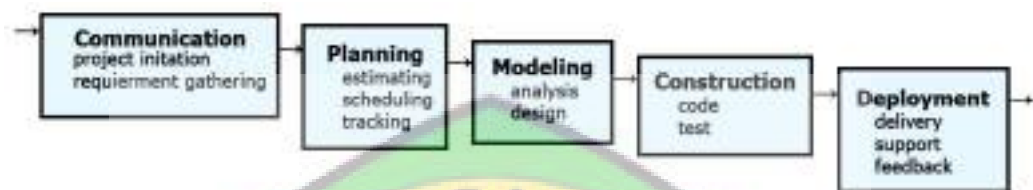
$\sum W' \in V W'_{ct}$ = Jumlah total W dari keseluruhan *term* yang berada di kelas c .

B' = Jumlah W kata unik (nilai *idf* tidak dikali dengan *tf*) pada seluruh dokumen.

2.8 Metode Pengembangan Perangkat Lunak *Waterfall*

Menurut Pressman (dalam Setiawan, 2015), nama lain dari Model *Waterfall* adalah Model Air Terjun kadang dinamakan siklus hidup klasik (*classic life cycle*), dimana hal ini menyiratkan pendekatan yang sistematis dan berurutan (*sekuensial*) pada pengembangan perangkat lunak. Pengembangan perangkat lunak dimulai dari spesifikasi kebutuhan pengguna dan berlanjut melalui tahapan-tahapan

perencanaan (*planning*), pemodelan (*modeling*), konstruksi (*construction*), serta penyerahan sistem perangkat lunak ke para pelanggan/pengguna (*deployment*), yang diakhiri dengan dukungan berkelanjutan pada perangkat lunak yang dihasilkan. Berikut ini adalah penjelasan dari tahap-tahap yang dilakukan di dalam Model *Waterfall* menurut Pressman (2015):



Gambar 2.1. *Waterfall Model* Pressman, 2015

1. *Communication*

Langkah pertama diawali dengan komunikasi kepada konsumen/pengguna. Langkah awal ini merupakan langkah penting karena menyangkut pengumpulan informasi tentang kebutuhan konsumen/pengguna.

2. *Planning*

Setelah proses *communication* kemudian menetapkan rencana untuk pengerjaan *software* yang meliputi tugas-tugas teknis yang akan dilakukan, risiko yang mungkin terjadi, sumber yang dibutuhkan, hasil yang akan dibuat, dan jadwal pengerjaan.

3. *Modeling*

Pada proses *modeling* ini menerjemahkan syarat kebutuhan sebuah perancangan perangkat lunak yang dapat diperkirakan sebelum

dibuat *coding*. Proses ini berfokus pada rancangan struktur data, arsitektur *software*, representasi *interface*, dan detail (algoritma) prosedural.

4. *Construction*

Construction merupakan proses membuat kode (*code generation*). *Coding* atau pengkodean merupakan penerjemahan desain dalam bahasa yang bisa dikenali oleh komputer. Programmer akan menerjemahkan transaksi yang diminta oleh *user*. Tahapan inilah yang merupakan tahapan secara nyata dalam mengerjakan suatu *software*, artinya penggunaan komputer akan dimaksimalkan dalam tahapan ini. Setelah pengkodean selesai maka akan dilakukan testing terhadap sistem yang telah dibuat. Tujuan *testing* adalah menemukan kesalahan-kesalahan terhadap sistem tersebut untuk kemudian bisa diperbaiki.

5. *Deployment*

Tahapan ini bisa dikatakan final dalam pembuatan sebuah *software* atau sistem. Setelah melakukan analisis, desain dan pengkodean maka sistem yang sudah jadi akan digunakan *user*. Kemudian *software* yang telah dibuat harus dilakukan pemeliharaan secara berkala.

2.9 Pengujian

2.8.1 Akurasi

Sistem yang melakukan klasifikasi diharapkan dapat melakukan klasifikasi semua data set dengan benar. Oleh karena itu, sebuah sistem klasifikasi juga harus diukur kinerjanya dan keakuratannya. Umumnya cara mengukur kinerja klasifikasi

menggunakan kinerja klasifikasi menggunakan matriks confusion (Prasetyo, 2014).

Matriks confusion merupakan tabel yang mencatat hasil kerja klasifikasi.

Tabel 2.2 Matriks Confusion Untuk Klasifikasi 2 Kelas

f_{ii}		Kelas hasil klasifikasi (j)	
		Kelas = 1	Kelas = 0
Kelas asli (i)	Kelas = 1	f_{11}	f_{10}
	Kelas = 0	f_{01}	f_{00}

Tabel 2.2 adalah contoh matriks confusion, yang melakukan klasifikasi masalah biner (dua kelas) untuk dua kelas, misalnya kelas 0 dan 1. Setiap sel f_{ij} dalam matriks menyatakan jumlah data dari kelas i yang hasil klasifikasinya masuk ke kelas j . Misalnya f_{11} adalah jumlah data dalam kelas 1 yang secara benar diklasifikasi ke kelas 1, dan f_{10} adalah data dalam kelas 1 yang diklasifikasi secara salah ke dalam kelas 0.

Berdasarkan isi matriks confusion, maka dapat diketahui jumlah data dari masing-masing kelas yang diprediksi secara benar, yaitu $(f_{11} + f_{00})$, dan data yang diklasifikasikan secara salah, yaitu $(f_{10} + f_{01})$. Dengan mengetahui jumlah data yang diklasifikasikan secara benar, maka dapat diketahui akurasi hasil klasifikasi, maka. Untuk menghitung akurasi, dengan formula sebagai berikut :

$$\text{Akurasi} = \frac{\text{Jumlah data yang diprediksi secara benar}}{\text{Jumlah prediksi yang dilakukan}} \quad (2.13)$$

$$\text{Akurasi} = \frac{f_{11} + f_{00}}{f_{11} + f_{10} + f_{01} + f_{00}} \quad (2.14)$$

2.8.2 *Black Box Testing*

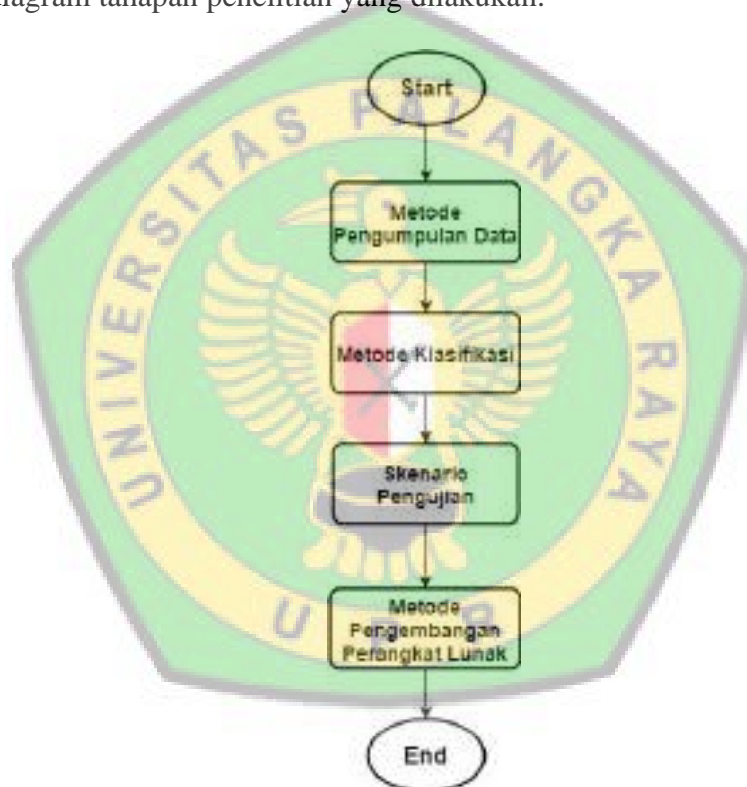
Black Box Testing merupakan pengujian perangkat lunak dari segi spesifikasi fungsional tanpa menguji desain dan kode program. Pengujian dimaksudkan untuk mengetahui apakah fungsi-fungsi, masukan, dan keluaran dari perangkat lunak sesuai dengan spesifikasi yang dibutuhkan. Pengujian ini dilakukan dengan membuat kasus uji yang bersifat mencoba semua fungsi dengan memakai perangkat lunak apakah sesuai dengan spesifikasi yang dibutuhkan. Kasus uji yang dibuat untuk melakukan pengujian ini harus dibuat dengan kasus benar dan kasus salah (Rosa & Salahudin, 2016).



BAB III

METODOLOGI PENELITIAN

Bab III berisi metodologi yang digunakan dalam penelitian ini, meliputi metode pengumpulan data, metode klasifikasi, pengujian menggunakan skenario yang sudah ditentukan, dan metode pengembangan perangkat lunak. Gambar 3.1. adalah diagram tahapan penelitian yang dilakukan.



Gambar 3.1. Diagram Alur Pengerjaan

3.1 Metode Pengumpulan Data

Metode pengumpulan data adalah sebuah metode tentang bagaimana dalam mengumpulkan data-data yang digunakan. Dalam penelitian ini metode

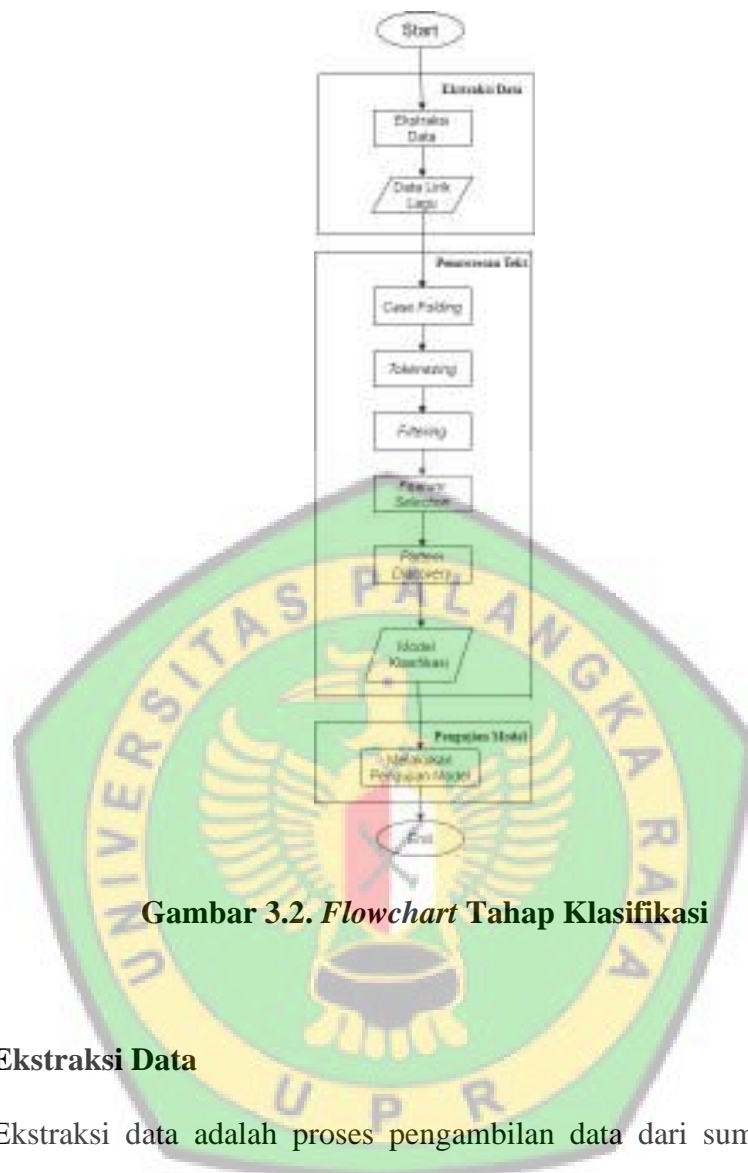
pengumpulan yang digunakan adalah studi pustaka dan literatur sejenis untuk mencari referensi yang berhubungan dengan penelitian yang dilakukan.

Pengumpulan data lirik lagu yang digunakan untuk melakukan klasifikasi emosi dalam penelitian ini berasal dari situs *website lirik.kapanlagi.com* yang dikumpulkan secara manual. Selain itu, terdapat data tambahan yang digunakan sebagai data pendukung untuk sistem yang dibuat, yaitu data lagu yang diambil dari situs *website stafabandplanetwapka.wapsing.com* dan *downloadLagu321.net*. Serta video yang diambil dari situs *youtube.com*.

3.2 Metode Klasifikasi

Dalam melakukan klasifikasi emosi pada lirik lagu di penelitian ini, akan dilakukan beberapa tahap, yaitu ekstraksi data, pemrosesan teks yang meliputi *text preprocessing* (*case folding* dan *tokenizing*), *text transformation* (*filtering*), *feature selection*, serta *pattern discovery*, dan pengujian model klasifikasi. Gambar 3.2. adalah *flowchart* tahap klasifikasi.





Gambar 3.2. Flowchart Tahap Klasifikasi

3.2.1 Ekstraksi Data

Ekstraksi data adalah proses pengambilan data dari sumber data untuk digunakan dalam pemrosesan (Kuncoro, Anon. 2015). Pada penelitian ini, data lirik lagu diperoleh dari situs *website lirik.kapanlagi.com* sebagai salah satu *website* pencarian lirik lagu Indonesia yang lengkap. Lirik lagu yang telah dikumpulkan akan diberi label sesuai dengan empat kelas emosi yaitu senang, sedih, takut, dan marah. Proses pelabelan data dilakukan menggunakan metode *majority voting*.

Menurut Lukito, Yuan dan Antonius Rachmat C (2015), *Majority voting*, yaitu beberapa responden memberi label pada suatu data yang sama, kemudian

keputusan akhir diambil dari jumlah suara terbanyak. *Voting* ini dilakukan oleh responden terhadap 242 data lirik lagu yang diunggah di *google form*. Berikut adalah tampilan *google form* yang akan diisi oleh responden.



Pelabelan Emosi Lirik Lagu XX

* Wajib

Nama Lengkap : *

Jawaban Anda

Berikutnya

Gambar 3.3. Tampilan *Google Form* Data Diri

Ketika responden akan melabelkan emosi pada lirik lagu yang ada di *google form* yang telah disediakan, responden akan mengisi data diri berupa nama lengkap terlebih dahulu. Setelah itu, responden dapat melabelkan lirik lagu yang ada. Berikut adalah tampilan *google form* untuk responden melabelkan lirik lagu yang telah ditentukan.

Pelabelan Emosi Lirik Lagu XX

* Wajib

Lirik Lagu

01. T2 - Semuanya Dusta *

"Inikah rasanya terluka
 Di kala saat putus cinta
 Perihnya ku relakan semua
 Sampai hati pun terluka
 Ternyata engkau menipuku
 Bercinta kau di belakangku
 Tak lagi ku mau denganmu
 Walau hatiku pun cinta
 Biar biarlah kau pergi saja
 Lelaki tak tahu diri itu
 Adalah dirimu penipu pendusta
 Kau bilang kepadaku
 Cinta mati semuanya dusta
 Ku benci dirimu saat ini
 Jangan kau kembali lagi padaku
 Kau lelaki buaya
 Taklah pantas mendampingiku
 Biar (biar) biarlah (biarlah)
 Kau pergi saja
 Lelaki tak tahu diri itu
 Adalah dirimu penipu pendusta
 Kau bilang kepadaku
 Cinta mati semuanya dusta
 Ku benci dirimu saat ini
 Jangan kau kembali lagi padaku
 Kau lelaki buaya
 Taklah pantas mendampingiku
 Semuanya dusta, semuanya dusta "

- Senang (Gembira, bahagia, cinta, suka, riang, sayang, takjub, kagum, damai)
- Sedih (Pilu, duka, lara, kecewa, hampa, merana, putus asa, galau, frustrasi, rindu)
- Takut (Cemas, rindu, ngeri, malu-malu, ragu, khawatir, merinding, gelisah)
- Marah (Jengkel, jijik, dendam, dongkol, geram, kesal, sebal, benci, muak)

Gambar 3.4. Tampilan Google Form Bagian Isi

Gambar 3.4. menunjukkan halaman *google form* yang berisikan lirik lagu yang akan dilabelkan dimana lirik tersebut akan dilabelkan ke dalam empat kelas emosi, yaitu senang, sedih, takut, dan marah. Dalam satu *google form*, masing-masing responden bisa melabelkan 10-15 data lagu untuk ke dalam empat kelas emosi yang telah ditentukan. Setelah itu, akan dilihat hasil dari lirik lagu yang telah dilabelkan oleh responden. Emosi dengan persentase tertinggi dari setiap lirik lagu akan dipilih sebagai emosi yang menyiratkan lagu tersebut. Berikut adalah contoh dari hasil pelabelan lirik lagu pada *google form*.



Gambar 3.5. Tampilan *Google Form* Bagian Hasil Responden

Gambar 3.5. menunjukkan bahwa data lirik lagu memiliki persentase tertinggi pada kelas marah. Jadi, lagu tersebut akan dilabelkan sebagai lagu dengan emosi marah. Dari keseluruhan *google form* yang telah dibuat, berikut adalah *link google form* yang digunakan dalam penelitian.

Tabel 3.1. Daftar Google Form

No.	Nama Google Form	Link Google Form
1.	Pelabelan Emosi Pada Lirik Lagu I	https://forms.gle/XJ6BAHEr1VZqZvPJ6
2.	Pelabelan Emosi Pada Lirik Lagu II	https://forms.gle/Y3JSwApRPicdFFBw6
3.	Pelabelan Emosi Pada Lirik Lagu III	https://forms.gle/8ao7AToXQ2uVwPfr9
4.	Pelabelan Emosi Pada Lirik Lagu IV	https://forms.gle/wAjkKsXEqBnN6zed8
5.	Pelabelan Emosi Pada Lirik Lagu V	https://forms.gle/d4nUFTbbeqmsZWRd6
6.	Pelabelan Emosi Pada Lirik Lagu VI	https://forms.gle/dst1Djdn8zkDxpPn6
7.	Pelabelan Emosi Pada Lirik Lagu VII	https://forms.gle/5zs93NY51VDV5aa8A
8.	Pelabelan Emosi Pada Lirik Lagu VIII	https://forms.gle/tSxSeLpzvQ6LavnM7
9.	Pelabelan Emosi Pada Lirik Lagu IX	https://forms.gle/tt7FgYugphqAdmS46
10.	Pelabelan Emosi Pada Lirik Lagu X	https://forms.gle/BfzcyVaSpNPibHPQ8

Tabel 3.1. Daftar Google Form (Lanjutan)

No.	Nama Google Form	Link Google Form
11.	Pelabelan Emosi Pada Lirik Lagu XI	https://forms.gle/JJ8H1rR5xpnohwEo8
12.	Pelabelan Emosi Pada Lirik Lagu XII	https://forms.gle/c4KSuZUJW3ai8ZSx6
13.	Pelabelan Emosi Pada Lirik Lagu XIII	https://forms.gle/21JcrvFjyJRfaKWg6
14.	Pelabelan Emosi Pada Lirik Lagu XIV	https://forms.gle/NMgbBAGUos5d699J9
15.	Pelabelan Emosi Pada Lirik Lagu XV	https://forms.gle/XMZ5CvdTxPXLNqyo7
16.	Pelabelan Emosi Pada Lirik Lagu XVI	https://forms.gle/b3pA5D8h9H2tkhqk8
17.	Pelabelan Emosi Pada Lirik Lagu XVII	https://forms.gle/PtajCCYVLtqCCW7cA
18.	Pelabelan Emosi Pada Lirik Lagu XVIII	https://forms.gle/7TDSZLqB2JyBLvJv7
19.	Pelabelan Emosi Pada Lirik Lagu XIX	https://forms.gle/Sh4m6TrJrJBRBnZF7
20.	Pelabelan Emosi Pada Lirik Lagu XX	https://forms.gle/KdAwkdeGWnyKso9B8

Tabel 3.1. tersebut merupakan tabel yang berisikan *link* untuk masing-masing *google form* yang digunakan. Dari *google form* yang telah dibuat tersebut, berikut adalah jumlah responden dan jumlah data untuk masing-masing *form* yang telah dibuat. Untuk rincian dari masing-masing *google form* tersebut dapat dilihat pada lampiran.

Tabel 3.2. Jumlah Responden *Google Form*

No.	Nama <i>Google Form</i>	Jumlah Data	Total Responden
1.	Pelabelan Emosi Pada Lirik Lagu I	10	30
2.	Pelabelan Emosi Pada Lirik Lagu II	10	24
3.	Pelabelan Emosi Pada Lirik Lagu III	10	23
4.	Pelabelan Emosi Pada Lirik Lagu IV	10	23
5.	Pelabelan Emosi Pada Lirik Lagu V	10	21
6.	Pelabelan Emosi Pada Lirik Lagu VI	10	14
7.	Pelabelan Emosi Pada Lirik Lagu VII	10	18
8.	Pelabelan Emosi Pada Lirik Lagu VIII	10	19
9.	Pelabelan Emosi Pada Lirik Lagu IX	10	18
10.	Pelabelan Emosi Pada Lirik Lagu X	10	16
11.	Pelabelan Emosi Pada Lirik Lagu XI	15	18
12.	Pelabelan Emosi Pada Lirik Lagu XII	15	20
13.	Pelabelan Emosi Pada Lirik Lagu XIII	15	20
14.	Pelabelan Emosi Pada Lirik Lagu XIV	15	18
15.	Pelabelan Emosi Pada Lirik Lagu XV	15	16

Tabel 3.2. Jumlah Responden *Google Form*

No.	Nama <i>Google Form</i>	Jumlah Data	Total Responden
16.	Pelabelan Emosi Pada Lirik Lagu XVI	15	16
17.	Pelabelan Emosi Pada Lirik Lagu XVII	15	5
18.	Pelabelan Emosi Pada Lirik Lagu XVIII	15	5
19.	Pelabelan Emosi Pada Lirik Lagu XIX	15	5
20.	Pelabelan Emosi Pada Lirik Lagu XX	7	5

Tabel 3.2. merupakan tabel yang berisikan jumlah *google form* yang dibuat untuk melabelkan data lirik lagu yang telah dikumpulkan dimana jumlah *google form* yang dibuat untuk melabelkan 242 data lagu tersebut berjumlah 20 *google form*. Rincian data responden dan hasil dari pelabelan data tersebut dapat dilihat pada lampiran. Dari hasil pelabelan lirik lagu terhadap 242 data tersebut maka didapatkan hasil untuk setiap kelas emosi pada Tabel 3.3. berikut.

Tabel 3.3. Data Lirik Lagu Hasil *Majority Voting*

Kelas Emosi	Jumlah Data
Senang	70
Sedih	56
Marah	61
Takut	55

Pada tabel 3.3. dapat dilihat bahwa hasil *majority voting* yang didapatkan, terdapat kelas emosi senang dan marah yang memiliki lebih banyak data

dibandingkan dengan kelas yang lainnya. Sedangkan pada saat proses pembentukan model dan pengujian model diperlukan data dengan distribusi yang sama rata agar pada saat pelatihan proses pembelajaran dilakukan secara merata pada setiap kelasnya untuk menghasilkan model yang general maka pendistribusian data dibagi secara merata, sehingga untuk memenuhi 200 *dataset* pada penelitian ini, pada setiap kelasnya diambil 50 data dari hasil *majority voting* yang dilakukan. Berikut adalah jumlah data untuk masing-masing kelas.

Tabel 3.4. Data Lirik Lagu

Kelas Emosi	Jumlah Data
Senang	50
Sedih	50
Marah	50
Takut	50

Tabel 3.4. merupakan jumlah data yang akan digunakan dalam melakukan klasifikasi emosi dimana distribusi untuk setiap kelasnya sama rata, yaitu 50 data. Data tersebut dibagi menjadi *data training* dan *data testing*. *Data training* akan melalui tahap *preprocessing* untuk mendapatkan sebuah model dan *data testing* akan melalui tahap *preprocessing* untuk mencocokkan *data testing* dengan model untuk mengetahui hasil klasifikasi.

3.2.2 Pemrosesan Teks

Menurut Sulhaerati, dkk (2017) tahapan-tahapan proses untuk mendapatkan pengetahuan dalam *text mining* adalah *text preprocessing*, *text transformation*, *feature selection*, dan *pattern discovery*.

1. *Text Preprocessing*

Tahap *pre-processing* akan melakukan proses pengolahan teks menjadi data yang akan digunakan dalam proses selanjutnya. Operasi yang dilakukan pada tahap ini adalah sebagai berikut.

i. *Case folding*

Case folding adalah tahapan yang dilakukan untuk mengkonversi atau mengubah seluruh teks dalam dokumen menjadi huruf kecil atau *lower case* yang ada di dalam dokumen. Tabel 3.5. memberikan ilustrasi penggunaan *case folding* dalam teks.

Tabel 3.5. Case Folding

Kategori	Senang (D1)
Dokumen	<p>Bilang papa mu Ku takkan buat kau berubah menjadi anak yang nakal Bilang mama mu Ku cinta padamu dan aku tak pernah main main Biarkanlah saja dulu kita jalan berdua Mereka pun pernah muda... pernah muda... Bilang papa mu berhenti urusin semua urusan kau dan aku Bilang mama mu tak perlu kuatir atau pun curiga kepadaku Biarkanlah saja dulu kita jalan berdua Mereka pun pernah muda...Saatnya kau dan aku sekarang</p>
Case Folding	<p>bilang papa mu ku takkan buat kau berubah menjadi anak yang nakal bilang mama mu ku cinta padamu dan aku tak pernah main main biarkanlah saja dulu kita jalan berdua mereka pun pernah muda... pernah muda... bilang papa mu berhenti urusin semua urusan kau dan aku bilang mama mu tak perlu kuatir atau pun curiga kepadaku biarkanlah saja dulu kita jalan berdua mereka pun pernah muda...saatnya kau dan aku sekarang</p>

Tabel 3.5. Case Folding (Lanjutan)

Kategori	Marah(D2)
Dokumen	<p>Kata katanya selangit Tersenyum penuh misteri Matanya membikin ngeri Semua gadis gadis remaja Gaya lelaki katanya semua begitu Dia bilang aku cantik Diapun bilang ku menarik Dia bilang body ku asik Hingga dada ini deg deg an Mulut lelaki katanya selalu begitu Apalagi dia yang masih suka Dengan gadis remaja Engkau lupakan anak cucumu Hanya demi kenikmatan Harga dirimu bahkan terbuang Yang ada hanya rayuan Mengaku bujangan Kepada tiap wanita Ternyata cucunya segudang</p>
Case Folding	<p>kata katanya selangit tersenyum penuh misteri matanya membikin ngeri semua gadis gadis remaja gaya lelaki katanya semua begitu dia bilang aku cantik diapun bilang ku menarik dia bilang body ku asik hingga dada ini deg deg an mulut lelaki katanya selalu begitu apalagi dia yang masih suka dengan gadis remaja engkau lupakan anak cucumu hanya demi kenikmatan harga dirimu bahkan terbuang yang ada hanya rayuan mengaku bujangan kepada tiap wanita ternyata cucunya segudang</p>

Table 3.5. Case Folding (Lanjutan)

Kategori	Sedih (D3)
Dokumen	<p>Dirimu... Tak pernah menyadari Semua... Yang telah kau miliki Kau buang aku, tinggalkan diriku Kau... hancurkan aku seakan ku tak pernah ada Reff: Aku kan bertahan Meski takkan mungkin Menerjang kisahny Walau perih... walau perih... Salahkah... Aku terlalu cinta Berharap.. Semua kan kembali Kau buang aku,tinggalkan diriku Kau.. hancurkan aku Seakan ku tak pernah ada Back to Reff 2x</p>
Case Folding	<p>dirimu... tak pernah menyadari semua... yang telah kau miliki kau buang aku, tinggalkan diriku kau... hancurkan aku seakan ku tak pernah ada reff: aku kan bertahan meski takkan mungkin menerjang kisahny walau perih... walau perih... salahkah... aku terlalu cinta berharap.. semua kan kembali kau buang aku,tinggalkan diriku kau.. hancurkan aku seakan ku tak pernah ada back to reff 2x</p>

Tabel 3.5. *Case Folding* (Lanjutan)

Kategori	Takut (D4)
Dokumen	<p>Ku tak percaya kau ada di sini Menemaniku di saat dia pergi Sungguh bahagia kau ada di sini Menghapus semua sakit yang kurasa Mungkinkah kau merasakan Semua yang ku pasrahkan Kenanglah kasih.. Reff : Ku suka dirinya, mungkin aku sayang Namun apakah mungkin, kau menjadi milikku Kau pernah menjadi , menjadi miliknya Namun salahlah aku, bila ku pendam rasa ini Na nana nanana nana nanana Back to Reff</p>
Case Folding	<p>ku tak percaya kau ada di sini menemaniku di saat dia pergi sungguh bahagia kau ada di sini menghapus semua sakit yang kurasa mungkinkah kau merasakan semua yang ku pasrahkan kenanglah kasih.. reff : ku suka dirinya, mungkin aku sayang namun apakah mungkin, kau menjadi milikku kau pernah menjadi , menjadi miliknya namun salahlah aku, bila ku pendam rasa ini na nana nanana nana nanana back to reff</p>

Tabel 3.5. merupakan contoh dari data yang belum diproses menggunakan *case folding* dan data yang telah diproses menggunakan *case*

folding dimana data tersebut telah dirubah menjadi huruf kecil atau *lowercase* seluruhnya.

ii. *Tokenizing*

Tokenizing adalah tahapan yang dilakukan pemotongan *string* terhadap teks dalam dokumen berdasarkan setiap kata yang menyusunnya dan menghilangkan seluruh tanda baca yang ada. Hasil dari pemotongan tersebut dinamakan token. Berikut adalah hasil tahap *tokenizing* dari contoh dokumen pada Tabel 3.6.

Tabel 3.6. *Tokenizing*

Token			
Senang	Marah	Sedih	Takut
bilang	Kata	dirimu	ku
papa	katanya	Tak	tak
mu	selangit	pernah	percaya
ku	tersenyum	menyadari	kau
takkan	pernah	semua	ada
buat	misteri	yang	di
kau	matanya	telah	sini
berubah	membikin	kau	menemaniku
menjadi	negeri	miliki	di
anak	semua	kau	saat
yang	gadis	buang	dia

Tabel 3.6. *Tokenizing (Lanjutan)*

Token			
Bahagia	Marah	Sedih	Takut
nakal	gadis	aku	pergi
bilang	remaja	tinggalkan	sungguh
mama	gaya	diriku	bahagia
cinta	semua	aku	di
padamu	begitu	seakan	sini
dan	dia	ku	menghapus
tak	aku	pernah	sakit
pernah	cantik	ada	yang
main	diapun	reff	kurasa
main	bilang	aku	mungkinkah
reff	ku	kan	kau
biarkanlah	menarik	bertahan	merasakan
saja	dia	meski	semua
dulu	bilang	takkan	yang
kita	body	mungkin	ku
jalan	ku	menerjang	ku
berdua	asik	kisahny	pasrahkan
mereka	hingga	walau	kenanglah

Tabel 3.6. *Tokenizing (Lanjutan)*

Token			
Bahagia	Marah	Sedih	Takut
pun	dada	perih	reff
pernah	ini	walau	ku
muda	deg	perih	suka
pernah	degan	salahkah	dirinya
muda	mulut	aku	mungkin
bilang	lelaki	terlalu	aku
papa	katanya	cinta	sayang
mu	selalu	berharap	namun
berhanti	begitu	semua	apakah
urusin	apalagi	kan	mungkin
semua	dia	kembali	kau
urusan	yang	kau	menjadi
kau	masih	buang	milikku
dan	suka	aku	kau
aku	dengan	tinggalkan	pernah
bilang	gadis	diriku	menjadi
mama	remaja	kau	menjadi
tak	engaku	hancurkan	miliknya
perlu	lupakan	Aku	namun

Tabel 3.6. Tokenizing (Lanjutan)

Token			
Bahagia	Marah	Sedih	Takut
kuatir	anak	Seakan	salahlah
atau	hanya	ku	aku
pun	demi	tak	bila
curiga	kenikmatan	pernah	ku
kepadaku	harga	ada	pendam
biarkanlah	dirimu	back	rasa
saja	bahkan	to	ini
dulu	terbuang	reff	back
kita	yang	x	to
jalan	ada		reff
berdua	hanya		
mereka	rayuan		
pun	mengaku		
pernah	bujanga		
muda	kepada		
saatnya	tiap		
kau	wanita		
dan	ternyata		
aku	cucunya		

Tabel 3.6. *Tokenizing* (Lanjutan)

Token			
Senang	Marah	Sedih	Takut
sekarang	segudang		

Tabel 3.6. merupakan hasil dari pemotongan *string* terhadap teks dari dokumen yang telah melalui tahap *case folding* berdasarkan kata penyusunnya,

2. *Text Transformation*

Tahap ini dilakukan untuk mendapatkan representasi dokumen yang diharapkan. Pada tahap ini akan dilakukan proses *filtering* yaitu tahapan untuk mengambil kata-kata penting hasil dari token dan membuang kata-kata yang dianggap kurang penting menggunakan *stopword*. *Stopword* adalah kata-kata yang tidak memiliki pengaruh apapun dalam klasifikasi teks, seperti kata sambung (dan, dengan, dan lain-lain), kata ganti orang (dia, aku, kamu, mereka, dan lain-lain), kata penunjuk waktu (siang, malam, pagi, sore, dan lain-lain), kata tanya, dan kata lainnya yang dianggap kurang penting.

Pada penelitian ini *stopword* yang digunakan diambil dari situs www.kaggle.com/oswinrh/indonesian-stoplist yang dibuat oleh Hartono, Oswin Rahadiyan (2016). Setelah dilakukan evaluasi terhadap *stopword* yang telah ada, terdapat tambahan *stopword* seperti kata *back*, *to*, *reff*, dan kata lainnya yang biasanya terdapat di dalam struktur lirik lagu. Selain itu, terdapat penghapusan kata

dari daftar *stopword* yang telah ada seperti kata tidak, tanpa, dan bukan. Di dalam klasifikasi teks untuk kategori artikel kata-kata tersebut dianggap tidak penting sehingga dimasukkan dalam *stopword*, namun dalam klasifikasi teks emosi kata-kata tersebut menjadi sangat berarti dan tidak boleh dihilangkan (Rizki, Mochamad dan Akuwan Saleh, 2011). Daftar *stopword* yang digunakan dapat dilihat pada lampiran. Berikut hasil dari *filtering* token pada Tabel 3.7.

Tabel 3.7. Filtering

Token			
Senang	Marah	Sedih	Takut
berubah	katanya	menyadari	percaya
nakal	selangit	miliki	menemaniku
cinta	tersenyum	buang	pergi
main	misteri	tinggalkan	bahagia
main	matanya	hancurkan	menghapus
biarkanlah	membikin	seakan	sakit
jalan	ngeri	bertahan	kurasa
muda	gaya	menerjang	merasakan
muda	katanya	kisahnya	pasrahkan
berhenti	bilang	perih	kenanglah
urusin	cantik	perih	kasih
urisan	bilang	salahkah	suka
khawatir	bilang	cinta	sayang
curiga	body	berharap	milikku

Tabel 3.7. Filtering (Lanjutan)

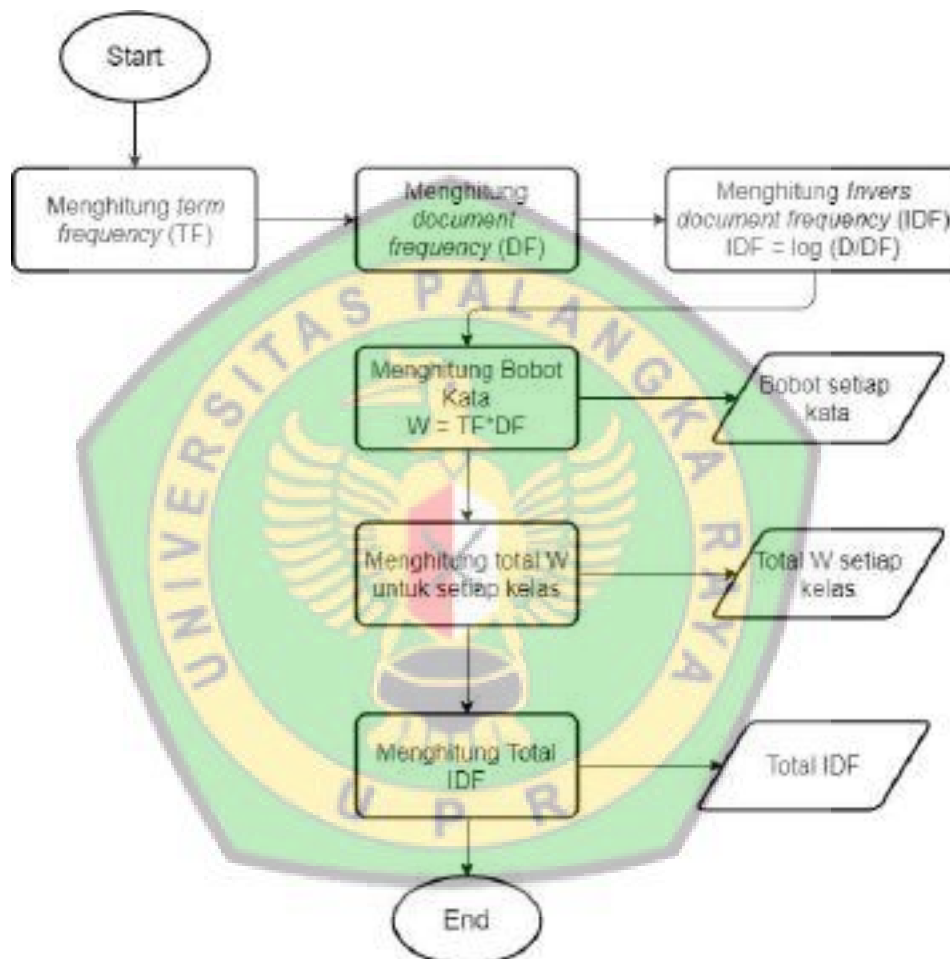
Token			
Senang	Marah	Sedih	Takut
biarkanlah	asik	kembali	miliknya
jalan	dada	buang	salahlah
	kenikmatan		
	harga		
	terbuang		
	rayuan		
	bujangan		
	segudang		

Tabel 3.7. merupakan tabel hasil dari proses *filtering* atau penghilangan kata yang dianggap kurang penting berdasarkan *stopword* yang telah dibuat. Tahap ini dilakukan terhadap hasil dari pemotongan pada proses *tokenizing* sebelumnya,

3. *Feature Selection*

Tahap dari proses ini adalah untuk memilih *term* atau kata apa saja yang dapat dijadikan sebagai wakil penting untuk kumpulan dokumen lirik lagu yang akan dianalisis dan dapat mewakili karakter seluruh dokumen. Dokumen yang akan dianalisis dapat dijadikan sebagai sebuah kumpulan vektor. Pada setiap elemen kata harus memiliki suatu nilai atau bobot. Oleh karena itu sebelum melakukan pemilihan kata yang dapat mewakili suatu dokumen harus mengetahui bobotnya

agar dapat mengetahui kata mana yang lebih menonjol dan mana yang tidak. Pembobotan yang digunakan adalah TF-IDF (*Term Frequency-Inverse Document Frequency*). Pada gambar 3.6. berikut berisikan tahapan dalam melakukan *TF-IDF*.



Gambar 3.6. Flowchart Perhitungan TF-IDF

Pada tahap pertama dilakukan perhitungan *term frequency* atau berapa kali sebuah kata muncul pada sebuah dokumen lirik lagu. Pada tahap kedua, dilakukan perhitungan *document frequency* atau menghitung jumlah dokumen yang mengandung setiap kata yang dimaksud. Setelah itu dilakukan *inverse document frequency* menggunakan rumus (2.2). *IDF* menunjukkan hubungan ketersediaan

sebuah kata dalam seluruh dokumen. Semakin sedikit jumlah dokumen yang mengandung kata yang dimaksud, maka nilai *IDF* akan semakin besar. Setelah mendapatkan nilai untuk *TF* dan *IDF*, maka akan dicari bobot dari setiap kata menggunakan rumus (2.3). Setelah itu, bobot dan *IDF* yang didapatkan akan dihitung total keseluruhannya. Jadi, dalam perhitungan probabilitas dalam klasifikasi *Naive Bayes* menggunakan total bobot untuk setiap kelas, total *IDF*, dan bobot setiap kata. Berikut adalah tabel hasil pembobotan kata dari tabel 3.8.



Tabel 3.8. *TF-IDF*

Term	Tf				Df	D/df	IDF Log(D/df)	W			
	D1	D2	D3	D4				D1	D2	D3	D4
berubah	1	0	0	0	1	4	0,60206	0,60206	0	0	0
nakal	1	0	0	0	1	4	0,60206	0,60206	0	0	0
cinta	1	0	1	0	2	2	0,30103	0,30103	0	0,30103	0
main	2	0	0	0	1	4	0,60206	1,20412	0	0	0
biarkanlah	2	0	0	0	1	4	0,60206	1,20412	0	0	0
jalan	1	0	0	0	1	4	0,60206	0,60206	0	0	0
muda	2	0	0	0	1	4	0,60206	1,20412	0	0	0
berhenti	1	0	0	0	1	4	0,60206	0,60206	0	0	0
urusin	1	0	0	0	1	4	0,60206	0,60206	0	0	0
urusan	1	0	0	0	1	4	0,60206	0,60206	0	0	0

Tabel 3.8. *TF-IDF* (Lanjutan)

Term	Tf				Df	D/df	IDF Log(D/df)	W			
	D1	D2	D3	D4				D1	D2	D3	D4
khawatir	1	0	0	0	1	4	0,60206	0,60206	0	0	0
curiga	1	0	0	0	1	4	0,60206	0,60206	0	0	0
berdua	1	0	0	0	1	4	0,60206	0,60206	0	0	0
selangit	0	1	0	0	1	4	0,60206	0	0,60206	0	0
tersenyum	0	1	0	0	1	4	0,60206	0	0,60206	0	0
misteri	0	1	0	0	1	4	0,60206	0	0,60206	0	0
matanya	0	1	0	0	1	4	0,60206	0	0,60206	0	0
membikin	0	1	0	0	1	4	0,60206	0	0,60206	0	0

Tabel 3.8. *TF-IDF* (Lanjutan)

Term	Tf				Df	D/df	IDF Log(D/df)	W			
	D1	D2	D3	D4				D1	D2	D3	D4
ngeri	0	1	0	0	1	4	0,60206	0	0,60206	0	0
gaya	0	1	0	0	1	4	0,60206	0	0,60206	0	0
bilang	0	1	0	0	1	4	0,60206	0	0,60206	0	0
cantik	0	1	0	0	1	4	0,60206	0	0,60206	0	0
body	0	1	0	0	1	4	0,60206	0	0,60206	0	0
asik	0	1	0	0	1	4	0,60206	0	0,60206	0	0
dada	0	1	0	0	1	4	0,60206	0	0,60206	0	0
mulut	0	1	0	0	1	4	0,60206	0	0,60206	0	0
suka	0	1	0	0	1	4	0,60206	0	0,60206	0	0

Tabel 3.8. *TF-IDF* (Lanjutan)

Term	Tf				Df	D/df	IDF Log(D/df)	W			
	D1	D2	D3	D4				D1	D2	D3	D4
lupakan	0	1	0	0	1	4	0,60206	0	0,60206	0	0
demi	0	1	0	0	1	4	0,60206	0	0,60206	0	0
kenikmatan	0	1	0	0	1	4	0,60206	0	0,60206	0	0
harga	0	1	0	0	1	4	0,60206	0	0,60206	0	0
terbuang	0	1	0	0	1	4	0,60206	0	0,60206	0	0
rayuan	0	1	0	0	1	4	0,60206	0	0,60206	0	0
bujangan	0	1	0	0	1	4	0,60206	0	0,60206	0	0
segudang	0	1	0	0	1	4	0,60206	0	0,60206	0	0
menyadari	0	0	1	0	1	4	0,60206	0	0	0,60206	0

Tabel 3.8. *TF-IDF* (Lanjutan)

Term	Tf				Df	D/df	IDF Log(D/df)	W			
	D1	D2	D3	D4				D1	D2	D3	D4
milik	0	0	1	0	1	4	0,60206	0	0	0,60206	0
buang	0	0	1	0	1	4	0,60206	0	0	0,60206	0
tinggalkan	0	0	2	0	1	4	0,60206	0	0	1,20412	0
hancurkan	0	0	2	0	1	4	0,60206	0	0	1,20412	0
seakan	0	0	2	0	1	4	0,60206	0	0	1,20412	0
bertahan	0	0	1	0	1	4	0,60206	0	0	0,60206	0
menerjang	0	0	1	0	1	4	0,60206	0	0	0,60206	0
kisahnya	0	0	1	0	1	4	0,60206	0	0	0,60206	0
perih	0	0	2	0	1	4	0,60206	0	0	1,20412	0

Tabel 3.8. *TF-IDF* (Lanjutan)

Term	Tf				Df	D/df	IDF Log(D/df)	W			
	D1	D2	D3	D4				D1	D2	D3	D4
salahkah	0	0	1	0	1	4	0,60206	0	0	0,60206	0
berharap	0	0	1	0	1	4	0,60206	0	0	0,60206	0
kembali	0	0	1	0	1	4	0,60206	0	0	0,60206	0
percaya	0	0	0	1	1	4	0,60206	0	0	0	0,60206
menemaniku	0	0	0	1	1	4	0,60206	0	0	0	0,60206
pergi	0	0	0	1	1	4	0,60206	0	0	0	0,60206
bahagia	0	0	0	1	1	4	0,60206	0	0	0	0,60206
menghapus	0	0	0	1	1	4	0,60206	0	0	0	0,60206
sakit	0	0	0	1	1	4	0,60206	0	0	0	0,60206

Tabel 3.8. *TF-IDF* (Lanjutan)

Term	Tf				Df	D/df	IDF Log(D/df)	W			
	D1	D2	D3	D4				D1	D2	D3	D4
kurasa	0	0	0	1	1	4	0,60206	0	0	0	0,60206
merasakan	0	0	0	1	1	4	0,60206	0	0	0	0,60206
pasrahkan	0	0	0	1	1	4	0,60206	0	0	0	0,60206
kenanglah	0	0	0	1	1	4	0,60206	0	0	0	0,60206
kasih	0	0	0	1	1	4	0,60206	0	0	0	0,60206
sayang	0	0	0	1	1	4	0,60206	0	0	0	0,60206
milikku	0	0	0	1	1	4	0,60206	0	0	0	0,60206
miliknya	0	0	0	1	1	4	0,60206	0	0	0	0,60206
salahlah	0	0	0	1	1	4	0,60206	0	0	0	0,60206

Tabel 3.8. *TF-IDF* (Lanjutan)

Term	Tf				Df	D/df	IDF Log(D/df)	W			
	D1	D2	D3	D4				D1	D2	D3	D4
pendam	0	0	0	1	1	4	0,60206	0	0	0	0,60206
rasa	0	0	0	1	1	4	0,60206	0	0	0	0,60206
Total							38,83287	9,33193	13,24532	10,23502	10,23502

Tabel 3.8. adalah *term* (kata) dari hasil *filtering* yang dibobotkan menggunakan *TF-IDF*. Pada kolom *TF* merupakan kolom yang berisikan frekuensi kemunculan kata dari setiap dokumen yang ada. Kolom *DF* merupakan kolom yang berisikan frekuensi kemunculan kata dari seluruh dokumen yang ada. Pada kolom *D/df* merupakan kolom yang berisikan jumlah dokumen dibagi dengan hasil dari nilai pada masing-masing kolom *DF*. Pada kolom *IDF* merupakan kolom yang berisikan hasil *inverse dokumen frequency (IDF)*. Pada kolom *W* merupakan kolom yang berisikan bobot untuk setiap kata dari setiap dokumen yang didapat dari hasil kali *TF* dan *IDF* seperti pada rumus (2.3). Masing-masing bobot dari setiap kelas dan nilai *IDF* akan ditotal. Bobot untuk setiap kata, total bobot setiap kelas, dan total dari *IDF* tersebut akan digunakan pada perhitungan *naïve bayes classifier*.

4. *Pattern Discovery*

Pattern discovery merupakan tahapan untuk menemukan pola atau tahap pelatihan sistem dimana tahap ini akan menghasilkan model yang akan digunakan untuk klasifikasi lirik lagu berbahasa Indonesia. Data yang dihasilkan dari pelatihan berupa kumpulan *term* atau kata yang memiliki nilai atau bobot yang dihasilkan dari perhitungan *TF-IDF* berdasarkan kelas emosi yang telah ditentukan.

Pada proses pelatihan digunakan data lirik lagu yang telah diberi label berdasarkan kelas emosinya sebagai data latih. Pada algoritma *Naive Bayes Classifier* akan dihitung probabilitas kemunculan kata untuk setiap kata menggunakan nilai bobot yang didapatkan dari perhitungan *TF-IDF*. Pada tahap

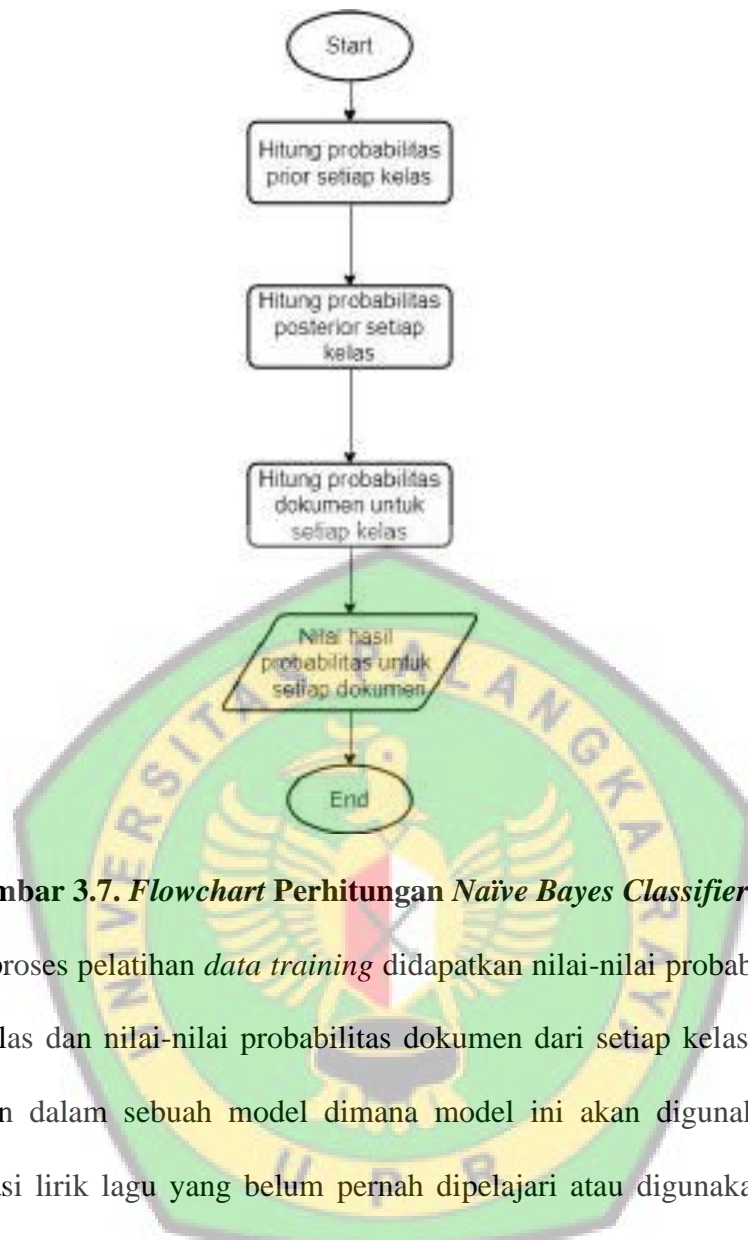
pertama akan dihitung probabilitas prior, yaitu menghitung probabilitas jumlah dokumen untuk setiap kelas dengan menggunakan rumas (2.11).

Model yang dihasilkan dari proses pelatihan akan digunakan sebagai sumber pengetahuan untuk sistem agar dapat mengklasifikasikan data baru. Data yang telah melalui tahap case folding, tokenizing, filtering, dan penghitungan bobot menggunakan TF-IDF akan disimpan sebagai kamus kata yang akan menjadi model klasifikasi dan digunakan dalam klasifikasi menggunakan naive bayes classifier.

3.2.3 Pengujian Model Klasifikasi

Klasifikasi emosi pada lirik lagu berbahasa Indonesia oleh sistem menggunakan algoritma *Naive Bayes Classifier*. Proses perhitungan pada algoritma ini menggunakan bobot kata, jumlah dokumen lirik lagu dalam setiap kelas, dan jumlah keseluruhan data latih.

Dataset yang digunakan menjadi data latih adalah data yang telah dilabelkan berdasarkan empat kelas emosi yang telah ditentukan. *Dataset* tersebut akan melalui tahap pemrosesan teks, yaitu *text preprocessing*, *text transformation*, *feature selection*, dan *pattern discovery*. Pada tahap *feature selection*, akan dilakukan pembobotan kata menggunakan algoritma *TF-IDF*. Dari pembobotan kata tersebut, maka akan didapatkan kata unik yang merepresentasikan atau mewakili masing-masing kelas. Kata-kata unik tersebut akan dipelajari oleh sistem menggunakan algoritma *Naive Bayes Classifier* dalam proses pelatihan. Berikut adalah *flowchart* perhitungan algoritma *Naive Bayes Classifier*.



Gambar 3.7. Flowchart Perhitungan Naïve Bayes Classifier

Pada proses pelatihan *data training* didapatkan nilai-nilai probabilitas kata dari setiap kelas dan nilai-nilai probabilitas dokumen dari setiap kelas. Nilai ini akan disimpan dalam sebuah model dimana model ini akan digunakan untuk mengklasifikasi lirik lagu yang belum pernah dipelajari atau digunakan sebagai *data testing*. Pada klasifikasi dari *data testing*, akan dilakukan perhitungan probabilitas kata dari setiap kelas menggunakan rumus (2.10).

Berikut ini adalah contoh penerapan *data testing*. *Data testing* merupakan data yang digunakan sebagai uji coba terhadap model yang terbentuk dari *data training*.

Kelas	Isi Dokumen
?	katamu cintaku berlebihan cemburuku tak beralasan membuat dirimu tak nyaman maafkan aku sayang reff: aku takut kehilangan dirimu aku takut takut kehilanganmu aku takut kehilangan cintamu aku takut hidup tanpa dirimu aku tak akan bisa, aku yang bisa gila bila kau pergi meninggalkanku

Data training tersebut akan melalui tahap *case folding*, *tokenizing*, dan *filtering* maka *data training* tersebut akan mendapat hasil sebagai berikut.

Tabel 3.9. Hasil untuk *Data Training*

cintaku	sayang	takut	pergi
berlebihan	takut	kehilanganmu	meninggalkanku
cemburuku	kehilangan	cintamu	
beralasan	takut	takut	
nyaman	takut	hidup	
maafkan	kehilanganmu	gila	

Tabel 3.9. merupakan hasil dari data uji yang telah melalui tahap *preprocessing* dan akan diklasifikasikan. Pada proses sebelumnya, telah dilakukan pelatihan data terhadap empat kelas emosi, yaitu senang, sedih, takut, dan marah dengan jumlah dokumen masing-masing kelas yaitu satu. Tahap pertama yang perlu

dilakukan adalah menghitung probabilitas masing-masing kelas dengan membagi jumlah dokumen disetiap kelas dengan jumlah seluruh dokumen yang digunakan pada proses pelatihan. Jadi, probabilitas untuk setiap kelas dokumen yang digunakan adalah sebagai berikut.

Tabel 3.10. Prior Probabilitas

Atribut Kelas	P(Kelas)
Senang	$P(\text{senang}) = \frac{1}{4} = 0,25$
Sedih	$P(\text{sedih}) = \frac{1}{4} = 0,25$
Takut	$P(\text{takut}) = \frac{1}{4} = 0,25$
Marah	$P(\text{marah}) = \frac{1}{4} = 0,25$

Tahap selanjutnya adalah menghitung probabilitas setiap kata. Perhitungan dilakukan dengan menghitung jumlah kata dari setiap kelas yang ada menggunakan rumus (2.12). Bobot kata dan IDF yang telah didapatkan pada proses pelatihan akan digunakan dalam perhitungan probabilitas posterior untuk setiap kata dalam data uji. Berikut adalah hasil perhitungan probabilitas posterior untuk data uji yang digunakan.

1. Probabilitas posterior di kelas senang.

$$P(\text{cintaku} | \text{senang}) = \frac{0 + 1}{9,33193 + 38,83287} = 0,02076$$

$$P(\text{berlebihan} | \text{senang}) = \frac{0 + 1}{9,33193 + 38,83287} = 0,02076$$

$$\begin{aligned}
 P(\text{cemburuku}|\text{senang}) &= \frac{0 + 1}{9,33193 + 38,83287} = 0,02076 \\
 P(\text{berasalan}|\text{senang}) &= \frac{0 + 1}{9,33193 + 38,83287} = 0,02076 \\
 P(\text{nyaman}|\text{senang}) &= \frac{0 + 1}{9,33193 + 38,83287} = 0,02076 \\
 P(\text{maafkan}|\text{senang}) &= \frac{0 + 1}{9,33193 + 38,83287} = 0,02076 \\
 P(\text{sayang}|\text{senang}) &= \frac{0 + 1}{9,33193 + 38,83287} = 0,02076 \\
 P(\text{takut}|\text{senang}) &= \frac{0 + 1}{9,33193 + 38,83287} = 0,02076 \\
 P(\text{takut}|\text{senang}) &= \frac{0 + 1}{9,33193 + 38,83287} = 0,02076 \\
 P(\text{takut}|\text{senang}) &= \frac{0 + 1}{9,33193 + 38,83287} = 0,02076 \\
 P(\text{takut}|\text{senang}) &= \frac{0 + 1}{9,33193 + 38,83287} = 0,02076 \\
 P(\text{takut}|\text{senang}) &= \frac{0 + 1}{9,33193 + 38,83287} = 0,02076 \\
 P(\text{kehilangan}|\text{senang}) &= \frac{0 + 1}{9,33193 + 38,83287} = 0,02076 \\
 P(\text{kehilanganmu}|\text{senang}) &= \frac{0 + 1}{9,33193 + 38,83287} = 0,02076 \\
 P(\text{kehilanganmu}|\text{senang}) &= \frac{0 + 1}{9,33193 + 38,83287} = 0,02076 \\
 P(\text{cintamu}|\text{senang}) &= \frac{0 + 1}{9,33193 + 38,83287} = 0,02076 \\
 P(\text{hidup}|\text{senang}) &= \frac{0 + 1}{9,33193 + 38,83287} = 0,02076 \\
 P(\text{gila}|\text{senang}) &= \frac{0 + 1}{9,33193 + 38,83287} = 0,02076 \\
 P(\text{meninggalkanku}|\text{senang}) &= \frac{0 + 1}{9,33193 + 38,83287} = 0,02076 \\
 P(\text{pergi}|\text{senang}) &= \frac{0 + 1}{9,33193 + 38,83287} = 0,02076
 \end{aligned}$$

2. Probabilitas kata di kelas sedih.

$$\begin{aligned}
 P(\text{cintaku}|\text{sedih}) &= \frac{0 + 1}{10,2305 + 38,83287} = 0,020381 \\
 P(\text{berlebihan}|\text{sedih}) &= \frac{0 + 1}{10,2305 + 38,83287} = 0,020381
 \end{aligned}$$

$$\begin{aligned}
 P(\text{cemburuku}|\text{sedih}) &= \frac{0 + 1}{10,2305 + 38,83287} = 0,020381 \\
 P(\text{berasalan}|\text{sedih}) &= \frac{0 + 1}{10,2305 + 38,83287} = 0,020381 \\
 P(\text{nyaman}|\text{sedih}) &= \frac{0 + 1}{10,2305 + 38,83287} = 0,020381 \\
 P(\text{maafkan}|\text{sedih}) &= \frac{0 + 1}{10,2305 + 38,83287} = 0,020381 \\
 P(\text{sayang}|\text{sedih}) &= \frac{0 + 1}{10,2305 + 38,83287} = 0,020381 \\
 P(\text{takut}|\text{sedih}) &= \frac{0 + 1}{10,2305 + 38,83287} = 0,020381 \\
 P(\text{takut}|\text{sedih}) &= \frac{0 + 1}{10,2305 + 38,83287} = 0,020381 \\
 P(\text{takut}|\text{sedih}) &= \frac{0 + 1}{10,2305 + 38,83287} = 0,020381 \\
 P(\text{takut}|\text{sedih}) &= \frac{0 + 1}{10,2305 + 38,83287} = 0,020381 \\
 P(\text{takut}|\text{sedih}) &= \frac{0 + 1}{10,2305 + 38,83287} = 0,020381 \\
 P(\text{kehilangan}|\text{sedih}) &= \frac{0 + 1}{10,2305 + 38,83287} = 0,020381 \\
 P(\text{kehilanganmu}|\text{sedih}) &= \frac{0 + 1}{10,2305 + 38,83287} = 0,020381 \\
 P(\text{kehilanganmu}|\text{sedih}) &= \frac{0 + 1}{10,2305 + 38,83287} = 0,020381 \\
 P(\text{cintamu}|\text{sedih}) &= \frac{0 + 1}{10,2305 + 38,83287} = 0,020381 \\
 P(\text{hidup}|\text{sedih}) &= \frac{0 + 1}{10,2305 + 38,83287} = 0,020381 \\
 P(\text{gila}|\text{sedih}) &= \frac{0 + 1}{10,2305 + 38,83287} = 0,020381 \\
 P(\text{meninggalkanku}|\text{sedih}) &= \frac{0 + 1}{10,2305 + 38,83287} = 0,020381 \\
 P(\text{pergi}|\text{sedih}) &= \frac{0 + 1}{10,2305 + 38,83287} = 0,020381
 \end{aligned}$$


3. Probabilitas kata di kelas takut.

$$\begin{aligned}
 P(\text{cintaku}|\text{takut}) &= \frac{0 + 1}{10,23502 + 38,83287} = 0,020379 \\
 P(\text{berlebihan}|\text{takut}) &= \frac{0 + 1}{10,23502 + 38,83287} = 0,020379
 \end{aligned}$$

$$\begin{aligned}
 P(\text{cemburuku}|\text{takut}) &= \frac{0 + 1}{10,23502 + 38,83287} = 0,020379 \\
 P(\text{berasalan}|\text{takut}) &= \frac{0 + 1}{10,23502 + 38,83287} = 0,020379 \\
 P(\text{nyaman}|\text{takut}) &= \frac{0 + 1}{10,23502 + 38,83287} = 0,020379 \\
 P(\text{maafkan}|\text{takut}) &= \frac{0 + 1}{10,23502 + 38,83287} = 0,020379 \\
 P(\text{sayang}|\text{takut}) &= \frac{0 + 1}{10,23502 + 38,83287} = 0,020379 \\
 P(\text{takut}|\text{takut}) &= \frac{0 + 1}{10,23502 + 38,83287} = 0,020379 \\
 P(\text{takut}|\text{takut}) &= \frac{0 + 1}{10,23502 + 38,83287} = 0,020379 \\
 P(\text{takut}|\text{takut}) &= \frac{0 + 1}{10,23502 + 38,83287} = 0,020379 \\
 P(\text{takut}|\text{takut}) &= \frac{0 + 1}{10,23502 + 38,83287} = 0,020379 \\
 P(\text{takut}|\text{takut}) &= \frac{0 + 1}{10,23502 + 38,83287} = 0,020379 \\
 P(\text{kehilangan}|\text{takut}) &= \frac{0 + 1}{10,23502 + 38,83287} = 0,020379 \\
 P(\text{kehilanganmu}|\text{takut}) &= \frac{0 + 1}{10,23502 + 38,83287} = 0,020379 \\
 P(\text{kehilanganmu}|\text{takut}) &= \frac{0 + 1}{10,23502 + 38,83287} = 0,020379 \\
 P(\text{cintamu}|\text{takut}) &= \frac{0 + 1}{10,23502 + 38,83287} = 0,020379 \\
 P(\text{hidup}|\text{takut}) &= \frac{0 + 1}{10,23502 + 38,83287} = 0,020379 \\
 P(\text{gila}|\text{takut}) &= \frac{0 + 1}{10,23502 + 38,83287} = 0,020379 \\
 P(\text{meninggalkanku}|\text{takut}) &= \frac{0 + 1}{10,23502 + 38,83287} = 0,020379 \\
 P(\text{pergi}|\text{takut}) &= \frac{0,60206 + 1}{10,23502 + 38,83287} = 0,032649
 \end{aligned}$$

4. Probabilitas kata di kelas marah.

$$\begin{aligned}
 P(\text{cintaku}|\text{marah}) &= \frac{0 + 1}{13,24532 + 38,83287} = 0,019201 \\
 P(\text{berlebihan}|\text{marah}) &= \frac{0 + 1}{13,24532 + 38,83287} = 0,019201
 \end{aligned}$$



$$\begin{aligned}
 P(\text{cemburuku}|\text{marah}) &= \frac{0 + 1}{13,24532 + 38,83287} = 0,019201 \\
 P(\text{berasalan}|\text{marah}) &= \frac{0 + 1}{13,24532 + 38,83287} = 0,019201 \\
 P(\text{nyaman}|\text{marah}) &= \frac{0 + 1}{13,24532 + 38,83287} = 0,019201 \\
 P(\text{maafkan}|\text{marah}) &= \frac{0 + 1}{13,24532 + 38,83287} = 0,019201 \\
 P(\text{sayang}|\text{marah}) &= \frac{0 + 1}{13,24532 + 38,83287} = 0,019201 \\
 P(\text{takut}|\text{marah}) &= \frac{0 + 1}{13,24532 + 38,83287} = 0,019201 \\
 P(\text{takut}|\text{marah}) &= \frac{0 + 1}{13,24532 + 38,83287} = 0,019201 \\
 P(\text{takut}|\text{marah}) &= \frac{0 + 1}{13,24532 + 38,83287} = 0,019201 \\
 P(\text{takut}|\text{marah}) &= \frac{0 + 1}{13,24532 + 38,83287} = 0,019201 \\
 P(\text{kehilangan}|\text{marah}) &= \frac{0 + 1}{13,24532 + 38,83287} = 0,019201 \\
 P(\text{kehilanganmu}|\text{marah}) &= \frac{0 + 1}{13,24532 + 38,83287} = 0,019201 \\
 P(\text{kehilanganmu}|\text{marah}) &= \frac{0 + 1}{13,24532 + 38,83287} = 0,019201 \\
 P(\text{cintamu}|\text{marah}) &= \frac{0 + 1}{13,24532 + 38,83287} = 0,019201 \\
 P(\text{hidup}|\text{marah}) &= \frac{0 + 1}{13,24532 + 38,83287} = 0,019201 \\
 P(\text{gila}|\text{marah}) &= \frac{0 + 1}{13,24532 + 38,83287} = 0,019201 \\
 P(\text{meninggalkanku}|\text{marah}) &= \frac{0 + 1}{13,24532 + 38,83287} = 0,019201 \\
 P(\text{pergi}|\text{marah}) &= \frac{0 + 1}{13,24532 + 38,83287} = 0,019201
 \end{aligned}$$

Tahap selanjutnya yaitu menentukan kelas dari dokumen yang digunakan pada data uji dengan mengalikan probabilitas prior yang didapatkan pada saat proses pelatihan dengan probabilitas kata untuk setiap kelas. Berikut adalah hasil perhitungan probabilitas kata untuk setiap kelas.

1. Kelas Senang

$$\begin{aligned}
P(D5 | senang) &= P(senang) \times P(cintaku | senang) \times P(berlebihan | senang) \\
&\times P(cemburuku | senang) \times P(beralasan | senang) \times P(nyaman | senang) \\
&\times P(maafkan | senang) \times P(maafkan | senang) \times P(sayang | senang) \times \\
&P(takut | senang) \times P(takut | senang) \times P(takut | senang) \times P(takut | senang) \\
&\times P(takut | senang) \times P(kehilangan | senang) \times P(kehilanganmu | senang) \\
&\times P(kehilanganmu | senang) \times P(cintamu | senang) \times P(hidup | senang) \times \\
&P(gila | senang) \times P(meninggalkanku | senang) \times P(pergi | senang)
\end{aligned}$$

$$\begin{aligned}
P(D5 | senang) &= 0,25 \times 0,02076 \times 0,02076 \times 0,02076 \times 0,02076 \times \\
&0,02076 \times 0,02076 \times 0,02076 \times 0,02076 \times 0,02076 \times 0,02076 \times 0,02076 \times \\
&0,02076 \times 0,02076 \times 0,02076 \times 0,02076 \times 0,02076 \times 0,02076 \times 0,02076 \times \\
&0,02076 \times 0,020505
\end{aligned}$$

$$P(D5 | senang) = 5,52696 \times 10^{35}$$

2. Kelas Sedih

$$\begin{aligned}
P(D5 | sedih) &= P(sedih) \times P(cintaku | sedih) \times P(berlebihan | sedih) \times \\
&P(cemburuku | sedih) \times P(beralasan | sedih) \times P(nyaman | sedih) \times \\
&P(maafkan | sedih) \times P(maafkan | sedih) \times P(sayang | sedih) \times P(takut | \\
&sedih) \times P(takut | sedih) \times P(takut | sedih) \times P(takut | sedih) \times P(takut | \\
&sedih) \times P(kehilangan | sedih) \times P(kehilanganmu | sedih) \times \\
&P(kehilanganmu | sedih) \times P(cintamu | sedih) \times P(hidup | sedih) \times P(gila | \\
&sedih) \times P(meninggalkanku | sedih) \times P(pergi | sedih)
\end{aligned}$$

$$P(\text{nyaman} \mid \text{marah}) \times P(\text{maafkan} \mid \text{marah}) \times P(\text{maafkan} \mid \text{marah}) \times$$

$$P(\text{sayang} \mid \text{marah}) \times P(\text{takut} \mid \text{marah}) \times P(\text{takut} \mid \text{marah}) \times P(\text{takut} \mid \text{marah})$$

$$\times P(\text{takut} \mid \text{marah}) \times P(\text{takut} \mid \text{marah}) \times P(\text{kehilangan} \mid \text{marah}) \times$$

$$P(\text{kehilanganmu} \mid \text{marah}) \times P(\text{kehilanganmu} \mid \text{marah}) \times P(\text{cintamu} \mid$$

$$\text{marah}) \times P(\text{hidup} \mid \text{marah}) \times P(\text{gila} \mid \text{marah}) \times P(\text{meninggalkanku} \mid$$

$$\text{marah}) \times P(\text{pergi} \mid \text{marah})$$

$$P(D5 \mid \text{marah}) = 0,25 \times 0,019201 \times 0,019201 \times 0,019201 \times 0,019201 \times$$

$$0,019201 \times 0,019201 \times 0,019201 \times 0,019201 \times 0,019201 \times 0,019201 \times$$

$$0,019201 \times 0,019201 \times 0,019201 \times 0,019201 \times 0,019201 \times 0,019201 \times$$

$$0,019201 \times 0,019201 \times 0,019201 \times 0,019201$$

$$P(D5 \mid \text{marah}) = 1,15989 \times 10^{35}$$

Dari hasil perhitungan probabilitas kata dari setiap kelas diatas dapat dilihat bahwa nilai untuk $P(D5 \mid \text{takut})$ lebih besar dari pada $P(D5 \mid \text{senang})$, $P(D5 \mid \text{sedih})$, $P(D5 \mid \text{marah})$. Jadi, dapat disimpulkan bahwa data uji tersebut masuk dalam kelas emosi takut.

3.3 Skenario Pengujian

Pengujian yang dilakukan dalam penelitian ini dibagi menjadi dua, yaitu pengujian terhadap fungsionalitas sistem menggunakan *black box* dan pengujian performa klasifikasi menggunakan metode akurasi. Berikut ini adalah skenario pengujian yang akan dilakukan.

3.3.1 Black Box Testing

Pada skenario pengujian sistem menggunakan *black box testing* akan menguji beberapa fungsi di dalam sistem sebagai berikut.

3.3.1.1 Admin

1. Admin dapat melakukan *login* ke halaman admin sehingga admin dapat mengelola sistem yang telah dibuat.
2. Sistem dapat menampilkan halaman lirik lagu yang dapat dikelola oleh admin sehingga fungsi kelola lirik lagu dapat berfungsi dengan baik.
3. Sistem dapat menampilkan halaman *stopword* yang dapat dikelola oleh admin sehingga fungsi kelola *stopword* dapat berfungsi dengan baik.
4. Sistem dapat menampilkan halaman tanda baca yang dapat dikelola oleh admin sehingga fungsi kelola tanda baca dapat berfungsi dengan baik.
5. Sistem dapat menampilkan halaman *dataset* yang dapat dikelola oleh admin sehingga fungsi kelola *dataset* dapat berfungsi dengan baik.
6. Admin dapat melakukan pelatihan terhadap *dataset* yang telah dibuat.
7. Admin dapat melihat hasil pelatihan terhadap *dataset* yang telah dilatih.
8. Admin dapat melakukan pengujian menggunakan data uji yang ada.

3.3.1.2 Pendengar

1. Sistem dapat menampilkan daftar lirik lagu berdasarkan emosinya.

2. Sistem dapat menampilkan isi dari daftar lirik lagu yang berupa lirik lagu, mp3, dan video sehingga fungsi sebagai tempat membaca lirik lagu dapat berjalan dengan baik.
3. Fungsi tambah lirik lagu baru dalam sistem dapat berfungsi dengan baik dan dapat melakukan klasifikasi terhadap lirik yang ditambahkan.

3.3.2 Akurasi

Pengujian metode akurasi digunakan untuk menguji performa klasifikasi yang hasilnya akan ditampilkan dalam tabel matriks confusion. Tabel matriks confusion ini berisikan jumlah lirik lagu yang diklasifikasikan dengan benar ke dalam kelas emosi yang telah ditentukan dan jumlah lirik lagu yang diklasifikasikan secara salah (tidak masuk ke dalam kelas emosi yang sesuai). Berikut adalah tabel matriks confusion yang akan digunakan.

Tabel 3.11. Matriks Confusion

Kelas		Kelas Hasil Klasifikasi			
		Senang (<i>a</i>)	Sedih (<i>b</i>)	Takut (<i>c</i>)	Marah (<i>d</i>)
Kelas Asli	Senang (<i>a</i>)	f_{aa}	f_{ab}	f_{ac}	f_{ad}
	Sedih (<i>b</i>)	f_{ba}	f_{bb}	f_{bc}	f_{bd}
	Takut (<i>c</i>)	f_{ca}	f_{cb}	f_{cc}	f_{cd}
	Marah (<i>d</i>)	f_{da}	f_{db}	f_{dc}	f_{dd}

Tabel 3.11. adalah jumlah data lirik lagu yang dipetakan berdasarkan data yang masuk ke dalam kelas yang sesuai dan data yang tidak masuk ke dalam kelas

yang sesuai. Misalnya, f_{aa} adalah jumlah lirik lagu dengan kelas emosi senang yang di petakkan dengan benar pada kelas emosi senang. f_{ab} adalah jumlah lirik lagu dengan emosi senang yang di petakkan secara salah pada kelas emosi sedih. Data yang lainnya akan didefinisikan seperti permasalahan tersebut.

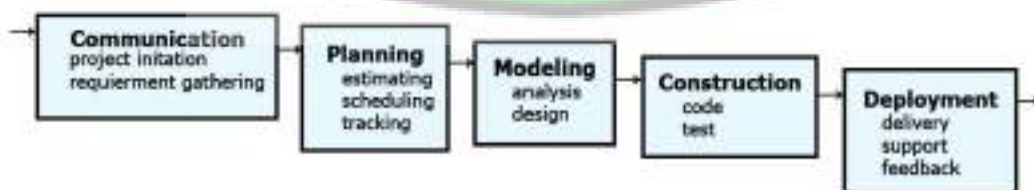
Berdasarkan isi matriks confusion, jumlah lirik lagu yang diklasifikasikan secara benar ke dalam kelasnya adalah $(f_{aa} + f_{bb} + f_{cc} + f_{dd})$. Maka akurasi pengklasifikasian yang dilakukan oleh sistem dapat dihitung sebagai berikut :

Akurasi =

$$\frac{f_{aa} + f_{bb} + f_{cc} + f_{dd}}{f_{aa} + f_{ab} + f_{ac} + f_{ad} + f_{ba} + f_{bb} + f_{bc} + f_{bd} + f_{ca} + f_{cb} + f_{cc} + f_{cd} + f_{da} + f_{db} + f_{dc} + f_{dd}}$$

3.4 Metode Pengembangan Perangkat Lunak

Metode pengembangan sistem dimana antara tahap satu ke tahap yang lainnya dilakukan secara berurutan. Dalam prosesnya, metode ini akan diselesaikan terlebih dahulu dari tahap yang pertama sebelum melanjutkan ketahap berikutnya. Pada penelitian ini menggunakan metode *Waterfall* menurut menurut Pressman (2015), berikut adalah tahapan dari metode ini.



Gambar 3.8. *Waterfall Model* Pressman, 2015

3.4.1 *Communication (Project Initiation & Requirements Gathering)*

Sebelum memulai pekerjaan yang bersifat teknis, diperlukan komunikasi untuk memahami sistem yang akan dibuat dengan dosen pembimbing. Hasil dari komunikasi tersebut yaitu analisis permasalahan yang dihadapi dan mengumpulkan data-data yang diperlukan.

Pengumpulan data-data diambil dari jurnal, buku, artikel, atau internet yang mendukung penelitian ini. Data yang dibutuhkan dalam penelitian ini, yaitu data lirik lagu, data musik dan video berdasarkan liriknya. Pengumpulan data lirik lagu dari situs website *lirik.kapanlagi.com* yang akan digunakan dalam penelitian ini. Jumlah data lirik lagu yang dikumpulkan berjumlah 200 lirik lagu, dimana data tersebut dibagi menjadi data latih dan data uji.

3.4.2 *Planning (Estimating, Scheduling, & Tracking)*

Tahap ini adalah tahapan perencanaan pengerjaan sistem yang akan dibangun yang meliputi tugas-tugas teknis yang akan dilakukan, sumber yang dibutuhkan, hasil yang akan dibuat, dan jadwal pengerjaan.

Pada penelitian ini akan dibuat sebuah sistem yang dapat melakukan klasifikasi terhadap lirik lagu ke dalam empat kelas emosi, yaitu senang, sedih, takut, dan marah. Berikut adalah kebutuhan fungsionalitas sistem yang akan dibuat dalam penelitian ini.

1. Fungsionalitas Admin

i. Kelola Lirik Lagu Baru

Sistem akan menampilkan data lirik lagu baru yang telah ditambahkan pendengar dan admin dapat mengelola lirik tersebut seperti menambahkan mp3, *link* video, atau menghapus data lirik tersebut.

ii. Kelola *Stopword*

Halaman yang berisikan *stopword* yang digunakan di dalam sistem, dimana admin dapat menambahkan, mengedit, atau menghapus kata yang ada.

iii. Kelola Tanda Baca

Halaman yang menampilkan daftar tanda baca yang digunakan, dimana admin dapat menambahkan, mengedit, dan menghapus tanda baca yang ada di dalam sistem.

iv. Kelola *Dataset*

Admin menambahkan *dataset* atau data lirik lagu yang akan digunakan untuk proses pelatihan atau pembuatan model. Admin juga akan menambahkan label untuk setiap data berdasarkan dengan label data yang telah ditentukan.

v. Pelatihan

Sistem akan melakukan pelatihan terhadap *dataset* yang telah ditambahkan oleh admin pada halaman kelola *dataset* untuk membentuk model yang akan digunakan dalam pengujian model klasifikasi.

vi. Hasil Pelatihan

Admin dapat melihat hasil pelatihan yang berisi jumlah kata dan informasi dari setiap kata yang telah diolah saat proses pelatihan.

vii. Pengujian

Admin dapat melakukan pengujian model klasifikasi terhadap data uji yang telah dilabelkan secara manual dan melihat hasil klasifikasi yang telah dilakukan oleh sistem.

2. Fungsionalitas Pendengar

i. Daftar Lirik Lagu

Sistem akan menampilkan daftar lirik lagu yang telah diklasifikasikan berdasarkan emosinya.

ii. Menampilkan Lirik Lagu, Mp3, dan Video

Sistem akan menampilkan lirik lagu berdasarkan lagu yang dipilih pada halaman daftar lirik lagu. Selain itu, terdapat lagu yang bisa diputar dan link yang menuju ke youtube untuk melihat video dari lirik lagu yang dipilih.

iii. Mengunggah Lirik Baru

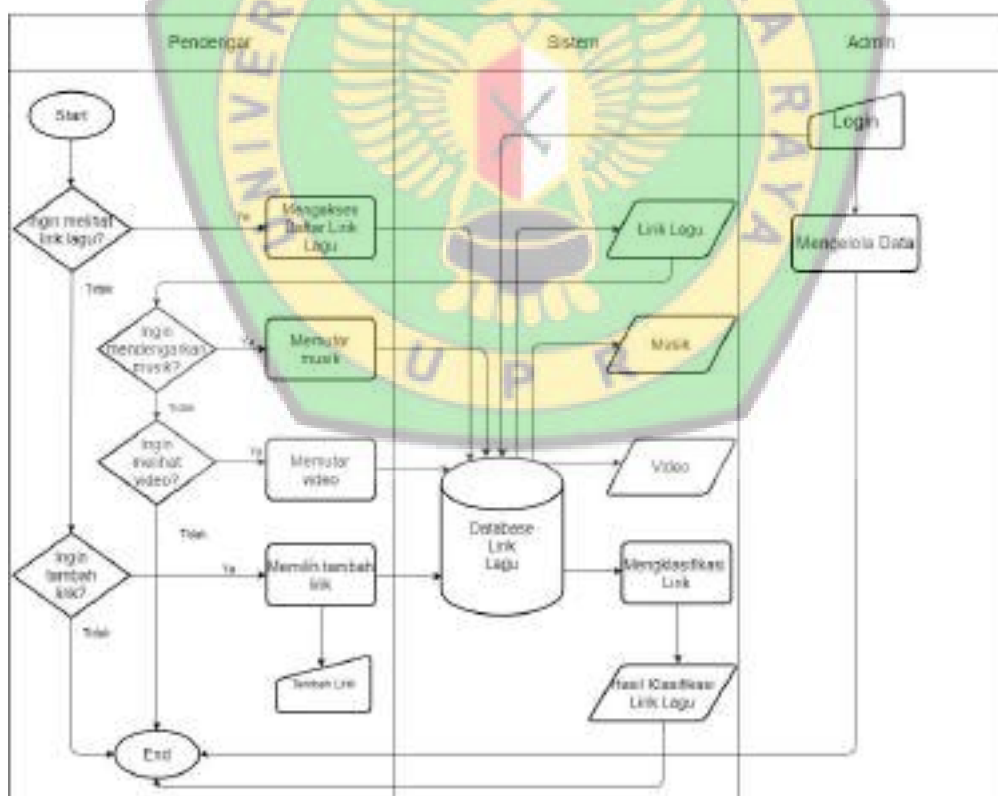
Pengguna dapat mengunggah lirik lagu ke dalam sistem dan sistem akan mengklasifikasikan lirik lagu tersebut ke dalam empat kelas emosi yang ada.

3.4.3 Modeling (Analysis & Design)

Tahap ini adalah proses pengumpulan kebutuhan dari sistem yang akan dibangun. Tahap ini dilakukan untuk pemecahan masalah dan bagaimana membangun sebuah sistem yang dapat mengklasifikasikan emosi pada lirik lagu. Pendeskripsian sistem yang dibuat dan analisis sistem menggunakan *flowchart* yang menjelaskan tentang alur suatu sistem bekerja. Tahap desain merupakan perancangan sistem dari hasil analisis kebutuhan. Desain sistem menggunakan *Unified Modelling Language (UML)*.

3.4.3.1 Analisis Sistem

Berikut ini merupakan *flowchart* proses bisnis dari sistem yang akan dibuat:



Gambar 3.9. Flowchart Sistem

Flowchart menggambarkan proses yang dapat dilakukan oleh pengguna pada bagian utama yang ditampilkan. Dimana ketika pengguna membuka website tersebut maka akan muncul daftar lirik lagu berdasarkan emosinya. Ketika pendengar memilih salah satu lagu dari daftar tersebut, maka sistem akan menampilkan lirik lagu, musik, dan link video yang dapat diputar. Pengguna dapat mengunggah lirik lagu baru pada situs *website* dan sistem akan mengklasifikasikan lirik lagu tersebut masuk ke kelas emosi senang, sedih, takut, atau marah.

3.4.3.2 Desain Sistem Menggunakan *Unified Modelling Language* (UML)

1. *Use Case Diagram*

Use case diagram merupakan pemodelan untuk kelakuan (behavior) sistem yang akan dibuat (Shalahuddin, M. dan Rosa A.S, 2016). *Use case* mendeskripsikan sebuah interaksi antara satu atau lebih aktor dengan sistem. Untuk mendefinisikan *scenario* dari penggunaan sistem maka kegiatan pengguna dibagi menjadi beberapa kegiatan yang dijabarkan dalam tabel-tabel berikut:

Tabel 3.12. Definisi Aktor

No.	Aktor	Deskripsi
1.	Pendengar dan Admin	Aktor yang akan menggunakan aplikasi.

Tabel 3.13. Definisi Case

No.	Use Case	Deskripsi
1.	Melihat daftar lirik lagu	<p><i>Pre-condition :</i></p> <p>Pendengar mengakses <i>website</i> dan pendengar ingin melihat daftar lirik lagu.</p> <p><i>Process :</i></p> <p>Menampilkan halaman daftar lirik lagu berdasarkan emosi.</p>
2	Memilih lirik lagu	<p><i>Pre-condition :</i></p> <p>Pendengar memilih lagu yang diinginkan.</p> <p><i>Process :</i></p> <p>Menampilkan daftar lirik lagu.</p>
3	Membaca lirik lagu	<p><i>Pre-condition :</i></p> <p>Pendengar mengklik lirik lagu yang ingin dibaca.</p> <p><i>Process :</i></p> <p>Menampilkan halaman yang berisi lirik lagu yang telah dipilih.</p>
4	Memutar lagu	<p><i>Pre-condition :</i></p> <p>Pendengar mengklik lirik lagu yang ingin dibaca.</p> <p><i>Process :</i></p>

Tabel 3.13. Definisi *Case* (Lanjutan)

No.	Use Case	Deskripsi
		Memutar lagu yang tersedia pada halaman lirik lagu yang telah dipilih.
5	Memutar Video	<p><i>Pre-condition :</i></p> <p>Pendengar mengklik lirik lagu yang ingin dibaca.</p> <p><i>Process :</i></p> <p>Memutar video melalui <i>link</i> yang telah disediakan pada halaman lirik lagu yang telah dipilih.</p>
6	Menambah lirik lagu baru	<p><i>Pre-condition :</i></p> <p>Pendengar mengakses <i>website</i> dan pengguna memilih menambah lirik lagu baru.</p> <p><i>Process :</i></p> <p>Menampilkan halaman tambah lirik lagu.</p>
7.	Admin Login	<p><i>Pre-condition :</i></p> <p>Admin mengakses <i>website</i>, admin login menggunakan <i>username</i> dan <i>password</i>.</p> <p><i>Process :</i></p> <p>Menampilkan halaman <i>login</i>.</p>

Tabel 3.13. Definisi Case (Lanjutan)

No.	Use Case	Deskripsi
8.	Admin mengelola data	<p><i>Pre-condition :</i></p> <p>Admin mengelola <i>database</i></p> <p><i>Process :</i></p> <p>Admin mengelola lirik lagu baru yang ditambahkan pendengar, <i>dataset</i>, data <i>stopword</i>, dan data tanda baca.</p>
9	Admin melakukan pelatihan	<p><i>Pre-condition :</i></p> <p>Admin mengelola data.</p> <p><i>Process :</i></p> <p>Admin melakukan pelatihan menggunakan data lirik lagu.</p>
10	Admin melihat hasil pelatihan	<p><i>Pre-condition :</i></p> <p>Admin telah menyelesaikan pelatihan</p> <p><i>Process :</i></p> <p>Admin memilih melihat hasil pelatihan</p>
11	Admin melakukan pengujian	<p><i>Pre-condition :</i></p> <p>Admin telah melakukan pelatihan.</p> <p><i>Process :</i></p> <p>Admin melakukan pengujian dengan melakukan klasifikasi terhadap data baru.</p>

Untuk mendefinisikan skenario dari penggunaan sistem tersebut, maka kegiatan pendengar dibagi menjadi beberapa bagian seperti pada gambar 3.10.

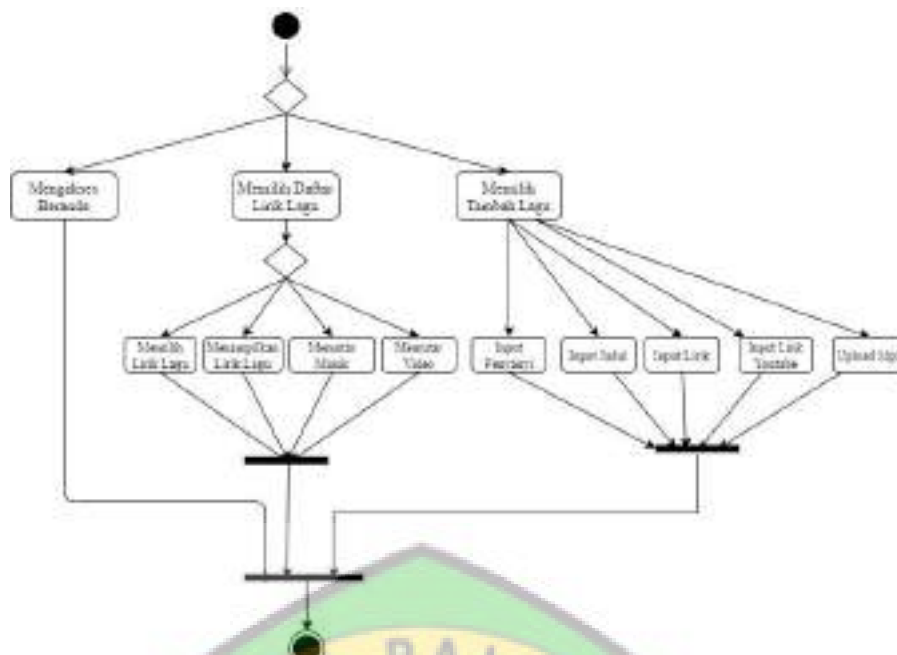


Gambar 3.10. Use Case Diagram

Gambar 3.10. merupakan gambar *use case diagram* yang menggambarkan interaksi yang dapat dilakukan oleh admin dan pendengar seperti yang telah diuraikan pada tabel 3.12. dan tabel 3.13.

2. Activity Diagram

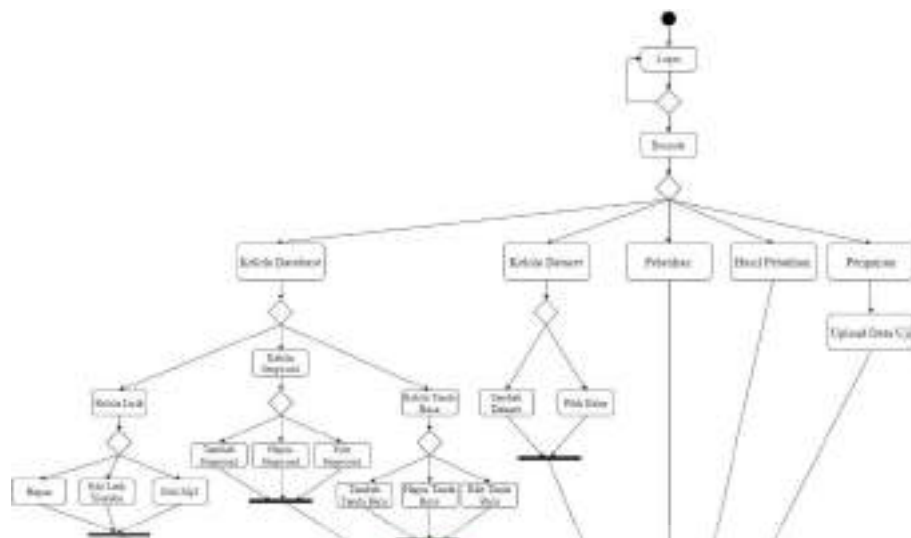
Activity diagram menggambarkan *workflow* (aliran kerja) atau aktivitas dari sebuah sistem atau proses bisnis atau menu yang ada pada perangkat lunak (Shalahuddin, M. dan Rosa A.S, 2016).



Gambar 3.11. Activity Diagram Pendengar

Gambar 3.11. adalah *activity diagram* yang terdapat pada sistem yang digunakan oleh pendengar.

1. Pada bagian awal pendengar akan melihat daftar lirik lagu.
2. Pendengar dapat memilih salah satu lirik lagu yang akan dilihat dari daftar yang tersedia.
3. Apabila lirik lagu yang diinginkan telah dipilih, maka *website* akan menampilkan halaman lirik lagu yang telah dipilih, musik, dan video.
4. Pendengar dapat memilih untuk menambahkan lirik lagu baru pada *website* dan sistem akan mengklasifikasikan lirik lagu tersebut berdasarkan empat kelas emosi yang tersedia.



Gambar 3.12. Activity Diagram Admin

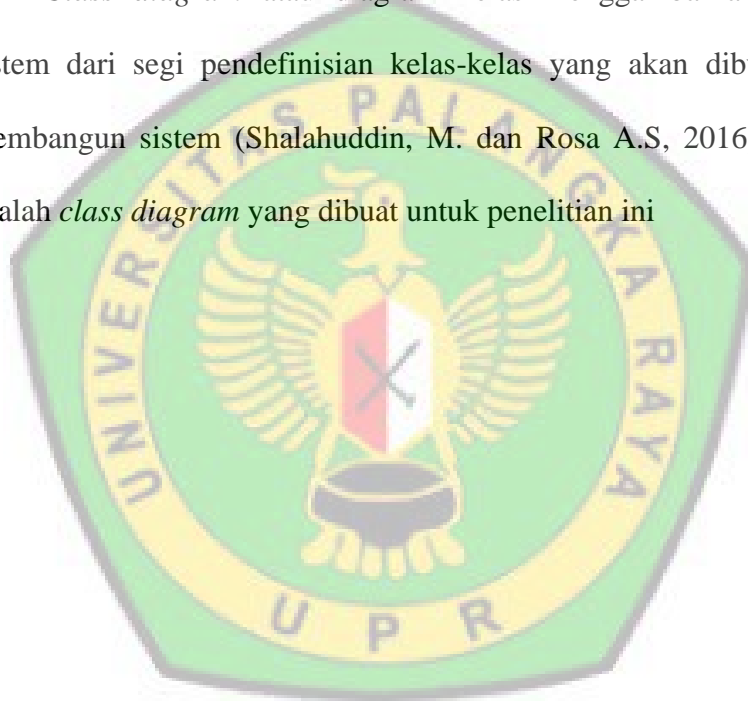
Gambar 3.12. adalah *activity diagram* yang terdapat dalam sistem yang digunakan oleh admin.

1. Pada bagian awal admin akan *login* terlebih dahulu pada sistem dengan memasukan *username* dan *password*.
2. Admin dapat melakukan kelola data, dimana data yang dikelola adalah lirik, *stopword*, dan tanda baca.
3. Admin dapat memilih menu kelola *dataset* untuk melihat data yang telah dilatih dan memberikan label pada data tersebut.
4. Admin dapat memilih menu pelatihan untuk melakukan pelatihan terhadap *dataset* yang ada.

5. Admin dapat melihat hasil pelatihan setelah selesai melakukan pelatihan pada menu hasil pelatihan.
6. Admin dapat memilih menu pengujian untuk melakukan pengujian terhadap data uji atau lirik lagu yang baru ditambahkan oleh pendengar.

3. *Class Diagram*

Class diagram atau diagram kelas menggambarkan struktur sistem dari segi pendefinisian kelas-kelas yang akan dibuat untuk membangun sistem (Shalahuddin, M. dan Rosa A.S, 2016). Berikut adalah *class diagram* yang dibuat untuk penelitian ini



Pada Gambar 3.13. menunjukkan *class diagram* yang akan diimplementasikan ke dalam sistem yang akan dibuat. Pada *class diagram* tersebut terdapat diagram UI yang digunakan untuk menampilkan sistem ke pemakai. Diagram main yang memiliki fungsi awal dieksekusi ketika sistem dijalankan. Diagram tabel yang berisikan tabel-tabel dari *database* yang akan digunakan di dalam sistem. Berikut adalah desain *database* yang berisikan struktur tabel-tabel yang digunakan dalam pembuatan sistem ini.

a. Tabel Admin

Tabel 3.14. Tabel Admin

Atribut	Tipe Data	Panjang Data	Keterangan
id_admin	Int	3	Primary Key
username	Varchar	10	
password	Varchar	8	

b. Tabel Lirik

Tabel 3.15. Tabel Lirik

Atribut	Tipe Data	Panjang Data	Keterangan
id_lirik	Int	3	Primary Key
lirik	Varchar	100	
id_kelas	Int	3	
status_latih	Varchar	10	

c. Tabel Kelas

Tabel 3.16. Tabel Kelas

Atribut	Tipe Data	Panjang Data	Keterangan
id_kelas	Int	3	Primary Key
kelas	Varchar	10	
prior	Float		

d. Tabel *Dataset***Tabel 3.17. Tabel *Dataset***

Atribut	Tipe Data	Panjang Data	Keterangan
id_dataset	Int	4	Primary Key
nama_dataset	Varchar	25	
id_lirik	Int	3	
id_kelas	Int	3	
total_dataset	Int	3	
idf	Float		
bobot	Float		

e. Tabel Lirik Uji

Tabel 3.18. Tabel Lirik Uji

Atribut	Tipe Data	Panjang Data	Keterangan
id_lirik_uji	Int	3	Primary Key

Tabel 3.18. Tabel Lirik Uji (Lanjutan)

Atribut	Tipe Data	Panjang Data	Keterangan
isi_lirik_uji	Varchar	255	
id_kelas	int	3	
created_at	Timestamp		

f. Tabel Lirik Baru

Tabel 3.19. Tabel Lirik Baru

Atribut	Tipe Data	Panjang Data	Keterangan
id_lirik_baru	Int	3	Primary Key
penyanyi	Varchar	100	
judul	Varchar	100	
lirik	Text		
filename	Varchar	100	
youtube	Varchar	100	
id_kelas	int	3	
waktu	Timestamp		

g. Tabel *Stopword***Tabel 3.20. Tabel *Stopword***

Atribut	Tipe Data	Panjang Data	Keterangan
id_stopword	Int	5	Primary Key
nama_stopword	Varchar	50	

h. Tabel Tanda Baca

Tabel 3.21. Tabel Tanda Baca

Atribut	Tipe Data	Panjang Data	Keterangan
id_tanda_baca	Int	3	Primary Key
simbol_tanda_baca	Varchar	5	
nama_tanda_baca	Varchar	50	

i. Tabel Kamus

Tabel 3.22. Tabel Tanda Baca

Atribut	Tipe Data	Panjang Data	Keterangan
id_kata	Int	10	Primary Key
kata	Varchar	50	

3.4.3.3 Design User Interface

Desain antarmuka dibuat berdasarkan kebutuhan dan kenyamanan pendengar dalam menggunakan situs *website*, dimana dalam desain antar muka akan mengedepankan sisi kesederhaan. Berikut merupakan desain antar muka situs *website* yang dibuat.

1. Halaman Admin

i. *Login*



ColourLyrics

LOGIN

Username

Password

Masuk

Gambar 3.14. Halaman *Login*

Pada halaman *login* admin akan memasukkan *username* dan *password* yang sudah terdaftar dan selanjutnya akan masuk pada halaman admin. Jika *username* dan *password* yang dimasukkan tidak sesuai dengan data yang sudah terdaftar maka tidak dapat masuk ke halaman admin dan tetap berada di halaman *login*.

ii. Halaman Utama Admin



Gambar 3.15. Halaman Utama Admin

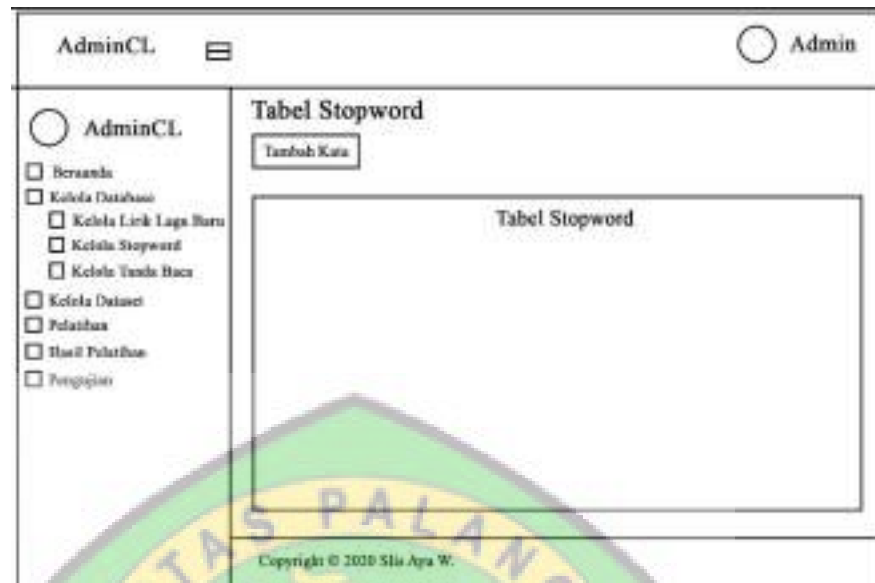
Halaman ini merupakan halaman utama admin setelah admin berhasil melakukan *login* ke sistem.

iii. Halaman Kelola Lirik Lagu Baru



Gambar 3.16. Halaman Kelola Lirik Lagu Baru

Halaman ini merupakan halaman kelola lirik lagu yang merupakan bagian dari menu kelola database. Pada halaman ini terdapat kumpulan data penyanyi, judul lagu, dan lirik lagu dari data yang ditambahkan pengguna. Pada halaman ini admin dapat menambahkan *link youtube* dan *upload* audio yang akan digunakan.

iv. Halaman Kelola *Stopword*

Gambar 3.17. Halaman Kelola *Stopword*

Pada halaman ini admin dapat melihat daftar *stopword* yang telah dimasukkan dan admin dapat mengelola data *stopword* tersebut. Admin dapat menambahkan kata *stopword* baru yang dibutuhkan dan dapat melakukan aksi hapus atau mengubah kata *stopword* yang ada pada daftar.

v. Halaman Kelola Tanda Baca



Gambar 3.18. Halaman Kelola Tanda Baca

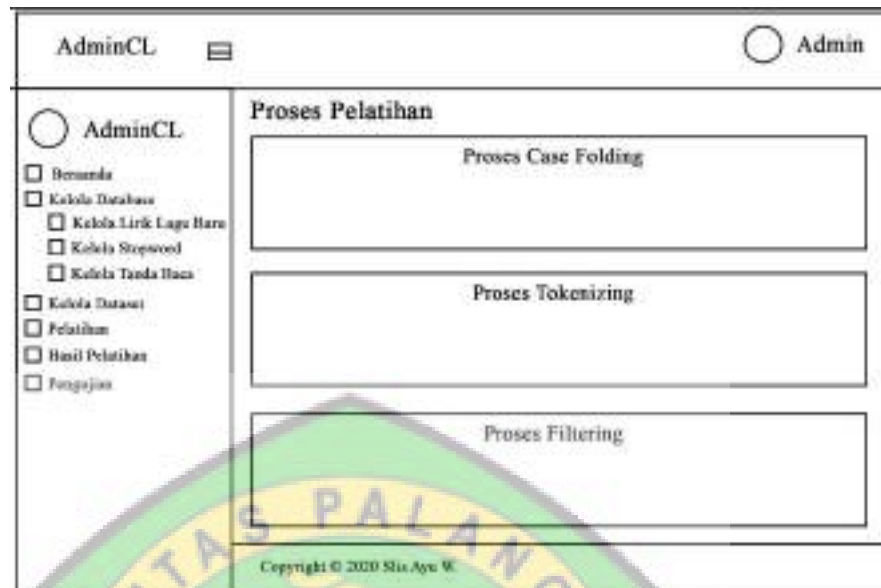
Pada halaman ini admin akan melihat daftar tanda baca yang telah ditambahkan sebelumnya. Admin dapat melakukan aksi hapus dan edit tanda baca. Selain itu, admin juga dapat menambahkan tanda baca yang belum ditambahkan.

vi. Halaman Kelola *Dataset*

Gambar 3.19. Halaman Kelola *Dataset*

Pada halaman ini akan ditampilkan *dataset* yang akan digunakan untuk proses pelatihan. Admin dapat melihat data lirik lagu yang sudah atau yang belum dilatih. Selain itu, admin dapat melakukan aksi pilih kelas untuk setiap data lirik lagu. Admin juga dapat menambahkan *dataset* yang akan digunakan, dimana ekstensi file yang digunakan yaitu *.txt* dan jumlah *multiple* data yang bisa ditambahkan maksimal 20 data lirik lagu.

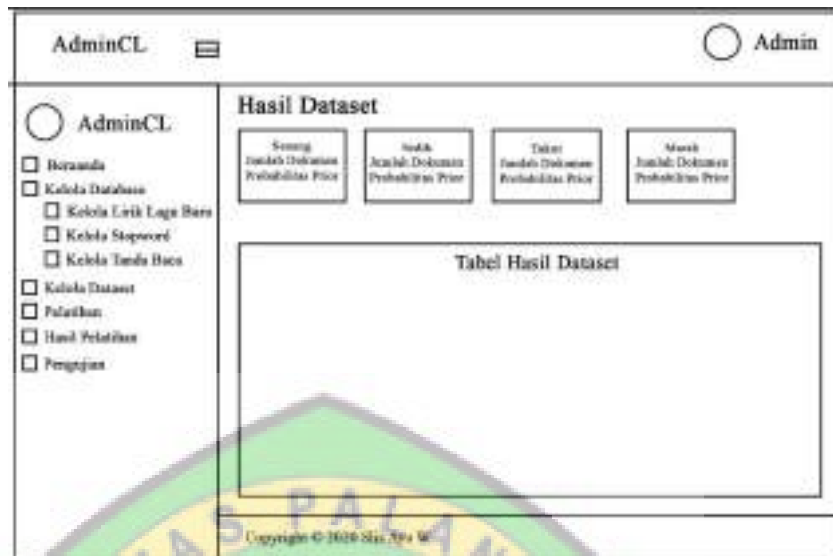
vii. Halaman Pelatihan



Gambar 3.20. Halaman Proses Pelatihan

Pada halaman ini akan ditampilkan proses pelatihan untuk *dataset* yang status latihnya belum. Proses yang akan ditampilkan pada halaman ini, yaitu *case folding*, *tokenizing*, dan *filtering*. Hasil dari proses pelatihan ini dapat dilihat pada halaman hasil pelatihan.


viii. Halaman Hasil Pelatihan



Gambar 3.21. Halaman Hasil Pelatihan

Pada halaman ini admin dapat melihat jumlah *dataset* yang masuk untuk setiap kelas emosi yang ada. Selain itu admin dapat melihat daftar kata hasil dari proses pelatihan beserta kelas untuk setiap kata, jumlah kemunculan kata (TF), nilai IDF , dan bobot untuk setiap kata yang akan digunakan pada proses pengujian.

ix. Halaman Pengujian



The screenshot shows the AdminCL web application interface. At the top, there is a header with 'AdminCL' on the left and a user profile icon labeled 'Admin' on the right. Below the header, the main content area is divided into two sections. On the left is a sidebar menu with a 'AdminCL' header and several menu items, each with a checkbox: 'Beranda', 'Kelola Database', 'Kelola Link Logo Baru', 'Kelola Stigword', 'Kelola Tenda Baca', 'Kelola Dataset', 'Pelatihan', 'Hasil Pelatihan', and 'Pengujian'. The main content area is titled 'Proses Pengujian' and contains a 'File Uji' section with 'Choose File' and 'Uji File' buttons. Below this are four stacked boxes representing the testing process: 'Proses Case Folding', 'Proses Tokenizing', 'Proses Filtering', and 'Proses Naive Bayes'. At the bottom of the main content area, there is a copyright notice: 'Copyright © 2020 Mix Aya W.'.

Gambar 3.22. Halaman Proses Pengujian

Pada halaman ini merupakan halaman proses pengujian. Untuk melakukan pengujian, admin akan menambahkan file yang akan diujikan dimana ekstensi file yang akan digunakan yaitu *.txt*. Admin dapat melihat proses *case folding*, *tokenizing*, *filtering*, dan proses perhitungan klasifikasi menggunakan algoritma *naïve bayes classifier*.

2. Halaman Pendengar

i. Halaman Beranda Pendengar



Gambar 3.23. Halaman Utama Pendengar

Halaman ini adalah halaman utama yang pertama kali diakses oleh pendengar ketika membuka *website*. Pendengar akan melihat deskripsi untuk *website*, menu daftar lirik lagu, dan tambah lirik lagu.

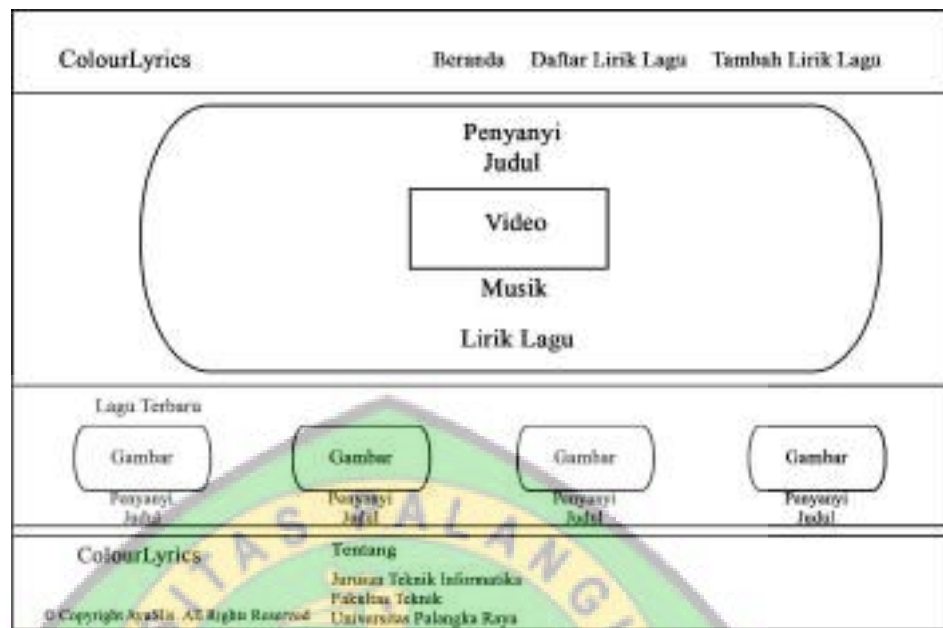
ii. Halaman Daftar Lirik Lagu



Gambar 3.24. Halaman Daftar Lirik Lagu

Halaman ini akan menampilkan daftar lirik lagu yang telah diklasifikasikan berdasarkan emosinya melalui tahap pemrosesan teks, pembobotan kata, dan klasifikasi menggunakan algoritma *Naïve Bayes Classifier*. Lirik lagu tersebut diklasifikasikan kedalam empat kelas, yaitu senang, sedih, takut, dan marah. Untuk melihat isi dari lagu yang dipilih pendengar dapat klik judul lirik lagu yang dipilih dan akan langsung masuk pada halaman isi lirik lagu.

iii. Halaman Isi Lirik



Gambar 3.25. Halaman Isi Lirik

Halaman ini adalah halaman yang ditampilkan ketika pengguna telah memilih salah satu lagu yang ingin dilihat pada halaman lirik lagu. Pada halaman ini akan ditampilkan lirik dari lagu yang dipilih. Selain itu, halaman ini akan menampilkan video yang terhubung ke situs *youtube* dan musik yang dapat diputar berikut adalah halaman lirik lagu. Pada bagian setelah isi lirik akan ditampilkan lirik lagu terbaru yang dapat diakses atau dipilih oleh pendengar.

iv. Halaman Tambah Lirik Lagu

Gambar 3.26. Halaman Tambah Lirik Lagu

Halaman ini merupakan halaman tambah lirik lagu oleh pendengar. Ketika menambahkan lirik lagu, pendengar akan memasukan penyanyi, judul, dan lirik pada *form* yang tersedia. Ketika pendengar klik tambah lirik, sistem akan melakukan proses klasifikasi pada lirik yang ditambahkan dan ketika selesai sistem akan menampilkan jendela dialog yang menampilkan informasi kelas emosi yang dimiliki oleh lirik tersebut.

3.4.4 Construction (Code & Test)

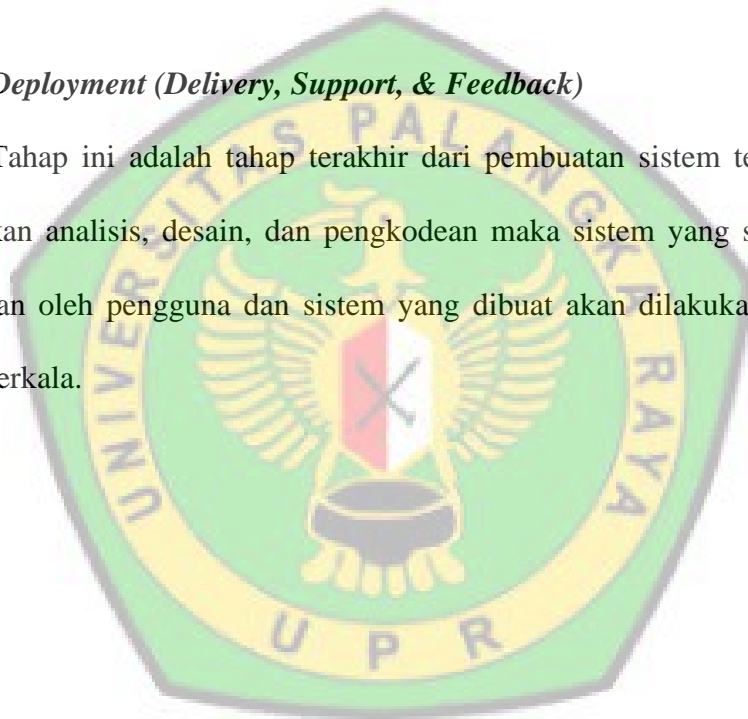
Tahap ini adalah pengimplementasian rancangan sistem yang sudah dibuat dan penerapan dari algoritma yang digunakan kedalam kode program untuk membangun sistem. Bahasa pemrograman yang digunakan adalah php menggunakan aplikasi sublime dan database menggunakan MySQL.

Penyatuan unit program kemudian diuji secara keseluruhan. Pengujian fungsionalitas sistem dilakukan menggunakan *black box testing* untuk mengetahui sistem dapat berjalan dan berfungsi dengan baik sebagaimana mestinya

Pengujian performa klasifikasi sistem menggunakan metode akurasi (*accuracy*) yang dibantu dengan matriks confusion untuk menghitung besarnya akurasi dari klasifikasi emosi pada lirik lagu yang dilakukan oleh sistem.

3.4.5 Deployment (Delivery, Support, & Feedback)

Tahap ini adalah tahap terakhir dari pembuatan sistem tersebut. Setelah melakukan analisis, desain, dan pengkodean maka sistem yang sudah jadi akan digunakan oleh pengguna dan sistem yang dibuat akan dilakukan pemeliharaan secara berkala.



BAB IV

HASIL DAN PEMBAHASAN

Pada tahap ini akan menjelaskan dan menampilkan hasil proses yang ada pada sistem berupa penggunaan program. Berikut ini adalah tahapan yang dilakukan.

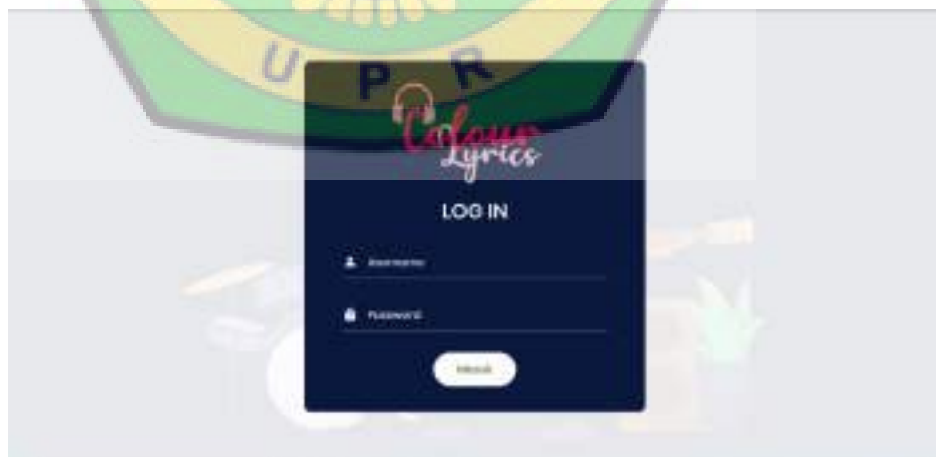
4.1 *Construction (Code & Test)*

Tahap *construction* adalah tahap pembuatan atau pengimplementasian rancangan aplikasi yang sudah dibuat dimana aplikasi yang dibuat berdasarkan dengan desain yang telah dibuat pada tahap sebelumnya. Hal-hal yang akan diimplementasikan pada aplikasi ini yaitu sebagai berikut.

4.1.1 Implementasi Halaman Admin

Berikut ini adalah halaman-halaman yang diimplementasikan dari desain *user interface* pada tahap sebelumnya.

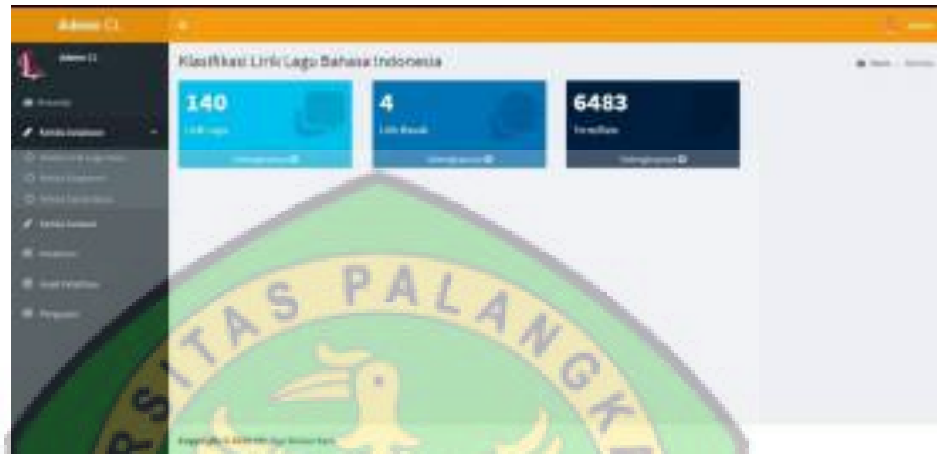
1. Implementasi Halaman *Login*



Gambar 4.1. Halaman *Login*

Halaman ini adalah tampilan halaman *login* admin pada saat memulai aplikasi pertama kali. Dalam *form* ini terdapat menu *login username* dan *password*.

2. Implementasi Halaman Utama Admin



Gambar 4.2. Halaman Utama Admin

Halaman utama admin adalah halaman yang muncul ketika admin berhasil melakukan *login*.

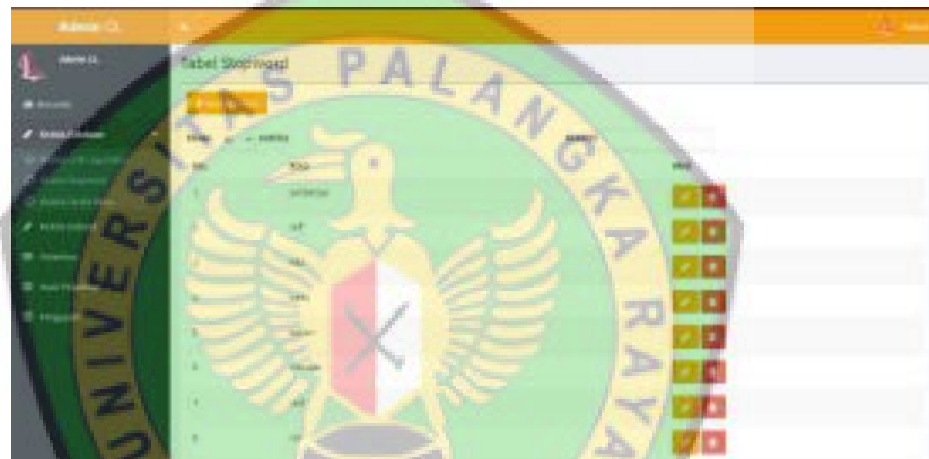
3. Implementasi Halaman Kelola Lirik Lagu Baru



Gambar 4.3. Halaman Kelola Lirik Lagu Baru

Halaman kelola lirik lagu baru adalah halaman yang berisikan data lirik lagu yang telah ditambahkan oleh pendengar berupa penyanyi, judul, dan lirik lagu. Setelah itu, admin akan menambahkan *link youtube* dan musik dari lirik lagu yang telah ditambahkan. Admin juga dapat melakukan aksi hapus data pada halaman tersebut.

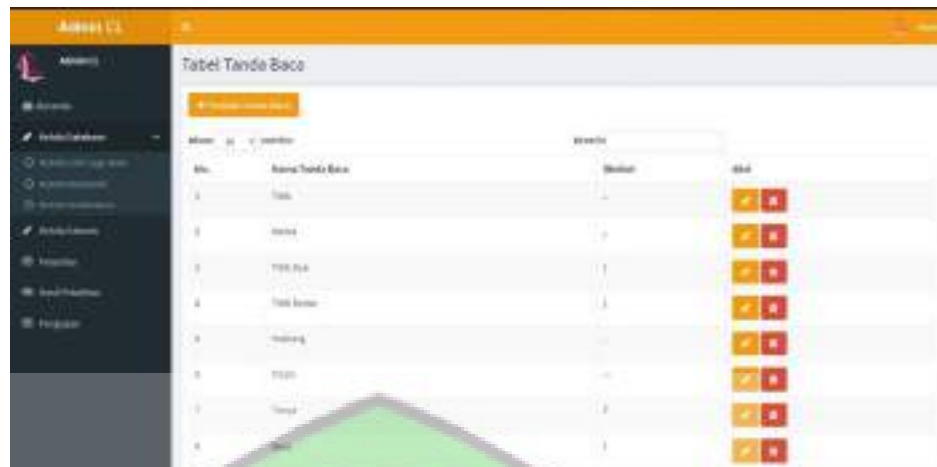
4. Implementasi Halaman Kelola *Stopword*



Gambar 4.4. Halaman Kelola Stopword

Halaman kelola *stopword* berisikan daftar kata *stopword*, dimana admin dapat mengelola kata *stopword* seperti menghapus dan mengubah kata *stopword*. Pada halaman ini admin juga dapat menambahkan kata *stopword* baru.

5. Implementasi Halaman Kelola Tanda Baca



Gambar 4.5. Halaman Kelola Tanda Baca

Halaman kelola tanda baca berisikan daftar tanda baca, dimana admin dapat mengelola tanda baca seperti menghapus tanda baca, mengedit tanda baca, dan menambah tanda baca baru yang ingin digunakan.

6. Implementasi Halaman Kelola *Dataset*



Gambar 4.6. Halaman Kelola Dataset

Halaman kelola *dataset* adalah halaman yang berisikan data lirik lagu yang dapat dikelola oleh admin, dimana admin dapat melebelkan data lirik lagu berdasarkan label lirik lagu yang telah dilabelkan secara manual. *Dataset* yang tersedia adalah data latih yang akan digunakan pada proses pelatihan.

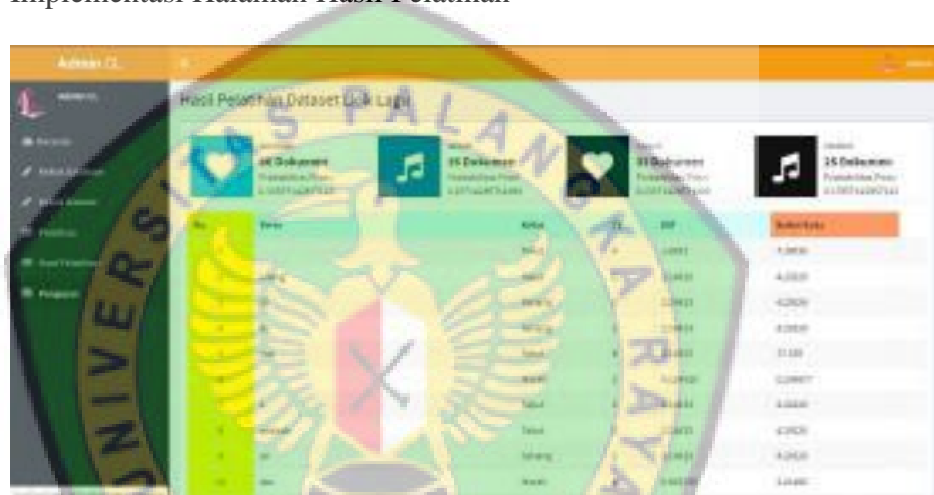
7. Implementasi Halaman Pelatihan



Gambar 4.7. Halaman Pelatihan

Halaman pelatihan adalah halaman yang menampilkan proses pelatihan, dimana data yang digunakan adalah data lirik pada halaman kelola *dataset* yang nilai status latihnya belum. Proses pelatihan yang ditampilkan pada halaman ini meliputi *case folding*, *tokenizing*, dan *filtering*.

8. Implementasi Halaman Hasil Pelatihan



Gambar 4.8. Halaman Hasil Pelatihan

Pada halaman ini admin dapat melihat jumlah *dataset* yang masuk untuk setiap kelas emosi yang ada. Selain itu admin dapat melihat daftar kata hasil dari proses pelatihan beserta kelas untuk setiap kata, jumlah kemunculan kata (*TF*), nilai *IDF*, dan bobot untuk setiap kata yang akan digunakan pada proses pengujian.

9. Implementasi Halaman Pengujian



Gambar 4.9. Halaman Pengujian

Pada halaman ini merupakan halaman proses pengujian. Untuk melakukan pengujian, admin akan menambahkan file yang akan diujikan dimana ekstensi file yang akan digunakan yaitu *.txt*. Admin dapat melihat proses *case folding*, *tokenizing*, *filtering*, dan proses perhitungan klasifikasi menggunakan algoritma *naïve bayes classifier*.

4.1.2 Implementasi Halaman Pendengar

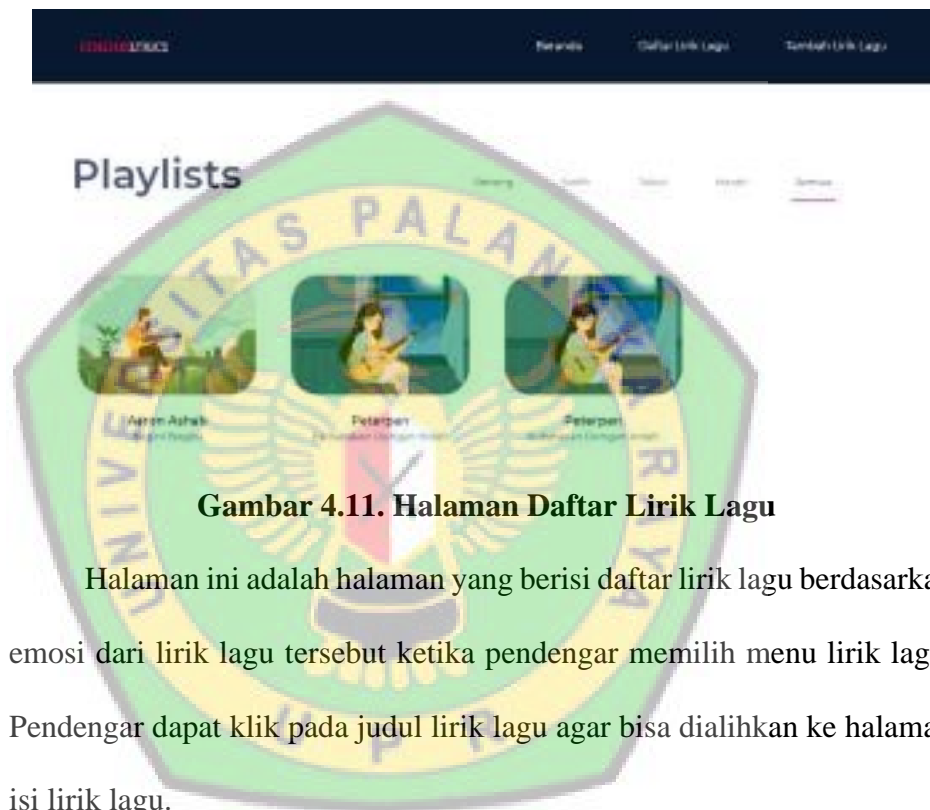
1. Implementasi Halaman Utama Pendengar



Gambar 4.10. Halaman Utama Pendengar

Halaman ini adalah halaman utama pendengar atau beranda dimana pada halaman ini pendengar dapat memilih menu daftar lirik lagu dan tambah lirik lagu.

2. Implementasi Halaman Daftar Lirik Lagu



3. Implementasi Halaman Lirik Lagu



Gambar 4.12. Halaman Lirik Lagu

Halaman ini adalah halaman yang akan tampil jika pendengar telah memilih lagu yang diinginkan. Halaman ini berisikan lirik lagu, mp3, dan video tentang lagu tersebut. Selain itu, akan ditampilkan lirik lagu terbaru pada *section* setelah isi dari lirik lagu tersebut.

4. Implementasi Halaman Tambah Lirik Lagu



Gambar 4.13. Halaman Tambah Lirik Lagu

Halaman ini adalah halaman yang berisikan form untuk menambahkan lirik lagu baru oleh pendengar dimana pendengar dapat menginputkan penyanyi, judul lagu, dan lirik lagu yang diinginkan dan jika ingin, pendengar dapat menambahkan *link youtube* dan musik dari lagu yang ingin ditambahkan. Ketika pendengar telah selesai mengisi *form* yang telah disediakan dan memilih *button* tambah lirik, maka sistem akan melakukan klasifikasi terhadap lirik tersebut dan akan menampilkan *section* yang berisikan informasi mengenai emosi dari lirik yang telah ditambahkan. Selanjutnya, lirik lagu tersebut akan masuk pada daftar lirik yang ada pada halaman daftar lirik lagu. Berikut adalah tampilan *section* yang berisikan informasi emosi dari lirik lagu yang ditambahkan.



Gambar 4.14. Section Informasi

4.1.3 Pengujian Sistem

Pengujian pada sistem ini menggunakan metode akurasi untuk menguji performa klasifikasi dan *black box* digunakan untuk menguji fungsionalitas sistem apakah berfungsi sebagaimana mestinya atau tidak.

4.1.3.1 Akurasi

Data yang digunakan dalam pengujian berupa file dengan format *teks.txt*. Total keseluruhan data lirik lagu adalah 200 dokumen lirik lagu dengan kelas emosi senang, sedih, takut, dan marah. Komposisi data yang digunakan dalam pengujian ini seperti pada Tabel 4.1, dilakukan empat kali pengujian untuk melihat pola data yang optimal untuk sistem tersebut.

Tabel 4.1. Pemabagian Data

Komposisi Data		Jumlah Data	
Data Latih	Data Uji	Data Latih	Data Uji
60%	40%	120	80
70%	30%	140	60
80%	20%	160	40
90%	10%	180	20

Hasil pengujian sistem klasifikasi emosi pada lirik lagu Berbahasa Indonesia dengan menggunakan algoritma *Naive Bayes Classifier* didapatkan dari hasil skenario pengujian pada subbab 3.3. Akurasi dihitung dari persentase jumlah data

yang benar terhadap keseluruhan data. Hasil perhitungan klasifikasi dapat dilihat pada tabel 4.2, 4.3, 4.4, dan 4.5.

Table 4.2. Matriks Confusion 60%-40%

Kelas		Kelas Hasil Klasifikasi			
		Senang	Sedih	Takut	Marah
Kelas Asli	Senang	6	9	2	3
	Sedih	1	13	4	2
	Takut	2	8	6	4
	Marah	2	8	0	10

$$\begin{aligned}
 \text{Akurasi} &= \frac{6 + 13 + 6 + 10}{80} \times 100\% \\
 &= \frac{35}{80} \times 100\% = 43,75\%
 \end{aligned}$$

Dari hasil pengujian diatas dapat diketahui kemampuan sistem untuk membantu mengklasifikasikan data teks sebesar 43,75%. Dengan masing-masing jumlah data uji setiap kelas, yaitu senang 20 data, sedih, 20 data, takut 20 data, dan marah 20 data.

Table 4.3. Matriks Confusion 70%-30%

Kelas		Kelas Hasil Klasifikasi			
		Senang	Sedih	Takut	Marah
Kelas Asli	Senang	8	5	1	10
	Sedih	1	10	4	0
	Takut	0	6	8	1
	Marah	2	2	0	10

$$Akurasi = \frac{8 + 10 + 8 + 10}{60} \times 100\%$$

$$= \frac{36}{60} \times 100\% = 60\%$$

Dari hasil pengujian diatas dapat diketahui kemampuan sistem untuk membantu mengklasifikasikan data teks sebesar 60% dengan jumlah data uji yang digunakan 15 data uji senang, 15 data uji sedih, 15 data uji takut, dan 15 data uji marah.

Table 4.4. Matriks Confusion 80%-20%

Kelas		Kelas Hasil Klasifikasi			
		Senang	Sedih	Takut	Marah
Kelas Asli	Senang	6	3	1	0
	Sedih	0	9	1	0
	Takut	1	2	6	1
	Marah	1	0	0	9

$$\begin{aligned}
 \text{Akurasi} &= \frac{6 + 9 + 6 + 9}{40} \times 100\% \\
 &= \frac{30}{40} \times 100\% = 75\%
 \end{aligned}$$

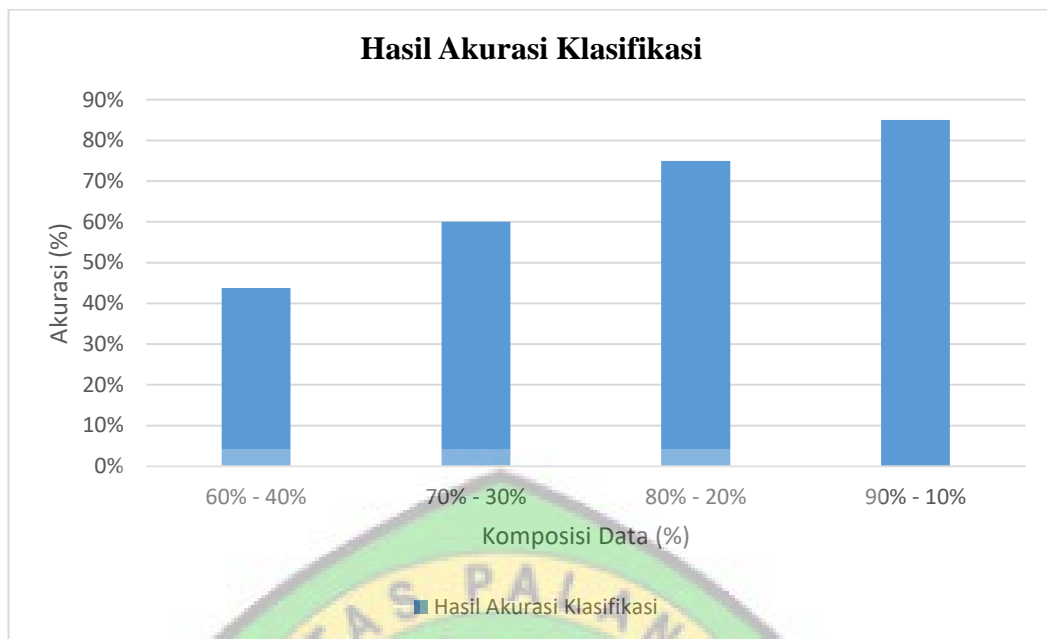
Dari hasil pengujian diatas dapat diketahui kemampuan sistem untuk membantu mengklasifikasikan data teks sebesar 75% dengan jumlah data uji yang digunakan 10 data uji senang, 10 data uji sedih, 10 data uji takut, dan 10 data uji marah

Table 4.5. Matriks Confusion 90%-10%

Kelas		Kelas Hasil Klasifikasi			
		Senang	Sedih	Takut	Marah
Kelas Asli	Senang	4	1	0	0
	Sedih	0	5	0	0
	Takut	0	1	4	0
	Marah	0	1	0	4

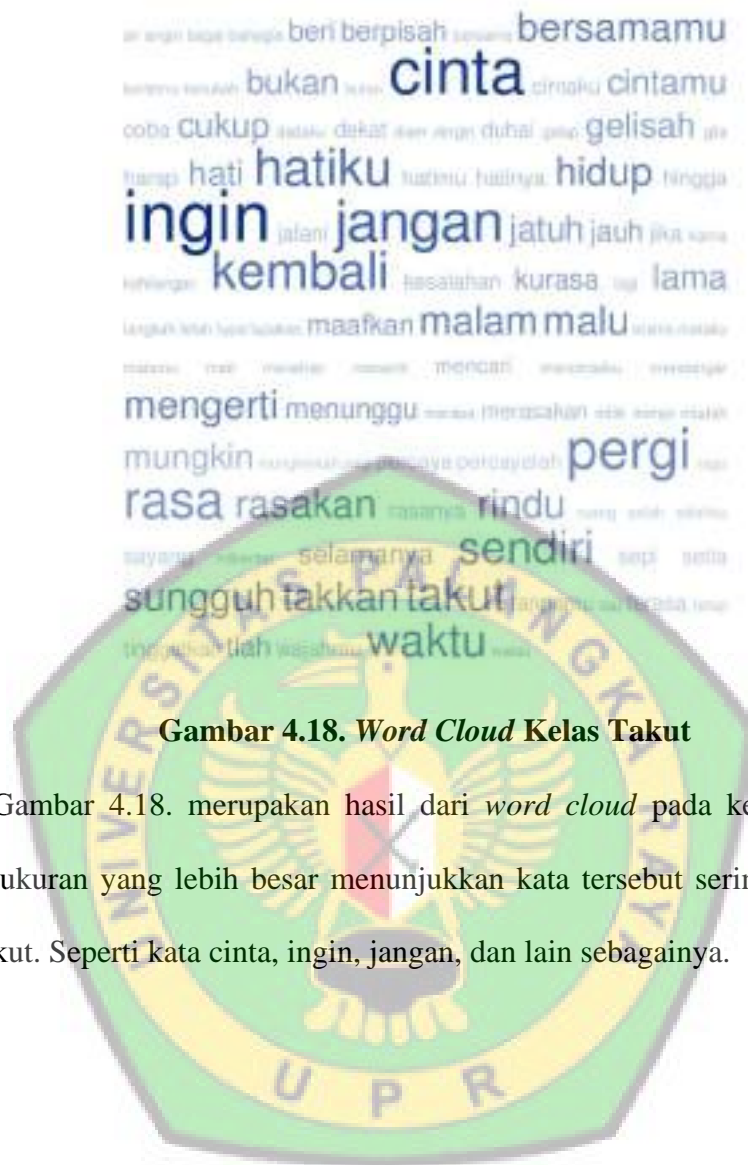
$$\begin{aligned}
 \text{Akurasi} &= \frac{4 + 5 + 4 + 4}{20} \times 100\% \\
 &= \frac{17}{20} \times 100\% = 85\%
 \end{aligned}$$

Dari hasil pengujian diatas dapat diketahui kemampuan sistem untuk membantu mengklasifikasikan data teks sebesar 85% dengan jumlah data uji yang digunakan 5 data uji senang, 5 data uji sedih, 5 data uji takut, dan 5 data uji marah.



Gambar 4.15. Grafik Hasil Akurasi Klasifikasi

Pada gambar 4.15 ditampilkan grafik dari hasil pengujian terhadap akurasi sistem. Dari pengujian yang dilakukan, dapat disimpulkan bahwa semakin banyak data latih yang digunakan maka semakin baik pula performansi sistem yang dihasilkan. Hal ini terjadi karena semakin banyak jumlah data latih yang digunakan maka semakin banyak pula jumlah kata yang disimpan untuk melakukan proses pembelajaran oleh sistem sehingga sistem memiliki *knowledge* yang banyak. Seperti pada penelitian yang telah dilakukan oleh Mahmudy, Wahyu Firdaus dan Agus Wahyu Widodo dengan topik Klasifikasi Artikel Berita Secara Otomatis Menggunakan Metode *Naïve Bayes Classifier* Yang Dimodifikasi juga menunjukkan hasil yang sama, yaitu dengan melakukan kombinasi dokumen latih dan dokumen uji dimana didapatkan hasil dengan kombinasi paling tinggi menghasilkan performa sistem yang lebih baik.



Gambar 4.18. Word Cloud Kelas Takut

Gambar 4.18. merupakan hasil dari *word cloud* pada kelas takut. Kata dengan ukuran yang lebih besar menunjukkan kata tersebut sering muncul pada kelas takut. Seperti kata cinta, ingin, jangan, dan lain sebagainya.

dalam suatu lirik lagu dicari berdasarkan kemungkinan kemunculan kata berdasarkan kata-kata penyusun dari setiap kalimatnya. Seperti yang terlihat dari *word cloud* yang ada, pada setiap kelas memiliki kemiripan atribut kata yang dihasilkan.

Berdasarkan hasil dari empat kali pengujian yang dilakukan, dapat diketahui kemampuan sistem untuk membantu mengklasifikasi data teks yang paling optimal, yaitu pada pengujian keempat dengan akurasi 85%, dimana komposisi data yang digunakan sebanyak 90% data latih dan 10% data uji.

4.1.3.2 *Black Box Testing*

Berikut ini adalah pengujian *black box* terhadap sistem yang telah dibuat.

1. *Black Box Testing* pada Halaman Admin

Pada Halaman Admin pengujian *Login* , Kelola Lirik Lagu , Kelola *Stopword*, Kelola Tanda Baca, Kelola *Dataset*, Pelatihan, dan Pengujian.

i. Pengujian Form *Login* Admin

Tabel 4.6. Pengujian *Form Login* Admin

No.	Kondisi Awal	Hasil Yang Diharapkan	Hasil Keluaran	Hasil Pengujian
1.	Melakukan <i>login</i> (memasukan	Gagal masuk ke halaman utama dan kembali ke halaman <i>login</i> .	Gagal masuk ke halaman utama dan kembali ke halaman <i>login</i> .	OK

Tabel 4.6. Pengujian *Form Login Admin* (Lanjutan)

No.	Kondisi Awal	Hasil Yang Diharapkan	Hasil Keluaran	Hasil Pengujian
	<i>username</i> dan <i>password</i> salah)			
2.	Melakukan <i>login</i> (memasukan <i>username</i> dan <i>password</i> benar)	Berhasil masuk ke halaman utama.	Berhasil masuk ke halaman utama.	OK

ii. Pengujian Halaman Kelola Lirik Lagu Baru

Tabel 4.7. Pengujian Halaman Kelola Lirik Lagu Baru

No.	Kondisi Awal	Hasil Yang Diharapkan	Hasil Keluaran	Hasil Pengujian
1.	Klik aksi hapus	Menampilkan notifikasi yakin hapus data?	Menampilkan notifikasi yakin hapus data?	OK
2.	Klik aksi hapus (pilih Hapus)	Menampilkan pesan data	Menampilkan pesan data	OK

Tabel 4.7. Pengujian Halaman Kelola Lirik Lagu Baru (Lanjutan)

No.	Kondisi Awal	Hasil Yang Diharapkan	Hasil Keluaran	Hasil Pengujian
		berhasil dihapus dan kembali ke halaman kelola lirik lagu.	berhasil dihapus dan kembali ke halaman kelola lirik lagu.	
3.	Klik aksi edit <i>link youtube.</i>	Menampilkan <i>form</i> untuk menambahkan <i>link youtbe.</i>	Menampilkan <i>form</i> untuk menambahkan <i>link youtbe.</i>	OK
4.	Klik simpan pada <i>form.</i>	Menampilkan <i>box</i> pesan data berhasil disimpan.	Menampilkan <i>box</i> pesan data berhasil disimpan.	OK
7.	Klik <i>button</i> Simpan.	Menampilkan <i>box</i> pesan data berhasil disimpan.	Menampilkan <i>box</i> pesan data berhasil disimpan.	OK

Tabel 4.7. Pengujian Halaman Kelola Lirik Lagu Baru(Lanjutan)

No.	Kondisi Awal	Hasil Yang Diharapkan	Hasil Keluaran	Hasil Pengujian
8.	Klik <i>button</i> Ok.	Menampilkan kelola halaman lirik lagu baru dan mp3 berhasil di perbarui.	Menampilkan kelola halaman lirik lagu baru dan mp3 berhasil di perbarui.	OK

iii. Pengujian Halaman Kelola *Stopword***Tabel 4.8. Pengujian Halaman Kelola *Stopword***

No.	Kondisi Awal	Hasil Yang Diharapkan	Hasil Keluaran	Hasil Pengujian
1.	Klik tambah kata	Menampilkan form tambah kata (masukan kata).	Menampilkan form tambah kata (masukan kata).	OK

Tabel 4.8. Pengujian Halaman Kelola *Stopword* (Lanjutan)

No.	Kondisi Awal	Hasil Yang Diharapkan	Hasil Keluaran	Hasil Pengujian
2.	Klik simpan pada form tambah lirik.	Menampilkan pesan data berhasil disimpan dan kembali ke halaman kelola <i>stopword</i> .	Menampilkan pesan data berhasil disimpan dan kembali ke halaman kelola <i>stopword</i> .	OK
3.	Klik aksi edit	Menampilkan form ubah data (kata yang akan diubah).	Menampilkan form ubah data (kata yang akan diubah).	OK
4.	Klik simpan pada form	Menampilkan pesan data berhasil diubah dan kembali ke halaman <i>stopword</i> .	Menampilkan pesan data berhasil diubah dan kembali ke halaman <i>stopword</i> .	OK
5.	Klik aksi hapus	Menampilkan notifikasi yakin hapus data?	Menampilkan notifikasi yakin hapus data?	OK

Tabel 4.8. Pengujian Halaman Kelola *Stopword* (Lanjutan)

No.	Kondisi Awal	Hasil Yang Diharapkan	Hasil Keluaran	Hasil Pengujian
6.	Klik aksi hapus (pilih Hapus)	Menampilkan pesan data berhasil dihapus dan kembali ke halaman <i>stopword</i> .	Menampilkan pesan data berhasil dihapus dan kembali ke halaman <i>stopword</i> .	OK

iv. Pengujian Halaman Kelola Tanda Baca

Tabel 4.9. Pengujian Halaman Kelola Tanda Baca

No.	Kondisi Awal	Hasil Yang Diharapkan	Hasil Keluaran	Hasil Pengujian
1.	Klik tambah tanda baca	Menampilkan form tambah tanda baca (masukan nama tanda baca dan simbol tanda baca).	Menampilkan form tambah tanda baca (masukan nama tanda baca dan simbol tanda baca).	OK

Tabel 4.9. Pengujian Halaman Kelola Tanda Baca (Lanjutan)

No.	Kondisi Awal	Hasil Yang Diharapkan	Hasil Keluaran	Hasil Pengujian
2.	Klik simpan pada form tambah lirik.	Menampilkan pesan data berhasil disimpan dan kembali ke halaman kelola tanda baca.	Menampilkan pesan data berhasil disimpan dan kembali ke halaman kelola tanda baca.	OK
3.	Klik aksi edit	Menampilkan form ubah data (nama tanda baca dan simbol tanda baca yang akan diubah).	Menampilkan form ubah data (nama tanda baca dan simbol tanda baca yang akan diubah).	OK
4.	Klik simpan pada form	Menampilkan pesan data berhasil diubah dan kembali ke halaman tanda baca.	Menampilkan pesan data berhasil diubah dan kembali ke halaman tanda baca.	OK

Tabel 4.9. Pengujian Halaman Kelola Tanda Baca (Lanjutan)

No.	Kondisi Awal	Hasil Yang Diharapkan	Hasil Keluaran	Hasil Pengujian
5.	Klik aksi hapus	Menampilkan notifikasi yakin hapus data?	Menampilkan notifikasi yakin hapus data?	OK
6.	Klik aksi hapus (pilih Hapus)	Menampilkan pesan data berhasil dihapus dan kembali ke halaman tanda baca.	Menampilkan pesan data berhasil dihapus dan kembali ke halaman tanda baca.	OK

v. Pengujian Halaman Kelola *Dataset***Tabel 4.10. Pengujian Halaman Kelola *Dataset***

No.	Kondisi Awal	Hasil Yang Diharapkan	Hasil Keluaran	Hasil Pengujian
1.	Klik aksi pilih kelas emosi.	Menampilkan form pilih kelas.	Menampilkan form pilih kelas.	OK
2.	Klik pilih kelas emosi.	Menampilkan daftar kelas emosi (senang, sedih, takut, dan	Menampilkan daftar kelas emosi (senang, sedih, takut, dan	OK

Tabel 4.10. Pengujian Halaman Kelola *Dataset* (Lanjutan)

No.	Kondisi Awal	Hasil Yang Diharapkan	Hasil Keluaran	Hasil Pengujian
		marah) dan memilih salah satu kelas.	marah) dan memilih salah satu kelas.	
3.	Klik simpan pada form.	Menampilkan pesan data berhasil diubah dan kembali ke <i>dataset</i> .	Menampilkan pesan data berhasil diubah dan kembali ke <i>dataset</i> .	OK

vi. Pengujian Halaman Pelatihan

Tabel 4.11. Pengujian Halaman Pelatihan

No.	Kondisi Awal	Hasil Yang Diharapkan	Hasil Keluaran	Hasil Pengujian
1.	Pilih halaman pelatihan.	Menampilkan proses pelatihan dari <i>dataset</i> yang status latihnya belum.	Menampilkan proses pelatihan dari <i>dataset</i> yang status latihnya belum.	OK

vii. Pengujian Halaman Hasil Pelatihan

Tabel 4.12. Pengujian Halaman Hasil Pelatihan

No.	Kondisi Awal	Hasil Yang Diharapkan	Hasil Keluaran	Hasil Pengujian
1.	Memilih halaman hasil <i>dataset</i> .	Menampilkan halaman berisikan prior dan jumlah dokumen setiap kelas dan menampilkan hasil pelatihan berupa kata yang telah dibobotkan.	Menampilkan halaman berisikan prior dan jumlah dokumen setiap kelas dan menampilkan hasil pelatihan berupa kata yang telah dibobotkan.	OK

viii. Pengujian Halaman Pengujian

Tabel 4.13. Pengujian Halaman Pengujian

No.	Kondisi Awal	Hasil Yang Diharapkan	Hasil Keluaran	Hasil Pengujian
1.	Pilih halaman pengujian.	Menampilkan halaman proses pengujian.	Menampilkan halaman proses pengujian.	OK

Tabel 4.13. Pengujian Halaman Pengujian (Lanjutan)

No.	Kondisi Awal	Hasil Yang Diharapkan	Hasil Keluaran	Hasil Pengujian
2.	Pilih <i>upload</i> data uji.	Menampilkan <i>form upload</i> data uji dan memilih <i>file</i> yang diupload, menampilkan <i>box</i> pesan data berhasil diinputkan.	Menampilkan <i>form upload</i> data uji dan memilih <i>file</i> yang diupload, menampilkan <i>box</i> pesan data berhasil diinputkan.	OK
3.	Klik Ok pada <i>box</i> pesan.	Menampilkan proses pengujian dan akan ditampilkan <i>box</i> pesan jika pengujian telah selesai dilakukan.	Menampilkan proses pengujian dan akan ditampilkan <i>box</i> pesan jika pengujian telah selesai dilakukan.	OK

Tabel 4.13. Pengujian Halaman Pengujian (Lanjutan)

No.	Kondisi Awal	Hasil Yang Diharapkan	Hasil Keluaran	Hasil Pengujian
4.	Klik Ok pada <i>Box</i> pesan.	Menampilkan kelas hasil pengujian.	Menampilkan kelas hasil pengujian.	OK

2. *Black Box Testing* Halaman Pendengar

i. Pengujian Halaman Daftar Lirik Lagu

Tabel 4.14. Pengujian Halaman Daftar Lirik Lagu

No.	Kondisi Awal	Hasil Yang Diharapkan	Hasil Keluaran	Hasil Pengujian
1.	Pendengar memilih halaman daftar lirik lagu.	Menampilkan halaman daftar lirik lagu.	Menampilkan halaman daftar lirik lagu.	OK
2.	Pendengar memilih salah satu kelas senang, sedih, takut dan marah.	Menampilkan daftar lirik lagu berdasarkan kelas senang, sedih, takut, dan marah.	Menampilkan daftar lirik lagu berdasarkan kelas senang, sedih, takut, dan marah.	OK

Tabel 4.14. Pengujian Halaman Daftar Lirik Lagu (Lanjutan)

No.	Kondisi Awal	Hasil Yang Diharapkan	Hasil Keluaran	Hasil Pengujian
3.	Pendengar memilih salah satu lagu.	Menampilkan halaman isi lirik lagu yang berisi musik, video, dan lirik.	Menampilkan halaman isi lirik lagu yang berisi musik, video, dan lirik.	OK
4.	Pendengar memutar musik dan video.	Memutar musik dan video.	Memutar musik dan video.	OK

ii. Halaman Tambah Lirik

Tabel 4.15. Pengujian Halaman Tambah Lirik

No.	Kondisi Awal	Hasil Yang Diharapkan	Hasil Keluaran	Hasil Pengujian
1.	Pendengar memilih menu tambah lirik lagu.	Menampilkan halaman tambah lirik lagu yang berisikan <i>form input</i> .	Menampilkan halaman tambah lirik lagu yang berisikan <i>form input</i> .	OK

Tabel 4.15. Pengujian Halaman Tambah Lirik (Lanjutan)

No.	Kondisi Awal	Hasil Yang Diharapkan	Hasil Keluaran	Hasil Pengujian
2.	Klik tambah lirik lagu.	Manampilkan hasil klasifikasi pada <i>box</i> pesan.	Manampilkan hasil klasifikasi pada <i>box</i> pesan.	OK

4.2 *Deployment (Delivery, Support, & Feedback)*

Tahapan *Deployment* merupakan tahapan implementasi *software* ke *customer*, pemeliharaan *software* secara berkala, perbaikan *software*, evaluasi *software*, dan pengembangan *software* berdasarkan umpan balik yang diberikan agar sistem dapat tetap berjalan dan berkembang sesuai dengan fungsinya.

Tahap ini adalah tahap akhir dalam pembuatan sistem. Setelah melakukan analisis, desain dan pengkodean maka sistem tersebut dapat digunakan oleh pengguna. Sistem yang dibuat akan di *onlinekan* atau di *hosting*. Pada penelitian ini, web akan di *hosting* di Niagahoster dan domain yang digunakan untuk website ini, yaitu *colourlyrics.com* dengan masa aktif penggunaan tiga bulan. Pada penelitian ini hanya dilakukan sampai tahap meng-*online* kan sistem agar dapat digunakan oleh pengguna, tidak sampai tahap pemeliharaan dan perbaikan sistem.

BAB V

PENUTUP

5.1 Kesimpulan

Berdasarkan penelitian yang telah dilakukan, dapat disimpulkan bahwa merancang dan membangun sistem pengklasifikasi emosi pada lirik lagu berbahasa Indonesia menggunakan metode pengembangan perangkat lunak *Waterfall Pressman*. Tahapan yang dilakukan, yaitu *communication, planning, modeling, construction, dan deployment*. Sumber data lirik lagu yang digunakan diambil dari situs *lirik.kapanlagi.com*. Sedangkan analisis sistem dibuat menggunakan *flowchart* untuk menjelaskan alur kerja suatu sistem dan pada bagian desain sistem menggunakan *Unified Modelling Language (UML)*. Bahasa pemrograman yang digunakan yaitu PHP dan database menggunakan MySQL. Pengujian yang dilakukan dalam pembuatan sistem ini menggunakan dua pengujian, yaitu *black box testing* untuk menguji fungsionalitas sistem dan akurasi untuk menguji performa klasifikasi.

Proses klasifikasi data lirik lagu hanya diberikan empat jenis kelas label emosi yaitu senang, sedih, takut, dan marah menggunakan algoritma *Naïve Bayes Classifier*. Tahap klasifikasi yang dilakukan dalam penelitian ini meliputi tahap *preprocessing* yang terdiri dari *case folding* dan *tokenizing, filtering* yang digunakan untuk menghilangkan kata yang dianggap kurang penting menggunakan *stopword, feature selection* dimana pada tahap ini akan dilakukan pembobotan

menggunakan *Term Frequency-Inverse Document Frequency (TF-IDF)*, terakhir yaitu klasifikasi menggunakan algoritma *Naïve Bayes Classifier*.

Pada skenario pengujian yang dilakukan, dapat disimpulkan bahwa semakin banyak data latih yang digunakan maka semakin baik pula performansi sistem yang dihasilkan. Hal ini terjadi karena semakin banyak jumlah data latih yang digunakan maka semakin banyak pula jumlah kata yang disimpan untuk melakukan proses pembelajaran oleh sistem sehingga sistem memiliki *knowledge* yang banyak. Selain itu, kesalahan klasifikasi yang terjadi pada sistem disebabkan oleh kata yang menjadi atribut atau fitur tidak merepresentasikan sebuah kelas emosi dengan cukup baik pada setiap kelas emosi, walaupun bobot atribut kata yang muncul telah diberikan nilai yang bersesuaian terhadap jumlah kemunculannya pada sebuah data disetiap kelasnya.

Hasil dari pengujian metode *black box* ini dari proses analisis hingga operasi diperoleh bahwa sistem ini dapat berfungsi sebagai sistem klasifikasi lirik lagu berbahasa Indonesia. Pengujian dengan metode akurasi mendapatkan hasil akurasi tertinggi 85% pada pengujian ke-4 dengan komposisi data 90% data latih dan 10% data uji.

5.2 Saran

Beberapa saran yang dirasa perlu untuk perbaikan dan peningkatan sistem yang diperlukan, yaitu :

1. Penambahan jumlah data agar dapat membantu memperbanyak kamus kata data dalam *data training*.

2. Pada penelitian selanjutnya agar dibuat pemrosesan teks menjadi sebuah kalimat agar sistem mendapatkan makna yang dicari dan mendapatkan hasil klasifikasi yang lebih baik.



DAFTAR PUSTAKA

- Abirawa, Irfan Dhiya, dkk. 2018. *Klasifikasi Emosi Pada Lirik Lagu Menggunakan Metode Support Vector Machine*. Jurnal e-Proceeding of Engineering. Vol.5 No.3. Bandung. Universitas Telkom.
- Akbar, Habibullah. 2017. *Warna-Warni Tipe Dan Jenis Data Buat Data Scientist*. Dikutip dari <https://mti.binus.ac.id/2017/12/05/warna-warni-tipe-dan-jenis-data-buat-data-scientist/>. Senin, 09 Maret 2020
- Aziz, Rifqi Abdul, dkk. 2016. *Klasifikasi Topik pada Lirik Lagu dengan Metode Multinomial Naïve Bayes*. Bandung. Universitas Telkom.
- C, Antonius Rahmat & Yuan Lukito. 2015. *Implementasi Sistem Crowdsourced Labelling Berbasis Web dengan Metode Weighted Majority Voting*. Jurnal ULTIMA. Yogyakarta. Universitas Kristen Duta Wacana.
- Ferdinan, Afif Hijra, dkk. 2013. *Klasifikasi Emosi Pada Lirik Lagu Menggunakan Metode K-Nearest Neighbor Method*. E-Proceeding of Engineering Vol 5 No.3. Bandung. Universitas Telkom.
- Hartono, Oswin Rahadiyan. 2016. *Indonesian Stoplist*. Dikutip dari <https://www.kaggle.com/oswinrh/indonesian-stoplist>. Selasa, 29 Juni 2020.
- Hudaya, Nova Farid. 2015. *Peningkatan Kemampuan Mengelola Emosi Marah Melalui Teknik Angermanagement Pada Siswa Kelas X Teknik Komputer dan Jaringan SMK Muhammadiyah 1 Moyudan*. Yogyakarta. Universitas Negeri Yogyakarta.
- Kristiyanti, Tri Junia. 2012. *Analisis Gaya Bahasa dan Pesan Moral pada Lirik Lagu Grup Band Nidji dalam Album Breakthru' dan Let's Play*. Purwokerto. Universitas Muhammadiyah Purwokerto.
- Kuncoro, Anon. 2015. *Data Warehouse : Extract, Transform, dan Load*. Dikutip dari <https://www.slideshare.net/AnonKuncoro1/data-warehouse-extract-transform-dan-load>. Senin, 09 Maret 2020
- Kurniawan, Bambang, dkk. 2012. *Klasifikasi Konten Berita Dengan Metode Text Mining*. Jurnal Dunia Teknologi Informasi Vol 1, No. 1. Medan. Universitas Sumatera Utara.
- Kusuma, Maulana Candra. 2017. *Klasifikasi Lirik Lagu Bahasa Indonesia Berdasarkan Emosi Menggunakan Metode Naive Bayes Classifier*. Skripsi. Semarang. Universitas Dian Nuswantoro
- Muhammad, Wayan Firdaus & Agus Wahyu Widodo. 2015. *Klasifikasi Artikerl Berita Secara Otomatis Menggunakan Metode Naïve Bayes Classifier Yang Dimodifikasi*. Jurnal Tekno. Malang. Universitas Brawijaya.

- Noviardini, Maftukhah Carolin, dkk. 2018. *Klasifikasi Emosi Pada Lirik Lagu Menggunakan Metode Naïve Bayes Classifier* . Jurnal e-Proceeding of Engineering. Vol.5 No.3. Bandung. Universitas Telkom
- Nurjannah, Musfiroh, dkk. 2013. *Penerapan Algoritma Term Frequency-Inverse Documen Frequency (TF-IDF) Untuk Text Mining*. Jurnal Informatika Mulawarman Vol.8 No.3. Samarinda. Universitas Mulawarman.
- Octaviana, Elsa Rika. 2015. *Analisis Sentimen Pada Lirik Lagu Menggunakan Metode Multinomial Naïve Bayes*. Skripsi. Yogyakarta. Universitas Sanata Dharma.
- Prasetyo, Eko. 2014. *Data Mining Mengolah Data Menjadi Informasi Menggunakan Matlab*. Yogyakarta: Andi.
- Ryansyah, Adi & Sri Andayani. 2017. *Implementasi Algoritma TF-IDF Pada Pengukuran Kesamaan Dokumen*. Jurnal Sistem I& Teknologi Informasi Komunikasi Vol.1 No.1. Palembang. Universitas Katolik Musi Charitas.
- S. Rosa A dan M. Salahudin. 2016. *Rekayasa Perangkat Lunak Terstruktur dan Berorientasi Objek*. Bandung: Informatika Bandung.
- Setiawan, Permadi, dkk. 2015. *Rancang Bangun Aplikasi Pengolahan Data Evaluasi Proses Belajar Mengajar Berbasis Web Pada STIKES Yayasan RS. Dr. Soetomo Surabaya*. Jurnal JSIKA. Vol.4 No.2. Surabaya. STMIK STIKOM Surabaya.
- Setiawan, Samhis.2019. "Emosi" Pengertian Menurut Para Ahli & Bentuk (Positif-Negatif). Dikutip dari <https://www.gurupendidikan.co.id/pengertian-emosi/>. Senin, 27 Februari 2020.
- Sulhaerati, dkk. 2017. *Analisis Tanggapan Pasar Terhadap Perusahaan Ritel Raksasa Indomart dan Alfamart*. Laporan Penelitian. Yogyakarta. Universitas Islam Indonesia.
- Suyanto.2017. *Data Mining untuk Klasifikasi dan Klasterisasi Data*. Bandung. Informatika Bandung.