

**KLASIFIKASI TEMA PADA LIRIK LAGU DENGAN
METODE *TRANSFORMED WEIGHT-NORMALIZED
COMPLEMENT NAIVE BAYES* (TWCNB)**

SKRIPSI

Laboratorium Komputasi Cerdas Dan Visualisasi

Diajukan untuk memenuhi sebagian persyaratan memperoleh gelar Sarjana Komputer



Disusun Oleh:

ASTRI TIKA PRATIWI

NIM. 0910683023

**KEMENTERIAN PENDIDIKAN DAN KEBUDAYAAN
UNIVERSITAS BRAWIJAYA
PROGRAM TEKNOLOGI INFORMASI DAN ILMU KOMPUTER**

MALANG

2014

LEMBAR PERSETUJUAN

**KLASIFIKASI TEMA PADA LIRIK LAGU DENGAN METODE
TRANSFORMED WEIGHT-NORMALIZED COMPLEMENT NAIVE BAYES
(TWCNB)**

SKRIPSI

Laboratorium Komputasi Cerdas Dan Visualisasi

Diajukan untuk memenuhi sebagian persyaratan memperoleh gelar Sarjana Komputer



Disusun Oleh:

ASTRI TIKA PRATIWI

NIM. 0910683023

Menyetujui :

Pembimbing I

Pembimbing II

Drs.Achmad Ridok, M.Kom

NIP. 196808251994031002

Indriati, ST., M.Kom.

NIK. 831013 06 1 2 0035

LEMBAR PENGESAHAN

**KLASIFIKASI TEMA PADA LIRIK LAGU DENGAN METODE
TRANSFORMED WEIGHT-NORMALIZED COMPLEMENT NAIVE BAYES
(TWCNB)**

SKRIPSI

Laboratorium Komputasi Cerdas Dan Visualisasi

Diajukan untuk memenuhi sebagian persyaratan memperoleh gelar Sarjana Komputer

Disusun oleh:

ASTRI TIKA PRATIWI

NIM. 0910683023

Skripsi ini telah diuji dan dinyatakan lulus pada tanggal 15 Januari 2014

Penguji I

Penguji II

Lailil Muflikhah, S.Kom., M.Sc
NIP. 19741113 200501 2 001

Novanto Yudistira, S.Kom., M.Sc
NIP. 831110 16 1 1 0425

Penguji III

Budi Darma Setiawan, S.Kom, M.Sc
NIP. 841015 06 1 1 0090

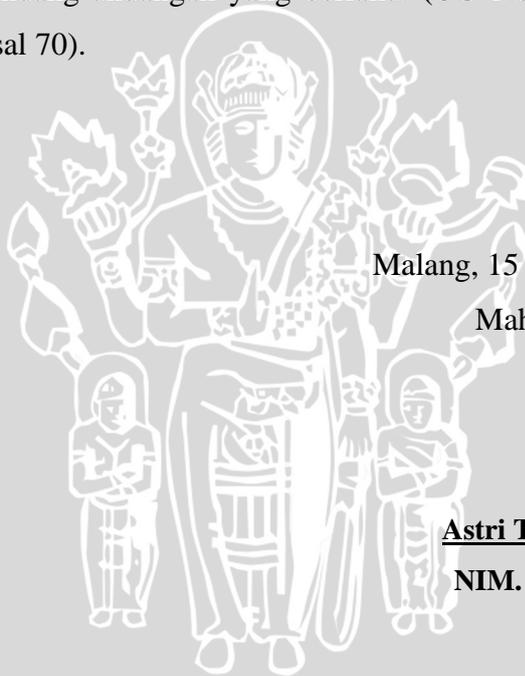
Mengetahui,
Ketua Program Studi Informatika / Ilmu Komputer

Drs. Marji, M.T.
NIP.19670801 199203 1 001

PERNYATAAN ORISINALITAS SKRIPSI

Saya menyatakan dengan sebenar-benarnya bahwa sepanjang pengetahuan saya, di dalam naskah SKRIPSI ini tidak terdapat karya ilmiah yang pernah diajukan oleh orang lain untuk memperoleh gelar akademik di suatu perguruan tinggi, dan tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis dikutip dalam naskah ini dan disebutkan dalam sumber kutipan dan daftar pustaka.

Apabila ternyata didalam naskah SKRIPSI ini dapat dibuktikan terdapat unsur-unsur PLAGIASI, saya bersedia SKRIPSI ini digugurkan dan gelar akademik yang telah saya peroleh (SARJANA) dibatalkan, serta diproses sesuai dengan peraturan perundang-undangan yang berlaku. (UU No. 20 Tahun 2003, Pasal 25 ayat 2 dan Pasal 70).



Malang, 15 Januari 2014

Mahasiswa,

Astri Tika Pratiwi

NIM. 0910683023

KATA PENGANTAR

Alhamdulillah rabbil 'alamin. Puji syukur penulis panjatkan kehadirat Allah SWT, karena atas segala rahmat dan limpahan hidayah-Nya, penulis dapat menyelesaikan skripsi yang berjudul “Klasifikasi Tema Pada Lirik Lagu Dengan Metode *Transformed Weight-Normalized Complement Naive Bayes* (TWCNB)”.

Dalam pelaksanaan dan penulisan skripsi ini penulis mendapatkan banyak bantuan dari berbagai pihak baik secara moral maupun material. Dalam kesempatan ini penulis ingin mengucapkan terima kasih yang sebesar – besarnya kepada :

1. Bapak Drs. Achmad Ridok, M.Kom dan Ibu Indriati, S.T., M.Kom selaku dosen pembimbing selama pelaksanaan skripsi.
2. Bapak Ir. Sutrisno, M.T, Bapak Ir. Heru Nurwasito, M.Kom, Bapak Himawat Aryadita, S.T, M.Sc, dan Bapak Eddy Santoso, S.Kom selaku Ketua, Wakil Ketua 1, Wakil Ketua 2 dan Wakil Ketua 3 Program Teknologi Informasi dan Ilmu Komputer Universitas Brawijaya.
3. Bapak Drs. Marji, M.T dan Bapak Issa Arwani, S.Kom, M.Sc selaku Ketua dan Sekretaris Program Studi Teknik Informatika Universitas Brawijaya.
4. Ayahanda, Ibunda, dan seluruh keluarga atas segenap dukungan, doa, dan kasih sayang yang telah diberikan.
5. Seluruh Dosen Teknik Informatika Universitas Brawijaya atas kesediaan membagi ilmunya kepada penulis.
6. Seluruh Civitas Akademika Teknik Informatika Universitas Brawijaya yang telah banyak memberi bantuan dan dukungan selama penulis menempuh studi di Teknik Informatika Universitas Brawijaya dan selama penyelesaian skripsi ini.
7. Teman-teman angkatan 2009 khususnya konsentrasi KCV, PAS's family+Pricil, UKM Skripsi, terima kasih atas kritik, saran, ilmu, waktu, dan kebersamaannya selama perkuliahan dan pengerjaan skripsi ini.

Penulis menyadari bahwa dalam penyusunan skripsi ini masih banyak kekurangan baik format penulisan maupun isinya. Oleh karena itu, saran dan

kritik membangun senantiasa diharapkan guna perbaikan bagi penelitian selanjutnya. Semoga skripsi ini dapat memberikan manfaat bagi semua pihak, Aamiin.

Malang, 15 Januari 2014

Penulis



ABSTRAK

Astri Tika Pratiwi. 2014. : **Klasifikasi Tema Pada Lirik Lagu Dengan Metode *Transformed Weight-Normalized Complement Naive Bayes* (TWCNB).** Skripsi Program Studi Informatika, Program Teknologi Informasi dan Ilmu Komputer, Universitas Brawijaya. Pembimbing: Drs.Achmad Ridok, M.Kom dan Indriati, ST., M.Kom.

Tema merupakan hal penting dalam sebuah lagu, karena dengan tema tersebut sebuah lirik dapat dirangkai sesuai dengan emosi yang tepat. Tema lagu yang tidak tepat dapat menimbulkan kesalahan. Untuk itu diperlukan pengklasifikasian lagu berdasarkan tema. Pada penelitian ini digunakan metode *Transformed Weight-Normalized Complement Naive Bayes* (TWCNB). Metode ini merupakan pengembangan dari Multinomial Naive Bayes (MNB). Evaluasi sistem yang dilakukan terhadap 29 data uji dan 224 data latih dengan perubahan jumlah data latih yang berbeda-beda pada 4 skenario menghasilkan nilai terbaik pada *precision* senilai 0,89, *recall* senilai 0,856 dan nilai *F-measure* senilai 0,857.

Kata Kunci : tema, lagu, teks, klasifikasi, *Transformed Weight-Normalized Complement Naive Bayes*

ABSTRACT

Astri Tika Pratiwi. 2014. : Klasifikasi Tema Pada Lirik Lagu Dengan Metode *Transformed Weight-Normalized Complement Naive Bayes* (TWCNB). Skripsi Program Studi Informatika, Program Teknologi Informasi dan Ilmu Komputer, Universitas Brawijaya. Pembimbing: Drs.Achmad Ridok, M.Kom dan Indriati, ST., M.Kom.

*Themes are important in a song, because they can be used to create a song lyric that based on the right emotion. Unsuitable themes song could make any error. Therefore the songs need to be classified by it's theme. This research is used *Transformed Weight-Normalized Complement Naive Bayes* (TWCNB) method. This method is a development of the *Multinomial Naive Bayes* (MNB). The evaluation that have performed by using 29 testing data and 224 training data with changes in the amount of training data that varies in 4 scenarios, produces precision of 0.89, recall of 0.856 and *F-measure* of 0.857 as the best value.*

Keywords: *theme, song, text, classification, *Transformed Weight-Normalized Complement Naive Bayes**

DAFTAR ISI

LEMBAR PERSETUJUAN.....	ii
LEMBAR PENGESAHAN	iii
PERNYATAAN ORISINALITAS SKRIPSI	iv
KATA PENGANTAR	v
ABSTRAK	vii
ABSTRACT.....	viii
DAFTAR ISI.....	ix
DAFTAR GAMBAR	xii
DAFTAR TABEL.....	xiii
DAFTAR SOURCECODE.....	xiv
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Batasan Masalah.....	3
1.4 Tujuan.....	3
1.5 Manfaat.....	3
1.6 Sistematika Penulisan.....	4
BAB II KAJIAN PUSTAKA DAN DASAR TEORI.....	6
2.1 Kajian Pustaka	6
2.1 Tema.....	7
2.3 Text Mining.....	7
2.4 Algoritma Stemming Arifin Setiono	8
2.5 Naive Bayes.....	11
2.5.1 Multinomial Naive Bayes.....	12
2.5.2 Transformed Weight-Normalized Complement Naive Bayes (TWCNB).....	13
2.6 <i>Precision, Recall, F-measure</i>	14
BAB III METODOLOGI DAN PERANCANGAN	17
3.1 Studi Literatur.....	18
3.2 Penyusunan Dasar Teori.....	18



3.3	Metode Pengumpulan Data	18
3.4	Analisis dan Perancangan.....	19
3.4.1	Kebutuhan Antar Muka	19
3.4.2	Kebutuhan Data.....	19
3.4.3	Kebutuhan Fungsional.....	19
3.5	Arsitektur Sistem.....	20
3.6	Diagram Alir Sistem.....	20
3.6.1	Desain Antar Muka.....	25
3.7	Perancangan Sistem Basis Data	26
3.8	Manualisasi Analisis Sentimen	28
3.9	Implementasi	40
3.10	Pengujian	40
3.11	Penulisan Laporan	41
BAB IV IMPLEMENTASI		42
4.1	Analisis dan Perancangan.....	42
4.1.1	Spesifikasi Perangkat Keras	42
4.1.2	Spesifikasi Perangkat Lunak	42
4.2	Batasan-Batasan Implementasi.....	43
4.3	Implementasi Algoritma.....	43
4.3.1	Proses Preprocessing	44
4.3.1.1	Proses Tokenizing	45
4.3.1.2	Proses Stopword Removal.....	45
4.3.1.3	Proses Stemming	46
4.3.2	Implementasi Metode TWCNB	46
4.4	Implementasi Antarmuka	49
4.4.1	Tampilan Halaman Data Training.....	49
4.4.2	Tampilan Halaman Klasifikasi.....	49
BAB V PENGUJIAN DAN ANALISIS.....		51
5.1	<i>Precision, Recall dan F-measure</i>	51
5.2	Analisis Hasil Pengujian	54
BAB VI PENUTUP		56
6.1	Kesimpulan.....	56

6.2 Saran..... 56
DAFTAR PUSTAKA 57



DAFTAR GAMBAR

Gambar 2.1 Stemming.....	8
Gambar 3.1 Desain Penelitian.....	17
Gambar 3.2 Arsitektur Perancangan Sistem.....	20
Gambar 3.3 Diagram Alir Sistem.....	21
Gambar 3.4 Diagram Alir Preprocessing.....	21
Gambar 3.5 Diagram Alir Filtering Stopword Removal.....	22
Gambar 3.6 Diagram Alir Data Latih TWCNB.....	23
Gambar 3.7 Diagram Alir Data Uji TWCNB.....	24
Gambar 3.8 Perancangan Desain Antar Muka Sistem.....	25
Gambar 3.9 Desain Antar Muka Halaman Hasil.....	26
Gambar 3.10 Tabel Database klasifikasi tema pada lirik lagu.....	27
Gambar 4.1 Halaman Data Training.....	49
Gambar 4.2 Halaman Klasifikasi.....	50
Gambar 5.1 Grafik Rata-Rata Hasil Pengujian.....	54

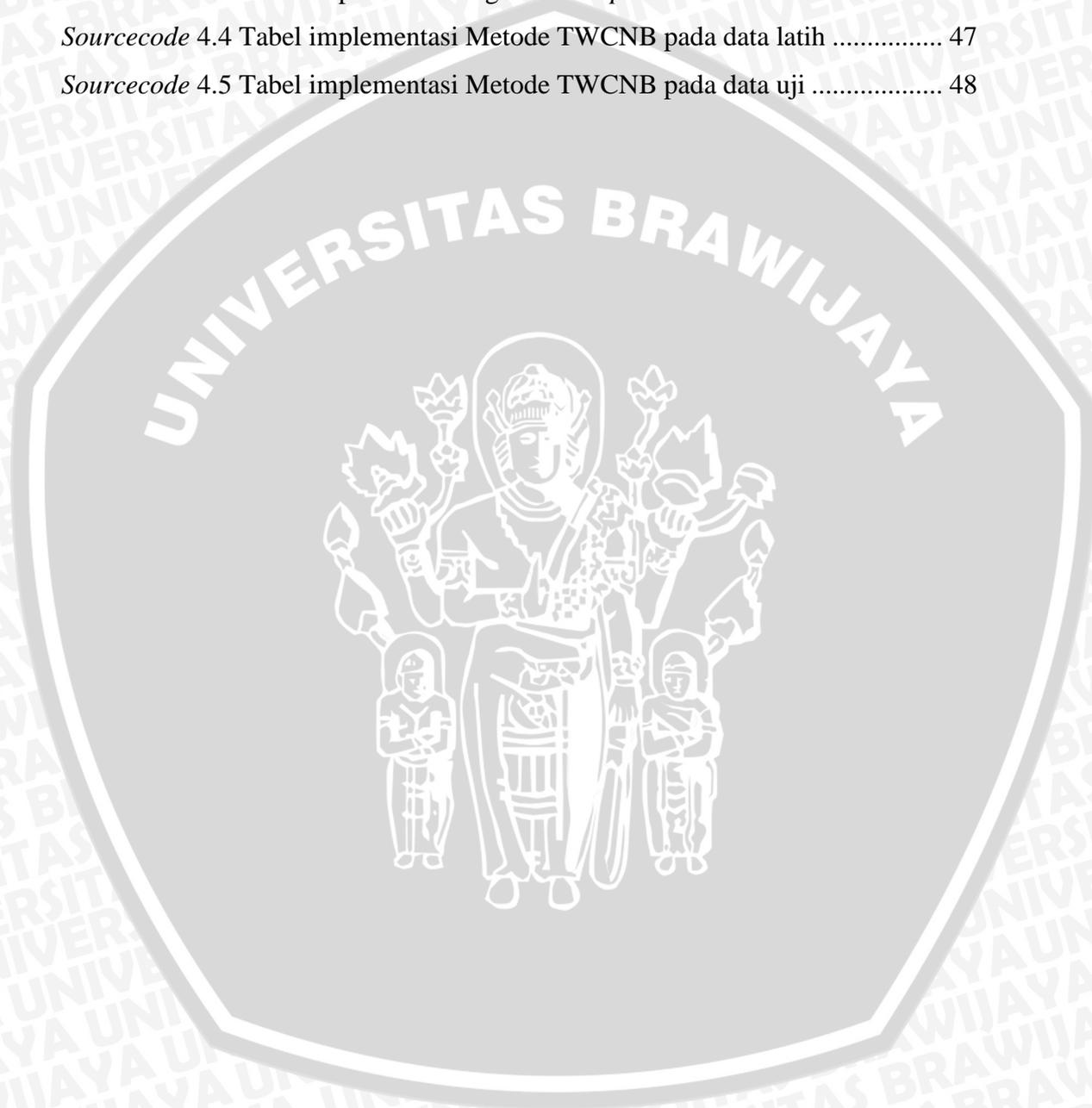
DAFTAR TABEL

Tabel 2.1 Tabel Kontigensi	15
Tabel 3.1 Contoh Masukan Lirik.....	28
Tabel 3.2 Contoh lirik hasil <i>tokenizing</i> dan <i>filtering</i>	30
Tabel 3.3 Contoh lirik hasil <i>stopword</i>	31
Tabel 3.4 Contoh lirik hasil <i>stemming</i>	32
Tabel 3.5 Tabel daftar kata hasil preprocessing	33
Tabel 3.6 Tabel daftar kata pada dokumen uji	35
Tabel 3.7 Tabel perhitungan bobot kata	38
Tabel 3.8 Contoh tabel pengujian.....	41
Tabel 4.1 Spesifikasi perangkat keras komputer	42
Tabel 4.2 Spesifikasi perangkat lunak.....	43
Tabel 4.3 Daftar Fungsi dan Variabel Pada Sistem.....	44
Tabel 5.1 Skenario Pengujian.....	51
Tabel 5.2 Hasil Pengujian oleh Sistem.....	52
Tabel 5.3 <i>Precision, Recall</i> dan <i>F-measure</i> pada Skenario 1	53
Tabel 5.4 <i>Precision, Recall</i> dan <i>F-measure</i> pada Skenario 2	53
Tabel 5.5 <i>Precision, Recall</i> dan <i>F-measure</i> pada Skenario 3	53
Tabel 5.6 <i>Precision, Recall</i> dan <i>F-measure</i> pada Skenario 4	53
Tabel 5.7 Rata - rata <i>Precision, Recall</i> dan <i>F-measure</i>	53



DAFTAR SOURCECODE

<i>Sourcecode 4.1</i> Tabel implementasi algoritma <i>preprocessing</i> pada dokumen.....	45
<i>Sourcecode 4.2</i> Tabel implementasi algoritma <i>tokenizing</i>	45
<i>Sourcecode 4.3</i> Tabel implementasi algoritma <i>stopword removal</i>	46
<i>Sourcecode 4.4</i> Tabel implementasi Metode TWCNB pada data latih.....	47
<i>Sourcecode 4.5</i> Tabel implementasi Metode TWCNB pada data uji.....	48



BAB I PENDAHULUAN

1.1 Latar Belakang

Musik adalah seni yang dikaruniai nilai kekuatan untuk menembus ke kedalaman jiwa. Seperti yang diungkapkan oleh seorang I Ching bahwa Musik memiliki kekuatan untuk mengikis kebalan nurani dari cengkraman emosi yang liar. Disadari atau tidak, ekspresi suara hati ada dalam lagu, tarian dan irama gerakan tubuh. [DJO-09].

Musik pada sebuah lagu menjelaskan makna atau tema yang dijabarkan melalui nada – nada. Banyak lagu yang digunakan sebagai pengiring sebuah kejadian atau momen. Namun penggunaannya sering tidak tepat karena hanya musiknya saja yang diperhatikan sedangkan tema nya terabaikan. Sehingga dapat menimbulkan kesalahpahaman dalam pengiringan momen tersebut.

Tema merupakan hal terpenting dalam sebuah lagu, karena dengan tema tersebut sebuah lirik dapat dirangkai sesuai dengan emosi yang tepat. Syair lagu yang baik bergantung pada makna mendalam yang diangkat dari tema lagu. [HAS-08] Berbagai – macam tema lagu telah banyak di masyarakat yaitu seperti percintaan, persahabatan, religi atau rohani, perjuangan, dan jenis tema lainnya. Tema – tema tersebut akan mudah diketahui apabila diklasifikasikan dengan baik. Salah satu metode dari *text mining* yang bisa digunakan untuk menyelesaikan masalah klasifikasi adalah *Naïve Bayes Classifier (NBC)*.

Naïve Bayes adalah algoritma pembelajaran yang sering digunakan untuk mengatasi masalah klasifikasi teks. Dari proses pengujian secara kualitatif disebutkan bahwa teks bisa diklasifikasikan dengan akurasi yang tinggi. [KFP-04]. Ada dua model Naïve Bayes yang umum digunakan, yaitu multivariate Bernoulli dan multinomial. Model multinomial sering disebut sebagai Multinomial Naive Bayes (MNB), yang melebihi performa pada multivariat dan juga teruji lebih menguntungkan dibandingkan dengan model yang lebih khusus [ELM-03]

Metode multinomial biasanya memiliki performa yang melebihi Bernoulli multivariat pada ukuran kosakata besar, dan hampir selalu mengalahkan Bernoulli multivariat bila ukuran kosakata yang dipilih secara optimal untuk keduanya.

Baru-baru ini, sebuah algoritma baru telah diusulkan, yang disebut "*Transformed Weight-Normalized Complement Naive Bayes*" (TWCNB), yang mudah untuk diterapkan dan memiliki waktu running yang baik. TWCNB adalah versi modifikasi dari MNB sendiri. Sebelumnya telah dilakukan eksperimen perbandingan antara Multinomial Naive Bayes, TWCNB dan Support Vector Machine (SVM) dari beberapa data set (Jason D. M. Rennie, Lawrence Shih, Jaime Teevan, dan David R. Karger, 2003). Performansi TWCNB jauh lebih baik daripada MNB dan hampir mendekati performansi SVM. [RST-03]

Saat ini penulis tertarik untuk melakukan penelitian dengan metode TWCNB pada pengklasifikasian tema pada lirik lagu yang sudah pernah dilakukan oleh penulis sebelumnya (Aprilliana Dharma Susanti, 2012) namun dengan metode yang berbeda. Penulis sebelumnya menggunakan metode *Improved KNN* dan mengambil data berupa lirik lagu berbahasa inggris, dengan menggunakan stemming Porter [SDA-12]. Sedangkan pada kasus kali ini penulis menggunakan data berupa lirik lagu berbahasa indonesia dan menggunakan algoritma stemming Arifin Setiono. Tema pada lirik lagu yaitu cinta, perjuangan, dan religi atau rohani. Berdasarkan latar belakang tersebut, penulis akan melanjutkan penelitian mengenai Klasifikasi Tema Pada Lirik Lagu Dengan Metode *Transformed Weight-Normalized Complement Naive Bayes* (TWCNB). Implementasi sistem menggunakan bahasa pemrograman PHP. Metode ini diharapkan dapat membantu pengguna dalam memilih jenis lagu berdasarkan tema dan memberi hasil yang baik dalam pengklasifikasiannya.

1.2 Rumusan Masalah

Berdasarkan pada permasalahan yang telah dijelaskan pada bagian latar belakang, maka rumusan masalah dapat disusun sebagai berikut:

- a. Bagaimana perancangan dari Klasifikasi Tema Pada Lirik Lagu Dengan Metode *Transformed Weight-Normalized Complement Naive Bayes* (TWCNB)
- b. Bagaimana implementasi proses *preprocessing* dan klasifikasi teks yang akan dibangun

- c. Berapa besar tingkat tingkat akurasi pada sistem perangkat lunak yang akan dibangun dengan menggunakan *precision*, *recall*, dan *F-measure*

1.3 Batasan Masalah

Agar permasalahan yang dirumuskan dapat lebih terfokus, maka pada penelitian ini dibatasi dalam hal:

1. Obyek yang diteliti pada perancangan sistem ini adalah lirik lagu yang berbahasa Indonesia
2. Penelitian ini hanya membahas dengan metode *Transformed Weight-Normalized Complement Naive Bayes* (TWCNB) sebagai algoritma untuk pengklasifikasian, sedangkan beberapa tema yang ditentukan yaitu cinta, perjuangan, dan religi atau rohani.
3. Aplikasi dibangun dengan menggunakan bahasa pemrograman PHP dan MySQL sebagai Database Management System (DBMS) nya.

1.4 Tujuan

Tujuan yang ingin dicapai dari pembuatan skripsi ini adalah menerapkan metode *Transformed Weight-Normalized Complement Naive Bayes* (TWCNB) dalam klasifikasi tema pada lirik lagu.

1.5 Manfaat

Penulisan skripsi ini diharapkan mempunyai manfaat yang baik dan berguna bagi pembaca dan penulis. Adapun manfaat yang diharapkan adalah sebagai berikut:

1. Bagi Penulis
 - a. Menerapkan ilmu yang telah diperoleh dari Program Teknologi Informasi dan Ilmu Komputer Universitas Brawijaya.
 - b. Mendapatkan pemahaman tentang perancangan dan pengembangan Klasifikasi Tema pada Lirik Lagu dengan Metode *Transformed Weight-Normalized Complement Naive Bayes* (TWCNB).
2. Bagi Pengguna

Manfaat penelitian ini adalah memudahkan pengguna untuk dapat menentukan lagu tersebut tergolong dalam lagu dengan tema tertentu sebelum mendengarkannya.

1.6 Sistematika Penulisan

Sistematika penulisan dalam skripsi ini sebagai berikut:

BAB I Pendahuluan

Memuat latar belakang, rumusan masalah, tujuan, batasan masalah, manfaat, metodologi pembahasan, dan sistematika penulisan.

BAB II Kajian Pustaka dan Dasar teori

Menguraikan tentang dasar teori dan referensi yang mendasari pembuatan Klasifikasi Tema pada Lirik Lagu dengan Metode *Transformed Weight-Normalized Complement Naive Bayes* (TWCNB).

BAB III Metode Penelitian

Membahas metode yang digunakan dalam penelitian yang terdiri dari studi literatur, perancangan sistem perangkat lunak, implementasi sistem perangkat lunak, pengujian dan analisis, serta penulisan laporan.

BAB IV Perancangan

Membahas analisis kebutuhan dan perancangan Klasifikasi Tema pada Lirik Lagu dengan Metode *Transformed Weight-Normalized Complement Naive Bayes* (TWCNB).

BAB V Implementasi

Membahas implementasi dari Klasifikasi Tema pada Lirik Lagu dengan Metode *Transformed Weight-Normalized Complement Naive Bayes* (TWCNB) yang sesuai dengan perancangan sistem yang telah dibuat.

BAB VI Pengujian dan Analisis

Memuat proses dan hasil pengujian terhadap sistem yang telah direalisasikan.

BAB VII Penutup

Memuat kesimpulan yang diperoleh dari pembuatan dan pengujian perangkat lunak yang dikembangkan dalam skripsi ini serta saran – saran untuk pengembangan lebih lanjut.



BAB II

KAJIAN PUSTAKA DAN DASAR TEORI

Pada bab ini dibahas mengenai dasar teori yang digunakan untuk menunjang penulisan penelitian mengenai implementasi metode *Transformed Weight-Normalized Complement Naive Bayes* (TWCNB) pada pengelompokan Tema Lirik Lagu. Dasar teori meliputi tema, *text mining*, tahapan *text mining*, dan metode *Transformed Weight-Normalized Complement Multinomial Naive Bayes* (TWCNB).

2.1 Kajian Pustaka

Kajian pustaka pada penelitian ini adalah membahas penelitian sebelumnya yang berjudul ‘Multinomial Naive Bayes for Text Categorization Revisited’. Penelitian ini mengkaji kinerja *Multinomial Naive Bayes* (MNB) dan beberapa cabangnya seperti *Transformed Weight-Normalized Complement Naive Bayes* (TWCNB) untuk kategorisasi teks berita. Penelitian menggunakan data dari <http://people.csail.mit.edu/people/jrennie/20Newsgroups> yang berjumlah lebih dari 18.000 dokumen berita dengan berbagai kategori. Untuk evaluasi kinerja algoritma dalam melakukan kategorisasi ditetapkan berdasarkan nilai presisi *recall* dengan menggunakan *toolkit* Bow. [KFP-04]

Perbedaan yang dibuat penulis pada penelitian ini adalah pada penggunaan implementasi klasifikasi tema pada lirik lagu berbahasa Indonesia. Proses stemming pada penelitian yang akan dibuat ini akan menggunakan stemming Arifin Setiono yang dikenal cukup efektif untuk proses stemming pada teks berbahasa Indonesia. Penelitian ini akan menggunakan metode *Transformed Weight-Normalized Complement Naive Bayes* (TWCNB) sebagai metode klasifikasi dengan *classifier sentimen* yang mampu menentukan sentimen positif dan negatif. TWCNB digunakan sebagai *classifier* untuk jenis tema pada lirik lagu yaitu berupa cinta, perjuangan, dan religi atau rohani. Pengujian pada penelitian ini menggunakan *precision*, *recall*, dan *F-measure*.

2.2 Tema

Tema berasal dari bahasa Yunani “*thithenai*”, berarti sesuatu yang telah diuraikan atau sesuatu yang telah ditempatkan. Tema merupakan amanat utama yang disampaikan oleh penulis melalui karangannya. Dalam karang mengarang, tema adalah pokok pikiran yang mendasari karangan yang akan disusun. Dalam tulis menulis, tema adalah pokok bahasan yang akan disusun menjadi tulisan. Tema ini yang akan menentukan arah tulisan atau tujuan dari penulisan artikel itu. Menentukan tema berarti menentukan apa masalah sebenarnya yang akan ditulis atau diuraikan oleh penulis.

Tema merupakan persoalan utama yang diungkapkan oleh pengarang dalam sebuah karya kesusteraan seperti cerpen atau novel. Biasanya tema diolah berdasarkan sesuatu motif tertentu yang terdiri dari pada objek, peristiwa kejadian dan sebagainya. Ada pendapat lain yang mengatakan bahwa tema sebagai satu gagasan, fikiran atau persoalan utama yang mendasari sesebuah karya sastra dan terungkap secara langsung (eksplisit) atau tidak langsung (implisit). Tema dalam sebuah cerita tidak dapat dilihat sepenuhnya sehingga cerita itu selesai dibaca. [ITA-10]

2.3 Text Mining

Text mining merupakan suatu proses yang melibatkan beberapa area teknologi. Namun secara umum proses-proses pada *text mining* mengadopsi proses data mining. Bahkan beberapa teknik dalam proses *text mining* juga menggunakan teknik-teknik data mining [EZY-02]. Dalam menganalisa sebahagian atau keseluruhan *unstructured text*, *text mining* mencoba untuk mengasosiasikan satu bagian teks dengan yang lainnya berdasarkan aturan-aturan tertentu [KAA-09]. Selain itu *text mining* juga diartikan sebagai kegiatan menambang data dari data yang berupa teks atau dokumen, dengan tujuan mencari kata-kata yang dapat mewakili apa yang ada dalam dokumen sehingga dapat dilakukan analisa keterhubungan antar dokumen.

Langkah-langkah yang dilakukan dalam *text mining* adalah sebagai berikut:

- *Tokenizing*

Proses ini memotong setiap kata dalam teks, dan mengubah semua huruf dalam dokumen menjadi huruf kecil. Hanya huruf „a“ sampai „z“ yang diterima, sedangkan karakter selain huruf dihilangkan. Jadi hasil dari proses *tokenizing* adalah kata kata yang merupakan penyusun kalimat/*string* yang dimasukkan.

- *Filtering*

Pada tahap ini dilakukan proses filter atau penyaringan kata hasil dari proses *tokenizing*, dimana kata yang tidak relevan dibuang. Proses ini menggunakan pendekatan *stoplist*. Yang termasuk *stoplist* adalah “yang”, “di”, “dari”, dan lain-lain.

- *Stemming*

Stemming adalah proses untuk menggabungkan atau memecahkan setiap varian-varian suatu kata menjadi kata dasar. *Stem* (akar kata) adalah bagian dari akar yang tersisa setelah dihilangkan imbuhnya (awalan dan akhiran).

- *Analizing*

Pada tahap ini dilakukan proses perhitungan bobot (w) dokumen agar diketahui seberapa jauh tingkat similaritas antara *keyword* yang dimasukkan dengan dokumen.

2.4 Algoritma Stemming Arifin Setiono

Tahap stemming adalah tahap mencari root kata dari tiap 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. Contoh dari tahapan ini pada teks adalah seperti gambar. [TRI-09]



Gambar 2.1 Stemming

Sumber: Perancangan

Algoritma ini didahului dengan pembacaan tiap kata dari file sampel. Sehingga input dari algoritma ini adalah sebuah kata yang kemudian dilakukan :

1. Pemeriksaan semua kemungkinan bentuk kata. Setiap kata diasumsikan memiliki 2 Awalan (prefiks) dan 3 Akhiran (sufiks). Sehingga bentuknya menjadi : Prefiks 1 + Prefiks 2 + Kata dasar + Sufiks 3 + Sufiks 2 + Sufiks 1. Seandainya kata tersebut tidak memiliki imbuhan sebanyak imbuhan di atas, maka imbuhan yang kosong diberi tanda x untuk prefiks dan diberi tanda xx untuk sufiks.
2. Pemotongan dilakukan secara berurutan sebagai berikut :

AW : AW (Awalan)

AK : AK (Akhiran)

KD : KD (Kata Dasar)

- a. AW I, hasilnya disimpan pada p1 (prefiks 1)
- b. AW II, hasilnya disimpan pada p2 (prefiks 2)
- c. AK I, hasilnya disimpan pada s1 (sufiks 1)
- d. AK II, hasilnya disimpan pada s2 (sufiks 2)
- e. AK III, hasilnya disimpan pada s3 (sufiks 3)

Pada setiap tahap pemotongan di atas diikuti dengan pemeriksaan di kamus apakah hasil pemotongan itu sudah berada dalam bentuk dasar. Kalau pemeriksaan ini berhasil maka proses dinyatakan selesai dan tidak perlu melanjutkan proses pemotongan imbuhan lainnya. Contoh pemenggalan kata “mempermainkannya” :

- a. Langkah 1 :

Cek apakah kata ada dalam kamus

Ya : Success

Tidak : lakukan pemotongan AW I

Kata = memainkannya

- b. Langkah 2 :

Cek apakah kata ada dalam kamus

Ya : Success

Tidak : lakukan pemotongan AW II

Kata = mainkannya

c. Langkah 3 :

Cek apakah kata ada dalam kamus

Ya : Success

Tidak : lakukan pemotongan AK I

Kata = mainkan

d. Langkah 4 :

Cek apakah kata ada dalam kamus

Ya : Success

Tidak : lakukan pemotongan AK II

Kata = main

e. Langkah 5 :

Cek apakah kata ada dalam kamus

Ya : Success

Tidak : lakukan pemotongan AK III. Dalam hal ini AK III tidak ada, sehingga kata tidak diubah.

Kata = main

f. Langkah 6

Cek apakah kata ada dalam kamus

Ya : Success

Tidak : "Kata tidak ditemukan"

3. Namun jika sampai pada pemotongan AK III, belum juga ditemukan di kamus, maka dilakukan proses kombinasi. KD yang dihasilkan dikombinasikan dengan imbuhan-imbuhan dalam 12 konfigurasi berikut:

a. KD

b. KD + AK III

c. KD + AK III + AK II

d. KD + AK III + AK II + AK I

e. AW I + AW II + KD

f. AW I + AW II + KD + AK III

g. AW I + AW II + KD + AK III + AK II

h. AW I + AW II + KD + AK III + AK II + AK I

- i. AW II + KD
- j. AW II + KD + AK III
- k. AW II + KD + AK III + AK II
- l. AW II + KD + AK III + AK II + AK I

Sebenarnya kombinasi a, b, c, d, h, dan l sudah diperiksa pada tahap sebelumnya, karena kombinasi ini adalah hasil pemotongan bertahap tersebut. Dengan demikian, kombinasi yang masih perlu dilakukan tinggal 6 yakni pada kombinasi-kombinasi yang belum dilakukan (e, f, g, i, j, dan k). Tentunya bila hasil pemeriksaan suatu kombinasi adalah 'ada' maka pemeriksaan apada kombinasi lainnya sudah tidak diperlukan lagi.

Pemeriksaan 12 kombinasi ini diperlukan, karena adanya fenomena overstemming pada algoritma pemotongan imbuhan. Kelemahan ini berakibat pada pemotongan bagian kata yang sebenarnya adalah milik kata dasar itu sendiri yang kebetulan mirip dengan salah satu jenis imbuhan yang ada. Dengan 12 kombinasi itu, pemotongan yang sudah terlanjur tersebut dapat dikembalikan sesuai posisinya. [ARS-00]

2.5 Naïve Bayes

Klasifikasi adalah salah satu tugas yang penting dalam data mining, dalam klasifikasi sebuah pengklasifikasi dibuat dari sekumpulan data latih dengan kelas yang telah di tentukan sebelumnya. Performa pengklasifikasi biasanya diukur dengan ketepatan (atau tingkat galat) [WAL-95]

Naïve Bayes merupakan salah satu metode *machine learning* yang menggunakan perhitungan probabilitas. Metode ini memanfaatkan teori yang dikemukakan oleh ilmuwan Inggris Thomas Bayes 8, yaitu memprediksi probabilitas di masa depan berdasarkan pengalaman di masa sebelumnya. Teorema Bayes adalah teorema yang digunakan dalam statistika untuk menghitung peluang untuk suatu hipotesis, Bayes Optimal Classifier menghitung peluang dari suatu kelas dari masing-masing kelompok atribut yang ada, dan menentukan kelas mana yang paling optimal [MAN-98].

2.5.1 Multinomial Naive Bayes

Model multinomial menangkap informasi frekuensi kata dalam dokumen sehingga dapat membantu proses klasifikasi. Dalam model multinomial diberikan permisalan seperti satu set kelas dilambangkan dengan C dan N menjadi ukuran kosakata. Kemudian MNB menetapkan dokumen tes t_i untuk kelas yang memiliki probabilitas tertinggi $Pr(c|t_i)$, yang dengan menggunakan aturan Bayes, ditentukan dengan:

$$Pr(c|t_i) = \frac{Pr(c)Pr(t_i|c)}{Pr(t_i)}, c \in C \quad (2.1)$$

Prioritas kelas $Pr(c)$ dapat diperkirakan dengan membagi jumlah dokumen milik kelas c dengan jumlah total dokumen. $Pr(t_i|c)$ adalah probabilitas untuk mendapatkan dokumen seperti t_i di kelas c dan dihitung sebagai:

$$Pr(t_i|c) = \left(\sum_n f_{ni} \right)! \prod_n \frac{Pr(w_n|c)^{f_{ni}}}{f_{ni}!} \quad (2.2)$$

Dimana f_{ni} adalah jumlah dari kata n dalam dokumen uji t_i dan $Pr(w_n|c)$ adalah probabilitas kata n yang diberikan oleh kelas c . Probabilitas terakhir diestimasi dari dokumen latih sebagai:

$$\bar{Pr}(w_n|c) = \frac{1 + F_{nc}}{N + \sum_{x=1}^N F_{xc}} \quad (2.3)$$

Dimana F_{xc} adalah jumlah kata x di seluruh dokumen latih pada kelas c . Faktor normalisasi $Pr(t_i)$ dalam persamaan (2.1) dapat dihitung dengan menggunakan

$$Pr(t_i) = \sum_{k=1}^{|C|} Pr(k) Pr(t_i|k) \quad (2.4)$$

Perhatikan bahwa istilah komputasi $(\sum_n f_{ni})!$ dan $\prod_n f_{ni}!$ dalam persamaan (2.2) dapat dihapus tanpa ada perubahan dalam hasil, karena tidak tergantung pada kelas c , dan persamaan (2.2) dapat ditulis sebagai:

$$Pr(t_i|c) = \alpha \prod_n Pr(w_n|c)^{f_{ni}} \tag{2.5}$$

dimana α adalah sebuah variabel konstan yang dibuang karena langkah normalisasi. [KFP-04]

2.5.2 Transformed Weight Normalized Complement Naive Bayes (TWCNB)

TWCNB dibangun berdasarkan MNB dan sangat mirip dengan MNB. Perbedaan utama antara TWCNB dan MNB yaitu TWCNB memperkirakan parameter kelas c dengan menggunakan data dari semua kelas selain dari c (yaitu menggunakan "komplemen"). Untuk tujuan ini, pada persamaan (2.3) disebut "bobot kata" daripada probabilitas dan didefinisikan dengan cara sebagai berikut:

$$w_{nc} = \log \left(\frac{1 + \sum_{k=1}^{|C|} F_{nk}}{N + \sum_{k=1}^{|C|} \sum_{x=1}^N F_{xk}} \right), k \neq c \wedge k \in C \tag{2.6}$$

Keterangan :

- F_{nk} = jumlah kata n pada dokumen latih
- w_{nc} = bobot kata
- C = satu set kelas
- N = jumlah seluruh kata pada data latih
- k = indeks untuk data latih
- n = tiap kata yang dihitung bobotnya
- x = indeks untuk jumlah seluruh kata pada data latih
- F_{xk} = jumlah seluruh kata pada kategori c.

Bobot Kata tersebut kemudian dinormalisasi untuk masing-masing kelas sehingga nilai mutlaknya berjumlah satu dan klasifikasi untuk pengujian dokumen t_i berdasarkan persamaan dibawah ini :

$$class(t_i) = argmax_c \left[\log(Pr(c)) - \sum_n (f_{ni}w_{nc}) \right] \tag{2.7}$$

f_{ni} adalah jumlah dari kata n dalam dokumen uji t_i dan $Pr(c)$ adalah prioritas kelas. Karena nilai dari $\log(Pr(c))$ biasanya diabaikan secara total, maka dapat disederhanakan menjadi :

$$class(t_i) = argmin_c \left[\sum_n (f_{ni} w_{nc}) \right] \quad (2.8)$$

Dimana F_{xc} adalah jumlah kata di seluruh dokumen latih pada kelas c . F_{nc} merupakan jumlah kata n pada dokumen latih. N adalah jumlah seluruh kata pada dokumen latih. Persamaan antara MNB dan TWCNB dapat dengan mudah diamati jika kita melihat aturan klasifikasi untuk MNB yang diberikan pada:

$$class(t_i) = argmax_c \left[\log(Pr(c)) + \sum_n f_{ni} \log \left(\frac{1 + F_{nc}}{N + \sum_{x=1}^N F_{xc}} \right) \right] \quad (2.9)$$

Aturan ini pada dasarnya sama dengan persamaan (2.1) jika kita menjatuhkan penyebut, melakukan log dan menggunakan persamaan (2.5) bukannya persamaan (2.2). [KFP-04]

2.6 Precision, Recall, dan F-measure

Distribusi dari training sample diasumsikan sama dengan distribusi dari testing sampel termasuk distribusi dari data baru (*unclassified sample*) yang ingin diduga kelas labelnya Untuk mendapat akurasi yang baik saat mengklasifikasi testing data maka training set harus mampu merepresentasikan keadaan dari data tes Jika tidak, maka akurasi biasanya rendah (kurang baik).

Precision dapat diartikan sebagai kepersisan atau kecocokan (antara permintaan informasi dengan jawaban terhadap permintaan itu). Jika seseorang mencari informasi di sebuah sistem, dan sistem menawarkan beberapa dokumen, maka kepersisan ini sebenarnya juga adalah relevansi. Artinya, seberapa persis atau cocok dokumen tersebut untuk keperluan pencari informasi, bergantung pada seberapa relevan dokumen tersebut bagi si pencari.

Recall adalah proporsi jumlah dokumen yang dapat ditemukan-kembali oleh sebuah proses pencarian di sistem IR (*Information Retrieval*). Rumusnya: Jumlah dokumen relevan yang ditemukan / Jumlah semua dokumen relevan di dalam

koleksi. Lalu, *precision* adalah proporsi jumlah dokumen yang ditemukan dan dianggap relevan untuk kebutuhan si pencari informasi. Rumusnya: Jumlah dokumen relevan yang ditemukan / Jumlah semua dokumen yang ditemukan. [PUT-08]

$$precision = \frac{|{\text{relevant documents}} \cap {\text{retrieved documents}}|}{|{\text{retrieved documents}}|} \tag{2.12}$$

$$recall = \frac{|{\text{relevant documents}} \cap {\text{retrieved documents}}|}{|{\text{relevant documents}}|} \tag{2.13}$$

Dengan kata lain, untuk menghitung *recall* dan *precision*-nya dapat digunakan rumus sebagai berikut:

Tabel 2.1 Tabel Kontigensi

	Classified Positive	Classified Negative
Actual Positive	TP	FN
Actual Negative	FP	TN

Sumber: [RIJ-79]

$$Precision = \frac{TP}{TP + FP} \tag{2.14}$$

$$Recall = \frac{TP}{TP + FN} \tag{2.15}$$

Keterangan:

- TN (True Negative) : dokumen salah dan diprediksi salah
- TP (True Positive) : dokumen benar dan diprediksi benar
- FN (False Negative) : dokumen benar dan diprediksi salah
- FP (False Positive) : dokumen salah dan diprediksi benar. [RIJ-79]



Kombinasi *precision* dan *recall* biasa dikombinasikan sebagai *harmonic mean*, biasa disebut *F-measure* yang mana dapat di formulasikan seperti persamaan berikut:

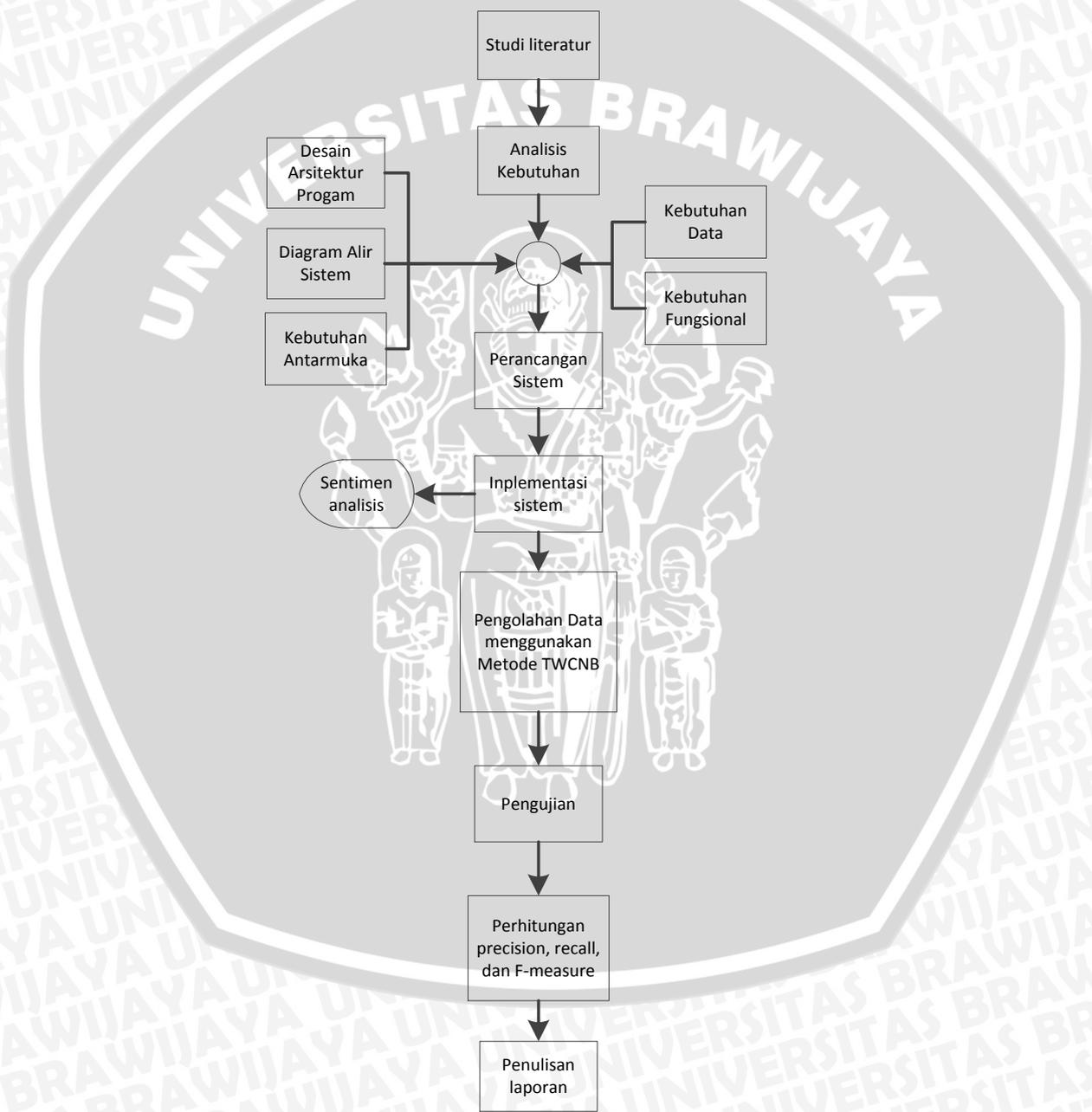
$$F - measure = 2 \times \frac{precision \times recall}{precision + recall} \quad (2.16)$$

F-measure biasa digunakan pada bidang sistem temu kembali informasi untuk mengukur klasifikasi pencarian dokumen dan performa query classification. Pada penelitian terdahulu *F-measure* lebih difokuskan untuk menghitung nilai, namun seiring dengan perkembangan mesin pencari dengan skala besar, kini *F-measure* lebih menekankan pada kinerja *precision* dan *recall* itu sendiri. Sehingga lebih bisa dilihat pada aplikasi secara keseluruhan [AHG-05].



BAB III METODOLOGI DAN PERANCANGAN

Bab ini menjelaskan langkah-langkah yang akan ditempuh dalam penyusunan skripsi, yaitu perancangan, implementasi dan pengujian dari aplikasi perangkat lunak yang akan dibuat. Kesimpulan dan saran disertakan sebagai catatan atas aplikasi dan kemungkinan arah pengembangan aplikasi selanjutnya.



Gambar 3.1 Desain penelitian
Sumber: Perancangan



3.1 Studi Literatur

Studi literatur mempelajari mengenai penjelasan dasar teori yang digunakan untuk menunjang penulisan skripsi. Teori-teori pendukung tersebut diperoleh dari buku, jurnal, *e-book*, penelitian sebelumnya, dan dokumentasi project. Referensi yang digunakan untuk mendukung penulisan skripsi ini meliputi:

1. Pemahaman tentang struktur dokumen penelitian
2. Pemahaman tentang *text mining*
3. Pemahaman tentang klasifikasi
4. Pemahaman tentang metode *Transformed Weight-Normalized Complement Naive Bayes* (TWCNB)
5. Bahasa pemrograman PHP
6. Basis data MySQL

3.2 Penyusunan Dasar Teori

Penyusunan dasar teori dilakukan setelah mendapatkan referensi yang tepat untuk mendukung penulisan penelitian ini. Teori-teori pendukung tersebut meliputi:

1. *Text Mining*

Meliputi tahap-tahap *cleansing*, *parsing*, *tokenizing*, *filtering/stopword removal*, dan *stemming*. untuk menghasilkan term yang nantinya akan di proses oleh program.

2. *Classification* (Klasifikasi)

Meliputi tahap-tahap *term weighting* dan *classification* untuk menghasilkan klasifikasi atau pengelompokan sesuai dengan metode yang telah dipilih oleh peneliti.

3.3 Metode Pengumpulan Data

Data yang dibutuhkan untuk penelitian ini adalah data lirik lagu berbahasa inggris yang akan diterjemahkan kedalam bahasa indonesia menggunakan Google Translate, dan lirik lagu berbahasa indonesia. Metode penentuan kelas sentimen data lirik lagu di dapat dari sumber – sumber kaset CD *official* lagu – lagu dari toko CD.

3.4 Analisis dan Perancangan

Analisis kebutuhan bertujuan untuk mendapatkan semua kebutuhan yang diperlukan oleh sistem yang akan dibangun. Analisis kebutuhan dilakukan dengan mengidentifikasi kebutuhan sistem dan siapa saja yang terlibat di dalamnya. Berikut analisis kebutuhan dalam aplikasi klasifikasi tema pada lirik lagu berikut ini.

3.4.1 Kebutuhan Antar Muka

Kebutuhan-kebutuhan untuk pengembangan perangkat lunak ini sebagai berikut:

1. Sistem yang akan dibangun harus memiliki tampilan yang familiar oleh user sehingga proses memasukkan teks menjadi mudah.
2. Program yang akan dibangun harus mampu menampilkan hasil analisis sentimen yang telah diproses sebelumnya.

3.4.2 Kebutuhan Data

Data yang diolah oleh perangkat lunak ini adalah:

1. Data berupa kata dasar dari lirik berbahasa indonesia, yang akan berfungsi pada saat proses *stemming*.
2. Data hasil analisis sentimen yang akan di simpan di dalam database.

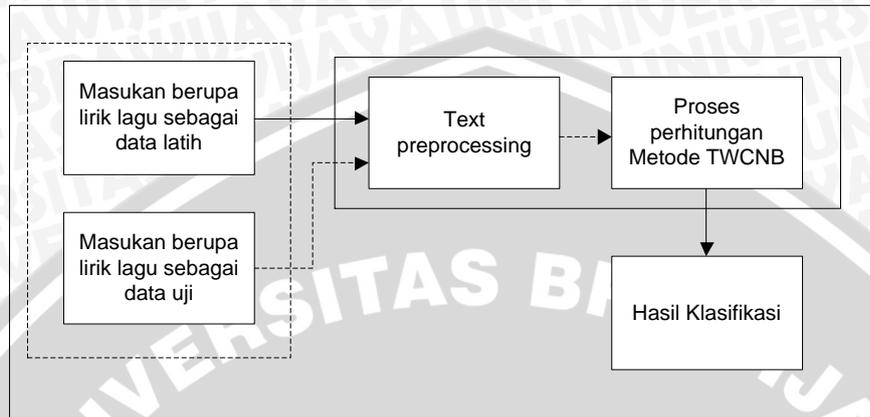
3.4.3 Kebutuhan Fungsional

Fungsi-fungsi yang dimiliki oleh perangkat lunak ini adalah:

1. Perangkat lunak harus mampu melakukan proses *preprocessing*, yakni berupa *parsing*, penghilangan kata *stopwords* dan *stemming*.
2. Perangkat lunak harus mampu melakukan proses pembobotan berdasarkan metode yang akan digunakan.
3. Perangkat lunak harus mampu menghasilkan hasil klasifikasi tema pada lirik lagu berdasarkan proses yang telah dilakukan.

3.5 Arsitektur Sistem

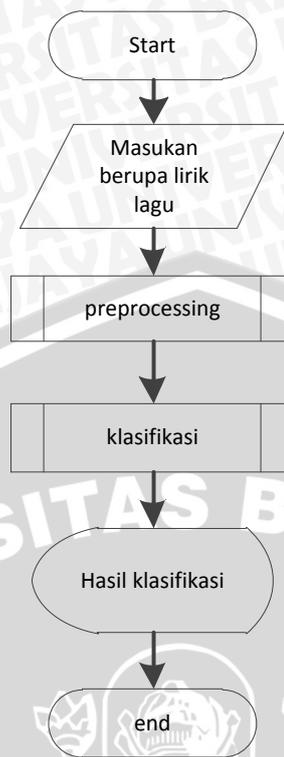
Arsitektur sistem menggambarkan kerangka dasar dari sistem yang akan dikembangkan. Tahapan proses klasifikasi tema terhadap lirik lagu dapat dilihat pada gambar berikut ini.



Gambar 3.2 Arsitektur Perancangan Sistem
Sumber: Perancangan

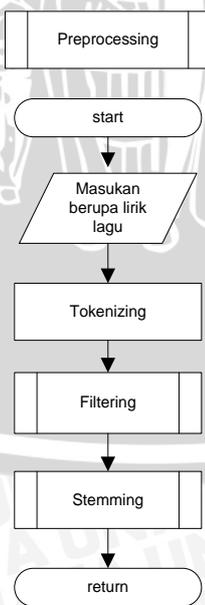
3.6 Diagram Alir Sistem

Diagram alir merupakan gambaran secara grafik yang terdiri dari simbol-simbol yang dibuatkan untuk memudahkan pemahaman proses yang terjadi dan memahami hasil akhir yang terjadi. Sistem ini dimulai dengan masukan berupa teks dari user, kemudian akan dilakukan *text preprocessing* dan pembobotan terhadap teks tersebut. Langkah akhir, hasil dari proses sebelumnya kemudian dilanjutkan dengan proses klasifikasi tema pada lirik lagu. Gambar 3.3 menunjukkan diagram alir sistem.



Gambar 3.3 Diagram Alir Sistem
Sumber: Perancangan

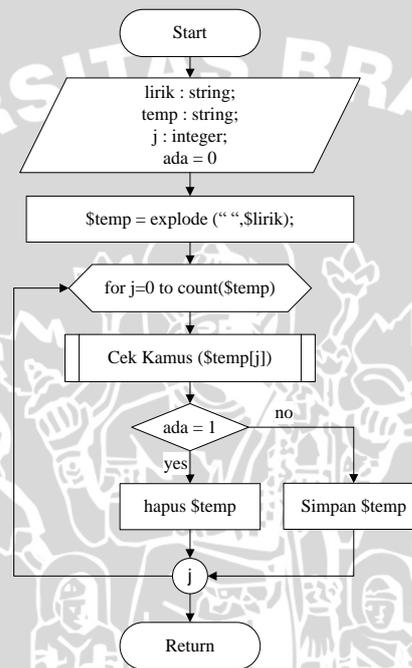
Preprocessing meliputi *tokenizing*, *filtering*, *stemming*. Proses – proses yang dilakukan pada tahap ini dilakukan pada gambar Diagram Alir tahap preprocessing.



Gambar 3.4 Diagram Alir Preprocessing
Sumber: Perancangan

Proses *Tokenizing* yaitu memotong setiap kata dalam teks, dan mengubah semua huruf dalam dokumen menjadi huruf kecil. Hanya huruf „a“ sampai „z“ yang diterima, sedangkan karakter selain huruf dihilangkan. Kemudian akan dipisahkan berdasarkan spasi menjadi sebuah kata – kata.

Pada tahap *Filtering* dilakukan proses filter atau penyaringan kata hasil dari proses *tokenizing*, dimana kata yang tidak relevan dibuang. Proses ini menggunakan pendekatan *stoplist*. Yang termasuk *stoplist* adalah “yang”, “di”, “dari”, dan lain-lain. Gambar 3.5 menunjukkan diagram alir *filtering*.

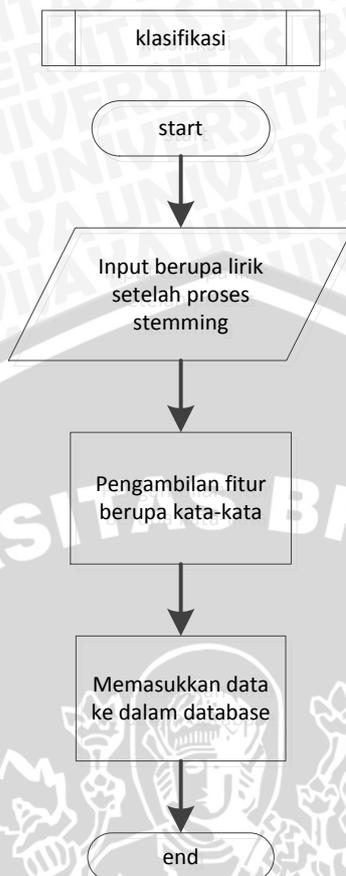


Gambar 3.5 Diagram Alir Filtering Stopword Removal

Sumber: Perancangan

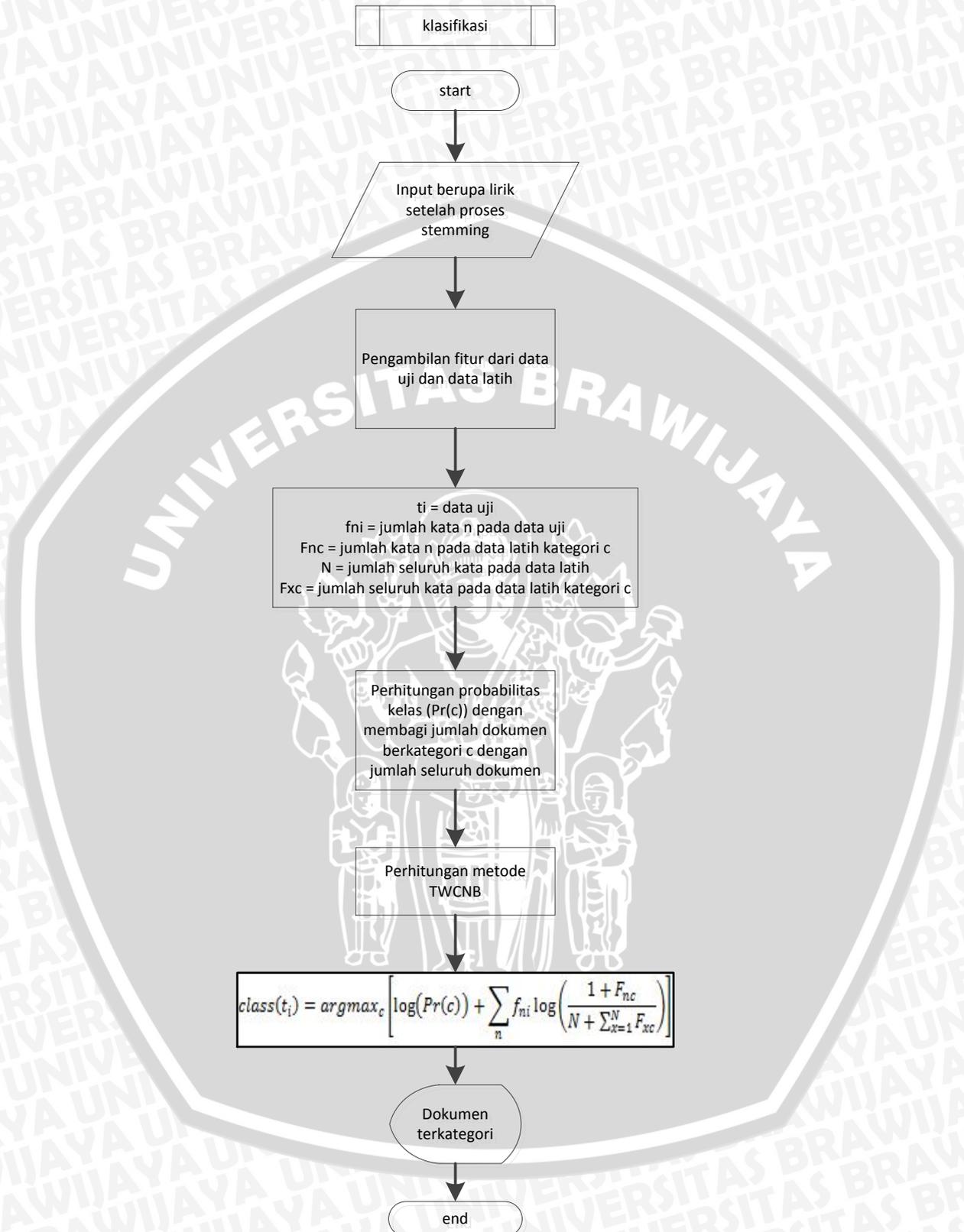
Stemming arifin-setiono dilakukan untuk mendapatkan kata dasar dari sebuah *term*. Proses ini digunakan untuk mencari kata dasar dari suatu *term*. Dalam penelitian ini stemming arifin-setiono yang digunakan merupakan hasil pengembangan dari Ahmad Afandi dalam skripsinya yang berjudul Implementasi Algoritma Jaro Winkler Distance Untuk Aplikasi Penilaian Otomatis Soal Essai Berbahasa Indonesia.

Proses klasifikasi pada metode TWCNB terbagi menjadi dua, yaitu proses latih dan uji. Untuk proses latih, maka lirik yang telah melalui proses *stemming* akan ditentukan kategorinya dan data-data nya tersimpan di dalam database.



Gambar 3.6 Diagram Alir Data Latih TWCNB
Sumber: Perancangan

Namun apabila proses uji, hasil preprocessing kemudian masuk ke dalam proses perhitungan metode yang nantinya akan menghasilkan bobot dari ketiga kategori. Bobot yang paling besar akan menentukan kelas dari dokumen uji tersebut.

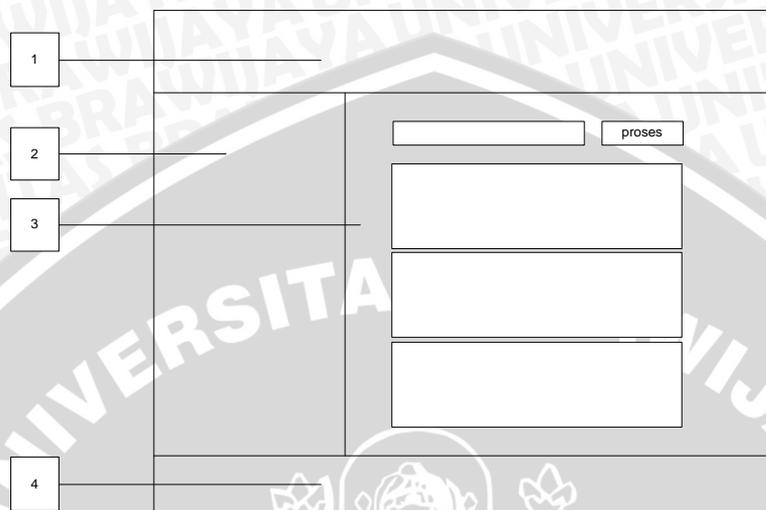


Gambar 3.7 Diagram Alir Data Uji TWCNB
Sumber: Perancangan



3.6.1 Desain Antar Muka

Perancangan desain antar muka sistem klasifikasi ini dapat dilihat pada Gambar 3.9. Desain ini dirancang untuk memudahkan dan memuaskan user dalam penggunaannya.



Gambar 3.8 Perancangan Desain Antar Muka Sistem
Sumber: Perancangan

Keterangan Gambar 3.9:

1. Header
2. Menu
3. Isi
4. Footer

Desain antar muka sistem ini terdiri dari empat komponen utama. Yaitu header, Menu, Isi, dan Footer. Isi terdiri dari sebuah *field*, sebuah tombol proses, dan 3 *text box*. *Field* digunakan untuk meng-*input* lirik yang akan di klasifikasikan, selanjutnya tombol proses ditekan untuk melakukan pengujian. Ketiga *text box* yang berada di bawah *field* merupakan postingan lirik – lirik yang telah diklasifikasi sebelumnya sesuai dengan kategorinya masing – masing.

Pada halaman hasil, akan ditampilkan sampel dokumen yang diuji dan perhitungan yang dihasilkan dari proses klasifikasi yang telah dilakukan. Gambar 3.10 menampilkan desain antar muka untuk halaman hasil.

The diagram shows a wireframe of a result page interface. It consists of a main rectangular area divided into four horizontal sections, each labeled with a number from 1 to 4 on the left side. Section 1 is a large text input field. Section 2 contains a button labeled 'proses'. Section 3 is a rectangular area representing a document sample. Section 4 is a larger rectangular area representing classification calculations. A vertical line separates the input field from the other sections. A small box labeled 'proses' is also shown to the right of the main area, with a line pointing to the button in section 2.

Gambar 3.9 Desain Antar Muka Halaman Hasil

Sumber: Perancangan

Keterangan gambar 3.10:

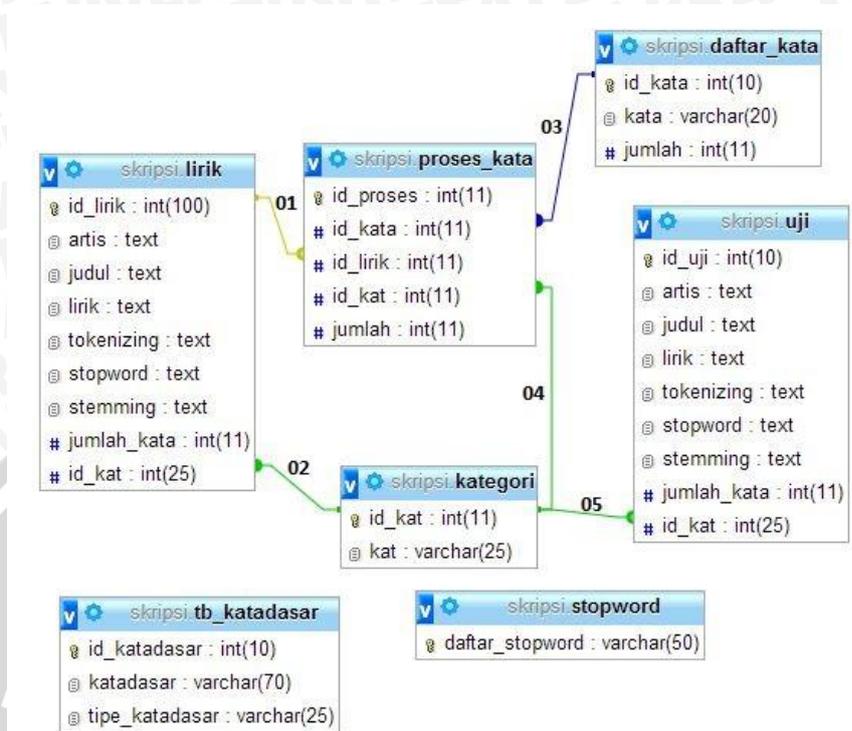
1. *Field* lirik
2. Tombol Proses
3. Sampel dokumen uji
4. Perhitungan klasifikasi

Desain antar muka sistem ini terdiri dari empat komponen yaitu yang *field* lirik, tombol proses, sampel dokumen uji, dan perhitungan klasifikasi. *Field* lirik merupakan text box untuk memasukkan lirik. Tombol proses merupakan tombol untuk memproses klasifikasi pada lirik. Sampel dokumen uji merupakan cuplikan dari lirik yang telah di klasifikasi, dan perhitungan klasifikasi yaitu hasil akhir yang menentukan tergolongnya lirik.

3.7 Perancangan Sistem Basis Data

Perancangan sistem basis data dilakukan dengan *Database Management System* MySQL. Hasil implementasi SQL pada *database* ini dimodelkan dalam

diagram konseptual *entity relationship* seperti pada gambar 3.9 dibawah ini.



Gambar 3.10 Tabel Database klasifikasi tema pada lirik lagu

Sumber: Perancangan

Relasi :

- Relasi 01 : relasi yang menghubungkan antara tabel lirik dengan tabel proses kata melalui id_lirik. Untuk satu id_lirik di tabel lirik bisa menjadi banyak di tabel kata sehingga hubungannya adalah 1:n.
- Relasi 02 : relasi yang menghubungkan tabel lirik dan tabel kategori melalui id_kat. Untuk satu id_kat pada tabel kategori bisa memiliki banyak lirik dengan kategori yang berbeda pada tabel lirik sehingga hubungannya adalah 1:n.
- Relasi 03 : menghubungkan tabel proses_kata dengan tabel daftar_kata melalui id_kata. Untuk satu id_kata pada tabel daftar_kata bisa memiliki banyak id_kata pada tabel proses_kata sehingga hubungannya adalah 1:n.
- Relasi 04 : menghubungkan tabel proses_kata dengan tabel kategori melalui id_kat. Untuk satu id_kat pada tabel kategori bisa memiliki banyak kata dengan kategori yang berbeda pada tabel proses_kata sehingga hubungannya adalah 1:n.

- Relasi 05 : relasi yang menghubungkan tabel uji dan tabel kategori melalui id_kat. Untuk satu id_kat pada tabel kategori bisa memiliki banyak lirik dengan kategori yang berbeda pada tabel uji sehingga hubungannya adalah 1:n.

Tabel – tabel yang tidak memiliki relasi seperti tabel db_katadasar dan stopword hanya berfungsi untuk menyimpan data tanpa perlu memiliki relasi dengan tabel lain. Untuk tabel db_katadasar berfungsi menyimpan data-data kata dasar yang dapat membantu disaat pengecekan kata dasar pada proses stemming. Untuk tabel stopword berfungsi menyimpan data-data kata stopword yang dapat membantu disaat penghapusan kata pada proses stopword removal.

3.8 Manualisasi Analisis Sentimen

Perhitungan manual berfungsi untuk memberukan gambaran umum perancangan sistem yang akan dibangun. Contoh manualisasi pelatihan sistem klasifikasi tema pada lirik lagu dengan metode TWCNB adalah sebagai berikut.

1. Input Lirik lagu

Tabel 3.1 Contoh Masukan Lirik

no	Lirik	Label	Jumlah Kata
1	Hari ini aku telah jatuh cinta Tak kan mampu aku menyangkalnya Jatuh cinta kepadamu Sosok yang sering menjengkelkan aku Sering mengganggu Kau memainkan rasa hatiku Namun kini aku berbalik Jatuh cinta dan bernyanyi lalala lalala lalala lalalalalalala lalala lalala lalala lalalalalalala Aku jatuh cinta kepada dirimu Orang yang tak pernah ku bayangkan Tak pernah ku mimpikan Untuk bisa menjadi pacarku Malam ini aku berniat Untuk menyatakan rasa cintaku Semoga tanganku berjodoh Untuk bertepuk dengan cintamu Aku jatuh cinta kepada dirimu Orang yang tak pernah ku bayangkan Tak pernah ku mimpikan Untuk bisa menjadi pacarku Jadi pacarku jadi pacarku Jadi pacarku jadi pacarku Aku jatuh cinta kepada dirimu Orang yang tak pernah ku bayangkan Tak pernah ku mimpikan Untuk bisa menjadi pacarku Aku jatuh cinta kepada dirimu Orang yang tak pernah ku bayangkan Tak pernah ku mimpikan Untuk bisa menjadi pacarku	cinta	140

2	Lihat lihatlah aku Dekat dekati aku Peluk peluklah aku Cium ciumlah aku oh kasih Teman temani aku Yakin yakin padaku Sayang sayangi aku Bawa bawalah aku Bersamamu apakah kau merasakannya getaran cinta Beri beri pertanda pertanda cinta Teman temani aku Yakin yakin padaku Sayang sayangi aku Bawa bawalah aku Teman temani aku Yakin yakin padaku Sayang sayangi aku Bawa bawalah aku Teman temani aku Yakin yakin padaku Sayang sayangi aku Bawa bawalah aku Teman temani aku Yakin yakin padaku Sayang sayangi aku Bawa bawalah aku	cinta	85
3	Berkibarlah benderaku Lambang suci gagah perwira Di seluruh pantai Indonesia Kau tetap pujaan bangsa Siapa berani menurunkan engkau Serentak rakyatmu membela Sang merah putih yang perwira Berkibarlah Slama-lamanya Kami rakyat Indonesia Bersedia setiap masa Mencurahkan segenap tenaga Supaya kau tetap cemerlang Tak goyang jiwaku menahan rintangan Tak gentar rakyatmu berkorban Sang merah putih yang perwira Berkibarkah Slama-lamanya	perjuangan	57
4	Demi cintaMu ya Allah Pada Muhammad nabiMu Ampuni dosaku Wujudkan harapanku Ya Rasulallah Siapa yang cinta pada nabinya Pasti bahagia dalam hidupnya Muhammadku Muhammadku Dengarlah seruanKu Aku rindu aku rindu Kepadamu Muhammadku Kau yang mengaku cinta kepada nabimu Kau yang mmengaku merindukan nabimu Jika kau benar cinta dan rindu kepada Muhammad nabimu Buktikan Taati perintahNya, tinggalkan laranganNya Teladani akhlakNya Niscaya kelak kau akan berjumpa dengan Rasulallah Niscaya kelak kau akan berjumpa dengan Rasulallah Muhammadku Muhammadku Dengarlah seruanKu Aku rindu aku rindu Kepadamu Muhammadku Kau jarkan hidup ini untuk saling mengasihi Ku tanamkan dalam hati ku amalkan sejak dini Engkaulah nabi pembawa cinta Kau bimbing kamu menuju cinta Muhammadku Muhammadku Dengarlah seruanKu Aku rindu aku rindu Kepadamu Muhammadku	religi / rohani	117
5	Ya Allah Ya Rabbi Ya Allah Ya Rabbi Ya Allah Ya Rabbi Ya Allah Ya Allah Ya Allah Ya Rabbi Ya Allah Ya Rabbi Ya Allah Ya Rabbi Ya Allah Ya Allah jangan kau coba aku Melebihi batas mampu dan sanggupku Ya Allah bila memang kau coba Aku percaya kau sayang padaku Ya Allah Ya Rabbi Ya Allah Ya Rabbi Ya Allah Ya Rabbi Ya Allah Ya Allah lindungilah diriku Dari yang menjahati menzhalimiku Ya Allah kaulah Maha segala Engkau pelindung hidup dan matiku Ya Allah Ya Rabbi Ya Allah Ya Rabbi Ya Allah Ya Rabbi Ya Allah Ya Allah jangan kau coba aku Melebihi batas mampu dan sanggupku	religi / rohani	107
6	Awalnya tak mengerti apa yang sedang kurasakan Segalanya berubah dan rasa rindu itu pun ada Sejak kau hadir di setiap malam di tidurku Aku tahu sesuatu sedang terjadi padamu Sudah sekian lama kualami pedih putus cinta Dan mulai terbiasa hidup sendiri tanpa asmara Dan hadirmu membawa cinta sembuhkan lukaku Kau berbeda dari yang ku kira Aku jatuh cinta kepada dirinya Sungguh-sungguh cinta oh apa adanya Tak pernah ku ragu namun tetap selalu menunggu Sungguh aku jatuh cinta kepadanya Coba-coba dengarkan apa yang ingin aku katakan Yang selama ini sungguh telah lama terpendam Aku tak percaya membuatku tak berdaya Tuk ungkapkan	xx	177

<p>apa yang ku rasa Aku jatuh cinta kepada dirinya Sungguh-sungguh cinta oh apa adanya Tak pernah ku ragu namun tetap selalu menunggu Sungguh aku jatuh cinta kepadanya Coba-coba dengarkan apa yang ingin aku katakan Yang selama ini sungguh telah lama terpendam Aku tak percaya membuatku tak berdaya Tuk ungkapkan apa yang ku rasa Kadang aku cemburu, kadang aku gelisah Seringnya ku tak tentu lalui hariku Tak dapat ku pungkiri hatiku yang terdalam Betapa aku jatuh cinta kepadanya</p>		
---	--	--

Sumber: Perancangan

2. Preprocessing

a. Tokenizing & filtering

Tabel 3.2 – Contoh lirik hasil tokenizing dan filtering

No	Lirik	Jumlah kata
1	<p>hari ini aku telah jatuh cinta tak kan mampu aku menyangkalnya jatuh cinta kepadamu sosok yang sering menjengkelkan aku sering mengganguku kau mainkan rasa hatiku namun kini aku berbalik jatuh cinta dan bernyanyi lalala lalala lalala lalalalalalala lalala lalala lalala lalalalalalala aku jatuh cinta kepada dirimu orang yang tak pernah ku bayangkan tak pernah ku mimpikan untuk bisa menjadi pacarku malam ini aku berniat untuk menyatakan rasa cintaku semoga tanganku berjodoh untuk bertepuk dengan cintamu aku jatuh cinta kepada dirimu orang yang tak pernah ku bayangkan tak pernah ku mimpikan untuk bisa menjadi pacarku jadi pacarku jadi pacarku jadi pacarku jadi pacarku aku jatuh cinta kepada dirimu orang yang tak pernah ku bayangkan tak pernah ku mimpikan untuk bisa menjadi pacarku aku jatuh cinta kepada dirimu orang yang tak pernah ku bayangkan tak pernah ku mimpikan untuk bisa menjadi pacarku</p>	140
2	<p>lihat lihatlah aku dekat dekati aku peluk peluklah aku cium ciumlah aku oh kasih teman temani aku yakin yakin padaku sayang sayangi aku bawa bawalah aku bersamamu apakah kau merasakannya getaran cinta beri beri pertanda pertanda cinta teman temani aku yakin yakin padaku sayang sayangi aku bawa bawalah aku teman temani aku yakin yakin padaku sayang sayangi aku bawa bawalah aku teman temani aku yakin yakin padaku sayang sayangi aku bawa bawalah aku teman temani aku yakin yakin padaku sayang sayangi aku bawa bawalah aku</p>	85
3	<p>berkibarlah benderaku lambang suci gagah perwira pantai indonesia tetap pujaan bangsa berani menurunkan engkau serentak rakyatmu membela sang merah putih perwira berkibarlah slama rakyat indonesia bersedia masa mencurahkan segenap tenaga tetap cemerlang goyang jiwaku menahan rintangan gentar rakyatmu berkorban sang merah putih perwira berkibarkah</p>	44



4	cintamu allah muhammad nabimu ampuni dosaku wujudkan harapanku rasulallah cinta nabinya bahagia hidupnya muhammadku muhammadku dengarlah seruanku rindu rindu kepadamu muhammadku mengaku cinta nabimu mmengaku merindukan nabimu benar cinta rindu muhammad nabimu buktikan taati perintahnya tinggalkan larangannya teladani akhlaknya niscaya kelak berjumpa rasulallah niscaya kelak berjumpa rasulallah muhammadku muhammadku dengarlah seruanku rindu rindu kepadamu muhammadku jarkan hidup mengasihi ku tanamkan hati ku amalkan engkaulah nabi pembawa cinta bimbing menuju cinta muhammadku muhammadku dengarlah seruanku rindu rindu kepadamu muhammadku	78
5	allah rabbi allah rabbi allah rabbi allah allah rabbi allah rabbi allah rabbi allah allah coba melebihi batas sanggupku allah coba percaya sayang padaku allah rabbi allah rabbi allah rabbi allah allah lindungilah diriku menjahati menzhalimiku allah kaulah maha engkau pelindung hidup matiku allah rabbi allah rabbi allah allah coba melebihi batas sanggupku	55
6	awalnya tak mengerti apa yang sedang kurasakan segalanya berubah dan rasa rindu itu pun ada sejak kau hadir di setiap malam di tidurku aku tahu sesuatu sedang terjadi padamu sudah sekian lama kualami pedih putus cinta dan mulai terbiasa hidup sendiri tanpa asmara dan hadirmu membawa cinta sembuhkan lukaku kau berbeda dari yang ku kira aku jatuh cinta kepada dirinya sungguh sungguh cinta oh apa adanya tak pernah ku ragu namun tetap selalu menunggu sungguh aku jatuh cinta kepadanya coba coba dengarkan apa yang ingin aku katakan yang selama ini sungguh telah lama terpendam aku tak percaya membuatku tak berdaya tuk ungkapkan apa yang ku rasa aku jatuh cinta kepada dirinya sungguh sungguh cinta oh apa adanya tak pernah ku ragu namun tetap selalu menunggu sungguh aku jatuh cinta kepadanya coba coba dengarkan apa yang ingin aku katakan yang selama ini sungguh telah lama terpendam aku tak percaya membuatku tak berdaya tuk ungkapkan apa yang ku rasa kadang aku cemburu kadang aku gelisah seringnya ku tak tentu lalui hariku tak dapat ku pungkiri hatiku yang terdalam betapa aku jatuh cinta kepadanya	181

Sumber: Perancangan

b. Stopword

Tabel 3.3 Contoh lirik hasil *stopword*

No	Lirik	Jumlah Kata
1	hari jatuh cinta menyangkalnya jatuh cinta kepadamu sosok menjengkelkan mengganguku memainkan rasa hatiku berbalik jatuh cinta bernyanyi lalala lalala lalala lalalalalalala lalala lalala lalala lalalalalalala jatuh cinta dirimu orang ku bayangkan ku mimpikan pacarku malam berniat menyatakan rasa cintaku semoga tanganku berjodoh bertepuk cintamu jatuh cinta dirimu orang ku bayangkan ku mimpikan pacarku jadi pacarku jadi pacarku jadi pacarku jadi pacarku jatuh cinta dirimu orang ku bayangkan ku mimpikan pacarku jatuh cinta dirimu orang ku bayangkan ku mimpikan pacarku	79
2	lihat lihatlah dekati peluk peluklah cium ciumlah oh kasih teman temani yakin yakin padaku sayang sayangi bawa bawalah bersamamu merasakannya getaran cinta beri beri pertanda pertanda cinta teman temani yakin yakin padaku sayang sayangi bawa bawalah teman temani yakin yakin padaku sayang sayangi bawa bawalah teman temani yakin yakin padaku sayang sayangi bawa bawalah teman temani yakin yakin padaku sayang sayangi bawa bawalah	63

3	berkibarlah benderaku lambang suci gagah perwira pantai indonesia tetap pujaan bangsa berani menurunkan engkau serentak rakyatmu membela sang merah putih perwira berkibarlah slama rakyat indonesia bersedia masa mencurahkan segenap tenaga tetap cemerlang goyang jiwaku menahan rintangan gentar rakyatmu berkorban sang merah putih perwira berkibarkah slama	45
4	cintamu allah muhammad nabimu ampuni dosaku wujudkan harapanku rasulallah cinta nabinya bahagia hidupnya muhammadku muhammadku dengarlah seruanaku rindu rindu kepadamu muhammadku mengaku cinta nabimu mmengaku merindukan nabimu benar cinta rindu muhammad nabimu buktikan taati perintahnya tinggalkan larangannya teladani akhlaknya niscaya kelak berjumpa rasulallah niscaya kelak berjumpa rasulallah muhammadku muhammadku dengarlah seruanaku rindu rindu kepadamu muhammadku jarkan hidup mengasihi ku tanamkan hati ku amalkan engkaulah nabi pembawa cinta bimbing menuju cinta muhammadku muhammadku dengarlah seruanaku rindu rindu kepadamu muhammadku	78
5	allah rabbi allah rabbi allah rabbi allah allah rabbi allah rabbi allah rabbi allah allah coba melebihi batas sanggupku allah coba percaya sayang padaku allah rabbi allah rabbi allah rabbi allah allah lindungilah diriku menjahati menzhalimiku allah kaulah maha engkau pelindung hidup matiku allah rabbi allah rabbi allah rabbi allah allah coba melebihi batas sanggupku	55
6	awalnya mengerti kurasakan berubah rasa rindu hadir malam tidurku tahu terjadi padamu sekian kualami pedih putus cinta mulai terbiasa hidup asmara hadirmu membawa cinta sembuhkan lukaku berbeda ku kira jatuh cinta sungguh sungguh cinta oh ku ragu tetap menunggu sungguh jatuh cinta coba coba dengarkan katakan sungguh terpendam percaya membuatku berdaya tuk ungkapkan ku rasa jatuh cinta sungguh sungguh cinta oh ku ragu tetap menunggu sungguh jatuh cinta coba coba dengarkan katakan sungguh terpendam percaya membuatku berdaya tuk ungkapkan ku rasa kadang cemburu kadang gelisah ku lalu hariku ku pungkiri hatiku terdalam betapa jatuh cinta	95

Sumber: Perancangan

c. Stemming

Tabel 3.4 Contoh Lirik hasil *stemming*

No	Lirik	Jumlah Kata
1	hari jatuh cinta sangkal jatuh cinta pada sosok jengkel ganggu main rasa hati balik jatuh cinta nyanyi lalala lalala lalala lalalalalalala lalala lalala lalala lalalalalalala jatuh cinta r orang ku bayang ku mimpi pacar malam niat tata rasa cinta moga tang jodoh tepuk cinta jatuh cinta r orang ku bayang ku mimpi pacar jadi pacar jadi pacar jadi pacar jadi pacar jatuh cinta r orang ku bayang ku mimpi pacar jatuh cinta r orang ku bayang ku mimpi pacar	79
2	lihat lihat dekat peluk luk ciu ciu oh kasih teman teman yakin yakin pada sayang sayang bawa bawa sama rasa getar cinta beri beri tanda tanda cinta teman teman yakin yakin pada sayang sayang bawa bawa teman teman yakin yakin pada sayang sayang bawa bawa teman teman yakin yakin pada sayang sayang bawa bawa teman teman yakin yakin pada sayang sayang bawa bawa	63



3	kibar bendera lambang suci gagah perwira pantai indonesia tetap puja bangsa berani turun engkau rentak rakyat bela sang merah putih perwira kibar slama rakyat indonesia dia masa curah segenap tenaga tetap cemerlang goyang jiwa tah rintang gentar rakyat korban sang merah putih perwira kibar slama	45
4	cinta allah muhammad nabi ampun dosa wujud harap rasulallah cinta nabi bahagia hidup muhammad muhammad dengar seru rindu rindu pada muhammad ka cinta nabi mmengaku rindu nabi benar cinta rindu muhammad nabi bukti taat perintah tinggal larang teladan akhlak niscaya kelak jumpa rasulallah niscaya kelak jumpa rasulallah muhammad muhammad dengar seru rindu rindu pada muhammad jarkan hidup kasih ku tanam hati ku amal engkau nabi bawa cinta bimbing tuju cinta muhammad muhammad dengar seru rindu rindu pada muhammad	78
5	allah rabbi allah rabbi allah rabbi allah allah rabbi allah rabbi allah rabbi allah allah coba lebih batas sanggup allah coba percaya sayang pada allah rabbi allah rabbi allah rabbi allah allah lindung r jahat menzhalikimu allah kau maha engkau lindung hidup mati allah rabbi allah rabbi allah rabbi allah allah coba lebih batas sanggup	55
6	awal erti rasa ubah rasa rindu hadir malam tidur tahu jadi pada kian alami pedih putus cinta mulai biasa hidup asmara hadir bawa cinta sembuh luka beda kira jatuh cinta sungguh sungguh cinta ragu tetap tunggu sungguh jatuh cinta coba coba dengar kata sungguh dam percaya buat daya ungkap rasa jatuh cinta sungguh sungguh cinta ragu tetap tunggu sungguh jatuh cinta coba coba dengar kata sungguh dam percaya buat daya ungkap rasa kadang cemburu kadang gelisah lalu hari pungkiri hati dalam betapa jatuh cinta	84

Sumber: Perancangan

Dari hasil preprocessing tersebut, setiap kata diambil fitur jumlahnya agar dapat digunakan dalam perhitungan metode. Berikut daftar kata dari tiap kategori.

Tabel 3.5 Tabel daftar kata hasil preprocessing

Kata	ID Lirik	Kategori	Jumlah
hari	1	cinta	1
jatuh	1	cinta	7
cinta	1	cinta	9
sangkal	1	cinta	1
pada	1	cinta	1
sosok	1	cinta	1
jengkel	1	cinta	1
ganggu	1	cinta	1
main	1	cinta	1
...
sayang	2	cinta	10
bawa	2	cinta	10
sama	2	cinta	1
rasa	2	cinta	1
getar	2	cinta	1

cinta	2	cinta	2
beri	2	cinta	2
tanda	2	cinta	2
kibar	3	perjuangan	3
bendera	3	perjuangan	1
lambang	3	perjuangan	1
segenap	3	perjuangan	1
tenaga	3	perjuangan	1
...
goyang	3	perjuangan	1
jiwa	3	perjuangan	1
tah	3	perjuangan	1
rintang	3	perjuangan	1
gentar	3	perjuangan	1
korban	3	perjuangan	1
cinta	4	religi	6
allah	4	religi	1
muhammad	4	religi	11
nabi	4	religi	6
ampun	4	religi	1
dosa	4	religi	1
wujud	4	religi	1
harap	4	religi	1
...
jahat	5	religi	1
menzhalimiku	5	religi	1
kau	5	religi	1
maha	5	religi	1
engkau	5	religi	1
hidup	5	religi	1
mati	5	religi	1

Sumber: Perancangan

3. Perhitungan Rumus

Pada metode TWCNB, proses perhitungan fitur untuk mendapatkan hasil klasifikasi dilakukan dengan menghitung bobot tiap kelas menggunakan rumus berikut :

$$\text{class}_{(t_i)} = \text{argmax}_c \left[\log(\text{Pr}_{(c)}) + \sum_n f_{ni} \log \left(\frac{1 + F_{nc}}{N + \sum_{x=1}^N F_{xc}} \right) \right]$$

t_i = dokumen uji

c = kelas kategori

Pr = prioritas kelas

f_{ni} = jumlah dari kata n dalam dokumen uji t_i

F_{nc} = jumlah kata n dalam dokumen latih berkategori c

N = jumlah seluruh daftar kata pada dokumen latih

F_{xc} = jumlah seluruh kata pada dokumen latih berkategori c .

Perhitungan dilakukan sebanyak kategori yang ditentukan. Nilai dari masing-masing kategori menunjukkan hasil klasifikasi dari metode TWCNB.

Berikut adalah daftar kata pada dokumen uji yang terdapat pada data latih.

Tabel 3.6 Tabel daftar kata pada dokumen uji

Kata	Dok Uji	Kat. Cinta	Kat. Religi	Kat. Perjuangan
awal	1	0	0	0
erti	1	0	0	0
rasa	4	3	0	0
ubah	1	0	0	0
rindu	1	0	8	0
hadir	2	0	0	0
malam	1	1	0	0
tidur	1	0	0	0
tahu	1	0	0	0
jadi	1	4	0	0
pada	1	6	4	0
kian	1	0	0	0
alami	1	0	0	0
pedih	1	0	0	0
putus	1	0	0	0
cinta	9	11	6	0
mulai	1	0	0	0
biasa	1	0	0	0
hidup	1	0	3	0
asmara	1	0	0	0
bawa	1	10	1	0
sembuh	1	0	0	0
luka	1	0	0	0
beda	1	0	0	0

kira	1	0	0	0
jatuh	5	7	0	0
sungguh	7	0	0	0
ragu	2	0	0	0
tetap	2	0	0	2
tunggu	2	0	0	0
coba	4	0	3	0
dengar	2	0	3	0
kata	2	0	0	0
dam	2	0	0	0
percaya	2	0	1	0
buat	2	0	0	0
daya	2	0	0	0
ungkap	2	0	0	0
sungguh	1	0	0	0
kadang	2	0	0	0
cemburu	1	0	0	0
gelisah	1	0	0	0
lalu	1	0	0	0
hari	1	1	0	0
pungkiri	1	0	0	0
hati	1	1	1	0
dalam	1	0	0	0
betapa	1	0	0	0

Sumber: Perancangan

Untuk memudahkan perhitungan, maka hal pertama yang harus kita hitung pada tiap kata adalah bobot kata yang menggunakan sebagian rumus dari rumus

utama, yaitu $\sum_n f_{ni} \log \left(\frac{1+F_{nc}}{N+\sum_{x=1}^N F_{xc}} \right)$

- Kategori cinta :

$$Pr = 2/5 = 0.4$$

$$N = 120$$

$$F_{xc} = 142$$

1. Rasa

$$f_{ni} = 2$$

$$F_{nc} = 3$$

$$4 \times \log \left(\frac{1+3}{120+142} \right) = -7.265$$

2. Rindu

$$f_{ni} = 8$$

$$F_{nc} = 0$$

$$8 \times \log\left(\frac{1+0}{120+142}\right) = -2.418$$

3. Malam

$$f_{ni} = 1$$

$$F_{nc} = 3$$

$$1 \times \log\frac{1+1}{120+142} = -2.117$$

- Kategori religi :

$$Pr = 2/5 = 0.4$$

$$F_{xc} = 133$$

1. Rasa

$$f_{ni} = 1$$

$$F_{nc} = 0$$

$$1 \times \log\frac{1+0}{120+133} = -9.612$$

2. Rindu

$$f_{ni} = 1$$

$$F_{nc} = 8$$

$$1 \times \log\frac{1+8}{120+133} = -1.449$$

3. Malam

$$f_{ni} = 1$$

$$F_{xc} = 1$$

$$1 \times \log\frac{1+1}{120+133} = -2.403$$

- Kategori perjuangan :

$$Pr = 1/5 = 0.2$$

$$F_{xc} = 45$$



1. Rasa

$$f_{ni} = 1$$

$$F_{nc} = 0$$

$$4 \times \log \frac{1 + 0}{120 + 45} = -8.870$$

2. Rindu

$$f_{ni} = 1$$

$$F_{nc} = 0$$

$$1 \times \log \frac{1 + 0}{120 + 45} = -2.217$$

3. Malam

$$f_{ni} = 1$$

$$F_{nc} = 0$$

$$1 \times \log \frac{1 + 0}{120 + 45} = -2.217$$

Perhitungan di atas dilakukan pada seluruh kata pada dokumen uji yang terdapat pada dokumen latih. Untuk lebih lengkapnya dapat dilihat pada tabel berikut:

Tabel 3.7 Tabel perhitungan bobot kata

Kata	Kategori 1	Kategori 2	Kategori 3
awal	-2.418	-2.403	-2.217
erti	-2.418	-2.403	-2.217
rasa	-7.265	-9.612	-8.870
ubah	-2.418	-2.403	-2.217
rindu	-2.418	-1.449	-2.217
hadir	-4.837	-4.806	-4.435
malam	-2.117	-2.403	-2.217
tidur	-2.418	-2.403	-2.217
tahu	-2.418	-2.403	-2.217
jadi	-1.719	-2.403	-2.217
pada	-1.573	-1.704	-2.217
kian	-2.418	-2.403	-2.217
alami	-2.418	-2.403	-2.217
pedih	-2.418	-2.403	-2.217
putus	-2.418	-2.403	-2.217
cinta	-12.052	-14.022	-19.957
mulai	-2.418	-2.403	-2.217
biasa	-2.418	-2.403	-2.217



hidup	-2.418	-1.801	-2.217
asmara	-2.418	-2.403	-2.217
bawa	-1.377	-2.102	-2.217
sembuh	-2.418	-2.403	-2.217
luka	-2.418	-2.403	-2.217
beda	-2.418	-2.403	-2.217
kira	-2.418	-2.403	-2.217
jatuh	-7.576	-12.016	-11.087
sungguh	-16.928	-16.822	-15.522
ragu	-4.837	-4.806	-4.435
tetap	-4.837	-4.806	-3.481
tunggu	-4.837	-4.806	-4.435
coba	-9.673	-7.204	-8.870
dengar	-4.837	-3.602	-4.435
kata	-4.837	-4.806	-4.435
dam	-4.837	-4.806	-4.435
percaya	-4.837	-4.204	-4.435
buat	-4.837	-4.806	-4.435
daya	-4.837	-4.806	-4.435
ungkap	-4.837	-4.806	-4.435
sungguh	-2.418	-2.403	-2.217
kadang	-4.837	-4.806	-4.435
cemburu	-2.418	-2.403	-2.217
gelisah	-2.418	-2.403	-2.217
lalu	-2.418	-2.403	-2.217
hari	-2.117	-2.403	-2.217
pungkiri	-2.418	-2.403	-2.217
hati	-2.117	-2.102	-2.217
dalam	-2.418	-2.403	-2.217
betapa	-2.418	-2.403	-2.217
Total	-183.410	-187.582	-186.013

Sumber: Perancangan

Setelah menghitung bobot tiap kata, kemudian dihitung bobot untuk tiap kelas dengan menggunakan rumus berikut ini :

$$class(t_i) = argmax_c \left[\log(Pr(c)) + \sum_n (f_{ni}w_{nc}) \right]$$

- Kategori Cinta

$$Pr = 2/5 = 0.4$$

$$class(t_i) = argmax_c [\log(0,4) + (-183.410)] = -183.012$$

- Kategori Religi

$$Pr = 2/5 = 0.4$$

$$class(t_i) = argmax_c [\log(0,4) + (-187.582)] = -187.185$$

- Kategori Perjuangan

$$Pr = 1/5 = 0.2$$

$$class(t_i) = argmax_c [\log(0,2) + (-186.013)] = -186.014$$

Dari ketiga nilai diatas, nilai tertinggi yaitu -183.012 merupakan nilai dari kategori cinta, sehingga dapat disimpulkan data uji tersebut terklasifikasi sebagai lagu bertema cinta.

3.9 Implementasi

Implementasi aplikasi klasifikasi ini dilakukan dengan mengacu pada perancangan sistem. Implementasi perangkat lunak dilakukan menggunakan bahasa pemrograman PHP dan database MySQL.

Implementasi aplikasi ini meliputi:

- Pembuatan *user interface* berupa halaman-halaman web yang menerima input dari user.
- Melakukan proses *preprocessing* pada teks yang di masukkan oleh pengguna.
- Menentukan klasifikasi pada lirik lagu
- Menampilkan hasil dari analisis klasifikasi.

3.10 Pengujian

Pengujian perangkat lunak pada skripsi ini dilakukan agar dapat menunjukkan bahwa perangkat lunak telah mampu bekerja sesuai dengan spesifikasi dari kebutuhan yang melandasinya. Pengujian yang dilakukan meliputi:

- Pengujian hasil program dengan cara membandingkan output program dengan output manual berdasarkan analisis pakar
- Pengukuran tingkat akurasi program yang akan di hitung menurut presisi, *recall*, dan *F-measure*.

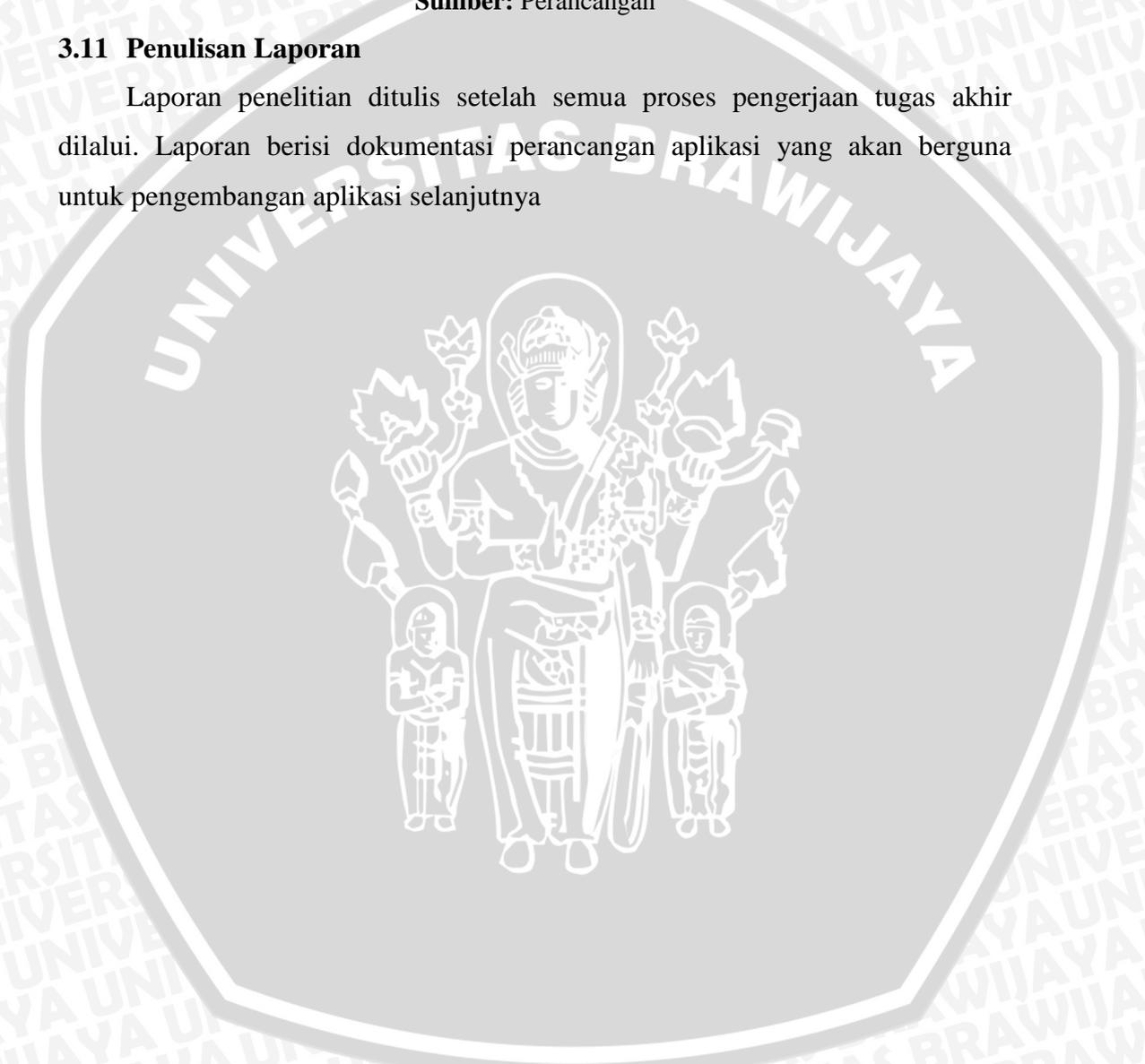
Tabel 3.8 Contoh tabel pengujian

Jumlah Dokumen Latih	Jumlah Dokumen Uji	akurasi		
		presisi	<i>recall</i>	<i>F-measure</i>

Sumber: Perancangan

3.11 Penulisan Laporan

Laporan penelitian ditulis setelah semua proses pengerjaan tugas akhir dilalui. Laporan berisi dokumentasi perancangan aplikasi yang akan berguna untuk pengembangan aplikasi selanjutnya



BAB IV IMPLEMENTASI

Bab ini membahas implementasi perangkat lunak berdasarkan hasil yang telah diperoleh dari analisis kebutuhan dan proses perancangan perangkat lunak yang dibuat pada bab III. Bab ini terdiri dari penjelasan tentang spesifikasi sistem, batasan-batasan dalam implementasi, implementasi algoritma pada program, implementasi antarmuka, dan implementasi metode.

4.1 Spesifikasi Sistem

Hasil analisis kebutuhan dan perancangan perangkat lunak yang telah diuraikan pada bab III menjadi acuan untuk melakukan implementasi menjadi aplikasi yang dapat berfungsi sesuai dengan kebutuhan. Spesifikasi aplikasi diimplementasikan pada spesifikasi perangkat keras dan perangkat lunak.

4.1.1 Spesifikasi Perangkat Keras

Pengembangan aplikasi klasifikasi tema pada lirik lagu ini menggunakan sebuah komputer dengan spesifikasi perangkat keras yang dijelaskan pada Tabel 4.1.

Tabel 4.1 Spesifikasi perangkat keras komputer

Nama Komponen	Spesifikasi
Prosesor	Intel(R) Core(TM) i3 CPU M 330 (2.13 GHz, 800 MHz, 1MB L2 cache)
Memori(RAM)	2GB DDR3
Hardisk	58.8 GB HDD
Kartu Grafis	ATI Mobility Radeon HD 4300 Series 1238 MB

Sumber: Pengujian

4.1.2 Spesifikasi Perangkat Lunak

Pengembangan aplikasi klasifikasi pada tema lagu ini menggunakan perangkat lunak dengan spesifikasi yang dijelaskan pada Tabel 4.2.

Tabel 4.2 Spesifikasi perangkat lunak

Nama	Spesifikasi
Sistem Operasi	Intel(R) Core(TM) i3 CPU M 330 (2.13 GHz, 800 MHz, 1MB L2 cache)
Bahasa Pemrograman	HTML dan PHP 5.3.1
<i>Tools</i> pemrograman	Aptana Studio 3 versi 3.3.2.201302081546 - 08022013154827
<i>Server localhost</i>	XAMPP versi 1.7.3
DBMS	MySQL versi 5.1.44
<i>Web Browser</i>	Google Chrome versi 26.0.1410.65

Sumber: Pengujian

4.2 Batasan-batasan Implementasi

Batasan implementasi adalah batasan proses yang dapat dilakukan sistem sesuai dengan perancangan awal sistem. Batasan implementasi ditampilkan agar penelitian ini memiliki ruang lingkup yang jelas dalam mengimplementasikan sistem. Beberapa batasan dalam mengimplementasikan aplikasi klasifikasi tema pada lirik lagu ini adalah sebagai berikut:

- Klasifikasi tema pada lirik lagu dirancang dan dijalankan menggunakan aplikasi web
- Metode penyelesaian masalah yang digunakan adalah *Transformed Weight-Normalized Complement Naive Bayes* (TWCNB).
- Dokumen yang digunakan sebagai data latih dan data uji adalah berasal dari CD/VCD *official* yang dapat dilihat pada situs www.disctarra.com dan dari beberapa situs website di internet.
- Dokumen yang digunakan merupakan dokumen berbahasa Indonesia.
- Output yang dikeluarkan berupa hasil klasifikasi yaitu lirik lagu bertema cinta, religi dan perjuangan.
- Penentuan tema lirik lagu berdasarkan pada frekuensi kemunculan kata, bukan pada struktur semantik.

4.3 Implementasi Algoritma

Aplikasi klasifikasi tema pada lirik lagu ini terdiri dari beberapa proses utama yaitu preprocessing dan klasifikasi. Data berupa lirik lagu akan diolah dengan preprocessing hingga didapatkan kata dasar unik yang disebut keywords,

kemudian dihitung bobot per kategori dan diklasifikasi temanya dengan menggunakan *Transformed Weight-Normalized Complement Naive Bayes* (TWCNB). Daftar fungsi pada aplikasi klasifikasi tema pada lirik lagu ini dapat dilihat pada Tabel 4.3.

Tabel 4.3 Daftar Fungsi dan Variabel Pada Sistem

No.	Proses	Fungsi/Variabel	Keterangan
1.	Preprocessing	function tokenizing()	Fungsi untuk menghilangkan tanda baca seperti koma, titik, petik dua, dan lain-lain, mengubah semua huruf menjadi huruf kecil, dan memecah kalimat menjadi kata.
		function stopword_removal()	Fungsi untuk menghilangkan kata yang tidak penting seperti 'yang', 'adalah', 'itu', dan lain – lain.
		function stemmingArifin()	Fungsi untuk memecah kata menjadi kata dasar
2.	TWCNB	Variabel weight1, weight2, weight3	Variabel untuk menyimpan bobot dari setiap kata berdasarkan tiap kategori
		Variabel fin1, fin2, fin3	Variabel untuk menyimpan hasil perhitungan klasifikasi dari tiap kategori

Sumber: Implementasi

4.3.1 Proses Preprocessing

Tahap ini terdiri dari pembersihan dokumen, *tokenizing*, *stopword removal*, dan *stemming*. Pembersihan dokumen berfungsi untuk menghilangkan tanda baca pada dokumen, *tokenizing* berfungsi untuk mengubah semua huruf menjadi huruf kecil dan memecah kalimat menjadi kata yg berdiri sendiri, *Stopword removal* berguna untuk menghilangkan kata-kata yang tidak penting dan mengambil kata penting menjadi *keyword*, dan *stemming* mengubah kata penting pada dokumen agar menjadi bentuk kata dasar. *Stemming* yang digunakan adalah *stemming arifin-setiono*.

Pemanggilan fungsi *text preprocessing* secara keseluruhan ditunjukkan pada

Sourcecode 4.1.

```

1 // cleansing, casefolding, tokenizing
2 $hasil_tok = tokenizing($kata);
3 // stopword removal
4 $hasil_stop = stopword_removal($hasil_tok);
5 // stemming
6 $hasil_stem = stemming($hasil_stop[$k], $kamus_kata);

```

Sourcecode 4.1 Implementasi Algoritma Preprocessing pada Dokumen

Sumber: Implementasi

4.3.1.1 Proses Tokenizing

Proses tokenizing mengubah semua huruf dalam dokumen menjadi huruf kecil, menghilangkan karakter yang tidak dibutuhkan seperti titik, koma, tanda seru dan sebagainya, dan mengubah kalimat menjadi kata.

```

1 function tokenizing($kata){
2     $kata= strtolower($kata);
3     $kata = preg_replace('/([^\a-z0-9]+)/i', ' ', $kata);
4     return $kata;
5 }

```

Sourcecode 4.2 Implementasi Algoritma *Tokenizing*

Sumber: Implementasi

4.3.1.2 Proses Stopword Removal

Tahap pertama proses *Stopword Removal* adalah mengambil tiap kata pada dokumen kemudian dibandingkan dengan kamus *stopword* pada *database*, jika ditemukan kata tidak penting maka akan dihapus, sehingga nantinya akan menghasilkan kata unik yang disebut sebagai *keyword*.

```

1 function stopword_removal($array){
2
3     $sql = "select * from stopword";
4     $sqlExe = mysql_query($sql);
5     $kamus_kata= array();
6
7     while($baris = mysql_fetch_array($sqlExe)){
8         $kamus_kata[] = $baris[0];
9     }
10
11     foreach ($array as $key => $data) {
12         if (in_array($data, $kamus_kata)){
13             unset($array[$key]);

```

```
14     }
15     }
16
17     return array_values(array_filter($array));
18 }
```

Sourcecode 4.3 Implementasi Algoritma Stopword Removal

Sumber: Implementasi

4.3.1.3 Proses Stemming

Setelah dilakukan proses *stopword_removal*, proses yang dilakukan selanjutnya adalah *stemming*. *Stemming* adalah suatu proses untuk mengubah bentuk kata menjadi kata dasar. *Stemming* yang digunakan dalam skripsi ini adalah *stemming* Arifin-Setiono. *Stemming* ini berdasarkan penelitian yang dilakukan oleh Arifin-Setiono.

4.3.2 Implementasi Metode TWCNB

```
1  $stem = explode(" ", $in_stem);
2  $b= array_count_values($stem);
3  $sql = "select kata from daftar_kata";
4
5  $sqlExe = mysql_query($sql) or die(mysql_error());
6  $tabel_kata= array();
7
8  while($baris = mysql_fetch_array($sqlExe)){
9
10     $tabel_kata[] = $baris[0];
11
12     foreach ($b as $key => $value) {
13         if (in_array($key, $tabel_kata)){
14             $sql1 = "UPDATE daftar_kata SET jumlah=jumlah+'$value'
15             WHERE kata='$key'";
16
17             $exe1=mysql_query($sql1);
18         }
19         else
20         {
21             $sql1 ="INSERT INTO daftar_kata
22             VALUES ('','$key','$value' )";
23
24             $exe1=mysql_query($sql1);
25         }
26
27     $idkata = mysql_result(mysql_query("SELECT id_kata FROM
28     daftar_kata where kata='$key'"),0);
29
30     $idlirik = mysql_result(mysql_query("SELECT
31     count(id_lirik)FROM lirik"),0);
```

```

29 $sql2 ="INSERT INTO proses_kata
VALUES ('','$idkata','$id lirik','$kat' , '$value')";
$exe2=mysql_query($sql2);
}

```

Sourcecode 4.4 Implementasi TWCNB Pada Data Latih

Sumber: Implementasi

Pada proses data latih, jumlah kata dari tiap kategori lagu akan dihitung dan dimasukkan kedalam database. Sebelumnya kata akan di cek terlebih dahulu ada atau tidak di dalam databaase, yaitu pada baris 1-20. Untuk kata yang baru, akan ditambahkan ke field baru pada tabel, sedangkan apabila kata nya sudah ada, maka jumlah nya ditambahkan. Jumlah kata ini kemudian akan digunakan dalam perhitungan metode TWCNB pada pengklasifikasian data uji.

```

1 $fxc1 = mysql_result(mysql_query("SELECT sum(jumlah) FROM
2 proses_kata where id_kat=1"),0);
3 $fxc2 = mysql_result(mysql_query("SELECT sum(jumlah) FROM
4 proses_kata where id_kat=2"),0);
5 $fxc3 = mysql_result(mysql_query("SELECT sum(jumlah) FROM
6 proses_kata where id_kat=3"),0);
7 $n = mysql_result(mysql_query("SELECT count(id_kata) FROM
8 daftar_kata"),0);
9
10 foreach ($b as $key => $value) {
11     foreach ($b as $key => $value) {
12         if (in_array($key, $tabel_kata)){
13             $idkata_u = mysql_result(mysql_query("SELECT id_kata
14 FROM daftar_kata where kata='$key'"),0);
15             $fni = $value;
16             $fnc1= mysql_result(mysql_query("SELECT sum(jumlah)
17 FROM proses_kata where id_kata='$idkata_u' and
18 id_kat=1"),0);
19             $fnc2= mysql_result(mysql_query("SELECT sum(jumlah)
20 FROM proses_kata where id_kata='$idkata_u' and
21 id_kat=2"),0);
22             $fnc3= mysql_result(mysql_query("SELECT sum(jumlah)
23 FROM proses_kata where id_kata='$idkata_u' and
24 id_kat=3"),0);
25
26             $weight1 = $fni*(log10((1+$fnc1)/($n+$fxc1)));
27             $weight2 = $fni*(log10((1+$fnc2)/($n+$fxc2)));
28             $weight3 = $fni*(log10((1+$fnc3)/($n+$fxc3)));
29             $we1[] = $weight1;
30             $we2[] = $weight2;
31             $we3[] = $weight3;

```

```
32     }
33
34     else {
35         $fni = $value;
36         $fnc=0;
37
38         $weight1 = $fni*(log10((1+$fnc)/($n+$fxc1)));
39         $weight2 = $fni*(log10((1+$fnc)/($n+$fxc2)));
40         $weight3 = $fni*(log10((1+$fnc)/($n+$fxc3)));
41         $we1[] = $weight1;
42         $we2[] = $weight2;
43         $we3[] = $weight3;
44     }
45 }
```

Sourcecode 4.5 Implementasi Metode TWCNB Pada Data Uji

Sumber: Implementasi

Untuk perhitungan dengan metode TWCNB setiap kata pada dokumen uji akan di cek keberadaannya pada dokumen latih. Kata – kata yang telah disaring kemudian dihitung bobotnya dengan menggunakan rumus TWCNB yang terdapat pada baris 26-31. bobot untuk tiap kata tersebut disimpan di dalam array \$we1, \$we2, dan \$we3 yang nantinya akan dijumlah keseluruhannya. Untuk perhitungan rumus dibutuhkan jumlah kemunculannya pada setiap dokumen latih, dapat dilihat pada baris no 1-8 dan 13-24 yaitu pencarian data pada database untuk dapat memenuhi syarat perhitungan rumus TWCNB. Kemudian pada rumus terakhir bobot per kategori dihitung kemudian dibandingkan. Nilai tertinggi pada suatu kategori menentukan jenis kategori pada dokumen uji tersebut.

4.4 Implementasi Antarmuka

Antarmuka klasifikasi tema pada lirik lagu ini digunakan oleh pengguna untuk berinteraksi dengan program.

4.4.1 Tampilan Halaman Data Training

Tampilan halaman data training akan menampilkan daftar data latih yang digunakan pada aplikasi ini. Gambar halaman data latih ditunjukkan pada Gambar 4.1.



klasifikasi tema pada lirik lagu

No.	Artis	Judul	Kategori
1.	Istid	lentang rasa	cinta
2.	ronald gustav	i say i love you	cinta
3.	Yovie & Nuno	manusia biasa	cinta
4.	Sheila On 7	Hari Bersamanya	cinta
5.	Judika	Setengah Mati Merindu	cinta
6.	Terry	Hanusna Kau Plain Aku	cinta
7.	Vania	Rahasia	cinta
8.	Elo	Benci Tapi Rindu	cinta
9.	Lala Karmela	Hasrat Cinta	cinta
10.	Glenn Fredy	Pada Satu Cinta	cinta
11.	Pinkan Mambo	Tak Mungkin Bersatu	cinta
12.	Cokelat	Drama	cinta
13.	Dieby romero	Kuasa Cinta	cinta
14.	Warna	Kisah Terindah	cinta
15.	Judika	Cinta Salukan Kita	cinta
16.	Nindy	Katakan Sejajarnya	cinta
17.	Audy	Bisa Saja	cinta
18.	Lala Karmela	Kamu Aku Cinta	cinta
19.	Yovie & Nuno	Merindu Lagi	cinta
20.	Nyawa	Ku Yakin Bisa	cinta
21.	Terry	Tepatnya Malam Minggu	cinta
22.	Elo	Takkan Ada Aku Lagi	cinta
23.	Tianga	O. Teganya	cinta
24.	Sheila On 7	Mari Ber cerita	cinta
25.	Rio Febrian	Wanita	cinta
26.	Cokelat	Bukan Hari Ini	cinta

Gambar 4.1 Halaman Data Training

Sumber: Implementasi

4.4.2 Tampilan Halaman Klasifikasi

Halaman klasifikasi akan menampilkan form untuk memasukkan data-data yang berkaitan dengan lirik lagu. Pengguna dapat memilih jenis dokumen dengan memberi cek pada radio button pada form tersebut. Halaman klasifikasi ditunjukkan pada Gambar 4.2.

klasifikasi tema pada lirik lagu

DAFTAR LAGU

TRAINING & UJI

Judul :

artis :

lirik :

Lath :

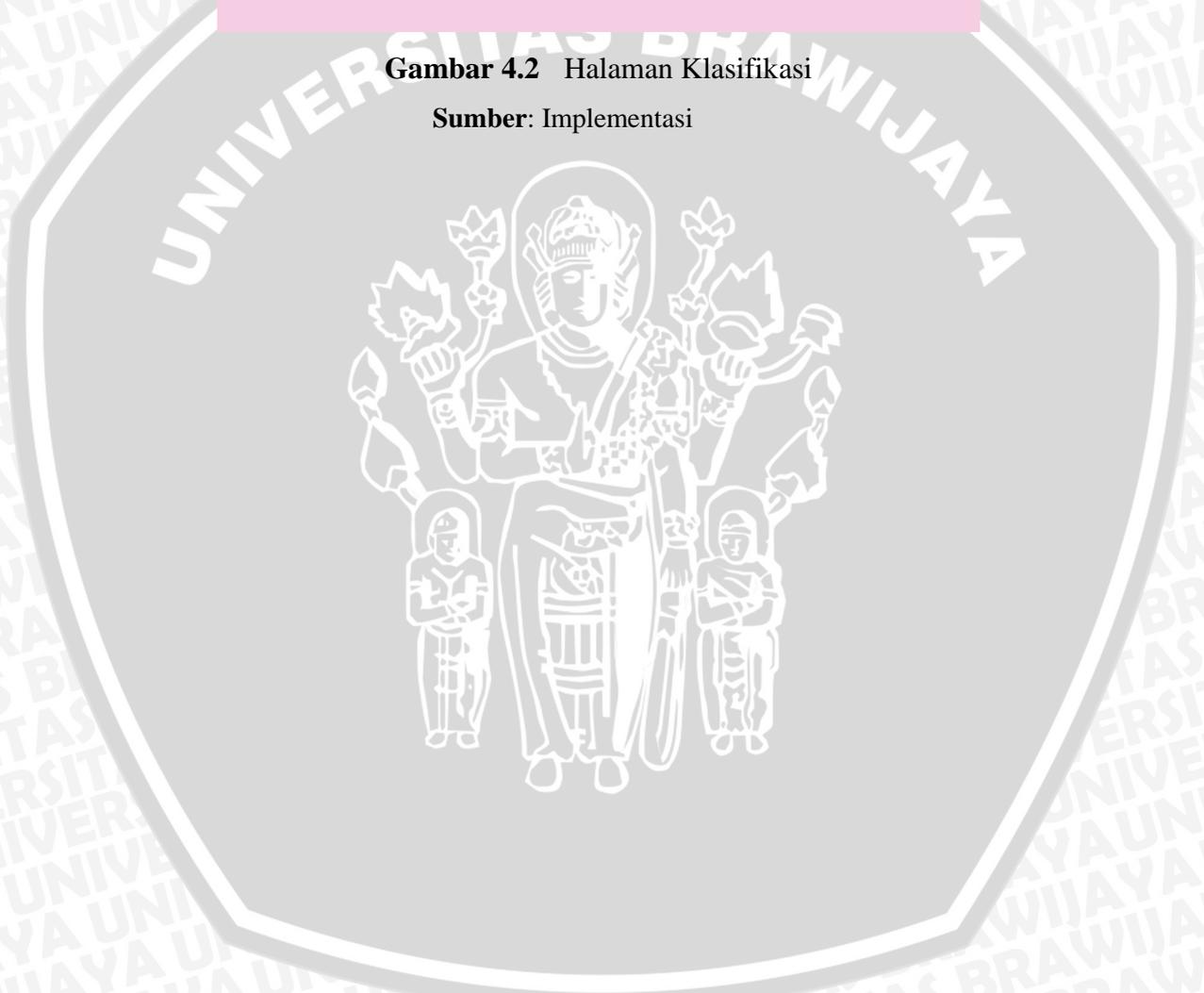
Uji

Submit

SKRIPSI | 2013

Gambar 4.2 Halaman Klasifikasi

Sumber: Implementasi



BAB V

PENGUJIAN DAN ANALISIS

Pada bab ini akan dipaparkan tentang pengujian terhadap hasil pengkategorian sistem dengan pengkategorian secara *manual*, pengujian efektifitas terhadap sistem ini akan menggunakan metode *Precision*, *Recall*, dan *F-measure* seperti yang telah dijelaskan pada bab II.

5.1 *Precision*, *Recall*, dan *F-measure*

Untuk mempelajari pengaruh jumlah data latih terhadap efektivitas sistem klasifikasi maka dilakukan pengujian dengan 4 skenario. Masing-masing skenario pengujian memiliki jumlah data latih yang berbeda dengan total 224 data dan menggunakan 29 data uji. Pengujian ini akan menguji lirik lagu berbahasa indonesia, untuk skenario lebih jelasnya akan dijelaskan pada tabel berikut:

Tabel 5.1 Skenario Pengujian

Skenario	Data Latih				Data Uji			
	Cinta	Religi	Perjuangan	Jumlah	Cinta	Religi	Perjuangan	Jumlah
1	35	36	29	100	10	10	9	29
2	44	42	34	160	10	10	9	29
3	73	73	44	190	10	10	9	29
4	90	90	44	224	10	10	9	29

Sumber: Pengujian

Pada tabel 5.1 diperlihatkan beberapa skenario pada pengujian sistem. Kategori cinta memiliki 10 data uji, kategori religi memiliki 10 data uji, dan kategori perjuangan memiliki 9 data uji. Untuk data latih memiliki perbedaan proporsi jumlah yang digunakan dalam melakukan pengklasifikasian.

Hasil dari pengujian semua skenario menghasilkan beberapa klasifikasi yang akan dibandingkan dengan data sebenarnya yang telah terklasifikasi. Pada tabel 5.2 diperlihatkan keseluruhan pengujian pada tiap skenario. Untuk data uji 1-10 merupakan data uji yang berkategori cinta(C), data uji 11-20 merupakan kategori religi(R) dan data uji 21-29 merupakan kategori perjuangan(P).

Tabel 5.2 Hasil Pengujian oleh Sistem

Data Uji	Skenario			
	1	2	3	4
1	C	C	C	C
2	C	C	C	C
3	C	C	C	C
4	C	C	C	C
5	C	C	C	C
6	C	R	C	C
7	C	C	C	C
8	C	C	C	C
9	C	C	C	C
10	C	C	C	C
11	C	C	C	C
12	R	R	R	R
13	R	R	R	R
14	R	R	R	R
15	R	R	R	R
16	R	R	R	R
17	R	R	R	R
18	R	R	R	R
19	R	R	R	R
20	R	R	R	R
21	P	P	P	P
22	P	P	P	P
23	P	P	P	P
24	P	P	P	P
25	P	P	P	P
26	P	P	P	P
27	C	C	C	C
28	C	C	C	C
29	R	R	R	R

Sumber: Pengujian

Dibawah ini merupakan hasil perhitungan *precision*, *recall*, dan *F-measure* yang dilakukan pada setiap skenario dan kategori. Perhitungan menggunakan rumus *precision* (2.14), *recall* (2.15), dan *F-measure* (2.16) yang telah tertulis pada bab II.

Tabel 5.3 *Precision, Recall* dan *F-measure* pada Skenario 1

Kategori	TP	FP	FN	<i>Precision</i>	<i>Recall</i>	<i>F-measure</i>
cinta	10	3	0	0,769	1	0,87
religi	9	1	1	0,9	0,9	0,9
perjuangan	6	0	3	1	0,67	0,8

Sumber: Pengujian

Tabel 5.4 *Precision, Recall* dan *F-measure* pada Skenario 2

Kategori	TP	FP	FN	<i>Precision</i>	<i>Recall</i>	<i>F-measure</i>
cinta	9	2	1	0,82	0,9	0,857
religi	9	2	1	0,82	0,9	0,857
perjuangan	6	0	3	1	0,67	0,8

Sumber: Pengujian

Tabel 5.5 *Precision, Recall* dan *F-measure* pada Skenario 3

Kategori	TP	FP	FN	<i>Precision</i>	<i>Recall</i>	<i>F-measure</i>
cinta	10	3	0	0,769	1	0,87
religi	9	1	1	0,9	0,9	0,9
perjuangan	6	0	3	1	0,67	0,8

Sumber: Pengujian

Tabel 5.6 *Precision, Recall* dan *F-measure* pada Skenario 4

Kategori	TP	FP	FN	<i>Precision</i>	<i>Recall</i>	<i>F-measure</i>
cinta	10	3	0	0,769	1	0,87
religi	9	1	1	0,9	0,9	0,9
perjuangan	6	0	3	1	0,67	0,8

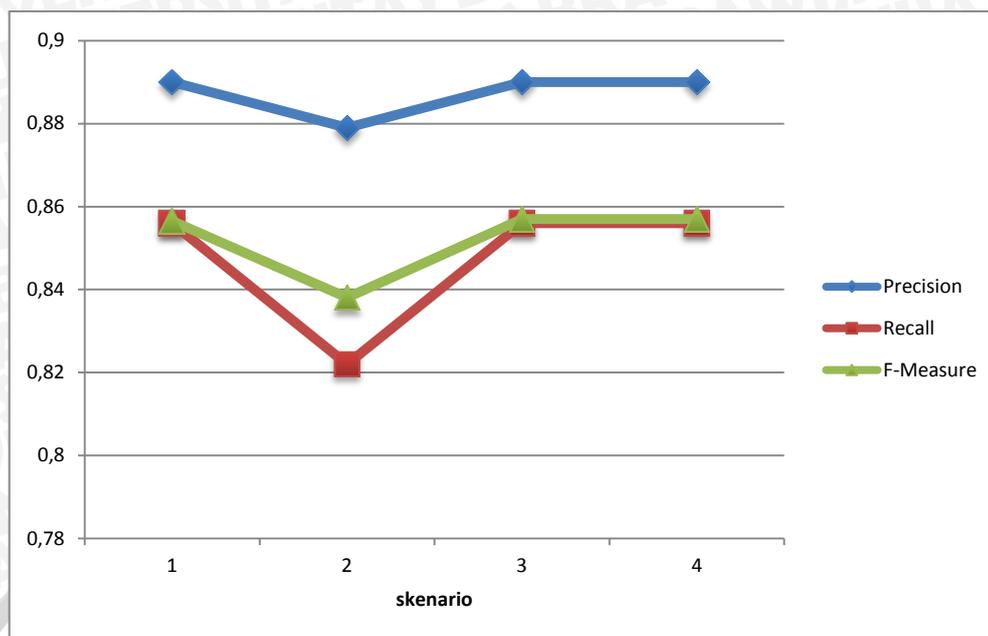
Sumber: Pengujian

Dari hasil pengujian ke 4 skenario maka didapatkan rata-rata untuk *precision, recall*, dan *F-measure*. Rata – rata *precision* terbaik didapatkan pada skenario 4 dengan 224 data latih dan 29 data uji.

Tabel 5.7 Rata - rata *Precision, Recall* dan *F-measure*

skenario	<i>precision</i>	<i>recall</i>	<i>f-measure</i>
1	0,890	0,856	0,857
2	0,879	0,822	0,838
3	0,890	0,856	0,857
4	0,890	0,856	0,857

Sumber: Pengujian



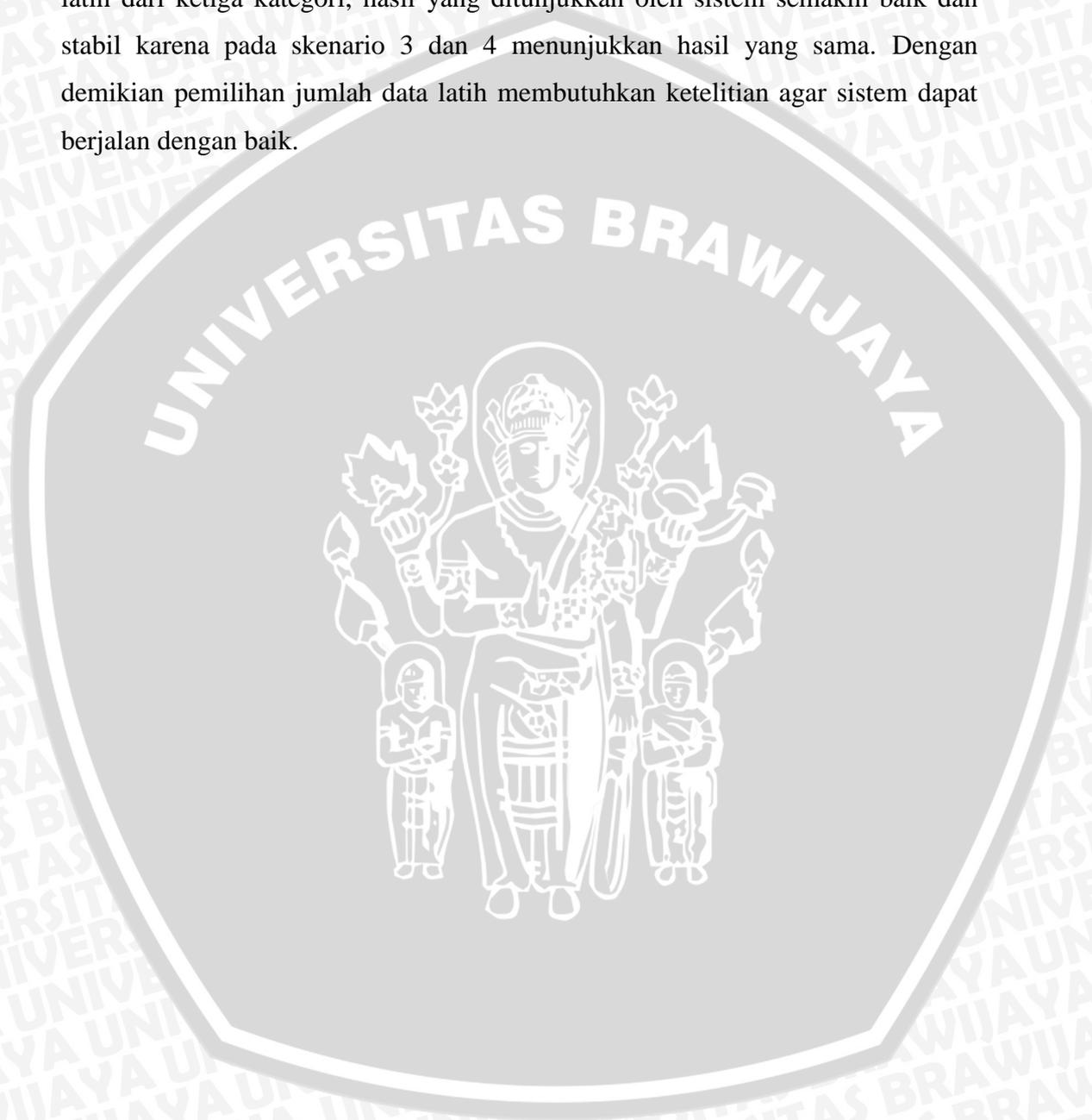
Gambar 5.1 Grafik Rata-Rata Hasil Pengujian

Sumber: Pengujian

5.2 Analisis Hasil Pengujian

Dari hasil pengujian, didapatkan jumlah kata pada data uji dan data latih mempengaruhi hasil presisi, *recall*, dan *F-measure* nya pada 29 data uji. Secara keseluruhan, dapat disimpulkan jumlah data latih pada skenario 3 dan 4 menunjukkan kestabilan nilai dari *precision*, *recall* dan *F-measure*. Pada skenario ke 2 terdapat penurunan pada nilai akurasi. Ini diakibatkan salah satu data uji mengalami kesalahan dalam klasifikasi. Setelah dianalisis terlihat jumlah kata pada data uji tersebut pada data latih, lebih banyak pada kategori religi daripada kategori cinta ataupun perjuangan. Sehingga data uji yang seharusnya berada di kategori cinta, menjadi kategori lagu religi. Pada skenario 3 dan 4 jumlah kata pada data uji tersebut lebih banyak ditemukan pada kategori cinta, sehingga hasilnya menjadi benar. Nilai *precision* terbaik didapat pada skenario 4 dengan nilai 0,856, sedangkan *precision* terburuk didapat pada skenario 2 dengan nilai 0,822. Tingkat ketepatan untuk tiap data uji menunjukkan nilai yang baik, dilihat dengan nilai *precision* yang tidak jauh berbeda antara skenario 1 dengan data latih terkecil sampai skenario 4 dengan data latih terbesar. Untuk *recall* nilai terbaik di dapat pada skenario 4 yaitu 0,911, sedangkan nilai *recall* terburuk berada pada skenario 2 yaitu 0,878. Hal ini menunjukkan bahwa keberhasilan sistem dalam

pengklasifikasian telah cukup baik. Untuk *F-measure* nilai terbaik nya berada pada skenario 4 yaitu 0,8697 dan nilai terburuk nya berada pada skenario 2 dengan nilai 0,838. Dapat dilihat pada skenario 1 nilai data latih paling sedikit, yaitu 100 data, pada skenario 2 yaitu 160 data. Seiring bertambahnya jumlah data latih dari ketiga kategori, hasil yang ditunjukkan oleh sistem semakin baik dan stabil karena pada skenario 3 dan 4 menunjukkan hasil yang sama. Dengan demikian pemilihan jumlah data latih membutuhkan ketelitian agar sistem dapat berjalan dengan baik.



BAB VI PENUTUP

Bab ini akan membahas mengenai kesimpulan dan saran yang dapat diambil dari pembuatan sistem Klasifikasi Tema Pada Lirik Lagu Dengan Metode *Transformed Weight-Normalized Complement Naive Bayes* (TWCNB)

6.1 Kesimpulan

Kesimpulan dari hasil skripsi Klasifikasi Tema Pada Lirik Lagu Dengan Metode *Transformed Weight-Normalized Complement Naive Bayes* (TWCNB), sebagai berikut:

1. Metode *TWCNB* dapat diimplementasikan dengan baik pada pengklasifikasian teks berbahasa Indonesia dengan studi kasus tema pada lirik lagu, dengan cara melakukan proses preprocessing yang terdiri dari *tokenizing*, *stopword removal*, dan *stemming*. Kemudian menghitung prior dari pengujian terhadap setiap kelas dan diambil nilai terbesarnya.
2. Pengujian Klasifikasi Tema Pada Lirik Lagu Dengan Metode *Transformed Weight-Normalized Complement Naive Bayes* (TWCNB) menghasilkan rata-rata *Precision* (akurasi) sebesar 0,887, rata-rata *Recall* (sensitifitas atau kemampuan sistem memilih hasil yang sesuai) sebesar 0,847, dan rata-rata *F-measure* sebesar 0,852.
3. Jumlah data latihan berpengaruh terhadap baik atau tidaknya hasil dari pengklasifikasian sistem.

6.2 Saran

Saran dari hasil klasifikasi tema pada lirik lagu menggunakan metode *Transformed Weight-Normalized Complement Naive Bayes* (TWCNB), untuk pengembangan lebih lanjut dapat memperhitungkan secara proporsional jumlah kata pada data latihan dan data uji sehingga dapat dihasilkan sistem yang lebih baik lagi.

DAFTAR PUSTAKA

- [AHG-05] Hripcsak, George and S Rothschild, Adam. 2005. *Agreement, The F-Measure, and Reliability in Information Retrieval*. *J Am Med Inform Assoc* 2005
- [ARS-00] Arifin, Agus Zainal dan Ari Setiono, Novan. 2000. *Klasifikasi Dokumen Berita Kejadian Berbahasa Indonesia dengan Algoritma Single Pass Clustering*. Institut Teknologi Sepuluh November (ITS). Surabaya
- [DJO-09] Djohan. 2009. *Psikologi Musik*. Yogyakarta: Penerbit Best Publisher.
- [ELM-03] Eyheramendy, S., Lewis, D.D., Madigan, D.: On the naive bayes text categorization. In: Ninth International Workshop on Artificial Intelligence and Statistic. (2003) 3-6.
- [EZY-02] Even-Zohar, Y. 2002. *Introduction to Text Mining, Supercomputing*.
- [HAS-08] Sulastianto, Harry, dkk. 2008. *Seni Budaya: untuk Kelas XI Sekolah Menengah Atas*. Bandung: Grafindo Media Pratama.
- [ITA-10] Nurisani, Devita. 2010. *Tema, Topik, dan Judul*. Online pada <http://ita-kyu-kiyut.blogspot.com/2010/01/ragam-bahasa-tugas-kelompok.html> diakses 17 April 2013.
- [KAA-09] Kunaifi, Aang. 2009. *Klasifikasi Email Berbahasa Indonesia Menggunakan Text Mining Dan Algoritma Kmeans*. Surabaya: Politeknik Elektronika Negeri Surabaya.
- [KFP-04] Kibriya Ashraf M., Frank Eibe, Pfahringer Bernhard, Holmes Geoffrey . 2004. *Multinomial Naïve Bayes for Text Categorization Revisited*. Australian joint conference on artificial intelligence No 17.
- [MAN-98] McCallum, A. and Nigam, K., 1998, *A comparison of event models for Naïve Bayes text classification*, online pada <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.46.1529> diakses 29 September 2012

- [MCN-98] McCallum, A., Nigam, K.: A Comparison of event model for naive bayes text classification. Technical report, American Association for Artificial Intelligence Workshop on Learning for Text Categorization (1998)
- [PUT-08] Pendit, Putu. 2008. *Recall & Precision*, online pada <http://iperpin.wordpress.com/2008/03/27/recall-precision/> diakses 26 maret 2013
- [REJ-04] Rennie, J.: Personal communication regarding WCNB. (2004)
- [RIJ-79] Rijsbergen C J V. 1979. *Information Retrieval*. Butterworths: University of Glasgow.
- [RST-03] Rennie, J.D.M., Shih, L., Teevan, J., Karger, D.R.: Tackling the poor assumptions of Naive Bayes text classifiers. In: Proceedings of the Twentieth International Conference on Machine Learning, AAAI Press (2003) 616-623
- [SDA-12] Susanti, Aprilliana Dharma. 2012. *Klasifikasi Mood Pada Musik Berbahasa Inggris Dengan Menerapkan Metode Improved K-NN Berbasis Clustering*. Universitas Brawijaya, Malang.
- [TRI-09] Triawati, Chandra. 2009. *Metode Pembobotan Statistical Concept Based untuk Klastering dan Kategorisasi Dokumen Berbahasa Indonesia*. Institut Teknologi Telkom. Bandung
- [WAL-95] Walpole, E. R., Myers, R. H. (1995). *Ilmu Peluang dan Statistika untuk Insinyur dan Ilmuan, Edisi ke-4*. Bandung, ITB.