



TUGAS AKHIR - KS091336

# **KLASIFIKASI ARTIKEL BERITA BERBAHASA INDONESIA BERBASIS *NAÏVE BAYES CLASSIFIER* MENGUNAKAN *CONFIX-STRIPPING STEMMER***

AMALIA ANJANI ARIFIYANTI  
NRP 5210 100 106

Dosen Pembimbing  
Prof. Ir. Arif Djunaidy, M.Sc, Ph.D  
Renny Pradina Kusumawardani, S.T, M.T

JURUSAN SISTEM INFORMASI  
Fakultas Teknologi Informasi  
Institut Teknologi Sepuluh Nopember  
Surabaya 2014

**FINAL PROJECT - KS091336**

**NEWS ARTICLES CLASSIFICATION  
IN INDONESIAN LANGUAGE  
BASED ON NAIVE BAYES CLASSIFIER  
USING CONFIX-STRIPPING STEMMER**

**AMALIA ANJANI ARIFIYANTI  
NRP 5210 100 106**

**Supervisor  
Prof. Ir. Arif Djunaidy, M.Sc, Ph.D  
Renny Pradina Kusumawardani, S.T, M.T**

**DEPARTMENT OF INFORMATION SYSTEM  
Faculty of Information Technology  
Institute of Technology Sepuluh Nopember  
Surabaya 2014**

# **KLASIFIKASI ARTIKEL BERITA BERBAHASA INDONESIA BERBASIS NAÏVE BAYES CLASSIFIER MENGUNAKAN CONFIX-STRIPPING STEMMER**

**Nama Mahasiswa : Amalia Anjani Arifiyanti**  
**NRP : 5210 100 106**  
**Jurusan : Sistem Informasi FTIf-ITS**  
**Dosen Pembimbing : Prof. Ir. Arif Djunaidy, M.Sc, Ph.D**  
**Renny Pradina Kusumawardani, S.T., M.T**

## ***Abstrak***

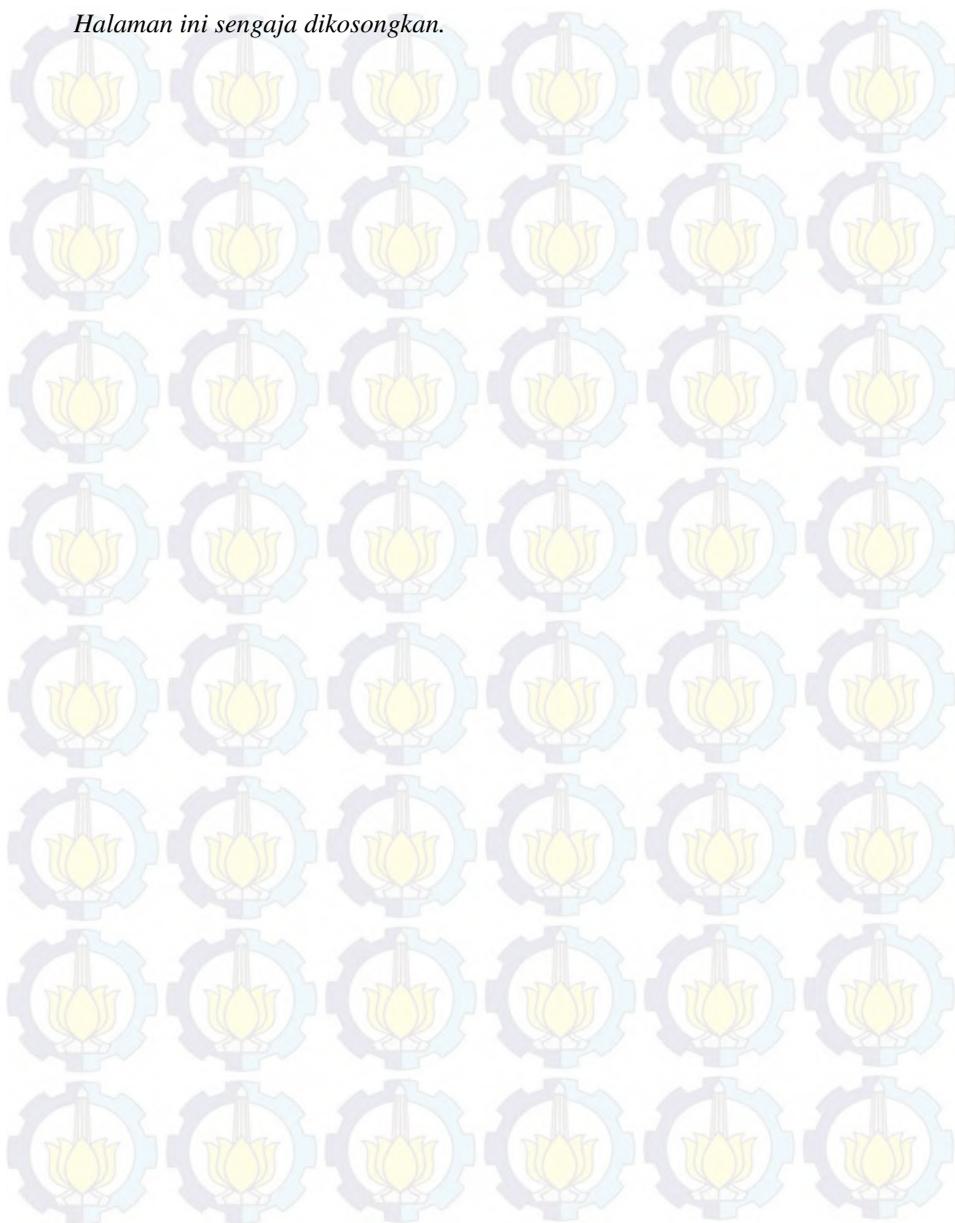
*Aliran berita yang dinamis dan mengalami perubahan yang cepat memerlukan pengklasifikasi artikel berita untuk mempermudah proses kategorisasi dokumen berita. Pada saat ini sudah banyak metode untuk melakukan kategorisasi berita, tetapi kebanyakan terbatas pada penggunaan korpus bahasa asing tertentu saja. Korpus berbahasa Indonesia memiliki tantangan tersendiri karena mempunyai susunan imbuhan yang berbeda dibandingkan dengan bahasa asing lainnya.*

*Dalam tugas akhir ini dibangun sebuah aplikasi pengklasifikasi artikel berita berbahasa Indonesia berbasis Naïve Bayes classifier dengan menggunakan confix-stripping stemmer. Sebelum proses klasifikasi dilakukan, praproses terhadap artikel berita dilakukan untuk menghapus stopwords dan memenggal artikel menjadi sejumlah token. Algoritma confix-stripping stemmer kemudian digunakan untuk mencari kata dasar dari setiap token yang dihasilkan. Aplikasi pengklasifikasi artikel berita berbasis Naïve Bayes classifier dibangun dengan menggunakan kata dasar yang diperoleh dari berbagai artikel berita yang mewakili semua kategori berita yang akan dikelola.*

*Aplikasi pengklasifikasi artikel berita yang telah berhasil dibangun dalam Tugas Akhir ini telah diujicobakan terhadap sejumlah sampel artikel berita yang terdiri dari 12 kategori. Hasil pengujian menunjukkan bahwa aplikasi yang dibangun dalam lingkungan sistem operasi Windows mampu menghasilkan pengklasifikasi dengan akurasi, precision, recall dan f-measure berturut-turut sebesar 87.675%, 88.21%, 87.65%, dan 87.69%.*

***Kata Kunci:artikel berita, confix-stripping stemmer, klasifikasi teks, naïve bayes classifier, stemming bahasa Indonesia.***

*Halaman ini sengaja dikosongkan.*



# NEWS ARTICLES CLASSIFICATION IN INDONESIAN LANGUAGE BASED ON NAIVE BAYES CLASSIFIER USING CONFIX-STRIPPING STEMMER

**Name** : Amalia Anjani Arifiyanti  
**NRP** : 5210 100 106  
**Departement** : Information System FTIF-ITS  
**Supervisor** : Prof. Ir. Arif Djunaidy, M.Sc, Ph.D  
Renny Pradina Kusumawardani, S.T., M.T

## ***Abstract***

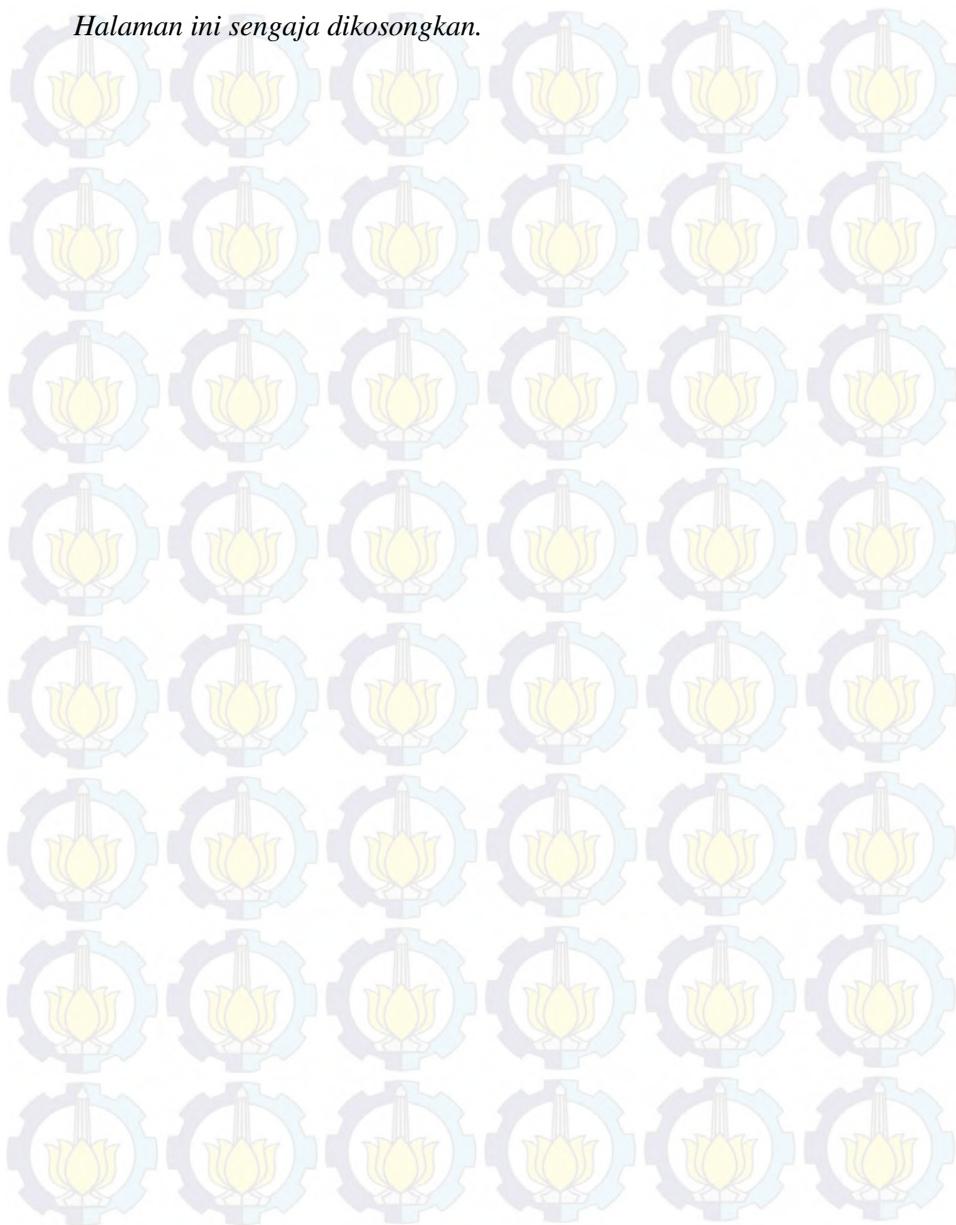
*Dynamic as well as rapid changing news stream needs news article classifier to ease the process of news document classification. Nowadays, there are already many methods for classifying news. Most of them, however, are limited only in the use of corpus for certain foreign languages. Indonesian language corpus has its own challenging section due to its different suffixes compared with other languages.*

*In this final project, an application for classifying news article in Indonesian language based on Naive Bayes classifier using confix-stripping stemmer is developed. Before classifying process is performed, a preprocessing step needs to be done to eliminate stopwords and split the article into a number of tokens. The confix-stripping stemmer algorithm is then applied to obtain the term corresponding to each token being produced. The news article classifier developed in this final project was modeled using all terms obtained from several news articles representing all news categories to be accommodated.*

*A news article classifier application that has been successfully developed in this final project was tested using a set of news articles consisting of 12 categories. The test results showed that the application, which is developed under the windows operating system, is capable of producing sufficiently good classifier with accuration, precision, recall, and f-measure of 87.675%, 88.21%, 87.65% and 87.69%, respectively.*

***Keywords:*** *confix-stripping stemmer, Indonesian language stemming, naïve bayes classifier, news articles, text classification.*

*Halaman ini sengaja dikosongkan.*



**KLASIFIKASI ARTIKEL BERITA BERBAHASA  
INDONESIA BERBASIS *NAÏVE BAYES CLASSIFIER*  
MENGUNAKAN *CONFIX-STRIPPING STEMMER***

**TUGAS AKHIR**

Disusun Untuk Memenuhi Salah Satu Syarat  
Memperoleh Gelar Sarjana Komputer  
pada

Jurusan Sistem Informasi  
Fakultas Teknologi Informasi  
Institut Teknologi Sepuluh Nopember

Oleh:

**AMALIA ANJANI ARIFIYANTI**  
**NRP. 5210 100 106**

Surabaya, Juli 2014

**Ketua Jurusan Sistem Informasi**



**Dr. Eng. FEBRILIYAN SAMOPA, S.Kom., M.Kom.**  
**NIP.19730219 199802 1 001**

**KLASIFIKASI ARTIKEL BERITA BERBAHASA  
INDONESIA BERBASIS *NAÏVE BAYES CLASSIFIER*  
MENGUNAKAN *CONFIX-STRIPPING STEMMER***

**TUGAS AKHIR**

Disusun Untuk Memenuhi Salah Satu Syarat  
Memperoleh Gelar Sarjana Komputer

pada

Jurusan Sistem Informasi  
Fakultas Teknologi Informasi  
Institut Teknologi Sepuluh Nopember

Oleh :

**AMALIA ANJANI ARIFIYANTI**

**NRP. 5210 100 106**

Disetujui Tim Penguji : Tanggal Ujian : 18 Juli 2014  
Periode Wisuda : September 2014

**Prof. Ir. Arif Djunaidi, M.Sc., Ph.D.**

  
**(Pembimbing I)**

**Renny Pradina Kusumawardani, S.T., M.T.**  **(Pembimbing II)**

**Wiwik Anggraeni, S.Si., M.Kom.**

  
**(Penguji I)**

**Retno Aulia Vinarti, S.Kom., M.Kom.**

  
**(Penguji II)**

## KATA PENGANTAR

Puji syukur kepada Allah SWT karena dengan rahmat dan ridho-Nya penulis dapat menyelesaikan laporan tugas akhir yang berjudul “KLASIFIKASI ARTIKEL BERITA BERBAHASA INDONESIA BERBASIS *NAÏVE BAYES CLASSIFIER* MENGGUNAKAN *CONFIX-STRIPPING STEMMER*” sebagai salah satu syarat kelulusan di Jurusan Sistem Informasi, Fakultas Teknologi Informasi, Institut Teknologi Sepuluh Nopember Surabaya. Dengan terselesaikannya tugas akhir ini, maka selesai pula masa studi penulis yang telah ditempuh selama empat tahun.

Penulis berharap tugas akhir ini dapat memberikan manfaat bagi pembaca, menjadi salah satu batu loncatan bagi penulis untuk terus berkarya, dan memberikan sumbangsih bagi ilmu pengetahuan.

Penulis sadar bahwa dalam proses pengerjaan hingga terselesaikannya tugas akhir ini dibutuhkan bantuan dari berbagai pihak dan penulis mengucapkan terima kasih yang sebesar-besarnya kepada pihak-pihak yang membantu pengerjaan tugas akhir ini, antara lain:

- Orang tua penulis yang senantiasa memberikan dukungan penuh dan telah berkorban banyak untuk kebahagiaan penulis. Terimakasih mama yang selalu mendoakan kelancaran studi penulis. Dhik Aghi dan dhik Alan yang selalu memberikan dorongan untuk terus mencapai yang terbaik.
- Seluruh Bapak dan Ibu dosen yang telah membagi banyak ilmu dan inspirasi kepada penulis. Terimakasih telah membagi ilmu-ilmu yang semoga selalu bermanfaat bagi penulis.
- Bapak Prof, Ir. Arif Djunaidy, M.Sc, Ph.D dan ibu Renny Pradina Kusumawardani, S.T., M.T yang telah dengan telaten dan sabar membimbing penulis dalam pengerjaan tugas akhir ini. Terimakasih atas waktu dan ilmu yang sudi dibagikan.

- Ibu Erma Suryani, S.T, M.T, Ph.D dan ibu Amalia Utamima, S.Kom, MBA yang telah memberikan saran-saran terkait pengerjaan tugas akhir ini.
- ‘Ksatria Baja Hitam’ alias ‘Ranger Merah’ yang selalu menyakinkan dan memberikan semangat serta dukungan untuk segera menyelesaikan tugas akhir ini. Terimakasih atas kesabaran, perhatian, dukungan, dan pengertiannya.
- Anggota BFF, Vinda dan Nizar, yang selalu memberikan semangat bahkan menolong dengan senang hati ketika penulis membutuhkan pertolongan segera. Terimakasih sudah menemani masa-masa kuliah ganda melalui jalur cepat. Tanpa kalian, kuliah ganda ini tidak mungkin terlewati dengan baik.
- Yani, Nurul, Denis, Bella, dan Husna yang selalu dengan senang hati memberikan tempat dan teman *ngobrol* yang nyaman. Terimakasih atas pemahaman kalian.
- Icha, Shinta, Muhammad, Piepie, Amalia, Izmi, Eja, Eka, Muci, Miya, Abdan, Ikar, Budhe Dita, Suvi, Alin, Amira, Tika, Putri, dan segenap warga SPK-IB lain yang tidak sempat penulis sebutkan. Terimakasih atas dukungan, doa, tawa, dan bantuan yang diberikan kepada penulis. Kita berjuang bersama, teman!.
- Djay, om Fahri, dan Bisma yang sudah meluangkan waktu untuk membantu pengerjaan teknis Tugas Akhir ini. Jika tidak ada bantuan dari kalian Tugas Akhir ini tidak akan mungkin dapat selesai.
- Imam, Ebi, Fakhri, Rio, Ayu, Fino, Vilat, Anin dan segenap anggota grup ++. Terimakasih atas dukungan kalian. Kalian selalu memberikan waktu-waktu bahagia kepada penulis.
- Mbak ifrina, mbak mitha, mbak anita, mas fathur, mas bonda, mas, cipto, mas rio, pak bayu, pak hasan, dan segenap teman-teman kuliah di magister bidang minat Sistem Informasi ITS 2013. Terimakasih atas doa dan

dukungan serta pengertiannya. Terimakasih untuk selalu mengingatkan untuk segera lulus kuliah jenjang sarjana.

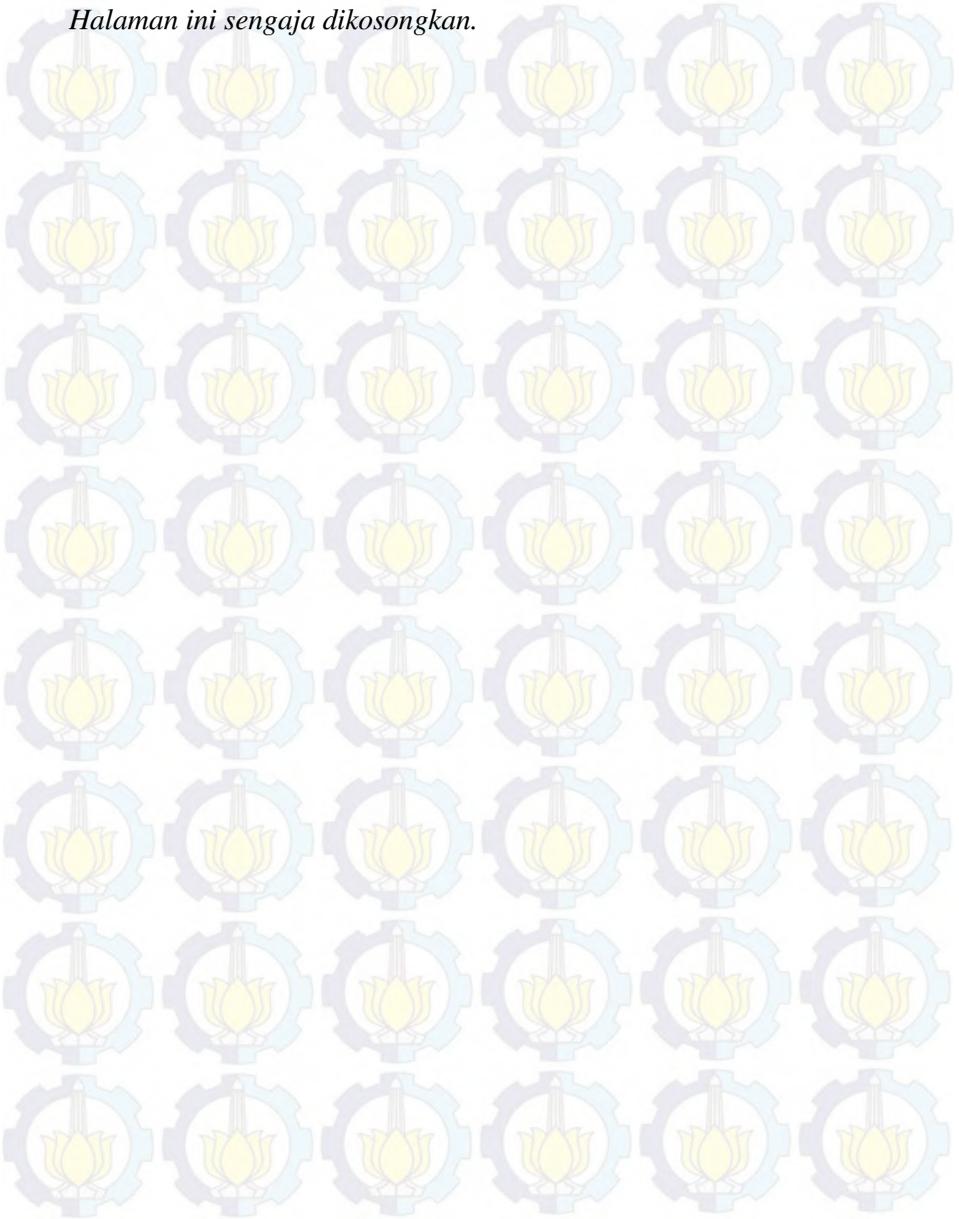
- Ozy, Damar, Bene dan Theo yang selalu memberikan waktu luang pada saat masa-masa pengerjaan Tugas Akhir ini. Terima kasih teman asisten praktikum abadi.
- Keluarga FOXIS yang selalu berjuang bersama-sama dan saling memberikan dukungan serta doa.
- Terimakasih kepada mas Diaz dan mbak Grandys yang secara tidak langsung memberikan semangat untuk terus maju dan tidak menyerah di tengah jalan.
- Terimakasih kepada pemerintah Indonesia, khususnya kepada Ditjen DIKTI yang melalui beasiswa bidik misi telah membantu penulis melanjutkan jenjang pendidikan hingga bangku kuliah ini. Semoga penulis dapat menjadi salah satu generasi bangsa yang bermanfaat.

Dan berbagai pihak yang tidak dapat penulis tuliskan namanya satu per satu. Semoga Tuhan membalas semua kebaikan yang telah dilakukan. Penulis menyadari banyaknya kekurangan yang dapat dikembangkan pada tugas akhir ini, oleh karena itu penulis menerima saran dan kritik yang membangun. Semoga tugas akhir ini dapat memberikan manfaat.

Surabaya, 10 Juli 2014

Penulis

*Halaman ini sengaja dikosongkan.*



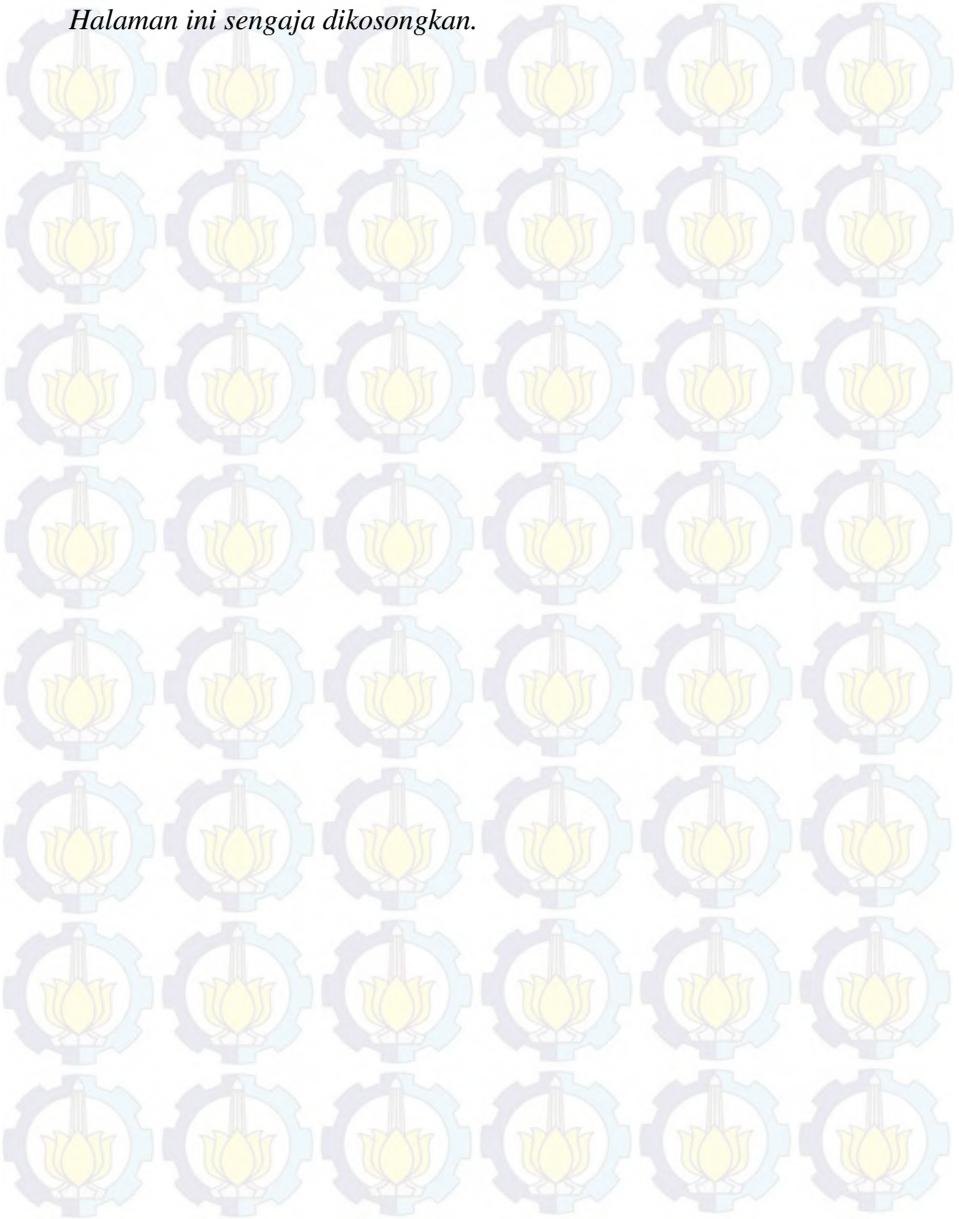
## DAFTAR ISI

DAFTAR ISI.....	xiii
DAFTAR GAMBAR .....	xvii
DAFTAR TABEL .....	xix
DAFTAR SEGMENT KODE PROGRAM.....	xxi
BAB I PENDAHULUAN .....	1
1.1. Latar Belakang Masalah .....	1
1.2. Rumusan Masalah.....	4
1.3. Batasan Tugas Akhir.....	4
1.4. Tujuan Tugas Akhir .....	5
1.5. Relevansi atau Manfaat Kegiatan Tugas Akhir .....	5
1.6. Sistematika Penulisan Tugas Akhir .....	6
BAB II TINJAUAN PUSTAKA .....	9
2.1. Klasifikasi Teks .....	9
2.1.1. Praproses Teks .....	9
2.1.2. <i>Nazief Stemmer</i> .....	10
2.1.3. <i>Confix-Stripping Stemmer</i> .....	14
2.2. Metode <i>Naïve Bayes Classifier</i> (NBC) .....	16
2.3. Pengukuran Performa.....	17
2.3.1. Akurasi .....	17
2.3.2. <i>Recall</i> .....	17
2.3.3. <i>Precision</i> .....	18
2.3.4. <i>F-Measure</i> .....	18
2.4. Data Sampel.....	19
2.4.1. Penentuan Sampel .....	19
2.4.2. <i>Systematic Random Sampling</i> .....	19
2.5. WEKA .....	20
BAB III METODOLOGI Pengerjaan Tugas Akhir.....	21
3.1. Penentuan Tujuan dan Studi Literatur .....	22
3.2. Penyiapan Data .....	22
3.3. Perancangan Desain Perangkat Lunak.....	23
3.4. Praproses Teks .....	24
3.5. Proses Klasifikasi.....	25

3.5.1.	Tahap Pelatihan .....	25
3.5.2.	Evaluasi Model.....	26
3.5.3.	Tahap Pengujian.....	27
3.5.4.	Evaluasi Hasil Klasifikasi .....	27
3.6.	Pembahasan dan Dokumentasi.....	28
<b>BAB IV DESAIN APLIKASI.....</b>		<b>29</b>
4.1.	Desain Aplikasi Klasifikasi Artikel Berita.....	29
4.1.1.	Desain Basis Data.....	29
4.1.2.	Diagram <i>Usecase</i> .....	30
4.1.3.	Diagram Sekuen .....	31
4.1.1.	Diagram Aktivitas .....	40
4.1.2.	Diagram Kelas.....	47
<b>BAB V IMPLEMENTASI PERANGKAT LUNAK .....</b>		<b>49</b>
5.1.	Persiapan dan Implementasi Data Awal .....	49
5.1.1.	Daftar <i>Stopwords</i> dan Daftar Kata Dasar .....	49
5.1.2.	Data Artikel Berita .....	50
5.2.	Implementasi Aplikasi Klasifikasi Artikel Berita .....	52
5.2.1.	Login .....	52
5.2.2.	Menambah Artikel Berita.....	53
5.2.3.	Praproses .....	56
5.2.4.	Latih Klasifikasi .....	65
5.2.5.	Uji Klasifikasi.....	69
5.2.6.	Mengklasifikasikan Artikel Berita .....	72
<b>BAB VI UJI COBA DAN ANALISIS HASIL .....</b>		<b>75</b>
6.1.	Data Uji Coba.....	75
6.2.	Lingkungan Uji Coba.....	75
6.3.	Skenario Uji Coba.....	76
6.3.1.	Uji Verifikasi.....	76
6.3.2.	Uji Validasi .....	77
6.4.	Pelaksanaan dan Hasil Uji Coba .....	77
6.4.1.	Uji Verifikasi.....	77
6.4.2.	Uji Validasi .....	81
6.5.	Analisis Hasil Uji Coba.....	85
6.5.1.	Uji Verifikasi.....	85
6.5.2.	Uji Validasi .....	86

BAB VI KESIMPULAN .....	89
DAFTAR PUSTAKA.....	91
RIWAYAT PENULIS.....	93
LAMPIRAN A .....	A-1
LAMPIRAN B.....	B-1
LAMPIRAN C.....	C-1
LAMPIRAN D .....	D-1

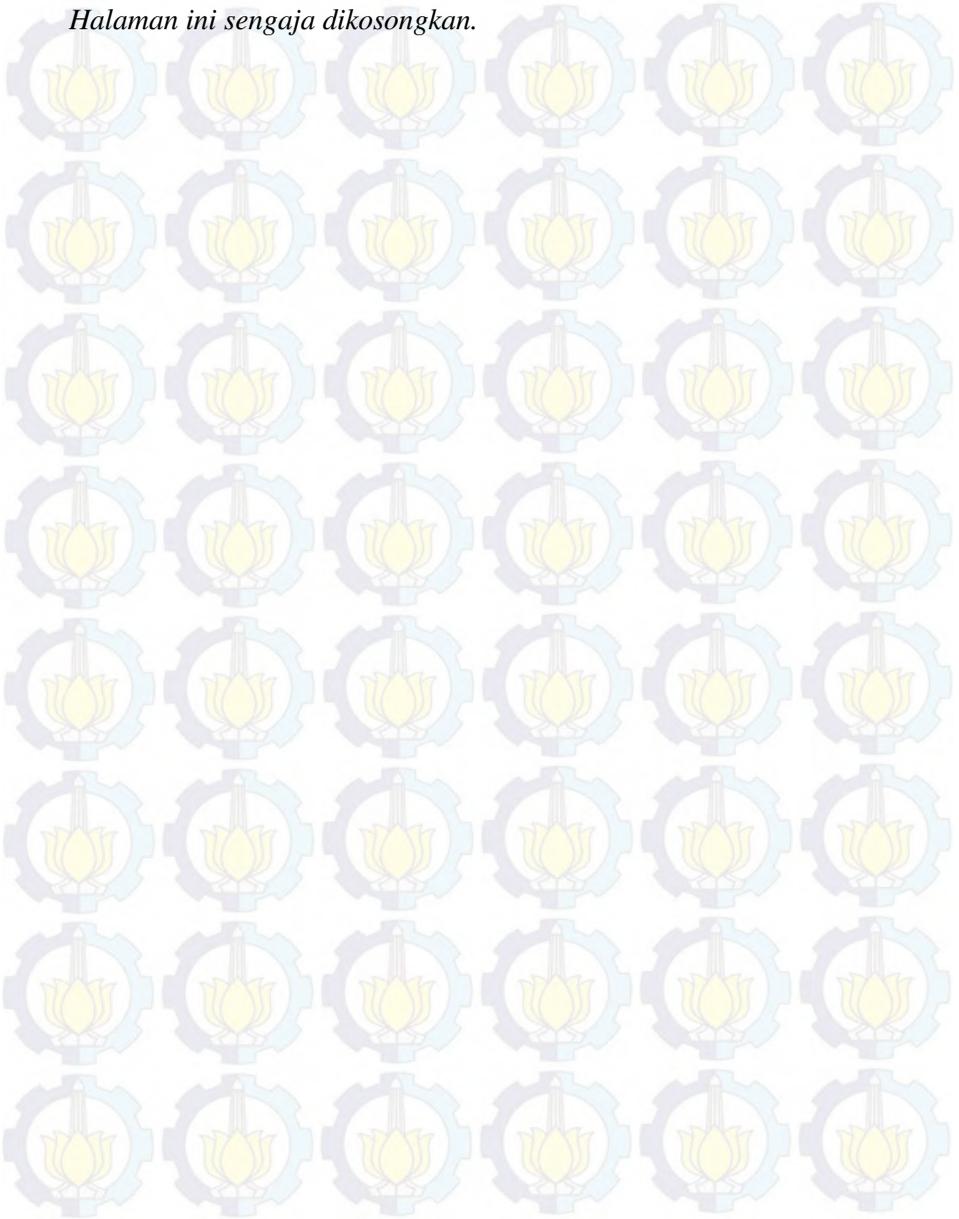
*Halaman ini sengaja dikosongkan.*



## DAFTAR TABEL

Tabel 2.1	Kombinasi awalan-akhiran yang dilarang .....	12
Tabel 2.2	Aturan Pemenggalan .....	13
Tabel 2.3	Aturan Pemenggalan (Lanjutan).....	14
Tabel 2.4	Modifikasi Aturan .....	15
Tabel 2.5	Tambahan Aturan .....	15
Tabel 5.1	Contoh Isi Tabel Stopword.....	49
Tabel 5.2	Daftar Kategori Berita .....	50
Tabel 5.3	Jumlah Sampel per Kategori .....	51
Tabel 6.1	Spesifikasi Lingkungan Uji Coba – Perangkat Keras.....	75
Tabel 6.2	Spesifikasi Lingkungan Uji Coba - Perangkat Lunak.....	76
Tabel 6.3	Hasil Uji Verifikasi Fungsional Aplikasi .....	78
Tabel 6.4	Hasil Uji Verifikasi Fungsional Aplikasi (Lanjutan).....	79
Tabel 6.5	Perbandingan Hasil Klasifikasi .....	80
Tabel 6.6	10 <i>cross-validation</i> .....	81
Tabel 6.7	Perbandingan Evaluasi Model Klasifikasi Aplikasi dan WEKA .....	82
Tabel 6.8	Hasil Klasifikasi Aplikasi.....	82
Tabel 6.9	Hasil Klasifikasi WEKA .....	83
Tabel 6.10	Tingkat Akurasi dari Hasil Klasifikasi Aplikasi .....	83
Tabel 6.11	Hasil Perhitungan <i>Precision</i> , <i>Recall</i> , dan <i>F-Measure</i> Hasil Klasifikasi .....	84
Tabel 6.12	Tingkat Galat per Kategori.....	85
Tabel 6.13	Tabel Kesalahan Klasifikasi.....	87

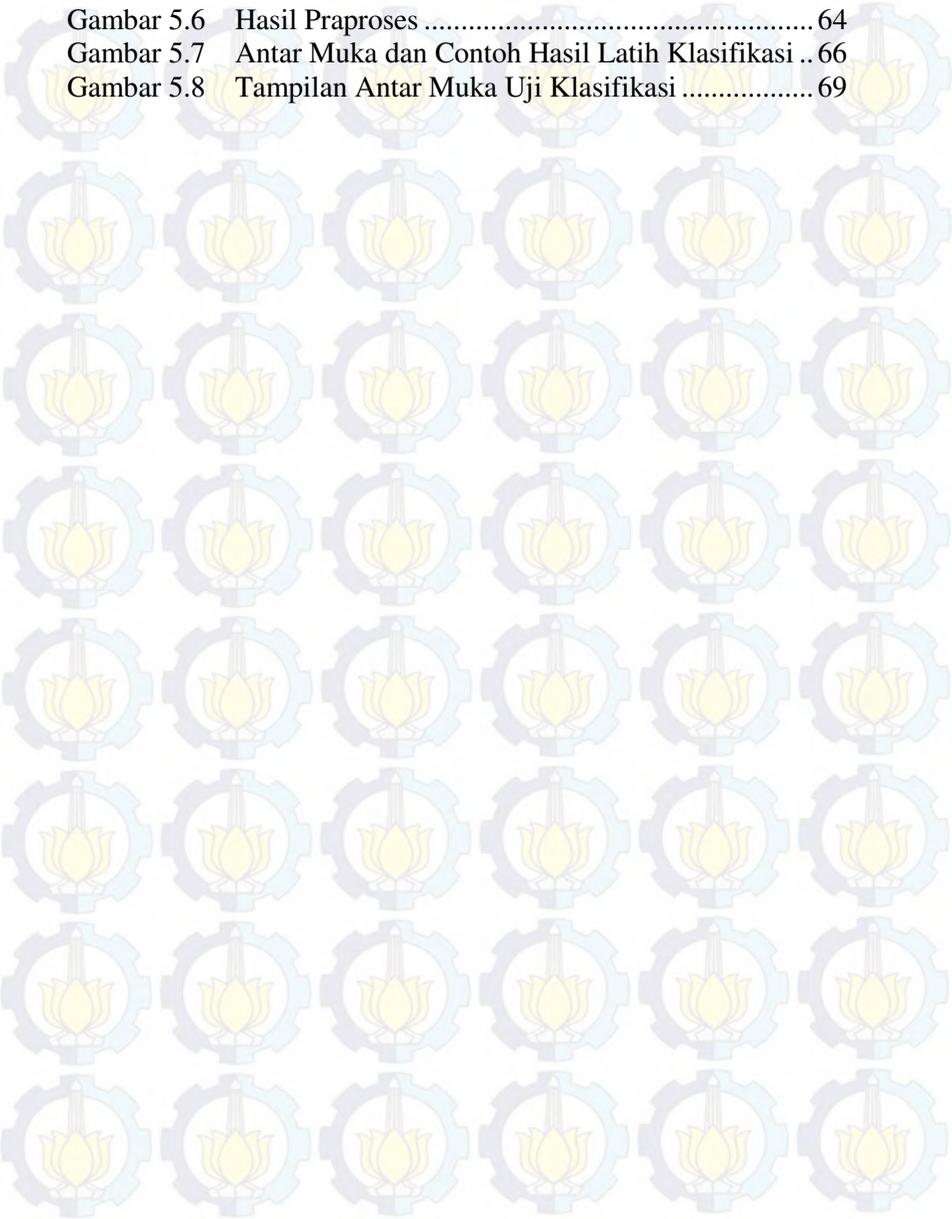
*Halaman ini sengaja dikosongkan.*



## DAFTAR GAMBAR

Gambar 3.1	Metodologi Pengerjaan Tugas Akhir.....	21
Gambar 3.2	Proses latih.....	26
Gambar 3.3	Proses Uji.....	28
Gambar 4.1	Desain Basis Data Klasifikasi.....	30
Gambar 4.2	Diagram <i>Usecase</i> .....	30
Gambar 4.3	Diagram Sekuen Login.....	32
Gambar 4.4	Diagram Sekuen Menambah Data Latih .....	33
Gambar 4.5	Diagram Sekuen Menambah Data Uji.....	34
Gambar 4.6	Diagram Sekuen Praproses Data Latih.....	35
Gambar 4.7	Diagram Sekuen Praproses Data Uji .....	37
Gambar 4.8	Diagram Sekuen Latih.....	38
Gambar 4.9	Diagram Sekuen Uji Klasifikasi .....	39
Gambar 4.10	Digram Sekuen mengklasifikasi Artikel Berita....	39
Gambar 4.11	Diagram Aktivitas Login.....	40
Gambar 4.12	Diagram Aktivitas Menambah Artikel Berita untuk Data Latih .....	41
Gambar 4.13	Diagram Aktivitas Menambah Artikel Berita untuk Data Uji .....	42
Gambar 4.14	Diagram Aktivitas Praproses Data Latih.....	43
Gambar 4.15	Diagram Aktivitas Praproses Data Uji .....	44
Gambar 4.16	Diagram Aktivitas Melakukan Latih Klasifikasi..	45
Gambar 4.17	Diagram Aktivitas Melakukan Uji Klasifikasi .....	46
Gambar 4.18	Diagram Aktivitas Mengklasifikasikan Artikel Berita .....	47
Gambar 4.19	Diagram Kelas .....	48
Gambar 5.1	Antar Muka Borang Login .....	52
Gambar 5.2	Antar Muka Tambah Artikel Berita untuk Data Latih.....	53
Gambar 5.3	Antar Muka Tambah Artikel Berita untuk Data Uji.....	54
Gambar 5.4	Antar Muka Praproses untuk Data Latih .....	56
Gambar 5.5	Antar Muka Praproses untuk Data Uji .....	57

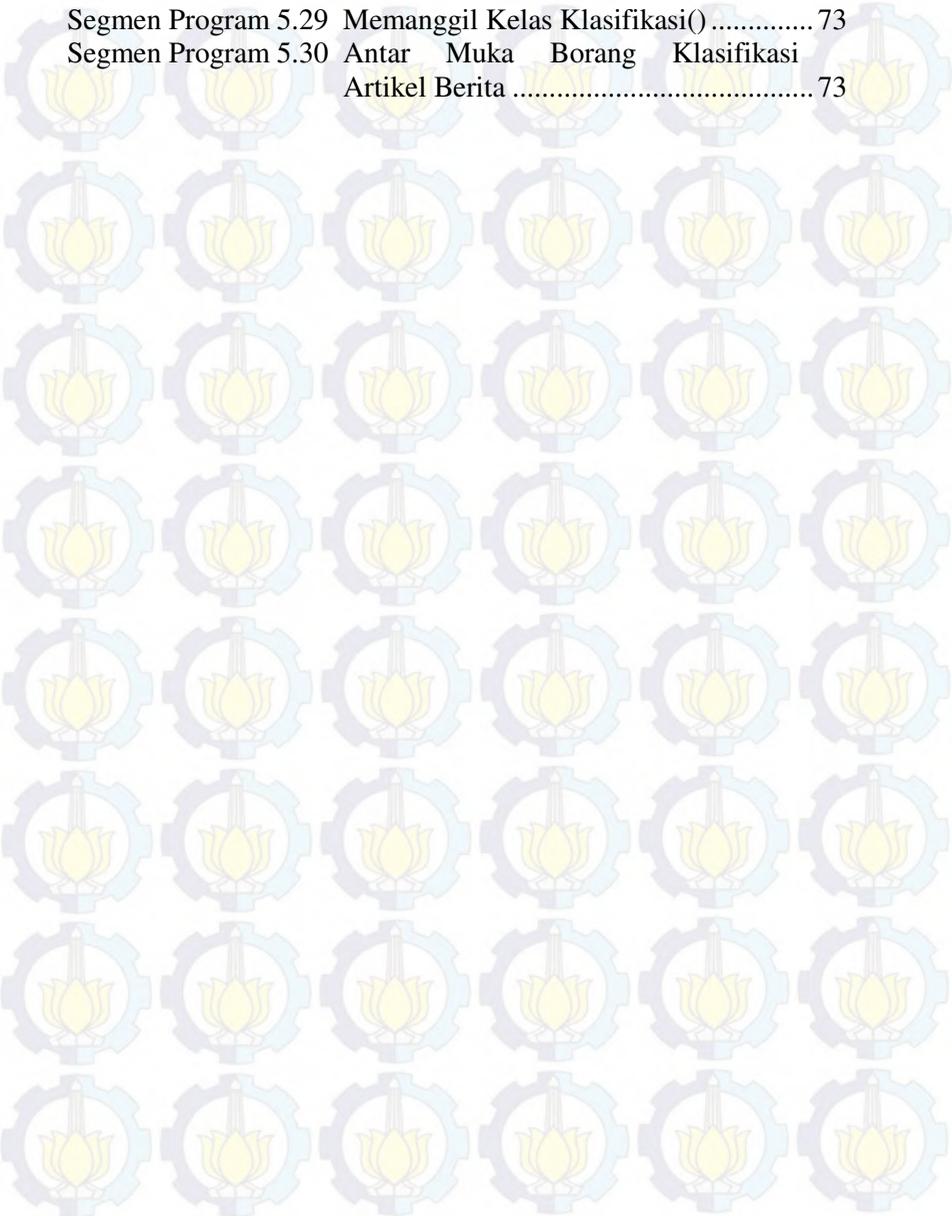
Gambar 5.6	Hasil Praproses .....	64
Gambar 5.7	Antar Muka dan Contoh Hasil Latih Klasifikasi ..	66
Gambar 5.8	Tampilan Antar Muka Uji Klasifikasi .....	69



## DAFTAR SEGMENT KODE PROGRAM

Segmen Program 5.1	Login.....	52
Segmen Program 5.2	Login (Lanjutan).....	53
Segmen Program 5.3	Tambah Artikel Berita untuk Data Latih.....	55
Segmen Program 5.4	Tambah Artikel Berita untuk Data Uji (Lanjutan) .....	55
Segmen Program 5.5	Tambah Artikel Berita untuk Data Uji... ..	56
Segmen Program 5.6	Menjalankan Praproses .....	57
Segmen Program 5.7	readKataInput untuk data latih.....	58
Segmen Program 5.8	readKataInput untuk data uji.....	59
Segmen Program 5.9	<i>Case Folding</i> dan <i>Tokenizing</i> .....	59
Segmen Program 5.10	Penghapusan <i>Stopwords</i> .....	60
Segmen Program 5.11	cekStopword .....	60
Segmen Program 5.12	Cek Kata Masukan dengan Kamus.....	61
Segmen Program 5.13	Hapus awalan.....	61
Segmen Program 5.14	<i>Stemming</i> .....	62
Segmen Program 5.15	<i>Stemming</i> Kata Ulang .....	63
Segmen Program 5.16	Penyimpanan Hasil Praproses terhadap Data Latih.....	64
Segmen Program 5.17	Penyimpanan Hasil Praproses terhadap Data Uji.....	65
Segmen Program 5.18	Baca Data Latih .....	67
Segmen Program 5.19	Evaluasi Data Latih.....	67
Segmen Program 5.20	Evaluasi Data Latih (Lanjutan).....	68
Segmen Program 5.21	Pembuatan Model Klasifikasi.....	68
Segmen Program 5.22	Baca Model Klasifikasi.....	69
Segmen Program 5.23	Baca Data Uji.....	70
Segmen Program 5.24	Pembuatan <i>Instance</i> .....	70
Segmen Program 5.25	Pembuatan <i>Instance</i> (Lanjutan) .....	71
Segmen Program 5.26	Penentuan Kategori.....	71
Segmen Program 5.27	Menampilkan Hasil Uji.....	72
Segmen Program 5.28	Memanggil Kelas Praproses().....	73

Segmen Program 5.29 Memanggil Kelas Klasifikasi()	73
Segmen Program 5.30 Antar Muka Borang Klasifikasi Artikel Berita	73



# **BAB I PENDAHULUAN**

Pada bab ini akan dijelaskan mengenai latar belakang, rumusan masalah, batasan pengerjaan, tujuan, dan manfaat dari pengerjaan tugas akhir.

## **1.1. Latar Belakang Masalah**

Informasi merupakan salah satu kebutuhan manusia. Penyebaran informasi dilakukan dalam berbagai macam cara salah satunya adalah melalui media maya. Seiring dengan perkembangan pesat teknologi informasi dan komunikasi, kemudahan akses internet juga mengalami peningkatan drastis. Menurut Asosiasi Penyelenggara Jasa Internet Indonesia (APJII) bersama Badan Pusat Statistik (BPS) pada tahun 2013 (Pitoyo, 2014), pengguna internet di Indonesia mencapai 71,19 juta pengguna. Seorang pengguna internet menghabiskan waktu rata-rata lebih dari 3 jam dalam sehari untuk mengakses internet, dan pengguna yang menghabiskan waktu lebih dari 3 jam tersebut telah mencapai 31,7 juta orang di tahun 2013. Kondisi ini mendukung perpindahan pembaca media massa cetak ke media online. Hal tersebut juga didukung dengan keseharian masyarakat yang begitu padat dengan pekerjaannya, sehingga mengakses media online lebih praktis dibandingkan membaca berita melalui media cetak.

Berita yang disampaikan pada media maya jumlahnya tidaklah sedikit. Hal tersebut dikarenakan artikel berita berbeda dengan koleksi dokumen. Aliran informasi berita terus menerus diperbaharui dalam waktu singkat sehingga koleksi berita terus bertambah dengan cepat sehingga perlu dilakukan pengorganisasian artikel berita. Salah satu metode pengorganisasian artikel berita adalah dengan cara mengklasifikasikan artikel tersebut ke beberapa kategori berita. Saat ini pengkategorian artikel-artikel berita tersebut masih

dilakukan secara manual. Seorang editor berita harus membaca artikel berita yang akan diterbitkan untuk mengetahui artikel berita tersebut termasuk ke dalam kategori tertentu. Hal tersebut tidak efisien karena seorang editor harus membaca artikel berita satu per satu. Oleh karena itu diperlukan suatu metode yang dapat mengorganisir dengan cara mengkategorikan artikel berita secara lebih mudah. Pengkategorian artikel berita yang lebih mudah dapat membantu para editor berita dan mempercepat tersampainya berita kepada para pembaca. Selain membantu editor berita, artikel berita yang telah dikategorikan sesuai dengan kategorinya juga membantu para pembaca untuk mencari dan membaca artikel berita sesuai dengan kategori berita yang diinginkannya.

Pengelompokkan artikel dapat dilakukan dengan cara penggalian data (*data mining*). Pada penggalian data dikenal dua metode untuk mengkategorikan data, yakni klasifikasi (*classification*) dan klustering (*clustering*). Klasifikasi merupakan suatu metode untuk memprediksi kategori atau kelas dari suatu item atau data. Pada klasifikasi, kategori atau kelas telah didefinisikan sebelumnya. (Oracle). Hal tersebut berbeda dengan klustering yang merupakan metode untuk mencari kluster dari data-data yang memiliki kesamaan. Suatu kluster merupakan hasil pengelompokkan data-data yang memiliki kesamaan ciri khas. (Oracle). Kluster ini tidak diketahui sebelumnya, sehingga terbentuknya kluster menyesuaikan dengan data-data yang ada.

Pengelompokkan artikel berita pada tugas akhir ini telah didefinisikan sebelumnya. Oleh karena itu, metode pengelompokkan yang dipilih adalah metode klasifikasi. Klasifikasi dokumen teks telah lama menjadi topik penelitian dengan berbagai teknik klasifikasi yang diantaranya adalah *Naïve Bayes Classifier* (NBC), *K-Nearest Neighbour*, *Support Vector Machines* (SVM), *Rule Learning Algorithm*, dan *Maximum Entropy*. Penelitian yang dilakukan tersebut sebagian besar masih menggunakan korpus yang berbahasa inggris maupun bahasa asing lainnya. Korpus merupakan himpunan karangan (Kamus

Besar Bahasa Indonesia) atau dengan kata lain merupakan kumpulan artikel berbentuk teks. Klasifikasi dengan metode NBC banyak digunakan dalam penelitian tersebut karena salah satu keunggulannya adalah metode ini hanya membutuhkan jumlah training data yang kecil untuk menentukan estimasi parameter yang diperlukan dalam proses pengklasifikasiannya. (Tan, 2006).

Penelitian sejenis dengan obyek teks berbahasa Indonesia belum sebanyak penelitian obyek berbahasa Inggris ataupun bahasa asing lainnya. Tantangan penggalan teks dengan menggunakan korpus berbahasa Indonesia adalah *stemming*. *Stemming* merupakan proses menghilangkan imbuhan pada suatu kata untuk mendapatkan kata dasar dari kata tersebut. Bahasa Indonesia berbeda dengan bahasa Inggris yang imbuhan hanya akhiran. Imbuhan pada bahasa Indonesia lebih rumit, yakni awalan, akhiran, sisipan, dan *confixes* (kombinasi awalan dan akhiran). Hal ini menyebabkan teknik *stemming* yang populer yakni, *Porter stemmer* tidak sesuai untuk digunakan pada bahasa Indonesia (Tala, 2003). Salah satu teknik *stemming* bahasa Indonesia yang mampu menghasilkan keakuratan yang baik adalah *stemmer* yang diperkenalkan oleh Nazief and Adriani (selanjutnya disebut *nazief stemmer*). Teknik *stemming* ini menghasilkan hasil yang lebih baik jika dibandingkan dengan teknik *stemming* bahasa Indonesia lainnya seperti pendekatan *stemming* milik Arifin dan Setiono serta pendekatan milik Vega (Asian, 2007). Namun *nazief stemming* memiliki beberapa kelemahan antara lain terdapat beberapa aturan, yakni jenis awalan, akhiran, dan peleburan dari awalan akhiran yang belum ada, dan *nazief stemming* tidak mencakup kata ulang. Kelemahan-kelemahan tersebut disempurnakan oleh teknik *stemming* yang diperkenalkan oleh Jelita Asian dan teknik tersebut dikenal sebagai *confix-stripping stemmer*. Sebagai contoh *confix-stripping stemmer* yang mampu menangani kata ulang misalnya 'berkirim-kiriman' yang kata dasarnya adalah kirim. Pada *stemmer* sebelumnya, yakni *nazief stemmer*, kata ulang tersebut akan menjadi kata dasar karena *stemmer* tersebut tidak menangani

kata ulang. Hal tersebut berbeda pada *confix-stripping stemmer* yang mampu mencari bentuk dasar dari kata ulang yakni 'kirim'. Jika penanganan kata dasar ini gagal maka kata 'berkirim-kiriman' akan dianggap berbeda dengan kata 'mengirim atau bahkan 'kirim' padahal ketiga kata tersebut sama yakni memiliki kata dasar 'kirim'. Selain itu, pada beberapa kata lain seperti bertaburan, teradili, melindungi, dan perlengkapi tidak mampu dicari kata dasarnya oleh nazief *stemmer* karena urutan pemecahan imbuhan pada kata. Hal ini mampu ditangani oleh *confix-stripping stemmer* yang mengubah urutan pemecahan imbuhan pada beberapa jenis imbuhan tertentu atau *rule precedence*.

Oleh karena itu, dalam tugas akhir ini digunakan teknik *confix-stripping stemmer* yang telah terbukti keandalannya dalam proses stemming. Untuk teknik klasifikasi artikel beritanya, akan digunakan metode NBC yang merupakan teknik klasifikasi yang sudah banyak digunakan untuk berbagai jenis pengklasifikasian dokumen teks. Hasil tugas akhir ini diharapkan dapat memberikan kontribusi dalam pengklasifikasian artikel berita, sehingga mempermudah dalam pengkategorian artikel berita. Selain itu, juga diharapkan dapat menjadi acuan dalam pengklasifikasian dokumen berbasis teks berbahasa Indonesia lainnya.

## **1.2. Rumusan Masalah**

Dari uraian pada bagian latar belakang, maka rumusan permasalahan pada pengerjaan tugas akhir ini adalah bagaimana membangun aplikasi pengklasifikasi artikel berita berbahasa Indonesia berbasis NBC menggunakan *confix-stripping stemmer* yang dapat dengan mudah dioperasikan oleh editor berita?

## **1.3. Batasan Tugas Akhir**

Pengerjaan tugas akhir ini memiliki beberapa batasan antara lain sebagai berikut:

- a. Data yang digunakan merupakan artikel-artikel berita berbahasa Indonesia yang didapatkan dari portal berita *online* [www.kompas.com](http://www.kompas.com).

- b. Data yang digunakan pada tahap latih dan uji merupakan sampel artikel berita dalam rentang waktu mulai pertama artikel berita tersimpan di kompas.com hingga akhir tahun 2013.
- c. Kategori yang digunakan sebanyak 12 kategori, yakni kategori berita nasional, internasional, olahraga, sains, edukasi, ekonomi, tekno, *entertainment*, otomotif, *health*, properti, dan travel.
- d. Kategori yang didefinisikan tidak bersifat hirarki atau bersusun.

#### **1.4. Tujuan Tugas Akhir**

Tujuan utama dari tugas akhir ini adalah membangun aplikasi pengklasifikasi artikel berita berbahasa Indonesia berbasis *naïve bayes classifier* menggunakan *confix-stripping stemmer* untuk membantu para editor berita dalam pengklasifikasian artikel berita.

#### **1.5. Relevansi atau Manfaat Kegiatan Tugas Akhir**

Pengerjaan tugas akhir ini diharapkan dapat memberikan manfaat bagi:

##### **a. Penulis**

Tugas akhir ini merupakan sumbangsih dalam bidang ilmu klasifikasi teks secara umum, dengan menggunakan metode *naïve baye classifiers* dan *confix-stripping stemmer* secara khususnya.

##### **b. Para pelaku di bidang jurnalistik**

Hasil tugas akhir ini diharapkan mampu membantu para pelaku di bidang jurnalistik dan para editor berita khususnya dalam pengkategorian artikel berita sehingga diharapkan dapat meningkatkan efisiensi kerja dan mampu mempercepat tersampainya berita.

**c. Para pembaca berita**

Secara tidak langsung, hasil dari klasifikasi berita akan dapat membantu para pembaca berita dalam pencarian artikel berita yang sesuai kategori yang diinginkannya.

**d. Ilmu pengetahuan dan teknologi**

Hasil tugas akhir ini diharapkan dapat menjadi dasar bagi pengembangan teknologi yang berkaitan dengan klasifikasi teks berbahasa Indonesia, misalnya pengembangan sistem informasi yang dapat mengklasifikasikan hasil pencarian dari berbagai artikel atau dokumen yang tersebar di dunia maya.

**1.6. Sistematika Penulisan Tugas Akhir**

Sistematika penulisan Laporan Tugas Akhir ini dibagi menjadi tujuh bab sebagai berikut :

**BAB I PENDAHULUAN**

Pada bab ini dijelaskan mengenai latar belakang, rumusan masalah, batasan pengerjaan tugas akhir, manfaat tugas akhir, tujuan tugas akhir dan sistematika penulisan yang diterapkan dalam memaparkan tugas akhir.

**BAB II TINJAUAN PUSTAKA**

Pada bab ini dijelaskan mengenai teori-teori terkait yang bersumber dari buku, jurnal, ataupun artikel yang berfungsi sebagai dasar dalam melakukan pengerjaan tugas akhir agar dapat memahami konsep atau teori penyelesaian permasalahan yang ada. Pada bab ini terdapat uraian mengenai klasifikasi teks yang mencakup penjelasan mengenai teknik *stemming* bahasa Indonesia, metode NBC, pengukuran performa klasifikasi, dan teknik *sampling* yang digunakan untuk pengerjaan tugas akhir.

### **BAB III METODE Pengerjaan TUGAS AKHIR**

Pada bab ini dijelaskan mengenai langkah–langkah pengerjaan tugas akhir beserta metode yang digunakan. Langkah–langkah pengerjaan dijelaskan dalam sebuah diagram alur yang sistematis dan akan dijelaskan tahap demi tahap.

### **BAB IV DESAIN APLIKASI**

Pada bab ini dijelaskan mengenai rancangan aplikasi yang akan dibangun. Rancangan aplikasi ini mencakup diagram *usecase*, skenario *usecase*, diagram sekuen, diagram aktifitas, dan kelas diagram. Rancangan aplikasi ini diharapkan dapat membantu dalam pembangunan aplikasi untuk kedepannya.

### **BAB V IMPLEMENTASI APLIKASI**

Pada bab ini akan dipaparkan mengenai cara pembangunan aplikasi berdasarkan desain aplikasi yang telah dibuat sebelumnya. Aplikasi yang dibangun menggunakan bahasa pemrograman JAVA.

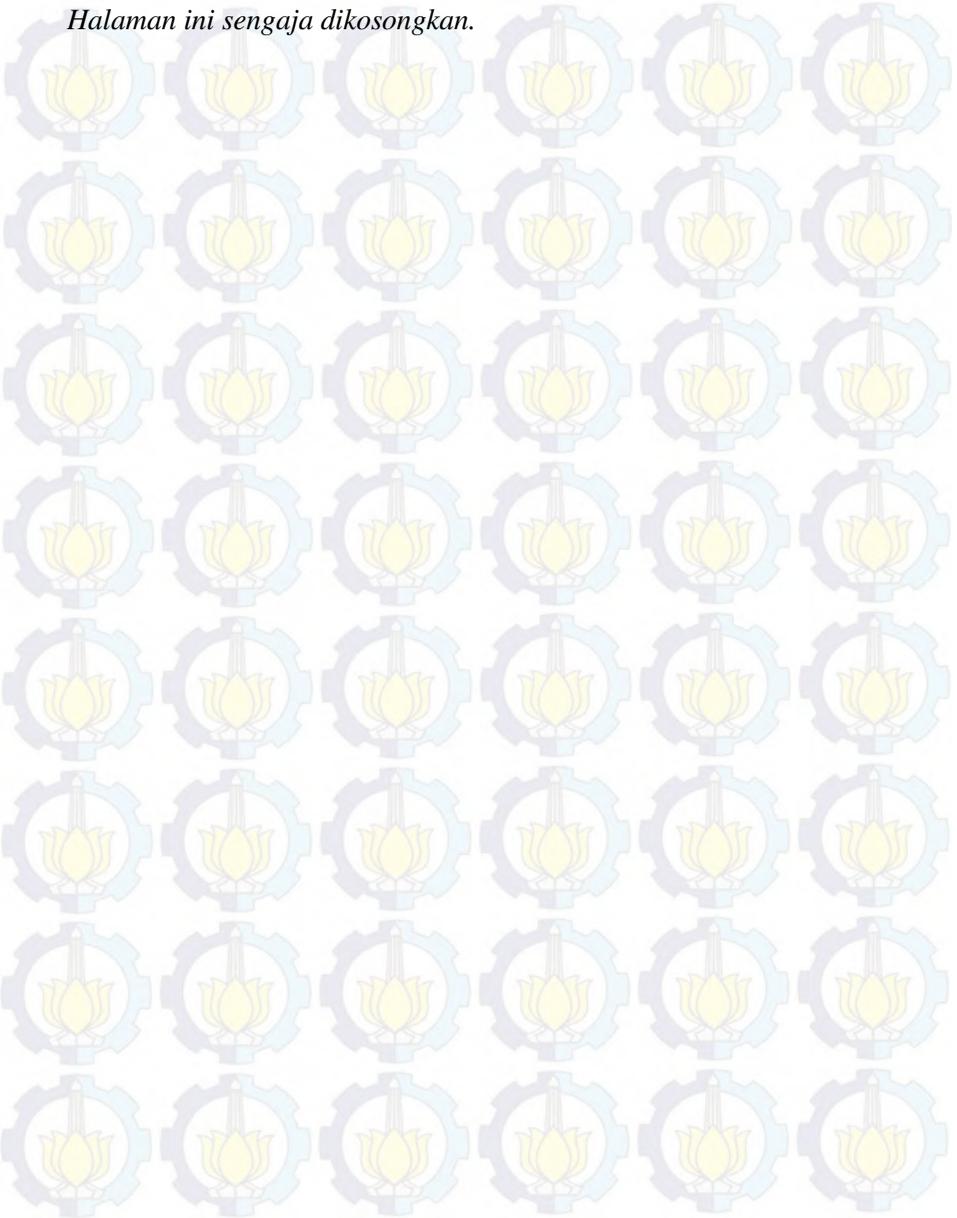
### **BAB VI UJI COBA DAN ANALISIS HASIL**

Pada bab ini akan dilakukan uji coba terhadap aplikasi yang terbagi menjadi dua skenario uji coba, yakni uji verifikasi dan uji validasi. Selain itu, juga dijelaskan mengenai analisis terhadap hasil keluaran aplikasi.

### **BAB VII KESIMPULAN DAN SARAN**

Bab ini berisi kesimpulan dari pengerjaan tugas akhir dan saran untuk pengembangan lebih lanjut.

*Halaman ini sengaja dikosongkan.*



## BAB II TINJAUAN PUSTAKA

Pada bagian bab ini akan dijelaskan mengenai berbagai teori yang digunakan sebagai dasar dan acuan dalam pengerjaan tugas akhir ini.

### 2.1. Klasifikasi Teks

Klasifikasi teks merupakan proses menemukan pola baru yang belum terungkap sebelumnya. Hal ini dilakukan dengan memproses dan menganalisa data dalam jumlah besar. Klasifikasi teks melibatkan struktur yang mungkin terdapat pada teks dan mengekstraks informasi yang relevan pada teks. Dalam menganalisis sebagian atau keseluruhan teks yang tidak terstruktur, klasifikasi teks mencoba mengasosiasikan sebagian atau keseluruhan satu bagian teks dengan yang lainnya berdasarkan aturan-aturan tertentu. Hasil yang diharapkan merupakan informasi baru yang tidak terungkap jelas sebelumnya. (Miller, 2005).

#### 2.1.1. Praproses Teks

Tahap praproses terhadap teks yang akan diproses baik clustering maupun klasifikasi merupakan tahapan yang penting. Tahapan ini perlu dilakukan sebelum proses klasifikasi.

Tahapan praproses teks secara garis besar adalah sebagai berikut:

- a. *Case Folding*, yakni mengubah semua teks menjadi huruf kecil (*toLowerCase*) dan hanya huruf 'a' hingga 'z' yang diterima, karakter selain huruf dihilangkan. Karakter ini termasuk angka dan tanda baca misalnya tanda baca titik (.), koma (,), dan karakter angka yang ada pada kata-kata tersebut. (Weiss, 2010).
- b. *Tokenizing*, yakni proses penguraian deskripsi yang semula berupa kalimat menjadi kata-kata atau dengan kata

lain melakukan pemotongan string masukan, yakni teks, berdasarkan tiap kata yang menyusunnya.

- c. Proses *stopping* berdasarkan *stoplist* yang berisi *stopwords* yang telah ditentukan sebelumnya. *Stopwords* merupakan kosakata yang bukan merupakan kata unik atau ciri pada suatu dokumen. Kosakata yang dimaksudkan adalah kata penghubung dan kata keterangan yang bukan merupakan kata unik misalnya “sebuah”, “oleh”, “pada”, dan sebagainya. (Dragut, 2009).
- d. *Stemming* pada kata-kata yang tersisa pada dokumen teks untuk mendapatkan kata dasar. *Stemming* merupakan proses pemetaan dan penguraian berbagai bentuk dari suatu kata menjadi bentuk kata dasarnya (Tala, 2003). Proses ini akan menghilangkan awalan, akhiran, sisipan, dan *confixes* (kombinasi dari awalan dan akhiran).

### 2.1.2. Nazief Stemmer

Teknik *stemming* telah banyak diperkenalkan dan teknik *Porter stemmer* merupakan pendekatan *stemming* yang banyak digunakan. Tetapi teknik ini memiliki banyak kelemahan jika diterapkan pada bahasa Indonesia. Oleh karena kebutuhan terhadap teknik *stemming* bahasa Indonesia, muncullah berbagai macam teknik *stemming*. Salah satunya adalah teknik *stemming* yang diperkenalkan oleh Nazief dan Adriani, yakni *nazief stemmer*. Algoritma *stemming* Nazief dan Adriani dikembangkan berdasarkan aturan bahasa Indonesia yang kata-katanya menggunakan imbuhan, yakni awalan (*prefix*), sisipan (*infix*), akhiran (*suffix*) dan kombinasi awalan dan akhiran (*confixes*). *Nazief stemmer* mengelompokkan imbuhan ke dalam beberapa kategori sebagai berikut:

- a. *Inflection Suffixes* : kelompok akhiran yang tidak mengubah bentuk kata dasar. Kelompok ini dibagi menjadi dua, yakni: (1) *Particle* (Partikel) yang didalamnya '-lah', '-

kah', '-tah', dan '-pun' ; (2) *Passive Pronoun* (Kata ganti kepemilikan) yang didalamnya '-ku', '-mu', '-nya'.

- b. *Derivation Suffixes* (Akhiran) : kumpulan akhiran yang secara langsung ditambahkan pada kata dasar.
- c. *Derivation Prefixes* (Awalan) : Kumpulan awalan yang dapat ditambahkan langsung pada kata dasar yang sudah mendapatkan penambahan sampai dua awalan.

Aturan-aturan dari berbagai jenis imbuhan tersebut dimodelkan sebagai berikut:

[[[AW +]AW +]AW +] Kata Dasar [[+ AK][+ KK][+ P]]

Dengan:

AW = Awalan  
AK = Akhiran

KK = Kata ganti kepunyaan  
P = Partikel

Langkah *nazief stemmer* adalah sebagai berikut:

- a. Kata yang belum di-*stemming* dicari ke dalam kamus kata dasar. Jika ditemukan maka kata tersebut diasumsikan sebagai kata dasar dan tahapan berhenti. Akan tetapi jika tidak ditemukan maka dilanjutkan pada tahap kedua.
- b. Hilangkan *Inflectional Particle* ('-lah', '-kah', '-tah', '-pun') dan kata ganti kepunyaan ('-ku', '-mu', '-nya').
- c. Hilangkan imbuhan akhiran ('-i', '-kah', '-an')
- d. Hilangkan imbuhan awalan ('di-', 'ke-', 'se-', 'me-', 'be-', 'pe-', 'te-') dengan iterasi maksimum adalah tiga kali:
  - 1) Langkah 'd' berhenti jika : (a) terjadi kombinasi awalan dan akhiran terlarang seperti pada tabel 2.1 ; (b) awalan yang dideteksi saat ini sama dengan awalan yang dihilangkan sebelumnya ; (c) tiga awalan telah dihilangkan.

Tabel 2.1 Kombinasi awalan-akhiran yang dilarang

Awalan	Akhiran yang tidak diperbolehkan
be-	-i
di-	-an
ke-	-i, -kan
me-	-an
se-	-i, -kan
te-	-an

- 2) Identifikasikan tipe awalan dan hilangkan. Awalan terbagi menjadi dua tipe:
    - a) standar = 'di-', 'ke-', 'se-' yang dapat langsung dihilangkan dari kata.
    - b) kompleks = 'me-', 'be-', 'pe-', 'te-' merupakan tipe awalan yang dapat bermorfologi sesuai kata dasar yang mengikutinya. Sehingga digunakan aturan pada tabel 2.2 untuk mendapatkan pemenggalan yang tepat.
  - 3) Cari kata yang telah dihilangkan awalnya dalam kamus. Jika tidak ditemukan maka seluruh tahap dihentikan.
- e. Jika setelah langkah 'd' kata dasar belum ditemukan, maka proses *recoding* dilakukan dengan mengacu pada aturan pada tabel 2.2. *Recoding* dilakukan dengan menambahkan karakter *recoding* di awal kata yang dipenggal. Pada tabel 2.2, karakter *recoding* adalah huruf kecil setelah tanda hubung dan berada sebelum tanda kurung. Sebagai contoh, kata 'menangkap' yang setelah dihilangkan awalan 'me-' menjadi 'nangkap', karena kata 'nangkap' tidak terdapat pada kamus kata dasar maka *recoding* dilakukan. Hasil

*recoding* ini adalah ‘tangkap’ yang dihasilkan dari penggantian karakter ‘n’ menjadi ‘t’.

- f. Jika semua langkah gagal maka input kata dianggap sebagai kata dasar.

Pada table 2.2, 2.3, 2.4, dan 2.5 simbol C merupakan huruf konsonan, simbol V merupakan vokal, simbol A merupakan vokal atau konsonan, dan simbol P merupakan partikel dari suatu kata, contohnya 'er'.

**Tabel 2.2 Aturan Pemenggalan**

Aturan	Format Kata	Pemenggalan
1	berV..	Ber-V..   be-rV..
2	berCAP..	Ber-CAP.. dimana C!- 'r' & P!-'er'
3	berCAerV..	Ber-CaerV.. dimana C!='r'
4	Belajar	Bel-ajar
5	beC <sub>1</sub> erC <sub>2</sub> ..	beC <sub>1</sub> erC <sub>2</sub> .. dimana C <sub>1</sub> !={ 'r'!'l' }
6	terV..	Ter-V..   te-rV..
7	terCerV..	Ter-CerV.. dimana C!='r'
8	terCP...	Ter-CP.. dimana C!='r' dan P!='er'
9	teC <sub>1</sub> erC <sub>2</sub> ..	Te-C <sub>1</sub> erC <sub>2</sub> ... dimana C <sub>1</sub> !='r'
10	Me{llrlwly}V...	Me-{llrlwly}V...
11	Mem{blflv}...	Mem-{blflv}...
12	Mempe{rll}...	Mem-pe..
13	Mem{rVIV}...	Me-m{rVIV}...   me-p {rVIV}...
14	Men{cldljz}...	Men-{cldljz}...
15	menV...	Me-nV.. lme-tV..
16	Meng{gghlq}...	Meng-{gghlq}...
17	mengV...	Meng-V... lmeng-kV...
18	menyV...	Meny-sV...
19	mempV...	mempV... dimana V!='e'
20	Pe{wly}V...	Pe-{wly}V...
21	perV...	Per-V...lpe-rV...

Tabel 2.3 Aturan Pemenggalan (Lanjutan)

Aturan	Format Kata	Pemenggalan
22	perCAP..	Per-CAP.. dimana C!= 'r' dan P!= 'er'
23	perCAerV...	Per-CAerV...dimana C!= 'r'
24	Pem{blfIV}..	Pem-{blfIV}..
25	Pem{rVIV}...	Pe-m{rVIV}...   pe-p{rVIV}...
26	Pen{cldljz}...	Pen-{cldljz}...
27	penV...	Pe-nV...lpe-IV..
28	Peng{ghlq}...	Peng-{ghlq}...
29	pengV...	Peng-V...l peng-kV...
30	penyV...	Peny-sV...
31	peIV...	Pe-IV... kecuali 'pelajar' yang menghasilkan 'ajar'
32	peCerV...	Per-erV...dimana C!={rlwlyllmln}
33	peCP...	peCP... dimana C!={rlwlyllmln} dan P!= 'er'

### 2.1.3. *Confix-Stripping Stemmer*

Jelita Asian mengembangkan algoritma nazief *stemmer*, dengan menambahkan beberapa perbaikan yang bertujuan untuk meningkatkan hasil *stemming* yang diperoleh (Asian, 2007). Algoritma ini dikenal sebagai *confix-stripping stemmer*. Perbaikan tersebut antara lain sebagai berikut:

- a. Menggunakan kamus kata dasar yang lebih lengkap
- b. Memodifikasi dan menambahkan aturan pemenggalan untuk tipe awalan yang kompleks (memodifikasi aturan pada tabel 2.2 sesuai modifikasi pada tabel 2.3 dan menambahkan aturan pada table 2.4 ke dalam tabel 2.2)
- c. Menambahkan aturan *stemming* untuk kata ulang dan bentuk jamak, misalnya kata (1) 'buku-buku' yang menjadi 'buku', (2) 'berkirim-kiriman' menjadi kirim. Hal ini

dilakukan dengan melakukan pemisahan kata tersebut menjadi dua kata yang masing-masing di-*stemming*. Jika *stemming* memberikan kata dasar yang sama, maka keluaran kata dasarnya adalah hasil *stemming* tersebut. Jika hasil *stemming* dua kata tersebut berbeda maka disimpulkan bahwa masukan adalah kata ulang semu dan tidak memiliki bentuk kata dasar lagi.

- d. Mengubah urutan *stemming* untuk beberapa kasus tertentu. Algoritma *stemmer* yang diperkenalkan oleh Nazief akan menghilangkan akhiran lebih dahulu, baru diikuti penghilangan awalan. Namun menurut Jelita, cara ini tidak selalu berhasil pada beberapa kata. Sehingga diberikan aturan yang akan mengubah urutan *stemming*, yang mana penghilangan awalan dilakukan terlebih dahulu kemudian diikuti oleh penghilangan akhiran. Aturan ini disebut *rule precedence* dan berlaku jika kata memiliki kombinasi awalan-akhiran 'be-lah', 'be-an', 'me-i', 'di-i', 'pe-i', atau 'te-i', misalnya 'bertaburan', 'melindungi', 'dilengkapi', dan 'teradili'.

Tabel 2.4 Modifikasi Aturan

Aturan	Format Kata	Pemenggalan
12	Mempe...	Mem-pe...
16	Meng{ghlqlk}	Meng-{ghlqlk}

Tabel 2.5 Tambahan Aturan

Aturan	Format Kata	Pemenggalan
34	terC <sub>1</sub> erC <sub>2</sub> ...	terC <sub>1</sub> erC <sub>2</sub> ... dimana C <sub>1</sub> != 'r'
35	peC <sub>1</sub> erC <sub>2</sub> ...	peC <sub>1</sub> erC <sub>2</sub> ... dimana C <sub>1</sub> != {r/w/y/l/l/m/n}

## 2.2. Metode *Naïve Bayes Classifier* (NBC)

Algoritma *naive bayes classifier* (NBC) merupakan algoritma yang digunakan untuk mencari nilai probabilitas tertinggi untuk mengklasifikasi data uji pada kategori yang paling tepat. Metode NBC menempuh dua tahap dalam proses klasifikasi teks, yakni tahap pelatihan dan tahap klasifikasi. Pada tahap pelatihan dilakukan proses analisis terhadap sampel dokumen berupa pemilihan kosa kata, yakni kata yang mungkin muncul dalam tahap pelatihan yang mungkin dapat menjadi representasi dokumen. Selanjutnya adalah penentuan probabilitas prior bagi tiap kategori berdasarkan sampel dokumen. Pada tahap klasifikasi ditentukan nilai kategori dari suatu dokumen berdasarkan *term* yang muncul dalam dokumen yang diklasifikasi.

Dalam algoritma NBC setiap dokumen direpresentasikan dengan pasangan atribut “ $a_1, a_2, a_3, \dots, a_n$ ” dimana  $a_1$  adalah kata pertama,  $a_2$  adalah kata kedua dan seterusnya. Sedangkan  $V$  adalah himpunan kategori berita. Pada saat klasifikasi algoritma akan mencari probabilitas tertinggi dari semua kategori dokumen yang diujikan ( $V_{MAP}$ ), dimana persamaannya adalah sebagai berikut :

$$V_{MAP} = \arg \max_{v_j \in V} P(v_j | a_1, a_2, \dots, a_n) \quad (2.1)$$

Karena nilai  $P(a_1, a_2, a_3, \dots, a_n)$  untuk semua  $v_1$  besarnya sama maka nilainya dapat diabaikan, sehingga persamaan di atas menjadi:

$$V_{MAP} = \arg \max_{v_j \in V} P(a_1, a_2, \dots, a_n | v_j) P(v_j) \quad (2.2)$$

Dengan asumsi bahwa setiap kata dalam  $\langle a_1, a_2, a_3, \dots, a_n \rangle$  adalah independen, maka  $P(a_1, a_2, a_3, \dots, a_n | v_1)$  dalam persamaan di atas dapat ditulis seperti persamaan 2.3 berikut ini:

$$P(a_1, a_2, \dots, a_n | v_j) = \prod_i P(a_i | v_j) \quad (2.3)$$

Sehingga persamaan tersebut dapat ditulis:

$$V_{MAP} = \arg \max_{v_j \in V} P(v_j) \prod_i P(a_i | v_j) \quad (2.4)$$

Nilai  $P(v_j)$  ditentukan pada saat pelatihan, yang nilainya didapat dengan :

$$P(v_j) = \frac{|doc\ j|}{|contoh|} \quad (2.5)$$

Dimana  $|doc\ j|$  merupakan banyak dokumen (dalam hal ini artikel berita) yang memiliki kategori  $j$  dalam pelatihan, sedangkan  $|contoh|$  merupakan banyak dokumen (dalam hal ini artikel berita) dalam contoh yang digunakan untuk pelatihan. Untuk nilai  $P(w_k | v_j)$ , yakni probabilitas kata  $w_k$  dalam kategori  $j$  ditentukan dengan :

$$P(w_k | v_j) = \frac{n_k + 1}{n + |kosakata|} \quad (2.6)$$

Dimana  $n_k$  adalah frekuensi munculnya kata  $w_k$  dalam dokumen yang berkategori  $v_j$ , sedangkan nilai  $n$  adalah banyaknya seluruh kata dalam dokumen berkategori  $v_j$  dan  $|kosakata|$  adalah banyaknya kata dalam contoh pelatihan.

## 2.3. Pengukuran Performa

Dalam klasifikasi, terdapat beberapa cara pengukuran performa klasifikasi. Namun cara yang paling sering digunakan adalah dengan perhitungan *precision*, *recall*, dan akurasi. (Jian Sheu, 2008).

### 2.3.1. Akurasi

Akurasi merupakan prosentase dari total dokumen yang benar diidentifikasi.

$$akurasi = \frac{\text{jumlah klasifikasi benar}}{\text{jumlah dokumen uji coba}} \times 100\% \quad (2.7)$$

### 2.3.2. Recall

*Recall* merupakan perbandingan jumlah materi relevan yang ditemubalikkan terhadap jumlah materi yang relevan.

$$recall = \frac{|{\{relevant\ doc\}} \cap {\{retrieved\ doc\}}|}{|{\{retrieved\ doc\}}|} \quad (2.8)$$

Keterangan :

*Recall* = tingkat keberhasilan

*Relevant doc* = kumpulan dokumen yang relevan

*Retrieved doc* = kumpulan dokumen yang ditemukan

### 2.3.3. Precision

*Precision* merupakan perbandingan jumlah materi relevan yang ditemubalikkan terhadap jumlah materi yang ditemubalikkan.

$$precision = \frac{|{relevant\ doc} \cap {retrieved\ doc}|}{|{retrieved\ doc}|} \quad (2.9)$$

Keterangan:

*Precision* = tingkat ketepatan

*Relevant doc* = kumpulan dokumen yang relevan

*Retrieved doc* = kumpulan dokumen yang ditemukan

### 2.3.4. F-Measure

*F-measure* merupakan parameter tunggal ukuran keberhasilan *retrieval* yang menggabungkan *recall* dan *precision*. (Rijsbergen, 1979).

$$F - measure = \frac{(\beta^2 + 1)PR}{\beta^2 P + R} \quad (2.10)$$

Keterangan:

$\beta$  = parameter kepentingan relative aspek *precision* dan *recall*

*P* = nilai *precision*

*R* = nilai *recall*

Jika nilai  $\beta > 1$ , maka akan memberikan bobot kepentingan *recall* lebih tinggi daripada *precision*. Jika nilai  $\beta = 2$  maka akan bobot *recall* dua kali lebih besar daripada *precision*. Jika nilai  $\beta = 0.5$  maka bobot *precision* dua kali lebih besar daripada *recall*. Tetapi jika *recall* dan *precision* memiliki bobot yang sama maka  $\beta = 1$ , dan parameter *F-measure* dituliskan pada persamaan 2.11.

$$F - measure = \frac{2PR}{P+R} \quad (2.11)$$

Suatu sistem dinyatakan efektif jika hasil perhitungan menunjukkan ketepatan (*precision*) yang tinggi sekalipun *recall*-nya rendah.

## 2.4. Data Sampel

Jika data yang akan diambil untuk penelitian jumlahnya sangat banyak dan sangat tidak memungkinkan untuk mendapatkan seluruh data tersebut maka perlu dilakukan pengambilan sampel dari seluruh data tersebut. Pengambilan sampel bertujuan untuk menghemat waktu dan tenaga, namun dapat mendapatkan hasil yang representatif dari keseluruhan data atau populasi yang menjadi obyek penelitian.

### 2.4.1. Penentuan Sampel

Terdapat banyak cara untuk menentukan sampel, namun pertama yang harus dipastikan adalah jumlah populasi dari obyek yang akan diteliti. Dari populasi yang ada, sampel dapat ditentukan besarnya. Salah satu cara untuk menentukan besar sampel adalah dengan menggunakan rumus Slovin (Ariola, 2006) sebagai berikut.

$$n = \frac{N}{1+Ne^2} \quad (2.12)$$

Keterangan:

- n = ukuran sampel
- N = ukuran populasi
- e = batas toleransi kesalahan

### 2.4.2. *Systematic Random Sampling*

Jika ukuran sampel telah diketahui, maka hal yang perlu ditentukan selanjutnya adalah menentukan metode penarikan sampel. Salah satu metode penarikan sampel adalah dengan metode *Systematic Random Sampling*. Metode pengambilan sampel ini dengan cara mengambil sampel dengan interval

tertentu. Berikut ini langkah umum pengambilan sampel dengan metode *systematic random sampling* (gunadarma):

- a. Daftar seluruh anggota populasi dan beri nomor secara acak.
- b. Tetapkan interval pengambilan satu sampel, dengan cara berikut:

$$k = \frac{N}{n} \quad (2.13)$$

- c. Pilih secara acak sebuah angka yang menjadi *random start* yang menjadi nilai awal sekaligus satuan sampel yang diambil pertama. Nilainya lebih besar 1 namun lebih kecil dari nilai  $k$  ( $1 < \text{random start} < k$ ).
- d. Satuan sampel lainnya dipilih secara sistematis sesuai dengan intervalnya, yakni dengan cara menambahkan angka yang terpilih dengan nilai  $k$ .

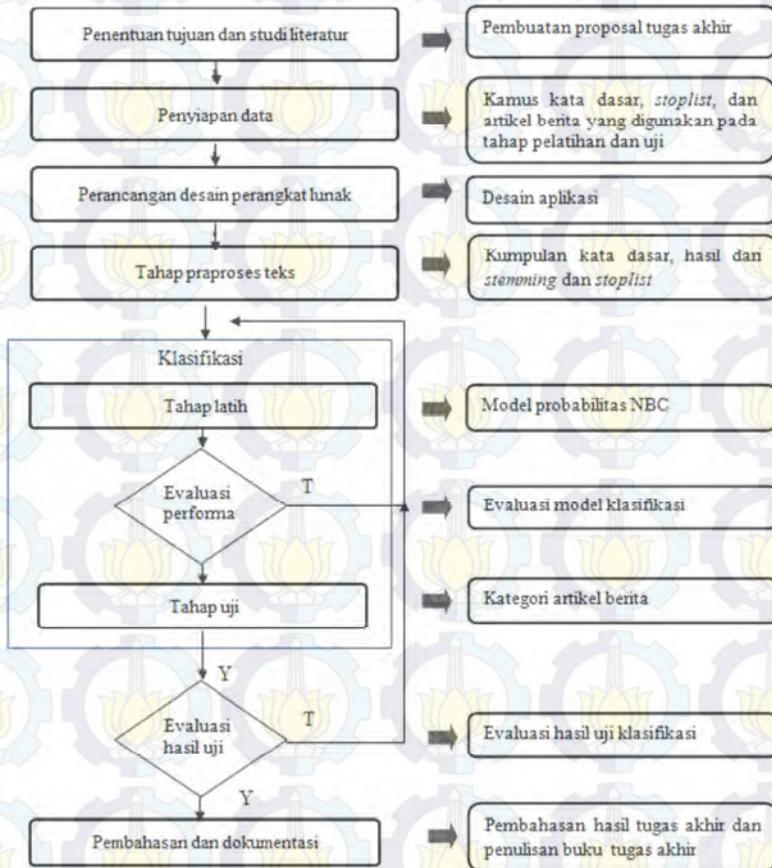
## 2.5. WEKA

WEKA (*Waikato Environment for Knowledge Analysis*) merupakan aplikasi perangkat lunak yang berbasis GUI (*Graphical User Interface*) yang memiliki koleksi berbagai macam algoritma *machine learning* untuk penggalian data maupun penggalian teks. WEKA terdiri dari berbagai *tools* untuk praproses data, klasifikasi, regresi, klustering, aturan asosiasi, dan visualisasi. (University of Waikato). Beberapa kelebihan aplikasi ini adalah kemudahan penggunaannya yang didukung GUI yang baik, *open-source*, dibangun dengan pemrograman JAVA sehingga dapat digunakan pada semua platform, dan koleksi algoritma yang lengkap.

Karena aplikasi WEKA bersifat *open-source* maka tim pengembang WEKA menyediakan *library* yang dapat digunakan oleh siapapun. Dan seperti yang diketahui bahwa WEKA dibangun dengan menggunakan bahasa pemrograman JAVA, maka untuk menggunakan *library* ini juga digunakan bahasa pemrograman JAVA.

### BAB III METODOLOGI Pengerjaan TUGAS AKHIR

Tugas akhir ini diselesaikan melalui berbagai tahap pengerjaan yang dapat dilihat pada gambar 3.1. Pada gambar tersebut juga dijelaskan keluaran dari masing-masing tahap yang terlihat pada bagian kanan dari alur pengerjaan.



Gambar 3.1 Metodologi Pengerjaan Tugas Akhir

Berikut ini merupakan penjelasan dari tahapan pengerjaan tugas akhir yang telah digambarkan pada gambar 3.1.

### 3.1. Penentuan Tujuan dan Studi Literatur

Tujuan tugas akhir ini berdasarkan permasalahan yang didapatkan dari membaca beberapa literatur mengenai klasifikasi dokumen. Banyak penelitian klasifikasi dokumen masih menggunakan korpus berbahasa Inggris atau bahasa asing lainnya. Penelitian dengan korpus bahasa Indonesia masih terbatas jumlahnya. Dari latar belakang tersebut munculah permasalahan, yakni mengimplementasikan klasifikasi dokumen pada artikel berita berbahasa Indonesia. Adapun batasan pengerjaan tugas akhir ini telah dijelaskan pada bagian 1.3 mengenai batasan pengerjaan tugas akhir.

Pengerjaan tugas akhir ini mengacu pada beberapa artikel jurnal, yakni "*Automatic Arabic Document Categorization Based on the Naïve Bayes Algorithm*" oleh M. Kourdi, dkk untuk panduan klasifikasi menggunakan NBC dan "*A Study of Stemming Effect on Information Retrieval in Bahasa Indonesia*" oleh F. Tala untuk daftar *stopwords* pada tahap praproses teks. "*Stemming Indonesian*" oleh J. Asian, dkk untuk proses *confix-stripping stemmer* yang merupakan pengembangan dari *stemmer* yang diperkenalkan oleh nazief.

Tugas akhir ini dibangun dengan menggunakan bahasa pemrograman Java. Hal ini dikarenakan pada proses klasifikasi akan digunakan *library* WEKA yang membantu proses klasifikasi pada tahap latih dan tahap uji serta perhitungan performa dan WEKA dibangun dengan bahasa pemrograman Java sehingga akan memudahkan dalam menggunakan *library* WEKA tersebut.

### 3.2. Penyiapan Data

Data-data yang digunakan untuk pengerjaan tugas akhir ini harus disiapkan sebelum aplikasi dibangun. Adapun data-data yang harus disiapkan adalah sebagai berikut:

a. Artikel berita

Artikel berita merupakan korpus berbahasa Indonesia yang didapatkan dari media *online* Kompas.com. Artikel berita yang diambil untuk tahap pelatihan dan pengujian merupakan sampel dari masing-masing kategori berita dari pertama kali berita pada kategori tersebut diunggah hingga akhir tahun 2013. Kategori artikel berita yang diambil sebanyak 12 kategori, yakni kategori berita nasional, internasional, olahraga, sains, edukasi, ekonomi, tekno, *entertainment*, otomotif, *health*, properti, dan travel. Jumlah sampel dari masing-masing kategori bervariasi tergantung dari hasil perhitungan sampel dengan menggunakan metode Slovin. Penarikan sampel artikel berita menggunakan metode *systematic random sampling*. Data sampel tersebut dibagi menjadi data latih dan data uji yang proporsinya 70:30 (Liu, Li, Lee, & Yu, 2004).

Pada tahap latih, artikel berita telah diketahui kategorinya. Hal tersebut berbeda dengan artikel berita pada tahap uji yang tidak diketahui kategorinya. Artikel berita pada tahap uji tersebut akan dikategorikan melalui model NBC yang didapatkan pada tahap latih.

b. Daftar *stopwords*

Daftar *stopwords* ini didapatkan pada tesis F. Tala yang berjudul “*A Study of Stemming Effect on Information Retrieval in Bahasa Indonesia*”.

c. Kamus kata dasar

Kamus kata dasar dibuat menyesuaikan dengan Kamus Besar Bahasa Indonesia.

### 3.3. Perancangan Desain Perangkat Lunak

Sebelum membangun aplikasi klasifikasi artikel berita, perlu dibuat desain dari aplikasi tersebut. Hal ini dilakukan sehingga semua kebutuhan yang akan diselesaikan oleh aplikasi tidak ada yang tertinggal. Desain ini mencakup desain basis

data, diagram *usecase*, skenario *usecase*, diagram sekuen, diagram aktivitas, dan kelas diagram.

### 3.4. Praproses Teks

Tahap ini dilakukan sebelum proses klasifikasi dilakukan. Tahap ini terbagi menjadi beberapa sub-tahapan yang telah dijelaskan pada 2.1.1 yang menjelaskan mengenai tahapan pada praproses, yakni sebagai berikut:

- a. Dilakukan proses *Case Folding*, yakni teks pada artikel berita akan diubah menjadi huruf kecil (*toLowerCase*) dan seluruh tanda baca dan karakter selain huruf 'a' hingga 'z' akan dihilangkan.
- b. Kemudian dilakukan proses *tokenizing* untuk memecah kalimat menjadi kata per kata.
- c. Proses *stemming* pada kata-kata yang tersisa pada dokumen teks untuk mendapatkan kata dasar. Pada tahap ini dilakukan *confix-stripping stemmer* yang akan memenggal kata per kata untuk mendapatkan kata dasar. Proses ini membutuhkan kamus kata dasar yang telah disiapkan sebelumnya untuk membandingkan hasil pemenggalan kata dengan kata dasar yang ada. Setelah proses *stemming* selesai dilakukan, dilakukan evaluasi terhadap hasil dari proses *stemming* tersebut. Pada evaluasi ini dicari apakah proses *stemming* yang dilakukan oleh sistem telah berhasil dengan baik.
- d. Kemudian dilakukan proses *stopping* berdasarkan *stoplist* yang berisi *stopwords* yang telah ditentukan sebelumnya. Kata-kata yang terdapat pada artikel berita akan dibandingkan dengan daftar *stopwords*, jika terdapat kata-kata yang terdapat pada *stopwords* maka kata tersebut akan dihapus dari artikel berita.
- e. Setelah proses *stopping* dilakukan, maka langkah selanjutnya adalah evaluasi hasil *stopping*. Evaluasi ini untuk mencari ada tidaknya *stopwords* yang tertinggal.

Jika ada, maka *stopwords* tersebut ditambahkan pada daftar *stopwords*.

### 3.5. Proses Klasifikasi

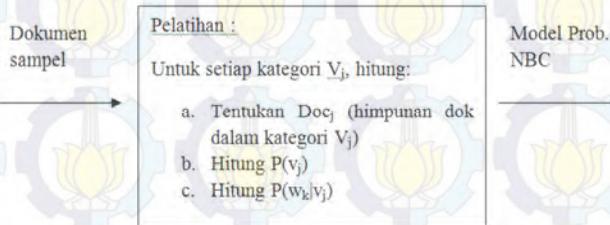
Pada tahap ini dilakukan proses klasifikasi dari artikel-artikel berita yang telah diolah sebelumnya pada tahap praproses teks. Proses klasifikasi akan menggunakan fungsi penggalian data yang dimiliki oleh *library* WEKA. Metode yang digunakan adalah NBC. Proses klasifikasi akan terbagi menjadi dua tahap, yakni tahap latih atau pembuatan model dan tahap uji yang merupakan tahap klasifikasi. Selain itu juga dilakukan evaluasi dari model yang telah dibuat pada tahap latih. Evaluasi ini juga dilakukan terhadap hasil klasifikasi tahap uji.

#### 3.5.1. Tahap Pelatihan

Pada tahap ini akan dilakukan tahap latih terhadap korpus atau artikel berita yang telah diketahui kategorinya, sehingga didapatkan model probabilitas NBC yang nantinya akan digunakan pada tahap uji. Tujuan dari tahap ini adalah untuk pembangunan model dasar yang digunakan pada saat tahap uji. Berikut ini merupakan tahapan pada proses latih:

- a. Artikel berita yang digunakan untuk tahap latih diambil dari basis data.
- b. Artikel berita tersebut akan menjalani tahap praproses teks, yakni proses *case folding*, *tokenizing*, *stoplist*, dan *stemming*.
- c. Kata-kata hasil praproses akan disimpan pada basis data sebagai kata fitur yang nantinya akan digunakan untuk proses klasifikasi.
- d. Menghitung probabilitas dari  $V_j$ , dimana  $V_j$  merupakan kategori berita, yakni  $j_1 =$  nasional,  $j_2 =$  ekonomi, dan sebagainya.
- e. Menghitung probabilitas kata  $w_k$  pada kategori  $v_j$ .
- f. Hasil keluaran, yakni model probabilitas NBC akan disimpan untuk digunakan pada tahap pengujian.

Proses latih tersebut dapat digambarkan pada gambar 3.2 (Hamzah, 2012).



**Gambar 3.2 Proses latih**

Rumus perhitungan telah dijelaskan pada bagian 2.2, yakni mengenai NBC. Pada pengerjaan tugas akhir ini, perhitungan tersebut akan dibantu dengan fungsi yang disediakan oleh *library* WEKA.

### 3.5.2. Evaluasi Model

Evaluasi ini dilakukan untuk mengetahui performa dari model yang dibuat dengan metode NBC pada tahap latih. Metode evaluasi yang digunakan adalah *cross-validation*. Dimana semua data yang terkumpul akan dibagi ke dalam sepuluh kelompok (10 *folds*) atau dapat disebut sebagai *10-cross-validation* (Refaeilzadeh, Tang, & Liu, 2009). Secara bergantian sembilan kelompok akan dijadikan data latih dan satu kelompok akan dijadikan data uji. Sedangkan evaluasi performa menggunakan akurasi, *precision*, *recall*, dan *F-Measure* untuk mengetahui seberapa akurat model klasifikasi dalam melakukan klasifikasi artikel berita ke dalam 12 kategori yang telah ditentukan sebelumnya. Rumus perhitungan akurasi, *precision*, *recall*, dan *F-Measure* telah dibahas pada bagian 2.3 yang membahas mengenai pengukuran performa. Hasil evaluasi performa yang dihasilkan merupakan hasil rata-rata dari *cross-validation*.

Setelah hasil evaluasi didapatkan selanjutnya hasil tersebut akan dianalisis untuk perbaikan pembuatan model klasifikasi.

Apabila hasil analisis tersebut menunjukkan hasil kurang baik maka akan dilakukan perbaikan pada pembuatan model klasifikasi.

### 3.5.3. Tahap Pengujian

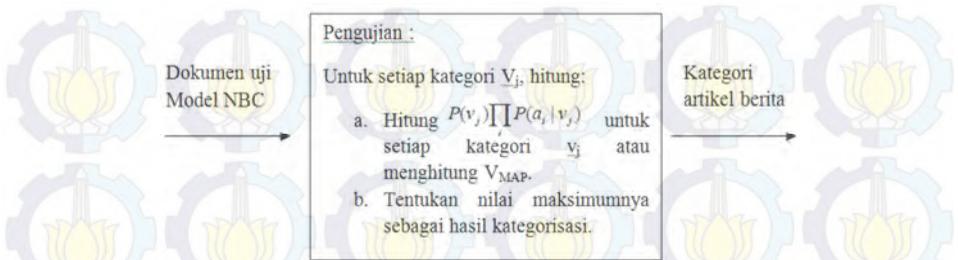
Setelah diketahui model probabilitas NBC pada tahap pelatihan, maka model tersebut akan digunakan pada tahap pengujian. Korpus yang digunakan pada tahap ini merupakan artikel berita berbahasa Indonesia yang belum diketahui kategorinya. Dengan mengacu pada model probabilitas yang merupakan keluaran pada tahap pelatihan, artikel berita yang belum diketahui ategorinya tersebut akan dikategorikan. Berikut ini merupakan proses pada tahap pengujian:

- a. Artikel berita yang digunakan untuk tahap pengujian diambil dari basis data.
- b. Artikel berita tersebut akan menjalani tahap praproses teks, yakni proses *case folding*, *tokenizing*, *stoplist*, dan *stemming*.
- c. Menghitung probabilitas tertinggi dari semua kategori yang diujikan ( $V_{MAP}$ ).
- d. Mencari nilai  $V_{MAP}$  paling maksimum, dan masukkan artikel berita tersebut pada kategori dengan  $V_{MAP}$ .

Proses pengujian tersebut dapat digambarkan pada gambar 3.3 (Hamzah, 2012). Rumus perhitungan telah dijelaskan pada bagian 2.2, yakni mengenai NBC. Pada pengerjaan tugas akhir ini, perhitungan tersebut akan dibantu dengan fungsi yang disediakan oleh WEKA.

### 3.5.4. Evaluasi Hasil Klasifikasi

Evaluasi ini dilakukan untuk mengetahui akurasi dari hasil keluaran pada tahap uji. Selain itu, juga dilakukan evaluasi *presision*, *recall*, dan *F-Measure* untuk mengetahui performa aplikasi pengklasifikasi dalam melakukan klasifikasi artikel berita ke dalam 12 kategori yang telah ditentukan sebelumnya. Rumus



**Gambar 3.3** Proses Uji

perhitungan akurasi, *precision*, *recall*, dan *F-Measure* telah dibahas pada bagian 2.3 yang membahas mengenai pengukuran performa.

### 3.6. Pembahasan dan Dokumentasi

Setelah aplikasi telah dibangun seluruhnya dan klasifikasi menghasilkan performa yang baik maka dilakukan analisis dan pembahasan mengenai hasil dari proses pengklasifikasian artikel berita. Setelah proses pembahasan dan evaluasi telah selesai dilakukan maka tahap selanjutnya adalah penulisan laporan tugas akhir yang merupakan dokumentasi dari pengerjaan tugas akhir. Dokumentasi ini mencakup keseluruhan proses pengerjaan tugas akhir dan analisis hasil pengerjaan tugas akhir.

## BAB IV DESAIN APLIKASI

Pada bab ini dipaparkan mengenai desain aplikasi dalam penerapan klasifikasi artikel berita dengan metode *confix stripping stemmer* dan NBC.

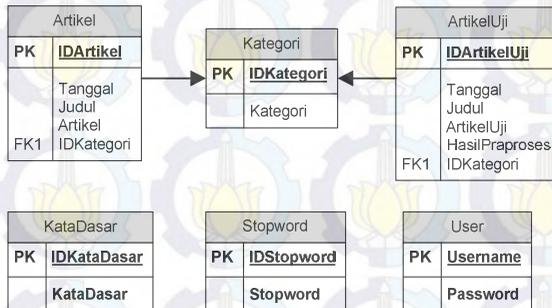
### 4.1. Desain Aplikasi Klasifikasi Artikel Berita

Sebelum pembangunan aplikasi, perlu dibuat perancangan desain aplikasi. Desain aplikasi terdiri dari diagram usecase, diagram diagram sekuen, diagram aktivitas, dan diagram kelas. Desain aplikasi ini diharapkan dapat membantu dalam pembangunan aplikasi yang dibuat.

#### 4.1.1. Desain Basis Data

Data-data masukan yang telah dijelaskan pada bagian penyiapan data disimpan dalam sebuah basis data MySQL. Nama basis data yang digunakan untuk menyimpan seluruh tabel adalah klasifikasi. Dalam basis data klasifikasi terdapat lima tabel yang digunakan dalam proses klasifikasi. Desain basis data dapat dilihat pada gambar 4.1. Terdapat dua relasi *oneToMany* antara tabel Kategori dan tabel Artikel serta tabel Kategori dan tabel ArtikelUji. Dua tabel lain, yakni tabel KataDasar dan tabel Stopword tidak memiliki relasi dengan tabel lainnya.

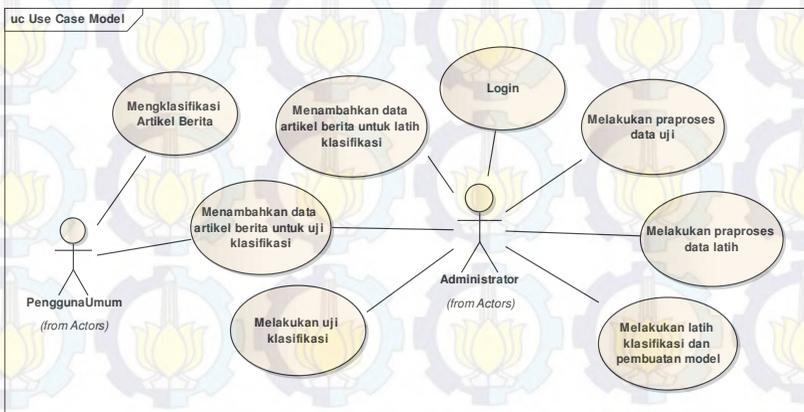
Tabel Artikel menyimpan semua artikel yang digunakan untuk data latih klasifikasi sedangkan tabel ArtikelUji digunakan untuk menyimpan data uji klasifikasi. Pada tabel Artikel terdapat lima atribut, yakni IDArtikel, Tanggal, Judul, ArtikelBerita, dan IDKategori yang merupakan *foreign key* dari tabel kategori. Pada tabel ArtikelUji terdapat enam atribut, yakni IDArtikelUji, Tanggal, Judul, ArtikelUji, HasilPraproses, dan IDKategori yang merupakan *foreign key* dari tabel kategori. Sedangkan seluruh kategori berita yang didefinisikan disimpan pada tabel kategori. Daftar kata dasar disimpan pada tabel KataDasar, sedangkan daftar stopword disimpan pada tabel Stopword.



Gambar 4.1 Desain Basis Data Klasifikasi

#### 4.1.2. Diagram Usecase

Diagram *usecase* ini digunakan untuk menggambarkan layanan apa saja yang disediakan oleh aplikasi. Dari diagram *usecase* ini dapat diketahui fungsionalitas dari aplikasi yang dibangun. Diagram *usecase* dari aplikasi pengklasifikasi artikel berita dapat dilihat pada gambar 4.2.



Gambar 4.2 Diagram Usecase

Pada aplikasi pengklasifikasi artikel berita ini, terdapat dua aktor yang terlibat, yakni pengguna umum dan administrator.

Pengguna memiliki enam fungsionalitas dalam aplikasi klasifikasi artikel berita. Keenam fungsionalitas tersebut dapat dilihat pada diagram *usecase* yang dapat dilihat pada gambar 4.2. Sedangkan skenario dari masing-masing *usecase* tersebut dapat dilihat pada lampiran A mengenai skenario *usecase*.

### 4.1.3. Diagram Sekuen

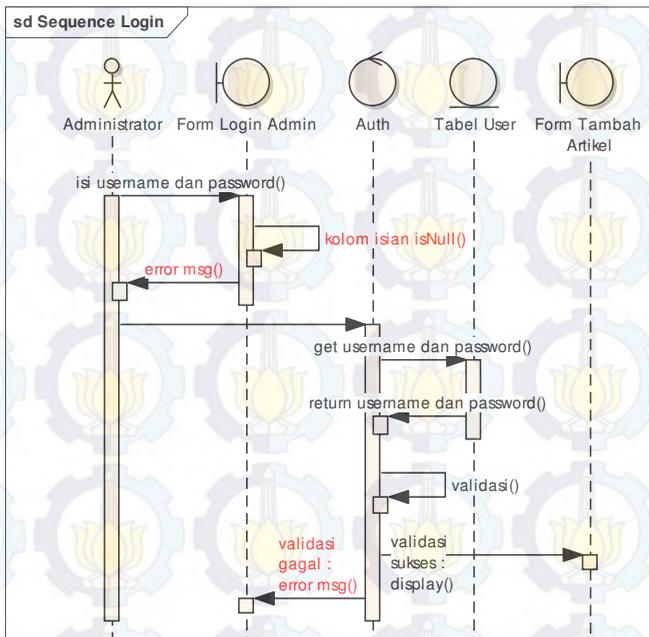
Aplikasi pengklasifikasi artikel berita memiliki enam diagram sekuen sesuai dengan jumlah *usecase*. Hal ini dikarenakan satu *usecase* menghasilkan satu diagram sekuen.

#### 4.1.3.1. Diagram Sekuen Login

Pengguna yang memerlukan login hanya administrator, hal ini terlihat pada gambar diagram *usecase* pada gambar 4.2. Administrator mengisi *username* dan *password* dan aplikasi akan melakukan pengecekan terhadap *username* dan *password* yang telah diisikan. Jika sesuai, maka aplikasi akan menampilkan borang Tambah Artikel Berita. Diagram sekuen login dapat dilihat pada gambar 4.3.

#### 4.1.3.2. Diagram Sekuen Menambah Artikel Berita untuk Data Latih.

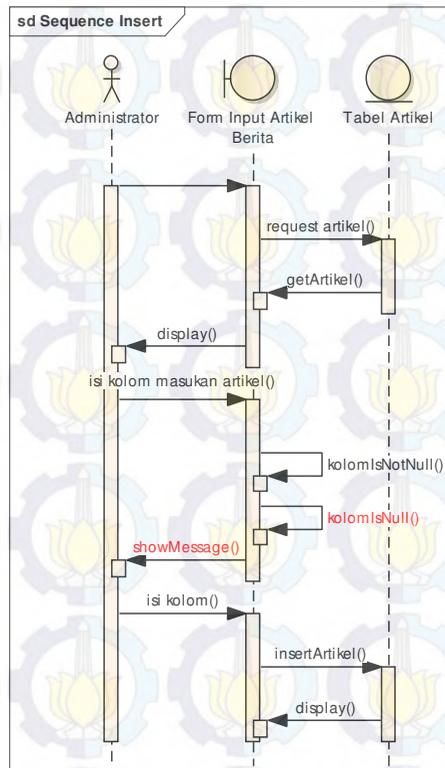
Pada diagram sekuen ini, borang yang diakses oleh pengguna adalah Tambah Artikel Berita. Pada borang tersebut, pengguna harus memilih pilihan Data Latih sehingga data yang diisikan akan ditambahkan untuk data latih. Kemudian pengguna mengisi seluruh kolom isian, jika tidak maka data masukan tidak dapat disimpan dan aplikasi akan menampilkan pesan kesalahan. Jika kolom isian telah terisi dengan benar, maka data isian akan disimpan pada tabel artikel. Aplikasi akan menampilkan pesan bahwa data telah sukses disimpan. Diagram sekuen ini dapat dilihat pada gambar 4.4.



Gambar 4.3 Diagram Sekuen Login

#### 4.1.3.3. Diagram Sekuen Menambah Artikel Berita untuk Data Uji

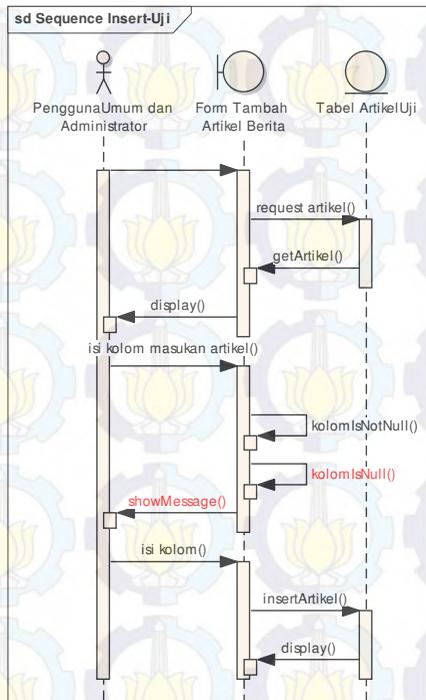
Pada diagram sekuen ini, borang yang diakses oleh pengguna adalah Tambah Artikel Berita. Pada borang tersebut, pengguna harus memilih pilihan Data Uji sehingga data yang diisikan akan ditambahkan untuk data uji. Kemudian pengguna mengisi seluruh kolom isian, jika tidak maka data masukan tidak dapat disimpan dan sistem akan menampilkan pesan kesalahan. Jika pengguna telah mengisi semua kolom isian dan menekan tombol untuk memulai proses, maka aplikasi akan menampilkan pesan bahwa data telah sukses disimpan. Diagram sekuen ini dapat dilihat pada gambar 4.5



Gambar 4.4 Diagram Sekuen Menambah Data Latih

#### 4.1.3.4. Diagram Sekuen Melakukan Praproses Klasifikasi pada Data Latih

Tahap praproses untuk data latih dirancang pada diagram sekuen yang dapat dilihat pada gambar 4.4. Pada diagram sekuen ini, pengguna pertama kali masuk pada borang Praproses Klasifikasi. Pengguna harus memilih pilihan Data Latih, sehingga praproses dilakukan terhadap data latih. Pengguna perlu mengisi lokasi file keluaran akan disimpan. Aplikasi akan melakukan pengecekan apakah semua kolom isian telah terisi, jika terdapat

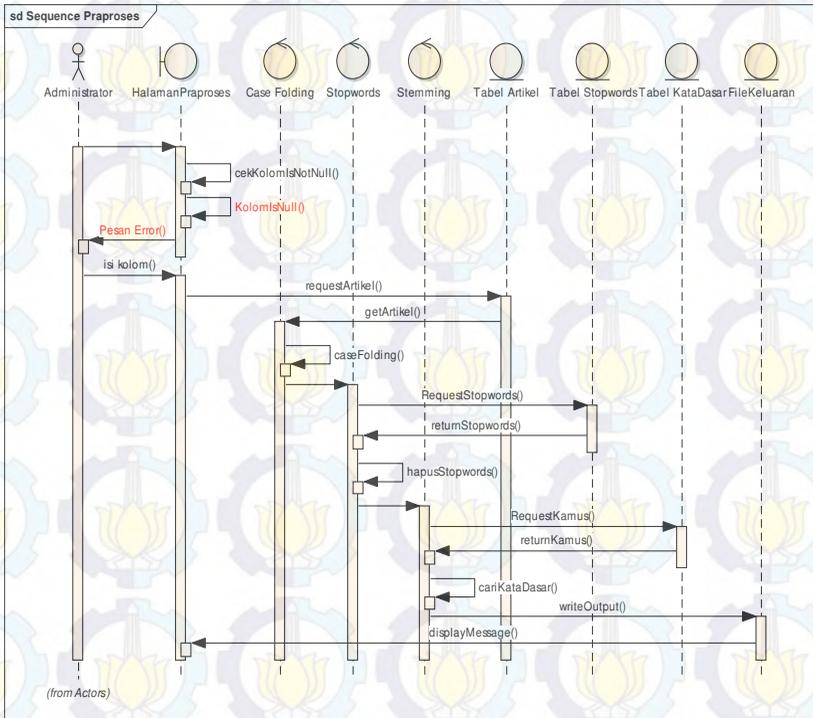


Gambar 4.5 Diagram Sekuen Menambah Data Uji

kolom isian yang kosong maka aplikasi akan menampilkan pesan kesalahan dan pengguna harus mengisi semua kolom isian.

Aplikasi mengambil artikel berita yang tersimpan di tabel artikel dan melakukan praproses. Tahap praproses telah dijelaskan pada 2.1.1 mengenai praproses teks yakni proses *case folding* dan *tokenizing*, yakni membuang karakter selain huruf, merubah seluruh huruf dalam artikel berita menjadi huruf kecil, dan memecah artikel menjadi per kata. Berikutnya adalah penghapusan *stopwords* yang terdapat dalam artikel berita. Daftar *stopwords* didapatkan dari tabel *stopword*. Setelah *stopwords* dihapus dari artikel berita, dilakukan pemecahan dari masing-masing kata untuk mencari bentuk kata dasarnya (*stemming*). Hal ini dibantu dengan membandingkan kata tersebut dengan daftar

kata dasar yang tersimpan di tabel kata dasar. Setelah proses tersebut selesai, sistem menyimpan hasil dari semua proses ke dalam sebuah file dengan format .arff sehingga dapat diolah lebih lanjut. Diagram sekuen dapat dilihat pada gambar 4.6.



Gambar 4.6 Diagram Sekuen Praproses Data Latih

#### 4.1.3.5. Diagram Sekuen Melakukan Praproses Klasifikasi pada Data Uji

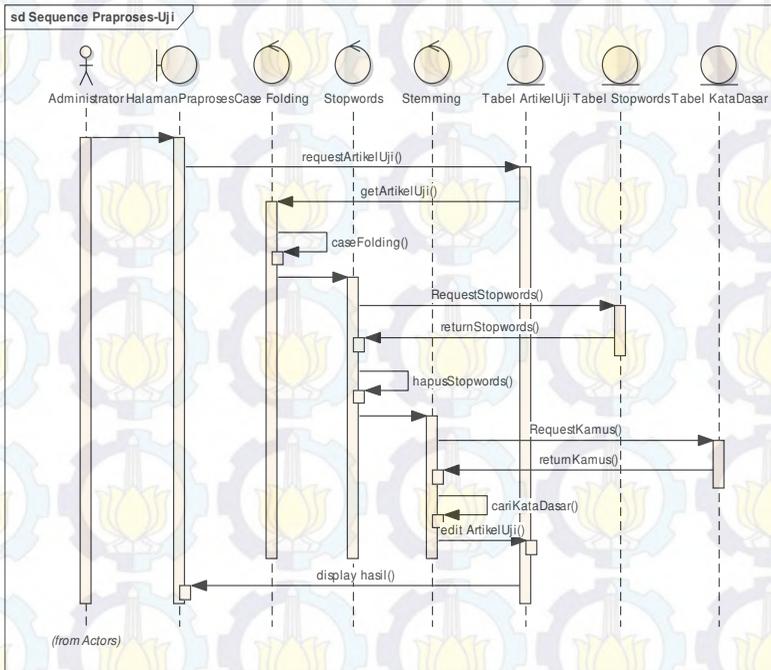
Tahap praproses untuk data latih dirancang pada diagram sekuen ini. Pada diagram sekuen ini, pengguna pertama kali masuk pada borang Praproses Klasifikasi. Pengguna harus memilih pilihan Data Uji, sehingga praproses dilakukan terhadap data uji dan kemudian menekan tombol OK. Proses selebihnya

sama seperti proses pada praproses terhadap data uji. Aplikasi akan melakukan pengecekan apakah semua kolom isian telah terisi, jika terdapat kolom isian yang kosong maka aplikasi akan menampilkan pesan kesalahan dan pengguna harus mengisikan semua kolom isian. Aplikasi mengambil artikel berita yang tersimpan dalam tabel artikel dan melakukan praproses. Tahap praproses telah dijelaskan pada 2.1.1 mengenai praproses teks yakni proses *case folding dan tokenizing*, yakni membuang karakter selain huruf, merubah seluruh huruf dalam artikel berita menjadi huruf kecil, dan memecah artikel menjadi per kata. Proses berikutnya adalah penghapusan *stopwords* yang terdapat dalam artikel berita. Daftar *stopwords* didapatkan dari tabel *stopword*. Setelah *stopwords* dihapus dari artikel berita, dilakukan pemecahan dari masing-masing kata untuk mencari bentuk kata dasarnya atau disebut sebagai proses *stemming*. Hal ini dibantu dengan membandingkan kata tersebut dengan daftar kata dasar yang tersimpan dalam tabel kata dasar. Setelah proses tersebut selesai, sistem menyimpan hasil dari semua proses ke dalam tabel ArtikelUji pada kolom hasilPraproses. Diagram sekuen dapat dilihat pada gambar 4.7.

#### **4.1.3.6. Diagram Sekuen Melakukan Latih Klasifikasi**

Pada tahap pengguna melakukan latih klasifikasi, pengguna harus membuka borang Latih Klasifikasi. Pengguna harus mengisikan lokasi file .arff tempat menyimpan hasil dari praproses. Selain itu, pengguna juga harus mengisikan lokasi tempat file model akan disimpan, file model ini berekstensi .dat. Jika kolom isian telah terisi, maka proses latih atau pembuatan model dapat dimulai. Aplikasi akan membaca isi dari file yang menyimpan data latih, kemudian melakukan evaluasi klasifikasi terhadap data latih tersebut. Setelah tahap evaluasi selesai, tahap berikutnya adalah membuat model klasifikasi berdasarkan data latih tersebut.

Setelah model selesai dibuat, aplikasi melakukan penyimpanan model klasifikasi tersebut pada file keluaran yang



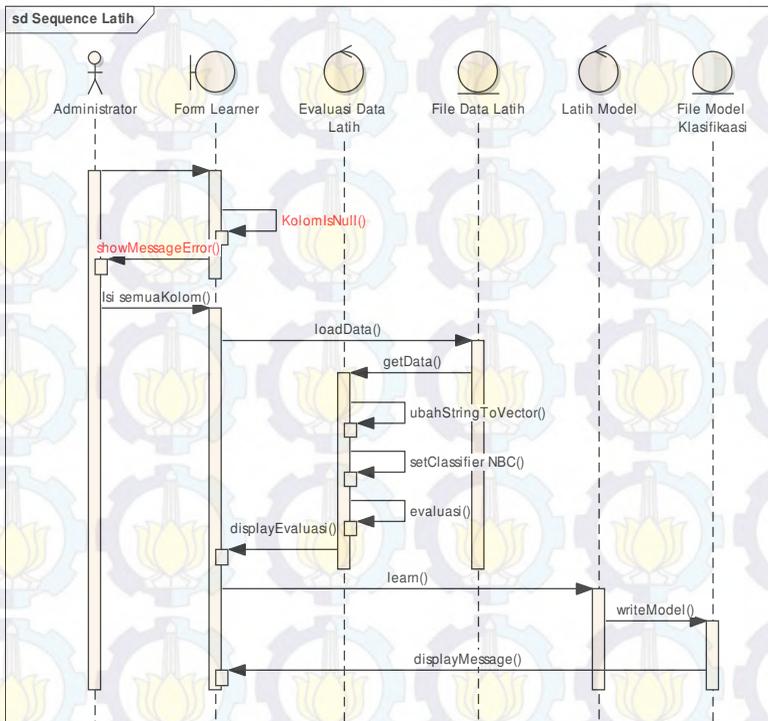
**Gambar 4.7 Diagram Sekuen Praproses Data Uji**

lokasi file-nya telah dimasukkan oleh pengguna. Diagram sekuen latihan klasifikasi dapat dilihat pada gambar 4.8.

#### 4.1.3.7. Diagram Sekuen Uji Klasifikasi

Pengguna mengisi lokasi file model klasifikasi pada borang Klasifikasi. Jika kolom isian telah terisi semua, maka sistem akan membaca data uji yang tersimpan dalam tabel artikeluji dan kemudian membaca model klasifikasi yang tersimpan dalam file model klasifikasi.

Masing-masing baris data uji akan diubah menjadi *instance* kemudian dibandingkan dengan model klasifikasi dan

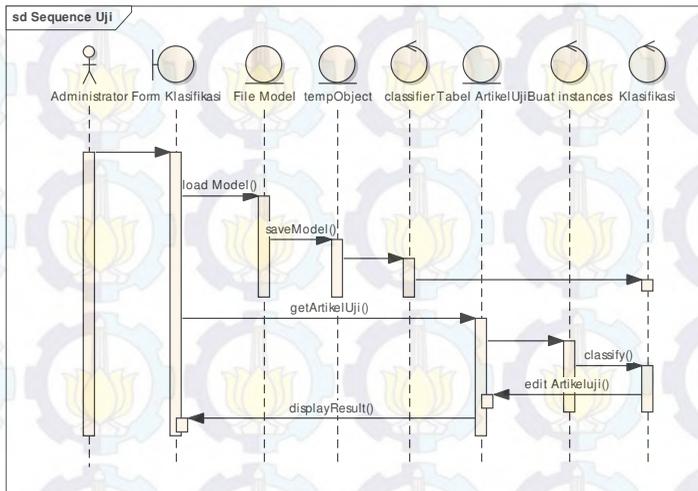


**Gambar 4.8 Diagram Sekuen Latih**

aplikasi dapat menentukan kategori dari data tersebut. Aplikasi akan menampilkan kategori yang merupakan hasil dari tahap uji klasifikasi. Diagram sekuen uji klasifikasi dapat dilihat pada gambar 4.9.

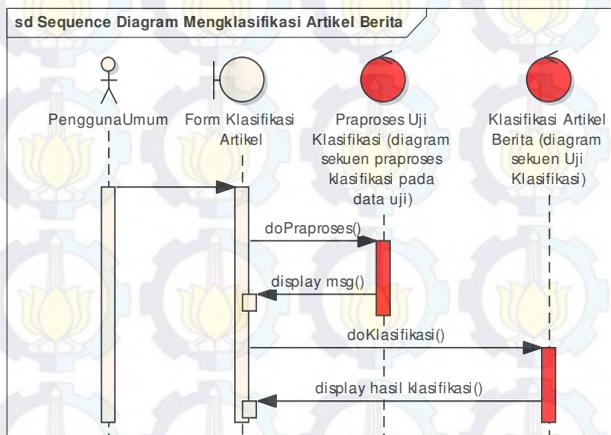
#### 4.1.3.8. Diagram Sekuen Mengklasifikasikan Artikel Berita

Pengguna yang terlibat adalah pengguna umum. Pengguna umum dapat melakukan klasifikasi dari artikel berita yang telah ditambahkan pada data uji. Pengguna umum menggunakan borang Klasifikasi Artikel Berita kemudian melakukan praproses dan mengklasifikasi artikel berita. Pada diagram sekuen ini terlibat dua aktivitas yang merujuk pada dua diagram sekuen lain, yakni diagram sekuen praproses klasifikasi



**Gambar 4.9 Diagram Sekuen Uji Klasifikasi**

pada diagram uji yang telah dijelaskan pada 4.1.3.4 dan diagram sekuen uji klasifikasi yang dijelaskan pada 4.1.3.7. Gambar diagram sekuen ini dapat dilihat pada gambar 4.10.



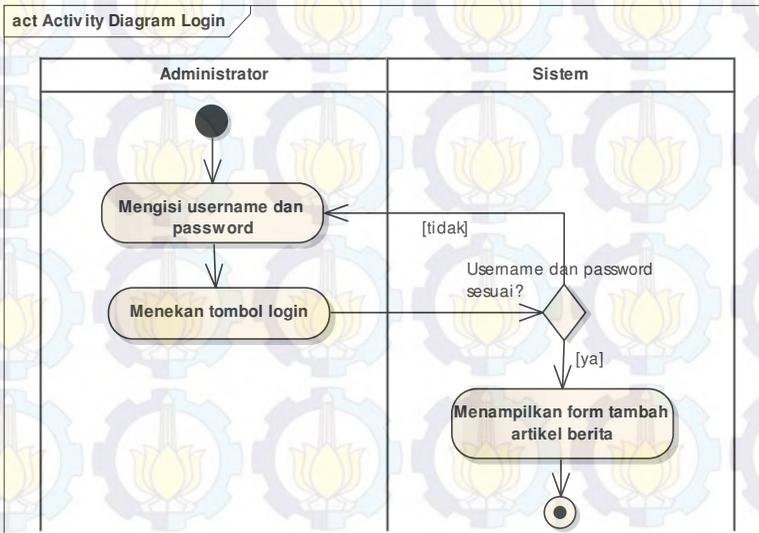
**Gambar 4.10 Digram Sekuen Mengklasifikasi Artikel Berita**

### 4.1.1. Diagram Aktivitas

Pada diagram aktivitas digambarkan mengenai alur proses dari sisi pengguna dan sistem. Masing-masing *usecase* digambarkan dengan satu diagram aktivitas.

#### 4.1.1.1. Diagram Aktivitas Login

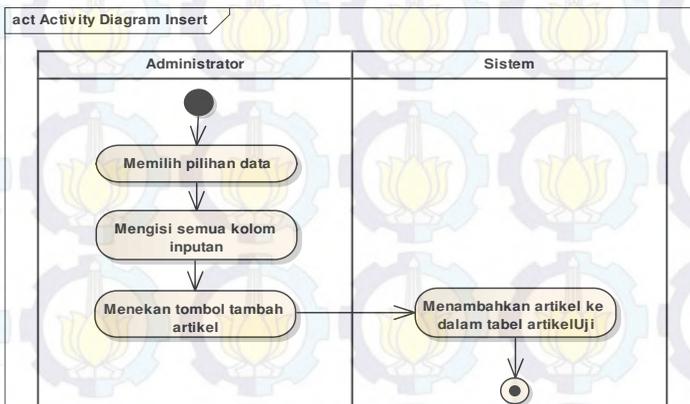
Administrator harus melakukan login untuk dapat melakukan penambahan artikel berita baik untuk data latih dan data uji, praproses, latih klasifikasi, dan uji klasifikasi. Untuk dapat melakukan login, administrator harus mengisi *username* dan *password*, kemudian aplikasi akan melakukan pengecekan terhadap *username* dan *password* tersebut. Jika sesuai, maka aplikasi akan menampilkan borang Tambah Artikel Berita. Alur aktivitasnya dapat dilihat pada gambar 4.11.



Gambar 4.11 Diagram Aktivitas Login

#### 4.1.1.2. Diagram Aktivitas Menambah Artikel Berita untuk Data Latih

Pengguna dapat menambahkan data artikel berita yang digunakan pada tahap latih melalui borang Tambah Artikel Berita. Alur prosesnya digambarkan pada diagram aktivitas menambah artikel berita. Pengguna memilih pilihan data yang akan ditambahkan kemudian mengisi kolom isian dan menekan tombol Tambah Artikel sehingga data artikel dapat disimpan dan ditambahkan pada tabel artikel. Diagram aktivitas menambah artikel berita dapat dilihat pada gambar 4.12.

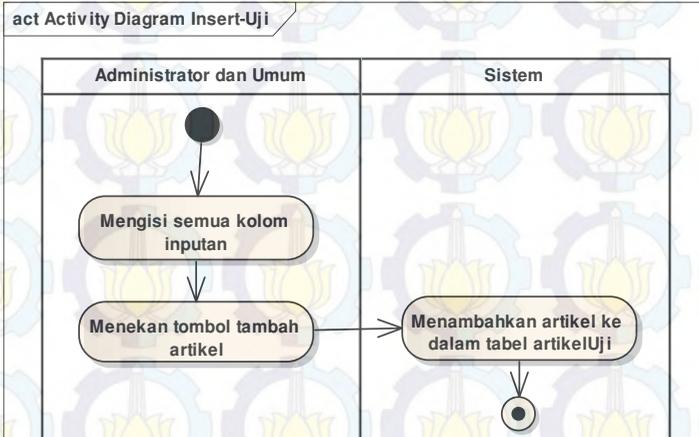


Gambar 4.12 Diagram Aktivitas Menambah Artikel Berita untuk Data Latih

#### 4.1.1.3. Diagram Aktivitas Menambah Artikel Berita untuk Data Uji

Pengguna, yakni administrator dan pengguna umum, dapat menambahkan data artikel berita yang akan digunakan pada tahap latih. Alur prosesnya digambarkan pada diagram aktivitas menambah artikel berita. Pada diagram ini terdapat prakondisi pilihan data telah dipilih yakni data uji. Pengguna memilih pilihan data yang akan ditambahkan kemudian mengisi kolom isian dan menekan tombol tambah artikel sehingga data artikel dapat

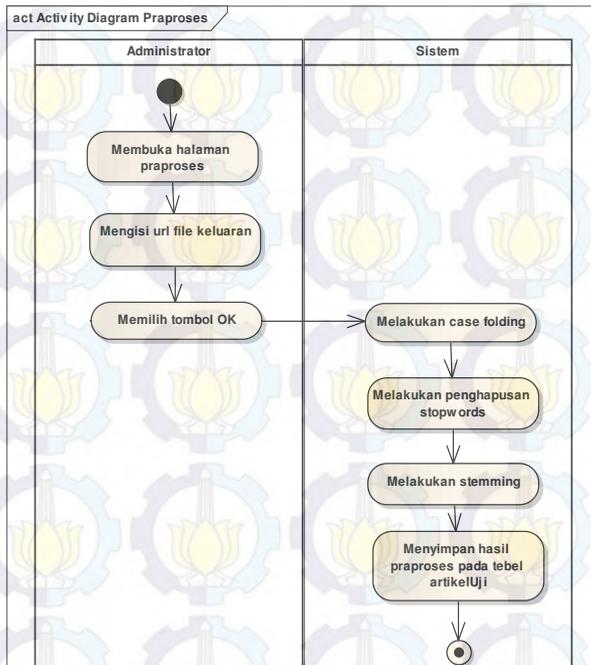
disimpan dan ditambahkan pada tabel `artikelUji`. Diagram aktivitasnya dapat dilihat pada gambar 4.13.



Gambar 4.13 Diagram Aktivitas Menambah Artikel Berita untuk Data Uji

#### 4.1.1.1. Diagram Aktivitas Melakukan Praproses Data Latih

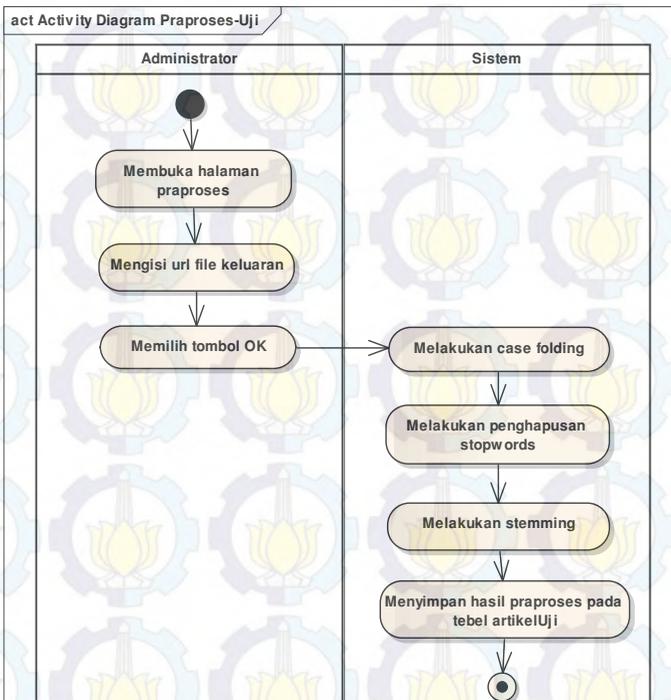
Artikel berita yang akan diolah untuk membuat model klasifikasi harus diolah terlebih dahulu. Proses pengolahan ini disebut dengan praproses klasifikasi yang di dalamnya terdapat beberapa proses seperti yang telah dijelaskan pada 2.1.1 mengenai praproses teks, yakni *case folding*, penghapusan *stopwords*, dan *stemming*. Aplikasi membaca isi dari tabel `KataDasar` dan tabel `stopword` untuk mendapatkan daftar kata dasar dan daftar *stopword*. Hasil keluaran dari praproses disimpan dalam satu file dengan format file `.arff`. Hal ini dikarenakan langkah pembuatan model hanya dapat mengolah data dari file dengan format `.arff`. Alur praproses ini digambarkan pada diagram aktivitas pada gambar 4.14.



Gambar 4.14 Diagram Aktivitas Praproses Data Latih

#### 4.1.1.1. Diagram Aktivitas Melakukan Praproses Data Uji

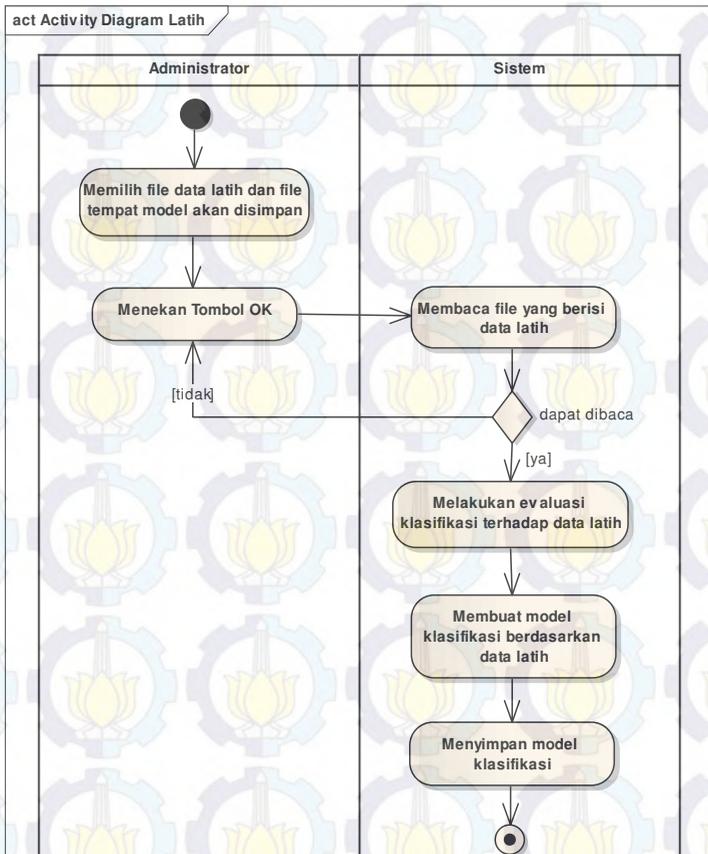
Artikel berita yang akan diolah untuk membuat model klasifikasi harus diolah terlebih dahulu. Proses pengolahan ini disebut dengan praproses klasifikasi yang di dalamnya terdapat beberapa proses, yakni *case folding*, penghapusan *stopwords*, dan *stemming*. Aplikasi membaca isi dari tabel KataDasar dan tabel *stopword* untuk mendapatkan daftar kata dasar dan daftar *stopword*. Hasil keluaran dari praproses disimpan dalam basis data. Alur praproses ini digambarkan pada diagram aktivitas pada gambar 4.15.



Gambar 4.15 Diagram Aktivitas Praproses Data Uji

#### 4.1.1.2. Diagram Aktivitas Melakukan Latih Klasifikasi

Untuk membuat model klasifikasi, data latih yang telah disimpan pada file .arff diolah pada proses latih klasifikasi. Pengguna dapat melakukan tahap latih melalui borang Latih Klasifikasi. Pengguna mengisi lokasi dari file yang berisi data latih dan lokasi dari file yang akan digunakan untuk menyimpan model klasifikasi. Kemudian setelah pengguna menekan tombol OK, maka aplikasi akan memproses data latih untuk mendapatkan model klasifikasi. Proses pembangunan model ini menggunakan metode NBC. Gambaran alurnya dapat dilihat pada gambar 4.16.

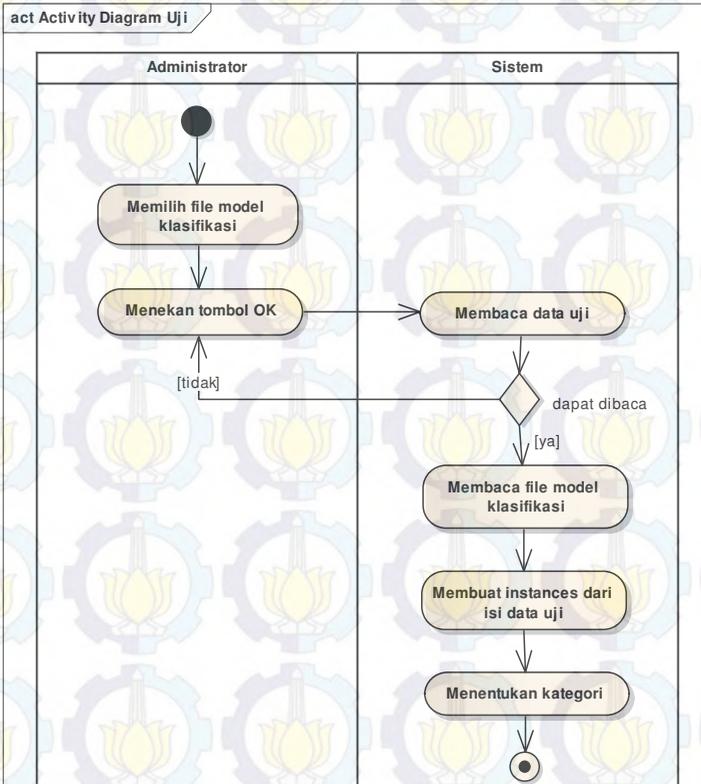


Gambar 4.16 Diagram Aktivitas Melakukan Latih Klasifikasi

#### 4.1.1.3. Diagram Aktivitas Melakukan Uji Klasifikasi

Administrator dapat mengetahui kategori dari suatu artikel berita yang menjadi data uji pada tahap ini, yakni dengan membuka borang Uji Klasifikasi dan mengisikan lokasi file uji, yakni artikel berita yang ingin diketahui kategorinya. Setelah pengguna menekan tombol mulai, aplikasi akan mencari kategori dari artikel berita. Alur proses uji klasifikasi dapat dilihat pada

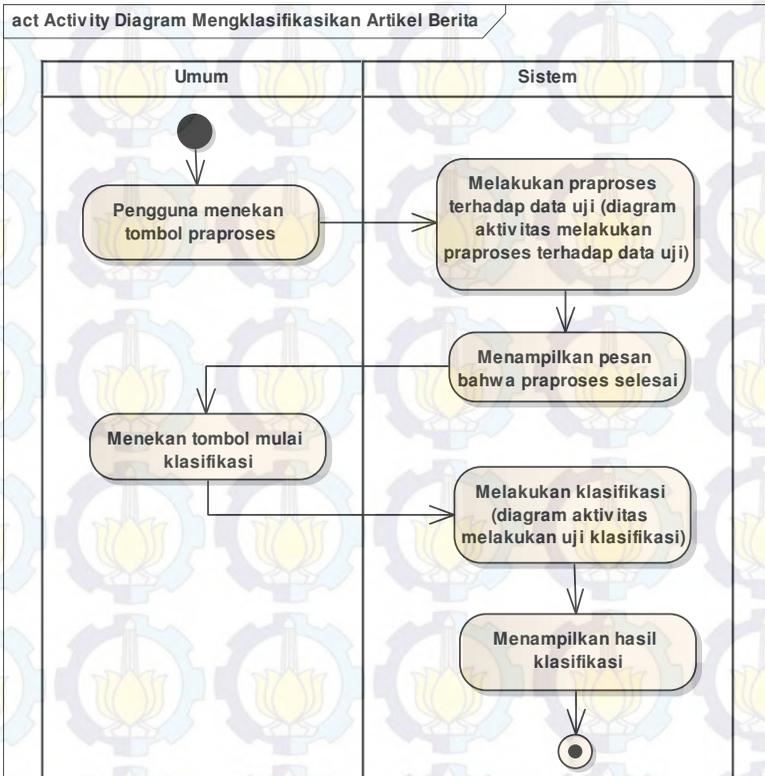
gambar 4.17 yang digambarkan dengan diagram aktivitas melakukan uji klasifikasi.



Gambar 4.17 Diagram Aktivitas Melakukan Uji Klasifikasi

#### 4.1.1.4. Diagram Aktivitas Mengklasifikasikan Artikel Berita

Pengguna umum mengklasifikasikan artikel berita melalui borang Klasifikasi Artikel Berita. Pengguna umum hanya perlu menekan tombol praproses yang akan melakukan proses praproses terhadap data uji, kemudian menekan tombol mulai klasifikasi yang akan melakukan proses klasifikasi artikel berita. Alur aktivitasnya dapat dilihat pada gambar 4.18.

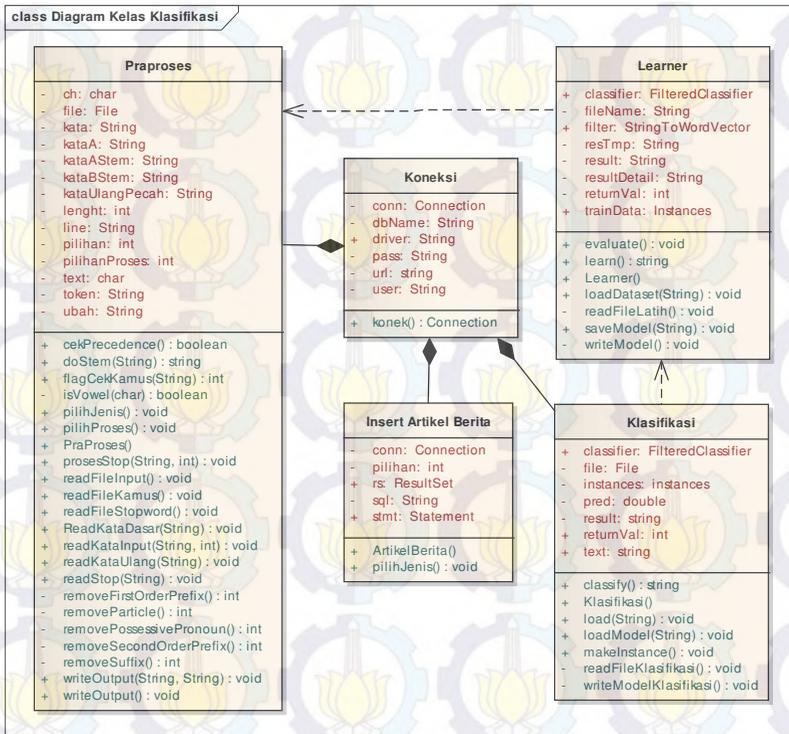


Gambar 4.18 Diagram Aktivitas Mengklasifikasikan Artikel Berita

#### 4.1.2. Diagram Kelas

Diagram kelas merupakan suatu diagram yang menggambarkan struktur dari aplikasi. Dalam diagram kelas digambarkan kelas-kelas apa saja yang berada dalam sistem aplikasi dan digambarkan juga hubungan antar kelas tersebut. Pada masing-masing kelas, dijabarkan atribut dan metode yang menjadi struktur kelas tersebut. Diagram kelas aplikasi pengklasifikasi artikel berita dapat dilihat pada gambar 4.19.

Pada aplikasi pengklasifikasi, terdapat lima kelas yang saling terkait sehingga aplikasi dapat berjalan dengan baik. Kelas-kelas yang membentuk klasifikasi artikel berita, yakni kelas koneksi yang membantu proses koneksi basis data dengan aplikasi, kelas praproses yang membantu untuk proses praproses klasifikasi, kelas *learner* merupakan kelas latih yang menghasilkan model klasifikasi dari data latih, dan kelas klasifikasi yang membantu proses tes klasifikasi.



Gambar 4.19 Diagram Kelas

## BAB V IMPLEMENTASI PERANGKAT LUNAK

Pada bab ini dijelaskan implementasi aplikasi pengklasifikasi artikel berbahasa Indonesia. Pada klasifikasi ini secara garis besar diterapkan dua metode, yakni metode *confix stripping stemmer* pada bagian praproses dan metode *Naïve Bayes Classifier* (NBC) pada bagian pembentukan model latih dan uji klasifikasi. Desain dari aplikasi telah dijelaskan pada bab sebelumnya. Pada bab ini dijelaskan mengenai implementasi dari desain sistem yang telah dibuat tersebut.

### 5.1. Persiapan dan Implementasi Data Awal

Data awal yang dibutuhkan pada aplikasi klasifikasi ini, yakni data artikel berita, data daftar *stopwords*, dan data daftar kata dasar bahasa Indonesia.

#### 5.1.1. Daftar *Stopwords* dan Daftar Kata Dasar

Untuk daftar *stopwords*, digunakan daftar *stopwords* yang telah didefinisikan dalam tesis F. Tala yang berjudul “*A Study of Stemming Effect on Information Retrieval in Bahasa Indonesia*”. Sedangkan untuk daftar kata dasar didapatkan dari kamus besar bahasa Indonesia. Kedua daftar kata-kata tersebut disimpan dalam dua tabel yang berbeda tabel *Stopword* dan tabel *Katadasar*.

**Tabel 5.1 Contoh Isi Tabel *Stopword***

ID <i>Stopword</i>	<i>Stopword</i>
1	ada
2	adalah
3	adanya
4	adapun
5	agak

### 5.1.2. Data Artikel Berita

Untuk masukan data artikel berita, data didapatkan dari portal berita *online* <http://www.kompas.com>. Artikel berita yang diambil merupakan hasil sampling dari masing-masing kategori. Kategori yang digunakan pada pengerjaan tugas akhir ini sebanyak 12 kategori berita dengan rentang waktu dari awal pengunggahan berita hingga akhir tahun 2013. Daftar kategori dan rentang waktu artikel berita dapat dilihat pada tabel 5.1.

**Tabel 5.2 Daftar Kategori Berita**

No	Kategori Berita	Rentang Waktu Berita
1	Nasional	4 Januari 2008 – 31 Desember 2013
2	Internasional	5 Januari 2008 – 31 Desember 2013
3	Olahraga	7 Januari 2008 – 31 Desember 2013
4	Sains	8 Januari 2008 – 31 Desember 2013
5	Edukasi	3 Januari 2011 – 31 Desember 2013
6	Ekonomi	7 Januari 2008 – 31 Desember 2013
7	Tekno	11 Februari 2008 – 31 Desember 2013
8	Entertainment	10 Januari 2008 – 31 Desember 2013
9	Otomotif	21 Januari 2008 – 31 Desember 2013
10	Health	21 Januari 2008 – 31 Desember 2013
11	Properti	1 Maret 2008 – 31 Desember 2013
12	Travel	2 Maret 2008 – 31 Desember 2013

Dari rentang waktu tersebut dihitung jumlah total artikel berita yang diunggah untuk mencari populasi artikel berita pada masing-masing kategori. Populasi dari masing-masing kategori perlu diketahui untuk mencari sampel artikel berita yang akan diambil. Sampel artikel berita pada masing-masing kategori dihitung dengan rumus Slovin yang telah dijelaskan pada 2.4.1 mengenai penentuan sampel. Hasil sampel artikel berita pada masing-masing kategori berita dapat dilihat pada tabel 5.2 dengan selang kepercayaan sebesar 90% atau dengan tingkat *error* sebesar 0.10. Jumlah sampel pada masing-masing kategori berbeda karena jumlah populasi pada masing-masing kategori

juga berbeda. Jumlah populasi ini dipengaruhi oleh rentang waktu artikel berita yang tersimpan yang dapat dilihat pada tabel 5.1 dan banyaknya artikel berita yang diunggah tiap harinya.

Hasil sampel dari masing-masing kategori dibagi menjadi dua, yakni data latih dan data uji. Proporsi pembagian data latih dan data uji adalah 70:30 pada masing-masing kategori. Jumlah sampel dan data latih serta data uji dapat dilihat pada tabel 5.2.

**Tabel 5.3 Jumlah Sampel per Kategori**

No	Kategori Berita	Jumlah Sampel	Jumlah Data Latih	Jumlah Data Uji
1	Nasional	100	70	30
2	Internasional	100	70	30
3	Olahraga	100	70	30
4	Sains	99	70	29
5	Edukasi	97	68	29
6	Ekonomi	100	70	30
7	Tekno	100	70	30
8	<i>Entertainment</i>	100	70	30
9	Otomotif	99	70	29
10	<i>Health</i>	100	70	30
11	Properti	100	70	30
12	Travel	100	70	30
TOTAL		1195	838	357

Penarikan sampel artikel berita pada masing-masing kategori menggunakan metode *systematic random sampling*. Sehingga diharapkan sampel dapat merepresentasikan keseluruhan artikel berita.

Komponen dari artikel berita yang disimpan adalah untuk data latih adalah tanggal, judul berita, isi berita, dan kategori berita. Artikel berita yang digunakan pada data latih telah diketahui kategori beritanya. Sedangkan komponen dari artikel berita yang disimpan untuk data uji adalah tanggal, judul berita,

dan isi berita. Hal ini dikarenakan pada data uji, kategori berita belum diketahui.

## 5.2. Implementasi Aplikasi Klasifikasi Artikel Berita

Setelah data telah disiapkan, maka tahap selanjutnya adalah pembangunan aplikasi klasifikasi artikel berita. Seluruh rancangan aplikasi telah dijelaskan pada bab desain aplikasi, dan berikut ini merupakan implementasinya.

### 5.2.1. Login

Proses login hanya diperuntukkan bagi administrator, pengguna umum tidak melakukan proses ini. Proses ini perlu dilakukan administrator untuk dapat melakukan semua aktivitas yang dapat dilakukan oleh administrator. Administrator mengisi *username* dan *password* pada borang Login seperti yang dapat dilihat pada gambar 5.1 berikut ini. Borang ini dibuat berdasarkan diagram sekuen 4.1.3.1 mengenai diagram sekuen login.



Gambar 5.1 Antar Muka Borang Login

Implementasi kode untuk menjalankan proses *otentikasi* login dapat dilihat pada segmen program 5.1 dan 5.2.

```
if(txtUser.getText().isEmpty() ||
txtPass.getPassword().toString().isEmpty()){
    JOptionPane.showMessageDialog(null, "Kolom isian tidak
boleh kosong", "Error", JOptionPane.ERROR_MESSAGE);
}else{
    String sql = "select * from user where username =" +
txtUser.getText() + "&& pass=" + txtPass.getPassword() + "";
```

Segmen Program 5.1 Login

```

ResultSet rs;
Statement st;
st = conn.createStatement();
rs = st.executeQuery(sql);
if (rs.wasNull()) {
    JOptionPane.showMessageDialog(null, "User dan password
salah", "Error", JOptionPane.ERROR_MESSAGE);
} else {
    this.dispose();
    new ArtikelBerita().setVisible(true);
}

```

### Segmen Program 5.2 Login (Lanjutan)

#### 5.2.2. Menambah Artikel Berita

Sebelum melakukan semua proses yang ada, data artikel berita yang merupakan data sampel harus disiapkan. Seluruh artikel berita yang akan digunakan sebagai data latih dan data uji disimpan dalam satu basis data. Untuk membantu pengguna menambahkan artikel berita ke dalam basis data telah disediakan satu borang, yakni borang Tambah Artikel Berita. Borang Tambah Artikel Berita dibuat berdasarkan diagram sekuen yang telah dibuat pada 4.1.3.2 dan 4.1.3.3. Tampilan antar muka pengguna untuk borang tambah artikel berita untuk data latih yang berdasarkan diagram sekuen yang dijelaskan pada 4.1.3.2, dapat dilihat pada gambar 5.2.

The screenshot shows a Java Swing window titled "Tambah Artikel Berita". The window contains a form with the following elements:

- Proses:** TAMBDAH ARTIKEL BERITA
- Pilihan Jenis Artikel:** A dropdown menu with "Data Latih" selected.
- Judul Artikel:** A text input field.
- Tanggal:** A date picker field with a mask of "\* yyyy-mm-dd".
- Nama Kategori:** A dropdown menu with "nasional" selected.
- Isi Berita:** A large text area for entering the article content.
- Buttons:** "Tambah Artikel" and "Reset" buttons at the bottom right.

Gambar 5.2 Antar Muka Tambah Artikel Berita untuk Data Latih

Pada borang tersebut terdapat beberapa kolom isian yang semuanya harus diisi oleh pengguna. Kolom isian ini menyangkut dari komponen artikel berita, yakni judul berita, tanggal yang ditulis dalam format ‘yyyy-mm-dd’, kategori berita, dan artikel berita. Selain itu juga terdapat pilihan jenis data yang akan ditambahkan. Jika pengguna akan menambahkan data latih maka aplikasi akan menampilkan pilihan kategori yang harus dipilih, namun jika pengguna akan menambahkan data uji maka pilihan kategori tidak akan ditampilkan. Antar muka borang Tambah Artikel sedikit berbeda jika pengguna memilih data uji, hal ini berdasarkan diagram sekuen yang dijelaskan pada 4.1.3.3. Tampilan dapat dilihat pada gambar 5.3 berikut ini.

The image shows a screenshot of a web application window titled "Tambah Artikel Berita". The window has a yellow title bar. Below the title bar, the text "Proses" is visible. The main content area is titled "TAMBAH ARTIKEL BERITA". There is a dropdown menu labeled "Pilihan Jenis Artikel:" with "Data Uji" selected. Below this are three input fields: "Judul Artikel:" (empty), "Tanggal:" (empty, with a placeholder "\* yyyy-mm-dd"), and "Isi Berita:" (a large empty text area). At the bottom right of the form area, there are two buttons: "Tambah Artikel" and "Reset".

**Gambar 5.3** Antar Muka Tambah Artikel Berita untuk Data Uji

Setelah pengguna mengisi semua isian yang tersedia, maka pengguna dapat menekan tombol Tambah Artikel. Jika terdapat kolom isian yang tidak terisi, maka aplikasi akan memberikan sebuah pesan kesalahan dan data yang diisikan tidak akan disimpan dalam basis data. Jika pengguna ingin mengosongkan kembali kolom isian terdapat tombol reset yang jika ditekan maka semua kolom isian akan dikosongkan. Penambahan data artikel ini telah digambarkan pada bab desain sistem. Implementasi

desain sistem untuk proses menambah artikel latih dapat dilihat pada segmen program 5.3.

```

if (txtJudul.getText().isEmpty() && txtBerita.getText().isEmpty() &&
txtTanggal.getText().isEmpty()) {
    JOptionPane.showMessageDialog(null, "Kolom inputan tidak
boleh kosong", "Error", JOptionPane.ERROR_MESSAGE);
} else {
    if (cmbPilihanJenis.getSelectedIndex() == 0) {
        conn = konek.konek();
        stmt = conn.createStatement();
        int pilihan = comboKategori.getSelectedIndex();
        String hitung = "select max(idArtikel)+1 from artikel";
        ResultSet rws = stmt.executeQuery(hitung);
        int max = 0;
        while (rws.next()) {
            max = rws.getInt(1);
        }
        String sql = "insert into artikel values (" + max + ", " +
txtTanggal.getText() + ", " + txtJudul.getText() + ", " +
+ "" + txtJudul.getText() + " " + txtBerita.getText() +
", " + (pilihan + 1) + ")";
        stmt.executeUpdate(sql);
        conn.close();
        JOptionPane.showMessageDialog(null, "Data telah
disimpan dalam basis data");
    }
}

```

#### Segmen Program 5.3 Tambah Artikel Berita untuk Data Latih

Sedangkan jika pengguna ingin menambahkan data artikel untuk data uji, implementasinya dapat dilihat pada segmen program 5.4 dan 5.5.

```

else if (cmbPilihanJenis.getSelectedIndex() == 1){
    conn = konek.konek();
    stmt = conn.createStatement();
    int pilihan = comboKategori.getSelectedIndex();
    String hitung = "select max(idArtikelUji)+1 from artikeluji";
    ResultSet rws = stmt.executeQuery(hitung);
}

```

#### Segmen Program 5.4 Tambah Artikel Berita untuk Data Uji (Lanjutan)

```

int max = 0;
while (rws.next()) {
    max = rws.getInt(1);
}
String sql = "insert into artikeluji (idartikeluji, tanggal, judul,
artikeluji) values "
            + "(" + max + "," + txtTanggal.getText() + "," +
txtJudul.getText() + ","
            + "" + txtJudul.getText() + "" + txtBerita.getText() + """;
stmt.executeUpdate(sql);
conn.close();
JOptionPane.showMessageDialog(null, "Data telah disimpan
dalam basis data");
}
}

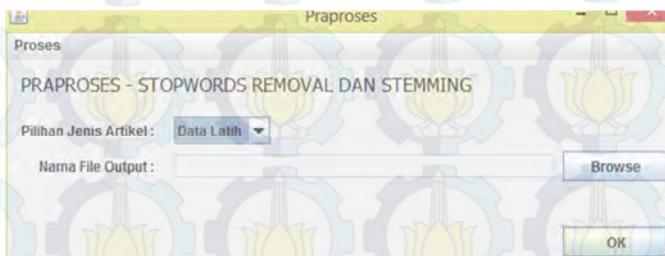
```

### Segmen Program 5.5 Tambah Artikel Berita untuk Data Uji

#### 5.2.3. Praproses

Setelah data masukan telah terkumpul dan tersimpan dalam basis data, maka tahap selanjutnya adalah praproses data yang telah disimpan tersebut. Data uji maupun data latih harus melalui tahap praproses. Hal ini dilakukan mengoptimalkan pembuatan model latih klasifikasi dan klasifikasi pada tahap uji. Tahap praproses ini terbagi menjadi empat tahap yang secara umum telah dijelaskan pada bab tinjauan pustaka pada bagian praproses.

Untuk membantu proses pada tahapan praproses dibuatlah aplikasi sederhana yang dibangun dengan bahasa pemrograman JAVA. Hasil tampilan antar muka pengguna untuk tahap praproses untuk data latih dapat dilihat 5.4 yang dibuat berdasarkan diagram sekuen yang dijelaskan pada 4.1.3.4.



Gambar 5.4 Antar Muka Praproses untuk Data Latih

Jika pengguna akan melakukan praproses terhadap data latih, maka pengguna harus mengisikan file keluaran yang akan menyimpan hasil praproses. Jika tidak maka sistem aplikasi akan menampilkan pesan kesalahan. Tetapi jika pengguna akan melakukan praproses terhadap data uji, maka aplikasi tidak akan menampilkan kolom isian untuk file keluaran. Hal ini dikarenakan hasil praproses untuk data uji akan disimpan dalam basis data. Antar muka pengguna untuk tahap praproses untuk data uji dapat dilihat 5.5 yang dibuat berdasarkan diagram sekuen yang dijelaskan pada 4.1.3.5.



**Gambar 5.5** Antar Muka Praproses untuk Data Uji

Tahap praproses untuk data latih maupun data uji tersebut dibangun berdasarkan kode program berikut ini.

```

if (cmbPilihanJenis.getSelectedIndex() == 0) {
    if (txtNamaFileOutput.getText().isEmpty()) {
        JOptionPane.showMessageDialog(null, "terdapat kolom yang
kosong", "Error", JOptionPane.ERROR_MESSAGE);
    } else {
        readKataInput(pilihan);
        //menuliskan hasilnya dalam file output
        writeOutput(txtNamaFileOutput.getText());
        JOptionPane.showMessageDialog(null, "praproses sukses");
    }
} else if (cmbPilihanJenis.getSelectedIndex() == 1) {
    readKataInput(pilihan);
    writeOutput();
    JOptionPane.showMessageDialog(null, "praproses sukses");
}

```

**Segmen Program 5.6** Menjalankan Praproses

Proses inti dari kedua jenis praproses adalah sama, yang berbeda hanyalah pada penyimpanan hasil praproses. Jika praproses terhadap data latih, hasil praproses akan disimpan dalam sebuah file keluaran. Hal ini berbeda pada praproses terhadap data uji yang hasil praprosesnya disimpan pada basis data, yakni pada tabel `artikelUji`.

Tahap praproses dimulai dengan memanggil *method* `readKataInput` yang fungsinya adalah membaca data artikel berita dari basis data. Jika praproses terhadap data latih maka data diambil dari tabel `artikel`, sedangkan jika praproses terhadap data uji maka data diambil dari tabel `artikelUji`. Aplikasi akan membaca tiap baris dan langsung melakukan tahap praproses.

Implementasi untuk praproses terhadap data latih dapat dilihat pada segmen program 5.5. Hasil keluaran dari praproses terhadap data latih dan kategorinya akan disimpan dalam *arrayList* `teksKeluaran`. Namun karena data yang diproses lebih dari satu baris dan setiap baris harus diolah terpisah, maka dibuatlah satu *arrayList* `daftarTeks`. *ArrayList* ini akan menyimpan semua *arrayList* `teksKeluaran`, sehingga hasil dari setiap baris data dapat ditempatkan dalam tempat yang berbeda. Implementasi *nested arrayList* ini juga dapat dilihat pada segmen program 5.7.

```
//baca data dari tabel alias bikin query
String sql = "select artikel,kategori from artikel,kategori where "
    + "idKategori = fkKategori ";
ResultSet rs;
Statement st;
st = conn.createStatement();
rs = st.executeQuery(sql);
while (rs.next()) {
    String tmp = rs.getString(1);
    teksKeluaran.add(rs.getString(2));
    prosesStop(tmp);
    daftarTeks.add((ArrayList) teksKeluaran.clone());
    teksKeluaran.clear();
}
```

**Segmen Program 5.7** `readKataInput` untuk data latih

Implementasi untuk praproses terhadap data latih dapat dilihat pada segmen program 5.8. Hasil keluaran dari praproses terhadap data latih dan `idArtikelUji` akan disimpan dalam `arrayList` `teksKeluaran`.

```
//baca data dari tabel alias bikin query
String sql = "select artikeluji,idartikeluji from artikeluji ";
ResultSet rs;
Statement st;
st = conn.createStatement();
rs = st.executeQuery(sql);
while (rs.next()) {
    String tmp = rs.getString(1);
    teksKeluaran.add(rs.getString(2));
    prosesStop(tmp);
    daftarTeks.add((ArrayList) teksKeluaran.clone());
    teksKeluaran.clear();
}
```

#### Segmen Program 5.8 `readKataInput` untuk data uji

Setiap baris dari data yang diambil akan melalui empat proses dalam praproses. Ketiga proses awal, yakni *tokenizing*, *casefolding*, dan *stopping* atau penghapusan *stopwords*, dilakukan pada *method* `prosesStop`. Untuk proses *tokenizing* dan *case folding*, implementasi kode programnya dapat dilihat pada segmen program 5.9.

```
kataInput = kataInput.toLowerCase();
StringTokenizer tokenizer = new StringTokenizer(kataInput, "
\t\n\r\f\"'\"'\"'^|,.;:?![]!@#$%&(){}<>+_0123456789/");
```

#### Segmen Program 5.9 *Case Folding* dan *Tokenizing*

Sedangkan untuk implementasi kode program proses penghapusan *stopwords* dapat dilihat pada segmen program 5.10. Hasil proses *tokenizing* dibagi menjadi dua jenis token, yakni token yang mengandung karakter ‘-’ atau kata ulang dan token yang tidak mengandung karakter tersebut. Hal ini dikarenakan, pemrosesan pada tahap *stemming* atau pencarian kata dasar dari kedua jenis token tersebut sedikit berbeda.

```

while (tokenizer.hasMoreTokens()) {
    int flag = 1;
    token = tokenizer.nextToken();
    if (token.contains("-")) {
        if (token.length() > 3) {
            readKataUlang(token);
        }
    } else {
        //cek daftar stopwords
        if (CekStopword(token) == true) {
            flag = 0;
        }
        if (flag == 1) {
            String teksOut = doStem(token);
            teksKeluaran.add(teksOut);
        }
    }
}

```

#### Segmen Program 5.10 Penghapusan *Stopwords*

Untuk mencari apakah suatu token termasuk *stopword* atau bukan dilakukan pencarian token pada tabel *stopwords* yang menyimpan daftar *stopwords*. Implementasi pencarian *stopwords* tersebut terdapat pada *method* *cekStopword*, yang kode programnya dapat dilihat pada segmen program 5.11.

```

String sql = "SELECT stopword FROM stopword WHERE stopword =" +
kata + " LIMIT 1 ";
st = conn.createStatement();
rs = st.executeQuery(sql);
while (rs.next()) {
    stop = rs.getString(1);
}
st.close();
rs.close();
if (stop != null) {
    return true;
}
return false;

```

#### Segmen Program 5.11 *cekStopword*

Proses selanjutnya adalah mengubah token atau kata menjadi bentuk kata dasarnya atau disebut sebagai proses *stemming*. Tahap *stemming* ini dilakukan berdasarkan algoritma *confix stripping stemmer* yang telah dijelaskan pada bab III

mengenai *confix stripping stemmer*. Penerapan algoritma tersebut diterapkan pada *method* `doStem` yang dapat dilihat pada segmen program 5.12, 5.13, dan 5.14.

Tahap pertama adalah memeriksa apakah kata masukan ada di dalam kamus kata dasar. Jika ada maka kata masukan tersebut merupakan kata dasar, dapat dilihat pada segmen program 5.12.

```
//cek kamus
if (flagCekKamus(kata) == 1) {
    return kataLama;
```

#### Segmen Program 5.12 Cek Kata Masukan dengan Kamus

Jika kata masukan tersebut tidak terdapat dalam kamus, maka langkah selanjutnya adalah menghilangkan imbuhan. Sebelum menghilangkan imbuhan, dilakukan pengecekan *rule precedence*. Jika terdapat *rule precedence* maka imbuhan awalan dihilangkan dahulu. Setiap kali menghilangkan imbuhan, hasilnya langsung dicek ke dalam kamus kata dasar, jika ada maka kata tersebut merupakan kata dasar. Kode program dapat dilihat pada segmen program 5.13.

```
} else {
char[] kataArray = kata.toCharArray();
boolean isPrecedence = cekPrecedence(kataArray, kataArray.length);
if (isPrecedence) {
//hilangkan awalanPertama
awalanPertama = removeFirstOrderPrefix(kataArray, kataArray.length);
for (int i = kataArray.length - 1; i >= awalanPertama; i--) {
    kataArray = ArrayUtils.remove(kataArray, i);
}
String awalPertama = String.valueOf(kataArray);
//cek kamus
if (flagCekKamus(awalPertama) == 1) {
return awalPertama;
}
//hilangkan awalanKedua
awalanKedua = removeSecondOrderPrefix(kataArray, kataArray.length);
for (int i = kataArray.length - 1; i >= awalanKedua; i--) {
    kataArray = ArrayUtils.remove(kataArray, i);
}
String awalKedua = String.valueOf(kataArray);
if (flagCekKamus(awalKedua) == 1) {
return awalKedua;
}
```

#### Segmen Program 5.13 Hapus awalan

Kemudian hapus imbuhan lain, yakni partikel, kata ganti milik, akhiran. Jika hasil semua pemotongan imbuhan masih tidak ditemukan pada kamus kata dasar, maka kata inputan tersebut merupakan kata dasar. Kode program dapat dilihat pada segmen program 5.14.

```
//hilangkan partikel
partikel = removeParticle(kataArray, kataArray.length);
for (int i = kataArray.length - 1; i >= partikel; i--) {
    kataArray = ArrayUtils.remove(kataArray, i);
}
String kataPartikel = String.valueOf(kataArray);
//cek kamus
if (flagCekKamus(kataPartikel) == 1) {
    return kataPartikel;
}
//hilangkan kata ganti kepemilikan, -ku, -mu, -nya
kataGantiMilik = removePossessivePronoun(kataArray,
kataArray.length);
for (int i = kataArray.length - 1; i >= kataGantiMilik; i--) {
    kataArray = ArrayUtils.remove(kataArray, i);
}
String kataGanti = String.valueOf(kataArray);
//cek kamus
if (flagCekKamus(kataGanti) == 1) {
    return kataGanti;
}
//hilangkan akhiran -i, -kan, -an
akhiran = removeSuffix(kataArray, kataArray.length);
for (int i = kataArray.length - 1; i >= akhiran; i--) {
    kataArray = ArrayUtils.remove(kataArray, i);
}
String akhir = String.valueOf(kataArray);
//cek kamus
if (flagCekKamus(akhir) == 1) {
    return akhir;
}
//jika tidak ditemukan maka kata inputan adalah kata dasar
return kataLama;
}
```

#### Segmen Program 5.14 *Stemming*

Jika tidak terdapat *rule precedence*, maka urutan penghapusan imbuhan adalah hapus partikel, kata ganti milik,

akhir, kemudian yang terakhir adalah awalan. *Method-method* yang digunakan sama seperti pada saat nilai *rule precedence* bernilai benar atau terdapat *rule precedence* pada kata masukan. Hanya urutan penghapusan imbuhan saja yang sedikit berbeda. Setiap kali selesai melakukan penghapusan, hasilnya akan dicari apakah kata tersebut terdapat di dalam kamus kata dasar. Jika ada maka kata tersebut merupakan kata dasar dan proses dihentikan. Tetapi jika hingga akhir penghapusan imbuhan tetap tidak ditemukan pada kamus, maka kata masukan dianggap sebagai kata dasar. Hal ini menyesuaikan aturan pada *confix-stripping stemmer* yang telah dijelaskan pada bab mengenai tinjauan pustaka.

Proses *stemming* tersebut dapat langsung dilakukan jika token masukan bukan kata ulang, namun jika kata ulang maka proses *stemming*-nya sedikit berbeda. Cara mencari bentuk kata dasar dari kata ulang telah dijelaskan pada bab Tinjauan Pustaka pada bagian *confix-stripping stemmer*. Berikut implementasi pencarian bentuk kata dasar dari kata ulang yang dapat dilihat pada segmen program 5.15.

```
public void readKataUlang(String kataUlang) {
    String kataAStem = null;
    String kataBStem = null;
    String[] kataUlangPecah = kataUlang.split("-");
    String kataA = kataUlangPecah[0];
    if (kataA.length() > 0) {
        kataAStem = doStem(kataA);
    }
    String kataB = kataUlangPecah[1];
    if (kataB.length() > 0) {
        kataBStem = doStem(kataB);
    }
    if (kataAStem != null && kataBStem != null) {
        if (kataAStem.equals(kataBStem)) {
            teksKeluaran.add(kataAStem);
        } else {
            teksKeluaran.add(kataUlang);
        }
    } else {
        teksKeluaran.add(kataA+kataB);
    }
}
```

**Segmen Program 5.15 Stemming Kata Ulang**

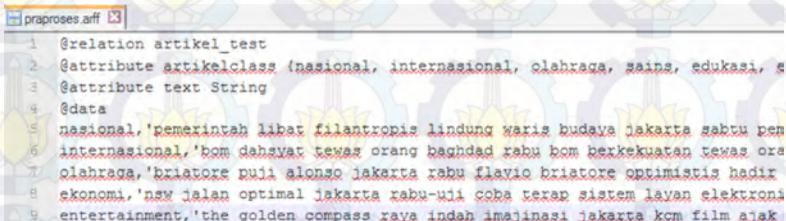
Setelah tahap praproses selesai, maka langkah berikutnya adalah menyimpan hasil keluaran tersebut. Untuk praproses terhadap data latih, hasil praproses disimpan dalam sebuah file. Segmen program 5.16 merupakan kode program untuk menuliskan hasil keluaran praproses ke dalam satu file arff. Pada file arff tersebut, hasil praproses juga disimpan beserta kategorinya. File keluaran ini akan digunakan sebagai masukan pada tahap latih klasifikasi.

```
FileOutputStream out = new FileOutputStream(fileOutput, true);
PrintStream print = new PrintStream(out);
for (int x = 0; x < daftarTeks.size(); x++) {
    teksKeluaran = daftarTeks.get(x);
    print.print(teksKeluaran.get(0));
    print.print(", ");

    for (int i = 1; i < teksKeluaran.size(); i++) {
        if (i < teksKeluaran.size() - 1) {
            print.print(teksKeluaran.get(i) + " ");
        } else {
            print.print(teksKeluaran.get(i));
        }
    }
    print.print(""" + "\n");
}
```

#### Segmen Program 5.16 Penyimpanan Hasil Praproses terhadap Data Latih

Untuk contoh isi dari file .arff yang menyimpan hasil praproses dapat dilihat pada gambar 5.6.



```
praproses.arff
1 @relation artikel_test
2 @attribute artikelclass (nasional, internasional, olahraga, sains, edukasi, s
3 @attribute text String
4 @data
5 nasional,'pemerintah libat filantropis lindung waris budaya jakarta sabtu pem
6 internasional,'bom dahsyat tewas orang baghdad rabu bom berkekuatan tewas ora
7 olahraga,'briatore puia alonso jakarta rabu flavio briatore optimistis hadir
8 ekonomi,'nsw jalan optimal jakarta rabu-uji coba terap sistem layanan elektroni
9 entertainment,'the golden compass raya indah imajinasi jakarta kcm film alak
```

Gambar 5.6 Hasil Praproses

Untuk praproses terhadap data uji, hasil praproses disimpan kembali dalam tabel `artikelUji` pada kolom `hasilPraproses`. Proses ini diimplementasikan pada *method* `writeOutput()` yang isinya dapat dilihat pada segmen program 5.17.

```
for (int x = 0; x < daftarTeks.size(); x++) {
    teksKeluaran = daftarTeks.get(x);
    String id = teksKeluaran.get(0);
    String teks = "";
    for (int i = 1; i < teksKeluaran.size(); i++) {
        if (i < teksKeluaran.size() - 1) {
            teks += teksKeluaran.get(i) + " ";
        }
    }
    String sqlupdate = "update artikeluji set hasilPraproses ="
        + teks + " where idartikeluji =" + id + """;
    Statement stmt = conn.createStatement();
    stmt.executeUpdate(sqlupdate);
}
```

**Segmen Program 5.17 Penyimpanan Hasil Praproses terhadap Data Uji**

#### 5.2.4. Latih Klasifikasi

Tahap latih klasifikasi merupakan tahap pertama dalam proses klasifikasi. Pada tahap ini dibuat model klasifikasi yang menjadi dasar dalam tahap uji klasifikasi. Model klasifikasi tersebut dibuat berdasarkan data latih yang telah diolah pada tahap praproses dengan menggunakan metode NBC. Pada sistem aplikasi klasifikasi, pembuatan model ini dilakukan melalui borang Latih Klasifikasi yang dibuat berdasarkan diagram sekuen yang dijelaskan pada bagian 4.1.3.6 mengenai latih klasifikasi. Tampilan tampilan antar muka pengguna untuk borang ini dapat dilihat pada gambar 5.7.

Pada borang Latih Klasifikasi, terdapat dua kolom isian, yakni lokasi file latih yang merupakan file hasil keluaran dari praproses dan lokasi file model yang merupakan file keluaran dari tahap ini. File model akan menyimpan model klasifikasi yang dibuat berdasarkan data latih dengan menggunakan metode NBC.

Pengguna harus mengisi semua kolom isian yang tersedia. Jika tidak, maka sistem akan menampilkan pesan kesalahan dan sistem akan menghentikan proses pembuatan model klasifikasi. Hasil evaluasi dari data latih akan ditampilkan pada kedua *textArea*. *TextArea* sebelah kiri akan menampilkan hasil evaluasi, dan *textArea* sebelah kanan akan menampilkan model klasifikasi.



**Gambar 5.7** Antar Muka dan Contoh Hasil Latih Klasifikasi

Seluruh proses pada tahap latih klasifikasi ini menggunakan library yang digunakan para pengembang aplikasi WEKA, yakni *weka.jar*. Dalam *library* ini terdapat kelas-kelas yang digunakan untuk memproses klasifikasi dengan metode NBC.

Proses pertama dalam latih klasifikasi adalah membaca data latih yang telah disiapkan. Data latih ini merupakan hasil keluaran dari praproses yang disimpan dalam satu file dengan format *.arff*. Sistem aplikasi akan membaca data latih tersebut dan menyimpannya dalam satu *object trainData*. Implementasinya dapat dilihat pada segmen program 5.18.

Tahap berikutnya adalah melakukan evaluasi terhadap data latih. Evaluasi harus dilakukan terlebih dahulu karena jika evaluasi dilakukan setelah pembuatan model maka hasil evaluasi tidak akan sesuai dengan data latih yang ada. Pada tahap evaluasi

ini, kategori pada data latih yang disimpan sebelumnya diatur untuk menjadi atribut pertama. Kemudian filter dibuat dan mengubah teks artikel berita menjadi *vector* fitur. Kategori dan *vector* fitur akan diolah dengan metode NBC dalam *classifier* yang dibuat.

```
public void loadDataset(String fileName) {
    try {
        BufferedReader reader = new BufferedReader(new
        FileReader(fileName));
        //baca file arff
        ArffReader arff = new ArffReader(reader);
        //data simpan dalam object
        trainData = arff.getData();
        hasilLatih.append("== Loaded dataset: " + fileName + " == " + "\n");
        reader.close();
    }
}
```

#### Segmen Program 5.18 Baca Data Latih

Evaluasi terhadap data latih tersebut dilakukan dengan menggunakan metode *cross validation* dengan jumlah *fold* adalah sepuluh *folds*. Hasil evaluasi dibagi menjadi dua, yakni hasil kesimpulan evaluasi secara keseluruhan dan hasil evaluasi dari masing-masing kategori. Implementasi kode programnya dapat dilihat pada segmen program 5.19 dan 5.120.

```
public void evaluate() {
    int folds = 10;
    try {
        trainData.setClassIndex(0);
        filter = new StringToWordVector();
        filter.setAttributeIndices("last");
        classifier = new FilteredClassifier();
        classifier.setFilter(filter);
        classifier.setClassifier(new NaiveBayes());
        // random data
        Random rand = new Random(1);
        Instances randData = new Instances(trainData);
        randData.randomize(rand);
        if (randData.classAttribute().isNominal()) {
            randData.stratify(folds);
        }
        Evaluation evalAll = new Evaluation(randData);
    }
}
```

#### Segmen Program 5.19 Evaluasi Data Latih

```

for (int n = 0; n < folds; n++) {
    Evaluation eval = new Evaluation(randData);
    Instances train = randData.trainCV(folds, n);
    Instances test = randData.testCV(folds, n);
    Classifier cls = Classifier.makeCopy(classifier);
    cls.buildClassifier(train);
    eval.evaluateModel(cls, test);
    evalAll.evaluateModel(cls, test);
    hasilLatih.append("==== Evaluation for fold: " + (n + 1) + "/" +
folds + "====" + "\n");
    hasilLatih.append(eval.toSummaryString() + "\n");
    hasilLatih.append(eval.toClassDetailsString() + "\n");
    hasilLatih.append("==== Evaluating on fold: " + (n + 1) + "/" +
folds + " done =====" + "\n");
}

hasilLatih.append("==== Evaluation summary for all folds
===== " + "\n");
hasilLatih.append(evalAll.toSummaryString() + "\n");
hasilLatih.append(evalAll.toClassDetailsString() + "\n");
hasilLatih.append("==== Evaluating on filtered (training) dataset
done =====" + "\n");

```

#### Segmen Program 5.20 Evaluasi Data Latih (Lanjutan)

Setelah evaluasi telah dilakukan maka tahap selanjutnya adalah melakukan proses pembuatan model klasifikasi dengan menggunakan metode NBC. Pembuatan model klasifikasi ini diproses melalui kode program pada segmen program 5.21.

Model klasifikasi yang dihasilkan disimpan dalam satu file yang nantinya akan digunakan dalam proses uji klasifikasi.

```

public String learn() {
    try {
        trainData.setClassIndex(0);
        filter = new StringToWordVector();
        filter.setAttributeIndices("last");
        classifier = new FilteredClassifier();
        classifier.setFilter(filter);
        classifier.setClassifier(new NaiveBayes());
        classifier.buildClassifier(trainData);
        resultDetail = classifier + "\n" + "==== Training on filtered
(training) dataset done =====";
        return resultDetail;
    }
}

```

#### Segmen Program 5.21 Pembuatan Model Klasifikasi

### 5.2.5. Uji Klasifikasi

Artikel berita yang tidak diketahui kategorinya, dapat ditentukan kategorinya pada tahap uji klasifikasi. Untuk melakukan proses uji klasifikasi ini, telah disediakan borang Uji Klasifikasi yang dibuat berdasarkan diagram sekuen yang dijelaskan pada bagian 4.1.3.7. Tampilan tampilan antar mukanya dapat dilihat pada gambar 5.8.



**Gambar 5.8 Tampilan Antar Muka Uji Klasifikasi**

Pada borang tersebut, pengguna hanya perlu menekan tombol mulai klasifikasi. Aplikasi akan melakukan klasifikasi terhadap data uji yang sebelumnya tidak diketahui kategori beritanya.

Secara umum proses yang dilakukan pada tahap ini adalah membaca model klasifikasi, membaca artikel berita yang akan diklasifikasikan, membuat *instance*, dan menentukan klasifikasi artikel berita tersebut. Aplikasi membaca model klasifikasi yang telah disimpan dalam satu file. Model yang dibaca oleh aplikasi disimpan dalam *temporary object*. Kode program untuk membaca model klasifikasi, dapat dilihat pada segmen program 5.20.

```
ObjectInputStream in = new ObjectInputStream(new
FileInputStream(fileName));
Object tmp = in.readObject();
classifier = (FilteredClassifier) tmp;
in.close();
```

**Segmen Program 5.22 Baca Model Klasifikasi**

Proses berikutnya adalah membaca artikel berita yang menjadi data uji. Aplikasi akan membaca data per baris dari tabel `artikelUji`. Setiap kali satu baris data selesai dibaca, maka aplikasi akan langsung membuat *instance*-nya dan menyimpan data yang telah dibaca ke dalam *instance* tersebut. Proses ini dapat dilihat pada segmen program 5.21.

```
//baca data dari tabel alias bikin query
String sql = "select idartikeluji, artikelUji from artikeluji ";
ResultSet rs;
Statement st;
st = conn.createStatement();
rs = st.executeQuery(sql);
while (rs.next()) {
    int id = rs.getInt(1);
    String tmp = rs.getString(2);
    makeInstance(tmp);
    classify();
    int kategori = kategori(result);
    String sqlupdate = "update artikeluji set kategori ="
        + kategori + " where idartikeluji =" + id + """;
    Statement stmt = conn.createStatement();
    stmt.executeUpdate(sqlupdate);
}
```

#### Segmen Program 5.23 Baca Data Uji

Pada tahap sebelumnya dalam pembuatan model klasifikasi terdapat dua atribut, yakni atribut nominal untuk kategori dan atribut string yang berisi teks. Pada proses pembuatan *instance* ini, dibuat sebuah *dataset* baru yang mana atribut pertamanya akan diklasifikasikan. Sehingga digunakan object *FastVector* dan ditambahkan jenis kategori yang telah didefinisikan. Kemudian dibuat *Instances object* dengan menggunakan *FastVector* untuk menyimpan dua atribut sebelumnya dan mengatur indexnya menjadi index 0. Kode program pembuatan *instance* dapat dilihat pada segmen program 5.22 dan 5.23 berikut ini.

```
public void makeInstance(String text) {
    // Membuat atribut, kelas, dan teks
    FastVector fvNominalVal = new FastVector(12);
```

#### Segmen Program 5.24 Pembuatan Instance

```

fvNominalVal.addElement("nasional");
fvNominalVal.addElement("internasional");
fvNominalVal.addElement("olahraga");
fvNominalVal.addElement("sains");
fvNominalVal.addElement("edukasi");
fvNominalVal.addElement("ekonomi");
fvNominalVal.addElement("tekno");
fvNominalVal.addElement("entertainment");
fvNominalVal.addElement("otomotif");
fvNominalVal.addElement("health");
fvNominalVal.addElement("properti");
fvNominalVal.addElement("travel");
Attribute attribute1 = new Attribute("class", fvNominalVal);
Attribute attribute2 = new Attribute("text", (FastVector) null);
// Membuat daftar dari instances dengan satu elemen
FastVector fvWekaAttributes = new FastVector(2);
fvWekaAttributes.addElement(attribute1);
fvWekaAttributes.addElement(attribute2);
instances = new Instances("Test relation", fvWekaAttributes, 1);
// Set class index
instances.setClassIndex(0);
// Membuat dan menambah instance
Instance instance = new Instance(2);
instance.setValue(attribute2, text);
instances.add(instance);

```

#### Segmen Program 5.25 Pembuatan Instance (Lanjutan)

Proses selanjutnya adalah menentukan kategori dari data uji yang telah dimasukkan oleh pengguna. Penentuan klasifikasi ini berdasarkan data model yang dihasilkan pada tahap latih klasifikasi. Karena *class index* atau kategori berada pada *index* ke-0 maka prediksi kategori diatur pada *index* 0. Kode program dapat dilihat pada segmen program 5.24.

```

public String classify() {
    try {
        double pred = classifier.classifyInstance(instances.instance(0));
        String resTmp;
        resTmp = instances.classAttribute().value((int) pred);
        result = resTmp;
        return resTmp;
    }
}

```

#### Segmen Program 5.26 Penentuan Kategori

Data uji pada awalnya tidak diketahui kategori beritanya, setelah tahap uji klasifikasi dapat diketahui kategori beritanya. Hasil keluaran ini disimpan dalam basis data pada tabel *artikelUji*. Penyimpanan hasil keluaran ini dilakukan dengan kode program yang dapat dilihat pada segmen program 5.23 pada bagian *'sql update'*. Setelah hasil keluaran disimpan, hasil ini juga ditampilkan pada borang klasifikasi dan kode programnya dapat dilihat pada segmen program 5.27.

```
String sql = "select idArtikelUji,judul,artikelUji,kategori.kategori "  
            + "from artikeluji,kategori where artikeluji.kategori=idkategori";  
ResultSet rs; Statement st;  
st = conn.createStatement();  
rs = st.executeQuery(sql);  
Vector judul = new Vector();  
judul.add("ID ArtikelUji");  
judul.add("Judul");  
judul.add("Artikel Uji");  
judul.add("Kategori");  
DefaultTableModel dtm = new DefaultTableModel(null, judul);  
dtm.setColumnIdentifiers(judul);  
//mengambil data per baris  
while (rs.next() ) {  
    Vector tampilTbl = new Vector();  
    tampilTbl.add(rs.getString(1));  
    tampilTbl.add(rs.getString(2));  
    tampilTbl.add(rs.getString(3));  
    tampilTbl.add(rs.getString(4));  
    dtm.addRow(tampilTbl);  
}  
tblHasil.setModel(dtm);
```

#### Segmen Program 5.27 Menampilkan Hasil Uji

### 5.2.6. Mengklasifikasikan Artikel Berita

Seperti yang telah dijelaskan pada desain aplikasi, pengguna umum dapat melakukan klasifikasi artikel berita melalui borang klasifikasi artikel berita. Untuk melakukan klasifikasi artikel berita pengguna umum menekan tombol praproses terlebih dahulu dan aplikasi akan melakukan tahap praproses terhadap data uji. Proses ini dilakukan dengan

memanggil kelas `PraProses()` yang akan melakukan pra-proses terhadap data uji. Seperti yang terlihat pada segmen program 5.28 berikut ini.

```
PraProses pra = new PraProses();
pra.readKataInput(1);
pra.writeOutput();
```

#### Segmen Program 5.28 Memanggil Kelas PraProses()

Setelah pra-proses selesai dilakukan, maka pengguna menekan tombol mulai klasifikasi untuk melakukan proses klasifikasi terhadap data uji. Proses klasifikasi ini dilakukan dengan memanggil kelas `Klasifikasi()` seperti yang dapat dilihat pada segmen program 5.29 berikut ini.

```
Klasifikasi classifier;
classifier = new Klasifikasi();
classifier.loadModel(fileModel);
classifier.load();
```

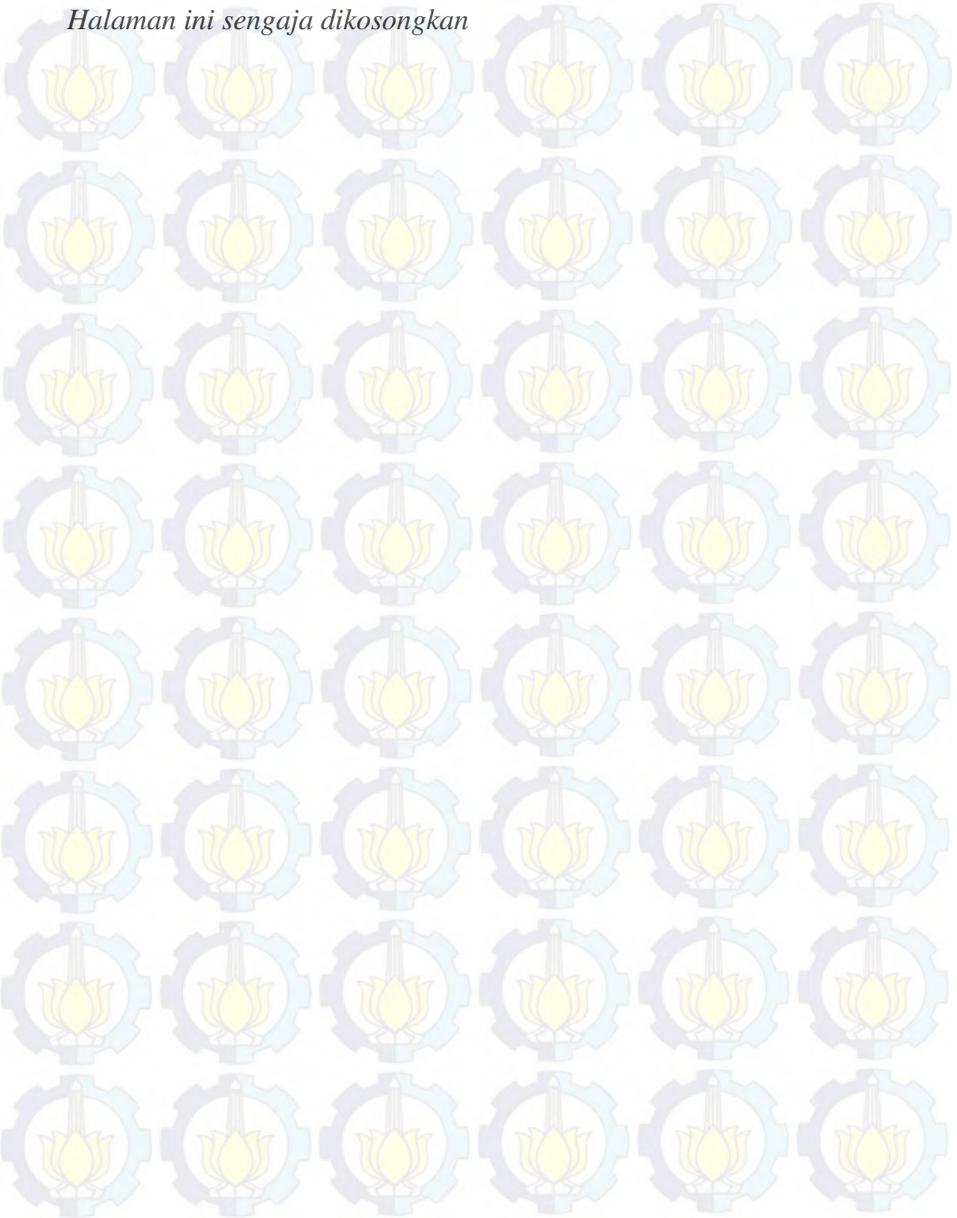
#### Segmen Program 5.29 Memanggil Kelas Klasifikasi()

Setelah tahap klasifikasi selesai, aplikasi akan menampilkan hasilnya melalui kode program 5.27. Sedangkan antar muka dari borang Klasifikasi Artikel berita yang dibuat berdasarkan diagram sekuen 4.1.3.8 dapat dilihat pada gambar 5.30 berikut ini.

ID ArtikelUji	Judul	Artikel Uji	Kategori

#### Segmen Program 5.30 Antar Muka Borang Klasifikasi Artikel Berita

*Halaman ini sengaja dikosongkan*



## BAB VI UJI COBA DAN ANALISIS HASIL

Bab ini menjelaskan mengenai proses uji coba terhadap kinerja aplikasi yang telah dibangun dan analisis dari hasil yang didapatkan. Proses uji coba meliputi lingkungan uji coba, data uji coba, skenario uji coba, pelaksanaan dan hasil uji coba, dan analisis hasil uji coba. Skenario uji coba yang dilakukan terbagi menjadi dua, yakni uji coba verifikasi dan uji coba validasi.

### 6.1. Data Uji Coba

Data artikel berita yang digunakan merupakan sampel artikel berita yang diambil dari portal berita online [www.kompas.com](http://www.kompas.com). Total sampel artikel berita yang diambil dapat dilihat pada tabel 5.2. Sampel data tersebut dibagi menjadi dua jenis data, yakni data latih dan data uji yang proporsinya sebesar 70:30 untuk masing-masing kategori berita. Data yang digunakan dalam tahap uji coba merupakan artikel berita yang digunakan pada tahap uji klasifikasi.

### 6.2. Lingkungan Uji Coba

Lingkungan uji coba merupakan kriteria perangkat pengujian yang digunakan dalam menguji sistem yang dibangun pada tugas akhir ini. Lingkungan uji coba terdiri dari perangkat keras dan perangkat lunak. Adapun perangkat keras yang digunakan ditunjukkan pada tabel 6.1.

**Tabel 6.1 Spesifikasi Lingkungan Uji Coba – Perangkat Keras**

<b>Perangkat Keras</b>	<b>Spesifikasi</b>
Jenis	Notebook
Processor	Intel(R) Core (TM) i3-2328M CPU @ 2.20GHz 2.20GHz
RAM	2 GB

Selain itu, terdapat beberapa perangkat lunak yang digunakan dalam pengujian sistem yang dibangun pada tugas akhir ini yang ditunjukkan pada tabel 6.2

**Tabel 6.2 Spesifikasi Lingkungan Uji Coba - Perangkat Lunak**

<b>Perangkat Lunak</b>	<b>Spesifikasi</b>
Sistem Operasi	Windows 8
Bahasa Pemrograman	JAVA
<i>Tools</i>	<ul style="list-style-type: none"> <li>• XAMPP Windows Ver 1.7.2</li> <li>• Netbeans IDE 6.9.1</li> <li>• Java SE Development Kit 7 Update 5</li> <li>• WEKA Library 3-6</li> </ul>

### **6.3. Skenario Uji Coba**

Skenario uji coba bertujuan untuk mengetahui apakah tujuan pembuatan aplikasi pengklasifikasi artikel berita telah tercapai. Uji coba yang dilakukan terbagi menjadi dua, yakni uji verifikasi dan uji validasi. Penjelasan lebih detil dijelaskan pada masing-masing bagian.

#### **6.3.1. Uji Verifikasi**

Uji verifikasi dilakukan pada sisi fungsional dan hasil keluaran aplikasi. Pengujian ini dilakukan untuk mengecek apakah fungsional dan hasil klasifikasi dari aplikasi pengklasifikasi artikel berita sesuai seperti yang diharapkan. Uji verifikasi pada sisi fungsional dilakukan berdasarkan dengan desain aplikasi yang telah dibuat pada bab IV mengenai desain aplikasi. Sedangkan uji verifikasi pada sisi hasil klasifikasi dilakukan dengan membandingkan hasil klasifikasi keluaran dari aplikasi pengklasifikasi artikel berita dengan hasil klasifikasi yang merupakan keluaran dari aplikasi WEKA yang merupakan aplikasi yang biasa digunakan untuk penggalian data. Selain itu,

juga dibandingkan dengan hasil klasifikasi yang dilakukan secara manual. Data artikel berita yang digunakan pada uji verifikasi merupakan data uji dijelaskan pada bagian 6.1 yang menjelaskan mengenai data uji coba.

### **6.3.2. Uji Validasi**

Evaluasi performa pada model klasifikasi dan evaluasi hasil klasifikasi yang merupakan hasil keluaran aplikasi pengklasifikasi artikel berita dan aplikasi WEKA yang dilakukan pada bagian uji validasi. Evaluasi yang dilakukan mencakup perhitungan akurasi, *precision*, *recall*, dan *f-measure*. Rumus perhitungan ini telah dijelaskan pada bagian 2.3 yang menjelaskan mengenai pengukuran performa.

## **6.4. Pelaksanaan dan Hasil Uji Coba**

Pada bagian ini dijelaskan mengenai pelaksanaan dari skenario uji coba. Uji coba yang dilakukan terbagi menjadi uji verifikasi dan uji validasi sesuai dengan skenario uji coba.

### **6.4.1. Uji Verifikasi**

Seperti yang telah dijelaskan pada skenario uji coba, uji verifikasi dibagi menjadi dua yakni uji fungsional aplikasi dan hasil klasifikasi. Berikut ini hasil dari uji verifikasi tersebut.

#### **6.4.1.1. Uji Verifikasi Fungsional Aplikasi**

Uji fungsional aplikasi dilakukan dengan melihat apakah semua fungsi pada aplikasi dapat berjalan sesuai dengan desain aplikasi yang dibuat pada bab IV mengenai desain aplikasi. Dilakukan uji coba pada seluruh fungsional aplikasi, baik alur fungsional secara normal maupun alur alternatif. Uji coba fungsional dilakukan berdasarkan skenario *usecase* yang telah dibuat yang dapat dilihat pada lampiran A. Hasil uji coba dapat dilihat pada tabel 6.3 dan 6.4.

Tabel 6.3 Hasil Uji Verifikasi Fungsional Aplikasi

Nama Fungsi	Skenario	Hasil	Status
Login	Administrator melakukan login dengan mengisi username dan password	Aplikasi menampilkan borang tambah artikel berita	Berhasil (Lampiran B.1.1)
	Administrator melakukan login namun tidak mengisi salah satu atau semua kolom isian	Aplikasi menampilkan pesan kesalahan	Berhasil (Lampiran B.1.2)
	Administrator melakukan login namun <i>username</i> atau <i>password</i> salah	Aplikasi menampilkan pesan kesalahan	Berhasil (Lampiran B.1.3)
Menambahkan data artikel untuk data latih	Administrator menambahkan artikel berita untuk data latih dengan memilih pilihan 'data latih' dan mengisi semua kolom	Artikel berita untuk data latih masuk ke dalam tabel artikel dan menampilkan pesan sukses	Berhasil (Lampiran B.2.1)
	Administrator menambahkan artikel berita namun terdapat kolom yang tidak terisi	Aplikasi menampilkan pesan kesalahan dan data yang diisikan tidak ditambahkan ke dalam tabel artikel	Berhasil (Lampiran B.2.2)
Menambahkan data artikel untuk data uji	Administrator/pengguna umum menambahkan artikel berita untuk data uji dengan memilih pilihan 'data uji' dan mengisi semua kolom	Artikel berita untuk data latih masuk ke dalam tabel artikelUji dan menampilkan pesan sukses	Berhasil (Lampiran B.3.1)
	Administrator/pengguna umum menambahkan artikel berita untuk data uji namun terdapat kolom yang tidak terisi	Aplikasi menampilkan pesan kesalahan dan data yang diisikan tidak ditambahkan ke dalam tabel artikelUji	Berhasil (Lampiran B.3.2)
Melakukan praproses data latih	Administrator memilih pilihan 'data latih' dan mengisi lokasi file yang akan menyimpan hasil keluaran kemudian menekan tombol OK	Aplikasi melakukan praproses terhadap data yang tersimpan pada tabel artikel dan menyimpan hasilnya pada file yang telah diisikan oleh pengguna	Berhasil (Lampiran B.4.1)

Tabel 6.4 Hasil Uji Verifikasi Fungsional Aplikasi (Lanjutan)

Nama Fungsi	Skenario	Hasil	Status
	Administrator tidak mengisi kolom mengenai lokasi file keluaran namun menekan tombol OK	Aplikasi akan menampilkan pesan kesalahan dan tidak melakukan praproses.	Berhasil (Lampiran B.4.2)
Melakukan praproses data uji	Administrator memilih pilihan 'data uji' dan menekan tombol OK	Aplikasi melakukan praproses terhadap data yang tersimpan pada tabel artikel dan menyimpan hasilnya pada file yang telah diisikan oleh pengguna	Berhasil (Lampiran B.5.1)
Melakukan latihan klasifikasi	Administrator mengisi lokasi file data latihan dan file keluaran, kemudian menekan tombol OK	Aplikasi melakukan proses latihan dan hasil model klasifikasi disimpan dalam file yang telah diisikan oleh pengguna	Berhasil (Lampiran B.6.1)
	Administrator tidak mengisi salah satu kolom isian atau semua kolom isian	Aplikasi menampilkan pesan kesalahan dan aplikasi tidak melakukan proses latihan klasifikasi serta model klasifikasi tidak dibuat	Berhasil (Lampiran B.6.2)
Melakukan uji klasifikasi	Administrator menekan tombol mulai untuk memulai klasifikasi	Aplikasi menentukan kategori berita dari artikel berita yang menjadi data uji berdasarkan model klasifikasi yang dibuat pada tahap latihan.	Berhasil (Lampiran B.7.1)
Mengklasifikasi artikel berita	Pengguna umum menekan tombol mulai praproses, setelah pesan muncul pengguna umum menekan tombol mulai klasifikasi	Aplikasi melakukan praproses terhadap data uji dan menentukan kategori berita dari artikel berita yang diisikan berdasarkan model klasifikasi.	Berhasil (Lampiran B.8.1)

#### 6.4.1.2. Uji Verifikasi Hasil Klasifikasi

Hasil klasifikasi atau hasil pada tahap uji klasifikasi menggunakan data uji yang telah dijelaskan pada bagian Data Uji Coba akan dibandingkan dengan hasil klasifikasi yang dilakukan secara manual dan hasil klasifikasi yang diproses dengan aplikasi WEKA. Klasifikasi secara manual dilakukan dengan melihat topik dari artikel berita dan kemudian ditentukan kategori berita dari artikel berita tersebut. Beberapa contoh hasil klasifikasi dengan menggunakan aplikasi dan manual dapat dilihat pada tabel 6.5 berikut ini.

**Tabel 6.5 Perbandingan Hasil Klasifikasi**

No	Judul Artikel	Kategori Manual	Kategori Aplikasi	Kategori WEKA
1	Mendagri Akan Konfirmasi Bupati Singkil	Nasional	Nasional	Nasional
2	Lima Tewas di Irak	Internasional	Internasional	Internasional
3	Stoner dan Pedrosa di Depan, Lorenzo Terpuruk	Olahraga	Olahraga	Olahraga
4	Inilah Planet Pertama di Luar Bima Sakti	Sains	Sains	Sains
5	Mahasiswa Indonesia Raih Prestasi Matematika Internasional	Edukasi	Edukasi	Edukasi
6	Bebas Bea Impor Kedelai Hanya Untungkan Importir	Ekonomi	Ekonomi	Ekonomi
7	November, Tablet Windows 8 Samsung Masuk Indonesia	Tekno	Tekno	Tekno
8	Jadi Produser, Anjasmara Diwantiwanti Istri	<i>Entertainment</i>	<i>Entertainment</i>	<i>Entertainment</i>
9	Terminal Konektor Khusus untuk Mobil Listrik	Otomotif	Otomotif	Otomotif
10	Menderita Kanker Tak Perlu Ditutupi	<i>Health</i>	<i>Health</i>	<i>Health</i>
11	BTN Salurkan KPR Dengan Fasilitas Likuiditas	Properti	Properti	Properti
12	Serabi Kocor dari Yogya Cuma Seribuan	Travel	Travel	Travel

## 6.4.2. Uji Validasi

Evaluasi model klasifikasi maupun evaluasi hasil uji klasifikasi dilakukan pada tahap ini. Evaluasi yang dilakukan merupakan evaluasi performa yang mencakup perhitungan akurasi, *precision*, *recall*, dan *F-Measure*.

### 6.4.2.1. Uji Validasi Model Klasifikasi

Pada bagian ini akan dipaparkan hasil evaluasi dari model klasifikasi yang telah dibuat oleh aplikasi dengan menggunakan metode NBC. Hasil evaluasi ini mencakup perhitungan akurasi, *precision*, *recall*, dan *f-measures* dari model klasifikasi. Metode evaluasi menggunakan metode 10 *cross-validation*. Hasil evaluasi tersebut dapat dilihat pada tabel 6.6 berikut ini.

Tabel 6.6 10 *cross-validation*

	Akurasi	<i>Precision</i>	<i>Recall</i>	<i>F-Measures</i>
K-1	86.904%	89.00%	86.90%	86.90%
K-2	88.095%	89.10%	88.10%	88.10%
K-3	91.666%	92.70%	91.70%	91.80%
K-4	83.333%	84.30%	83.30%	82.90%
K-5	86.904%	88.80%	86.90%	86.80%
K-6	86.904%	87.70%	86.90%	86.50%
K-7	89.285%	90.30%	89.30%	89.20%
K-8	86.747%	88.70%	86.70%	86.60%
K-9	80.723%	84.10%	80.70%	81.30%
K-10	86.747%	87.80%	86.70%	86.50%
Rata	86.738%	87.20%	86.70%	86.80%
StdDev	0.0299	0.0256	0.0301	0.0294

Sedangkan pada tabel 6.7 dapat dilihat perbandingan dari hasil evaluasi model klasifikasi yang dibuat oleh aplikasi dan evaluasi model klasifikasi yang dibuat oleh weka.

Tabel 6.7 Perbandingan Evaluasi Model Klasifikasi Aplikasi dan WEKA

	Aplikasi	WEKA
Akurasi	86.738%	<b>86.857%</b>
<i>Precision</i>	87.20%	<b>87.30%</b>
<i>Recall</i>	86.70%	<b>86.90%</b>
<i>F-Measure</i>	86.80%	<b>86.90%</b>

#### 6.4.2.2. Uji Validasi Hasil Uji Klasifikasi

Berbeda dengan model klasifikasi yang menggunakan data latih, evaluasi hasil uji klasifikasi menggunakan data uji sebagai masukan datanya. Evaluasi hasil dari tahap uji klasifikasi mencakup perhitungan akurasi, *precision*, *recall*, dan *f-measure*. Hasil klasifikasi dengan menggunakan aplikasi pengklasifikasi artikel berita dapat dilihat pada tabel 6.8. Pada tabel tersebut dapat dilihat hasil klasifikasi oleh aplikasi, yakni kelas prediksi dan kelas asli.

Tabel 6.8 Hasil Klasifikasi Aplikasi

Kelas Asli		Kelas Prediksi											
		a	b	c	d	e	f	g	h	i	j	k	l
a	nasional	25	0	0	0	2	2	0	1	0	0	0	0
b	internasional	1	28	0	0	0	0	0	1	0	0	0	0
c	olahraga	0	1	28	1	0	0	0	0	0	0	0	0
d	sains	1	1	0	21	1	0	0	0	1	2	1	1
e	edukasi	1	0	0	1	26	0	0	0	0	0	0	1
f	ekonomi	2	1	0	0	0	25	0	0	0	1	0	1
g	teknologi	1	0	0	2	3	0	24	0	0	0	0	0
h	<i>entertainment</i>	0	0	0	0	0	0	0	28	0	0	1	1
i	otomotif	0	0	0	0	1	0	0	0	27	0	0	1
j	<i>health</i>	1	0	0	1	0	1	0	0	0	26	1	0
k	properti	0	0	0	0	0	1	0	1	0	0	26	2
l	travel	0	0	0	0	1	0	0	0	0	0	0	29

Sedangkan hasil klasifikasi jika menggunakan aplikasi WEKA dapat dilihat pada tabel 6.9. Pada tabel tersebut dapat dilihat hasil klasifikasi oleh aplikasi, yakni kelas prediksi dan kelas asli.

**Tabel 6.9 Hasil Klasifikasi WEKA**

Kelas Asli		Kelas Prediksi											
		a	b	c	d	e	f	g	h	i	j	k	l
a	nasional	25	0	0	0	2	2	0	1	0	0	0	0
b	internasional	1	28	0	0	0	0	0	1	0	0	0	0
c	olahraga	0	1	28	1	0	0	0	0	0	0	0	0
d	sains	1	1	0	21	1	0	0	0	1	2	1	1
e	edukasi	1	0	0	1	26	0	0	0	0	0	0	1
f	Ekonomi	2	1	0	0	0	25	0	0	0	1	0	1
g	Tekno	1	0	0	2	3	0	24	0	0	0	0	0
h	<i>entertainment</i>	0	0	0	0	0	0	0	28	0	0	1	1
i	Otomotif	0	0	0	0	1	0	0	0	27	0	0	1
j	<i>Health</i>	1	0	0	1	0	1	0	0	0	26	1	0
k	Property	0	0	0	0	0	1	0	1	0	0	26	2
l	Travel	0	0	0	0	1	0	0	0	0	0	0	29

Berdasarkan hasil klasifikasi yang dihasilkan oleh aplikasi pengklasifikasi artikel berita dan WEKA, berikut ini merupakan nilai akurasi yang dapat dilihat pada tabel 6.10. Nilai akurasi dihitung berdasarkan rumus persamaan 2.7.

**Tabel 6.10 Tingkat Akurasi dari Hasil Klasifikasi Aplikasi**

	Aplikasi	Weka
Akurasi	87.675%	87.675%

Sedangkan nilai *precision*, *recall* dan *f-measures* dari hasil klasifikasi yang dihasilkan oleh aplikasi pengklasifikasi artikel berita dan WEKA dapat dilihat pada tabel 6.11.

Tabel 6.11 Hasil Perhitungan *Precision*, *Recall*, dan *F-Measure* Hasil Klasifikasi

Kategori	<i>Precision</i>		<i>Recall</i>		<i>F-measure</i>	
	Aplikasi	WEKA	Aplikasi	WEKA	Aplikasi	WEKA
nasional	78.13%	78.13%	83.33%	83.33%	80.65%	80.65%
internasional	90.32%	90.32%	93.33%	93.33%	91.80%	91.80%
olahraga	100.00%	100.00%	93.33%	93.33%	96.55%	96.55%
sains	80.77%	80.77%	72.41%	72.41%	76.36%	76.36%
edukasi	76.47%	76.47%	89.66%	89.66%	82.54%	82.54%
ekonomi	86.21%	86.21%	83.33%	83.33%	84.75%	84.75%
tekno	100.00%	100%	80.00%	80.00%	88.89%	88.89%
<i>entertainment</i>	90.32%	90.32%	93.33%	93.33%	91.80%	91.80%
otomotif	96.43%	96.43%	93.10%	93.10%	94.74%	94.74%
<i>health</i>	89.66%	89.66%	86.67%	86.67%	88.14%	88.14%
properti	89.66%	89.66%	86.67%	86.67%	88.14%	88.14%
travel	80.56%	80.56%	96.67%	96.67%	87.88%	87.88%
<b>Rata-rata</b>	<b>88.21%</b>	<b>88.21%</b>	<b>87.65%</b>	<b>87.65%</b>	<b>87.69%</b>	<b>87.69%</b>

Dari tabel kontingensi 6.8 maka tingkat kesalahan klasifikasi dapat digambarkan pada tabel 6.12 berikut ini.

**Tabel 6.12 Tingkat Galat per Kategori**

No	Kategori	Tingkat Galat
1	nasional	16.67 %
2	internasional	6.67 %
3	olahraga	6.67 %
4	sains	<b>27.59 %</b>
5	edukasi	10.34 %
6	ekonomi	16.67 %
7	tekno	20%
8	<i>entertainment</i>	6.67 %
9	otomotif	6.90 %
10	<i>health</i>	13.33 %
11	properti	13.33 %
12	travel	3.33 %

## 6.5. Analisis Hasil Uji Coba

Tahap uji coba telah dilaksanakan dan menghasilkan berbagai nilai. Analisis dari hasil uji coba dipaparkan pada masing-masing bagian berikut ini sesuai dengan jenis uji coba.

### 6.5.1. Uji Verifikasi

Berdasarkan hasil uji kebenaran pada fungsional aplikasi yang telah dilakukan pada bagian tahap uji verifikasi, dapat diketahui bahwa seluruh fungsional aplikasi telah berjalan sesuai desain yang telah dibuat pada bab IV mengenai desain aplikasi. Seluruh fungsi aplikasi pengklasifikasi artikel berita berjalan lancar dan dapat menangani skenario alternatif.

Sedangkan untuk hasil klasifikasi, aplikasi pengklasifikasi artikel berita yang telah dibangun dapat mengklasifikasikan artikel berita. Aplikasi dapat menentukan kategori berita dari artikel berita yang belum diketahui kategori beritanya. Hal ini dapat dibuktikan dengan beberapa contoh hasil klasifikasi yang

dilakukan oleh aplikasi telah sesuai dengan hasil klasifikasi yang dilakukan secara manual. Hasil klasifikasi dari aplikasi juga menghasilkan hasil yang sama dengan hasil klasifikasi yang dilakukan melalui aplikasi WEKA. Walaupun pada hasil klasifikasi masih terdapat artikel berita yang belum dapat diklasifikasikan dengan tepat. Hal ini dimungkinkan karena kata-kata yang ada pada artikel berita tersebut menjadi fitur dari kategori berita lain. Akurasi hasil klasifikasi ini dapat dilihat pada uji validasi terhadap hasil uji klasifikasi.

### 6.5.2. Uji Validasi

Hasil uji validasi terhadap model klasifikasi dengan metode NBC yang dibuat oleh aplikasi pengklasifikasi artikel berita mencakup perhitungan akurasi, *precision*, *recall*, dan *f-measures*. Metode evaluasinya menggunakan metode *cross-validation* dengan jumlah *folds* sebanyak 10 *folds*. Hasil evaluasi pada masing-masing *folds* dapat dilihat pada tabel 6.6. Rata-rata akurasi dari model yang dihasilkan adalah sebesar 86.97%. Dengan nilai akurasi tersebut dapat dinyatakan bahwa model klasifikasi yang dihasilkan aplikasi pengklasifikasi artikel berita memiliki akurasi yang cukup baik. Sedangkan nilai *precision* adalah sebesar 87.60%, yang dapat dinyatakan model klasifikasi memiliki tingkat keakuratan yang cukup baik. Nilai *recall* sebesar 87% yang berarti bahwa model klasifikasi memiliki tingkat sensitivitas cukup baik. Dan nilai *f-measures* sebesar 87.10% yang dapat dinyatakan bahwa model klasifikasi cukup baik dalam menggali informasi pada teks.

Perhitungan akurasi, *precision*, *recall*, dan *f-measure* dari model klasifikasi yang dibuat oleh aplikasi menghasilkan hasil yang sedikit berbeda dengan perhitungan *precision*, *recall*, dan *f-measure* dari model klasifikasi yang dibuat oleh aplikasi WEKA seperti yang dapat dilihat pada tabel 6.7. Hal ini dikarenakan kata atau fitur yang dihasilkan oleh aplikasi pengklasifikasi artikel berita jumlahnya berbeda dengan yang dihasilkan oleh aplikasi WEKA. Jumlah fitur yang dihasilkan oleh WEKA lebih banyak

daripada fitur yang dihasilkan oleh aplikasi yakni 7300 fitur. Sedangkan aplikasi hanya menghasilkan 7103 fitur.

Evaluasi terhadap hasil uji coba juga menggunakan perhitungan akurasi, *precision*, *recall*, dan *f-measure*. Hasil uji klasifikasi dengan menggunakan aplikasi pengklasifikasi artikel berita dan aplikasi WEKA memiliki hasil yang sama seperti yang terlihat pada tabel 6.8 dan 6.9. Tingkat akurasi dari hasil klasifikasi aplikasi pengklasifikasi artikel berita adalah sebesar 87.675% yang dapat dilihat pada tabel 6.10. Pada tabel tersebut juga terlihat bahwa akurasi hasil uji klasifikasi dengan menggunakan aplikasi WEKA juga sama. Sedangkan nilai *precision*, *recall*, dan *f-measure* dari hasil uji klasifikasi dengan menggunakan aplikasi pengklasifikasi berita dan aplikasi WEKA juga memiliki nilai yang sama, hal ini dapat dilihat pada tabel 6.11. Nilai *precision*, *recall*, dan *f-measure* rata-ratanya secara berurutan adalah 88.21%, 87.65%, dan 87.69%. Dari hasil tersebut dapat dinyatakan bahwa hasil uji klasifikasi memiliki tingkat presisi dan tingkat sensitifitas yang cukup baik. Dari hasil ini, dapat disimpulkan bahwa aplikasi dapat digunakan oleh pengguna untuk mencari kategori berita dari artikel berita yang belum diketahui kategori beritanya.

Pada tahap uji, aplikasi pengklasifikasi artikel berita menghasilkan beberapa kesalahan klasifikasi. Hasil prediksi baik prediksi yang benar maupun salah dapat dilihat pada tabel 6.8. Beberapa contohnya dapat dilihat pada tabel 6.12.

**Tabel 6.13 Tabel Kesalahan Klasifikasi**

No	Judul Artikel	Kelas Asli	Kelas Prediksi
1	Berduka, Megawati Tak Gelar "Open House" Lebaran	Nasional	<i>Entertainment</i>
2	ICW Apresiasi Sistem Baru Seleksi CPNS	Nasional	Edukasi

Kesalahan klasifikasi ini dapat terjadi karena nilai prediksi pada kategori lain lebih tinggi dibandingkan dengan nilai prediksi pada kategori asli. Nilai prediksi didapatkan dari hasil perhitungan kata/fitur dalam suatu artikel berita berdasarkan

model klasifikasi yang dibuat dari data latih. Hal ini dapat dilihat pada lampiran C. Sedangkan untuk nilai prediksi dari masing-masing data uji dapat dilihat pada lampiran D.

Pada tabel 6.12 contoh no 1, nilai prediksi artikel tersebut pada kategori nasional hanya sebesar 0.115, internasional sebesar 0.001, edukasi sebesar 0.006, sedangkan *entertainment* sebesar 0.877. Karena nilai prediksi paling tinggi adalah kategori *entertainment* maka artikel berita dimasukkan pada kategori *entertainment* walaupun artikel tersebut seharusnya dimasukkan pada kategori nasional. Pada contoh no 2, artikel yang seharusnya masuk pada kategori nasional namun dimasukkan ke dalam kategori edukasi. Pada artikel tersebut terdapat banyak fitur pada kategori edukasi. Nilai prediksinya pada kategori nasional hanya 0.005. Nilai tersebut lebih rendah dibandingkan nilai prediksi pada kategori edukasi yang sebesar 0.995.

Dari tabel 6.8 dan 6.12 dapat dilihat bahwa kategori sains paling banyak mengalami kesalahan klasifikasi kemudian diikuti oleh tekno, nasional, dan ekonomi. Kesalahan pada kategori sains menyebar pada hampir seluruh kategori berita lainnya. Hal ini berbeda dengan tekno yang juga memiliki jumlah kesalahan klasifikasi cukup tinggi, namun kesalahannya tidak terlalu menyebar yakni paling tinggi dimasukkan pada kategori edukasi dan sains. Sedangkan kategori berita travel paling sedikit kesalahan klasifikasinya. Pada tabel 6.8 juga dapat diketahui bahwa artikel berita yang salah paling tinggi dimasukkan ke dalam kategori berita edukasi, nasional, dan travel. Hal ini menunjukkan bahwa fitur pada ketiga kategori ini paling banyak memiliki kesamaan dengan kategori lainnya. Sedangkan kategori olahraga dan tekno paling sedikit mendapatkan artikel berita yang salah klasifikasi. Hal ini menunjukkan bahwa kedua kelas ini memiliki fitur yang cukup berbeda dengan kategori lainnya sehingga artikel berita jarang disalah-klasifikasikan pada kategori ini. Kesalahan klasifikasi yang terjadi dapat dikurangi dengan menambahkan data latih yang memiliki fitur lebih representatif, sehingga kata/fitur pada masing-masing kategori juga dapat lebih.

## BAB VI KESIMPULAN

Setelah menyelesaikan perancangan dan pembangunan aplikasi pengklasifikasi artikel berita berbahasa Indonesia berbasis *Naïve Bayes classifier* menggunakan *confix-stripping stemmer*, dapat diambil kesimpulan sebagai berikut:

- a. Aplikasi pengklasifikasi artikel berita yang telah berhasil dibangun dalam Tugas Akhir ini telah berhasil diujicobakan pada artikel berita yang diambil dari portal berita *online* <http://www.kompas.com> dengan total data sampel sebesar 1.195 yang terdiri dari 12 kategori. Total data sampel tersebut dibagi menjadi dua jenis data, yakni 70% untuk data latih dan 30% untuk data uji pada setiap kategori berita. Aplikasi pengklasifikasi artikel berita berhasil mengklasifikasikan artikel berita sesuai dengan kategori beritanya. Aplikasi ini dapat dengan mudah digunakan, sehingga memungkinkan untuk diimplementasikan dan digunakan untuk membantu para editor berita dalam melakukan klasifikasi artikel berita berbahasa Indonesia
- b. Dari hasil evaluasi kinerja aplikasi dapat disimpulkan bahwa:
  1. Model klasifikasi yang dibangun dengan menggunakan data latih menghasilkan klasifikasi dengan nilai akurasi, *precision*, *recall*, dan *f-measure* berturut-turut sebesar 86.738%, 87.20%, 86.70%, dan 86.80%. Kinerja model ini hampir sama dengan yang dihasilkan oleh aplikasi WEKA yang berturut-turut memberikan nilai akurasi, *precision*, *recall*, dan *f-measure* sebesar 86.857%, 87.30%, 86.90%, dan 86.90%. Perbedaan hasil yang sangat tidak signifikan ini mengindikasikan bahwa model pengklasifikasi yang dibangun dalam Tugas Akhir ini sudah valid.
  2. Hasil pengujian terhadap model pengklasifikasi yang telah berhasil dibangun dengan menggunakan data uji memberikan hasil klasifikasi dengan nilai akurasi,

*precision*, *recall*, dan *f-measure* berturut-turut sebesar 87.675%, 88.21%, 87.65%, dan 87.69%. Hasil uji klasifikasi ini persis sama dengan yang dihasilkan oleh aplikasi WEKA. Dengan demikian, aplikasi pengklasifikasi artikel berita yang berhasil diimplementasikan dalam lingkungan sistem operasi Windows dapat digunakan untuk membantu para editor berita dalam pengklasifikasian artikel berita berbahasa Indonesia.

## DAFTAR PUSTAKA

- Ariola, M. M. (2006). *Principles and methods of research*. Sta. Mesa Manila: Rex Printing Company.
- Asian, J. A. (2007). Stemming Indonesian : a confix-stripping approach. *ACM Transactions on Asian Language Information Processing (TALIP)*, 6(4), 1-33.
- Barlas, Y. (1994). Model validation in system dynamics. *International System Dynamics Conference* (pp. 1-10). Scotland: Sterling.
- Dragut, E. F. (2009). Stop word and related problems in web interface integration. *Proceedings of the VLDB Endowment*, 2(1), 349-360.
- gunadarma. (n.d.). *elearning.gunadarma.ac.id*. Retrieved Juni 17, 2014, from [http://elearning.gunadarma.ac.id/docmodul/risetbisnis.../05\\_bab\\_3\\_sampel.pdf](http://elearning.gunadarma.ac.id/docmodul/risetbisnis.../05_bab_3_sampel.pdf)
- Hamzah, A. (2012). Klasifikasi teks dengan Naive Bayes Classifier (NBC) untuk pengelompokkan teks berita dan abstract akademis. *Seminar Nasional Aplikasi Sains dan Teknologi Periode III*. Yogyakarta.
- Jian Sheu, J. (2008). An Efficient Two-Phase Spam Filtering Method. *Sciendirect*.
- Kamus Besar Bahasa Indonesia. (n.d.). *Korpus*. Retrieved Maret 14, 2014, from [kbbi.web.id: http://kbbi.web.id/korpus](http://kbbi.web.id/korpus)
- Liu, B., Li, X., Lee, W., & Yu, P. (2004). Text classification by labeling words. *Proceedings of the 19th national conference on Artificial intelligence* (pp. 425-430). AAAI Press.
- Miller, T. (2005). *Data and Tekx Mining A Business Application Approach*. New Jersey, USA: Prentice Hall.
- Oracle. (n.d.). *Classification*. Retrieved Maret 14, 2014, from [docs.oracle.com: http://docs.oracle.com/cd/B28359\\_01/datamine.111/b28129/classify.htm#i1005746](http://docs.oracle.com/cd/B28359_01/datamine.111/b28129/classify.htm#i1005746)

Oracle. (n.d.). *Clustering*. Retrieved Maret 14, 2014, from /docs.oracle.com:

[http://docs.oracle.com/cd/B28359\\_01/datamine.111/b28129/clustering.htm#DMCON008](http://docs.oracle.com/cd/B28359_01/datamine.111/b28129/clustering.htm#DMCON008)

Pitoyo, A. (2014, Januari 15 ). *Jumlah pengguna internet Indonesia capai 71,19 juta pada 2013*. Retrieved Maret 14, 2014, from merdeka.com:

<http://www.merdeka.com/teknologi/jumlah-pengguna-internet-indonesia-capai-7119-juta-pada-2013.html>

Refaeilzadeh, P., Tang, L., & Liu, H. (2009). Cross-validation. In *Encyclopedia of database systems* (pp. 532-538). USA: Springer.

Rijsbergen, C. (1979). *Information Retrieval Glasgow*. Information retrieval group, University of Glasgow .

Tala, a. Z. (2003). *A Study of Stemming Effects on Information Retrieval in Bahasa Indonesia*. Netherlands: Master of Logic Project. Institute for Logic, Language and Computation, Universiteit van Amsterdam.

Tan, S. &. (2006). *Introduction to Data Mining*. Pearson Addison Wesley.

University of Waikato. (n.d.). *Weka 3: Data Mining Software in Java*. Retrieved from [www.cs.waikato.ac.nz: http://www.cs.waikato.ac.nz/ml/weka/](http://www.cs.waikato.ac.nz/ml/weka/)

Weiss, S. M. (2010). *Text mining: predictive methods for analyzing unstructured information*. New York: Springer.

## RIWAYAT PENULIS

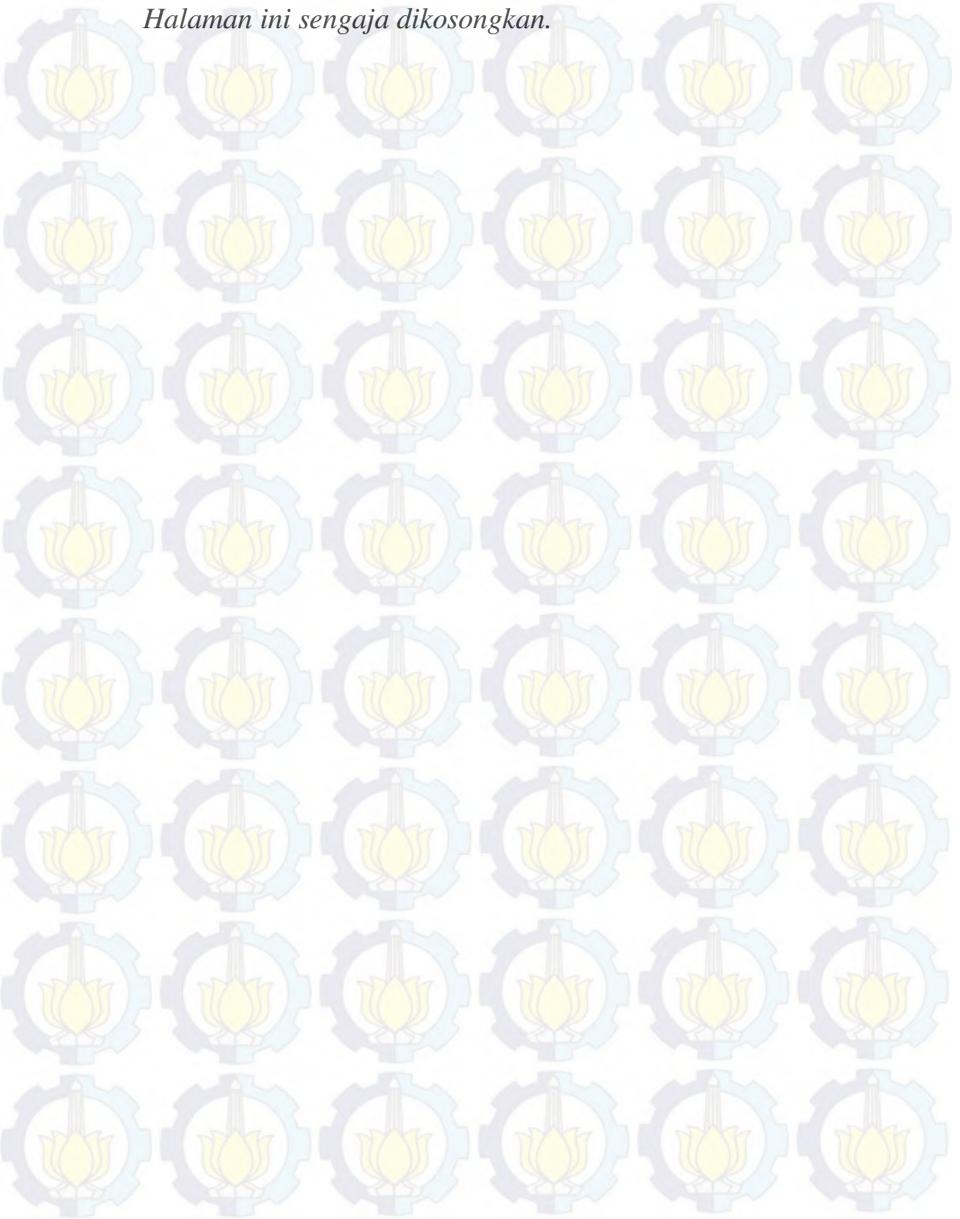


Amalia Anjani Arifiyanti dilahirkan di Blitar pada tanggal 12 Agustus 1992. Gadis yang disapa Amel ini menghabiskan waktu kecilnya di desa Turen Kabupaten Malang. Menempuh pendidikan dari TK hingga SMP di Turen, yakni SD Negeri 1 Turen, dan SMPN 1 Turen. Jenjang pendidikan berikutnya Amel memilih untuk menempuh di SMK Telkom SAndhy Putra Malang dan

lulus pada tahun 2010. Pada tahun yang sama, Amel melanjutkan pendidikan ke bangku perguruan tinggi di Jurusan Sistem Informasi Institut Teknologi Sepuluh Nopember Surabaya setelah dinyatakan lolos sebagai salah satu penerima bidik misi dengan NRP 5210100106.

Aktivitasnya selain kuliah cukup beragam dengan mengikuti keanggotaan organisasi mahasiswa ITS dan beberapa kepanitian dari berbagai kegiatan di ITS. Pada tahun keduanya Amel sempat menjadi staff di dua organisas yang berbeda, yakni departemen Hubungan Luar HMSI ITS dan kementrian Pengembangan Sumber Daya Mahasiswa BEM ITS. Dari pengalaman itulah Amel memutuskan menjadi staff departemen Pengembangan Sumber Daya Mahasiswa HMSI pada tahun kepengurusan berikutnya. Karena keinginannya untuk lebih mengasah kemampuan komunikasinya, Amel tergabung dengan kepeemanduan ITS mulai tahun 2011. Selain itu, selama menyandang gelar sebagai mahasiswa, Amel menyempatkan diri menjadi asisten praktikum beberapa matakuliah. Dan akhirnya Amel memilih Sistem Pendukung Keputusan sebagai bidang minatnya dalam pengerjaan tugas akhir, karena ketertarikannya dalam bidang pengolahan data dan informasi serta *machine learning*.

*Halaman ini sengaja dikosongkan.*



## LAMPIRAN A

Lampiran ini berisi mengenai skenario dari masing-masing *usecase* yang telah dibuat pada desain aplikasi.

### A.1. Skenario *Usecase* Menambahkan Data Artikel Berita untuk Latih Klasifikasi

UC-01 Menambahkan data artikel berita untuk latih klasifikasi	
Deskripsi	<i>Usecase</i> ini menggambarkan proses menambahkan artikel berita yang akan digunakan sebagai data latih klasifikasi.
Aktor	Pengguna
Pra-kondisi	Aktor telah membuka form klasifikasi dan aplikasi telah terhubung dengan basis data
Skenario Normal	<ol style="list-style-type: none"> <li>1. Aktor memilih pilihan 'data latih'</li> <li>2. Aplikasi menampilkan semua kolom isian</li> <li>3. Aktor mengisi semua kolom isian, kemudian menekan tombol tambah.</li> <li>4. Aplikasi menambahkan data yang diisikan aktor ke dalam basis data.</li> <li>5. Aplikasi menampilkan pesan sukses.</li> </ol>
Skenario Alternatif	Jika terdapat satu kolom isian yang kosong maka aplikasi akan menampilkan pesan kesalahan. Setelah menekan tombol OK, pengguna harus mengisi semua kolom kembali.

## A.2. Skenario *Usecase* Menambahkan Data Artikel Berita untuk Uji Klasifikasi

UC-02 Menambahkan data artikel berita untuk uji klasifikasi	
Deskripsi	<i>Usecase</i> ini menggambarkan proses menambahkan artikel berita yang akan digunakan sebagai data uji klasifikasi.
Aktor	Pengguna
Pra-kondisi	Aktor telah membuka form klasifikasi dan aplikasi telah terhubung dengan basis data
Skenario Normal	<ol style="list-style-type: none"> <li>1. Aktor memilih pilihan 'data uji'</li> <li>2. Aplikasi menampilkan semua kolom isian kecuali isian kategori</li> <li>3. Aktor mengisi semua kolom isian, kemudian menekan tombol tambah.</li> <li>4. Aplikasi menambahkan data yang diisikan aktor ke dalam basis data.</li> <li>5. Aplikasi menampilkan pesan sukses.</li> </ol>
Skenario Alternatif	Jika terdapat satu atau lebih kolom isian yang kosong maka aplikasi akan menampilkan pesan kesalahan. Setelah menekan tombol OK, pengguna harus mengisi semua kolom kembali.

### A.3. Skenario *Usecase* Melakukan Praproses Data Latih

UC-03 Melakukan praproses data latih	
Deskripsi	<i>Usecase</i> ini menggambarkan proses dari praproses pada data latih yang tersimpan dalam basis data
Aktor	Pengguna
Pra-kondisi	Aktor telah membuka form klasifikasi dan aplikasi telah terhubung dengan basis data
Skenario Normal	<ol style="list-style-type: none"> <li>1. Aktor memilih pilihan 'data latih'</li> <li>2. Aplikasi menampilkan kolom isian lokasi file keluaran</li> <li>3. Aktor mengisi kolom isian, kemudian menekan tombol praproses.</li> <li>4. Aplikasi melakukan praproses terhadap data latih dan menyimpan hasilnya pada file keluaran.</li> <li>5. Aplikasi menampilkan pesan sukses.</li> </ol>
Skenario Alternatif	Jika kolom isian tidak diisi maka aplikasi akan menampilkan pesan kesalahan. Setelah menekan tombol OK, pengguna harus mengisi kolom isian

### A.4. Skenario *Usecase* Melakukan Praproses Data Latih

UC-04 Melakukan praproses data uji	
Deskripsi	<i>Usecase</i> ini menggambarkan proses dari praproses pada data latih yang tersimpan dalam basis data
Aktor	Pengguna
Pra-kondisi	Aktor telah membuka form klasifikasi dan aplikasi telah terhubung dengan basis data
Skenario Normal	<ol style="list-style-type: none"> <li>1. Aktor memilih pilihan 'data uji'</li> <li>2. Aplikasi tidak menampilkan kolom isian</li> <li>3. Aktor menekan tombol praproses.</li> <li>4. Aplikasi melakukan praproses terhadap data latih dan menyimpan hasilnya pada basis data.</li> <li>5. Aplikasi menampilkan pesan sukses.</li> </ol>
Skenario Alternatif	-

### A.5. Skenario *Usecase* Melakukan Latih Klasifikasi

UC-05 Melakukan latihan klasifikasi	
Deskripsi	<i>Usecase</i> ini menggambarkan proses pembuatan model klasifikasi yang merupakan keluaran dari tahap latihan klasifikasi. Proses ini menggunakan hasil dari praproses data latih.
Aktor	Pengguna
Pra-kondisi	Aktor telah membuka form klasifikasi dan aplikasi telah terhubung dengan basis data
Skenario Normal	<ol style="list-style-type: none"> <li>1. Aktor mengisi semua kolom isian dan menekan tombol ok</li> <li>2. Aplikasi melakukan pembuatan model dan menyimpannya dalam file keluaran.</li> <li>3. Aplikasi menampilkan pesan sukses.</li> </ol>
Skenario Alternatif	Jika terdapat satu atau lebih kolom isian yang kosong maka aplikasi akan menampilkan pesan kesalahan. Setelah menekan tombol OK, pengguna harus mengisi semua kolom kembali.

### A.6. Skenario *Usecase* Melakukan Uji Klasifikasi

UC-06 Melakukan uji klasifikasi	
Deskripsi	<i>Usecase</i> ini menggambarkan proses klasifikasi dari data uji berdasarkan model klasifikasi yang telah dibuat.
Aktor	Pengguna
Pra-kondisi	Aktor telah membuka form klasifikasi dan aplikasi telah terhubung dengan basis data
Skenario Normal	<ol style="list-style-type: none"> <li>1. Aktor mengisi kolom isian, kemudian menekan tombol ok.</li> <li>2. Aplikasi melakukan praproses terhadap data latih dan menyimpan hasilnya pada file keluaran.</li> <li>3. Aplikasi menampilkan pesan sukses.</li> </ol>
Skenario Alternatif	Jika kolom isian tidak diisi maka aplikasi akan menampilkan pesan kesalahan. Setelah menekan tombol OK, pengguna harus mengisi kolom isian

## LAMPIRAN B

### B.1. Login

#### B.1.1 Alur Normal

The image shows two screenshots of a web application. The top screenshot is the login page, titled "APLIKASI PENGKLASIFIKASI ARTIKEL BERITA BERBAHASA INDONESIA NAIVE BAYES CLASSIFIER DAN CONFIX-STRIPPING STEMMER". It features a login form with "Username : admin" and "Password : \*\*\*" fields. Below the fields are two buttons: "LOGIN ADMIN" and "KELUAR". A blue box highlights the username and password fields. A blue arrow points from the "LOGIN ADMIN" button to the second screenshot. The second screenshot is the "Tambah Artikel Berita" page, titled "Proses TAMBAH ARTIKEL BERITA". It contains a form with a dropdown menu for "Pilihan Jenis Artikel" (set to "Data Latin"), a "Judul Artikel:" field, a "Tanggal:" field with a date picker, a "Nama Kategori:" dropdown menu (set to "nasionat"), and an "Isi Berita:" text area.

Lampiran B - 1 Login Alur Normal

#### B.1.2 Alur Alternatif 1

The image shows a screenshot of the login page from the application. The "Username : admin" and "Password : \*\*\*" fields are highlighted with a blue box. Below the fields is the "LOGIN ADMIN" button. A blue arrow points from the "LOGIN ADMIN" button to an error dialog box. The error dialog box is titled "Error" and contains the message "User dan password salah" (User and password are wrong) with an "OK" button.

Lampiran B - 2 Login Alur Alternatif 1

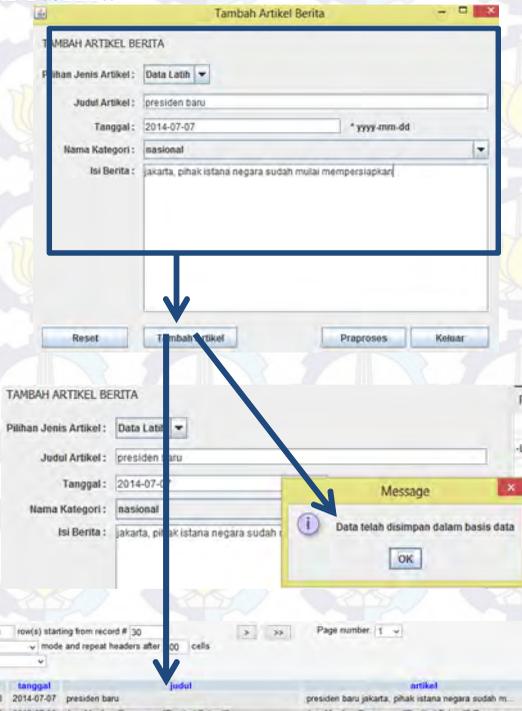
### B.1.3 Alur Alternatif 2



Lampiran B - 3 Login Alur Alternatif 2

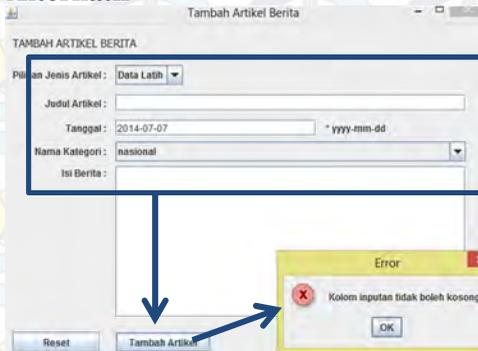
## B.2. Menambahkan Data Artikel Untuk Data Latih

### B.2.1 Alur Normal



Lampiran B - 4 Menambahkan Data Artikel Untuk Data Latih Alur Normal

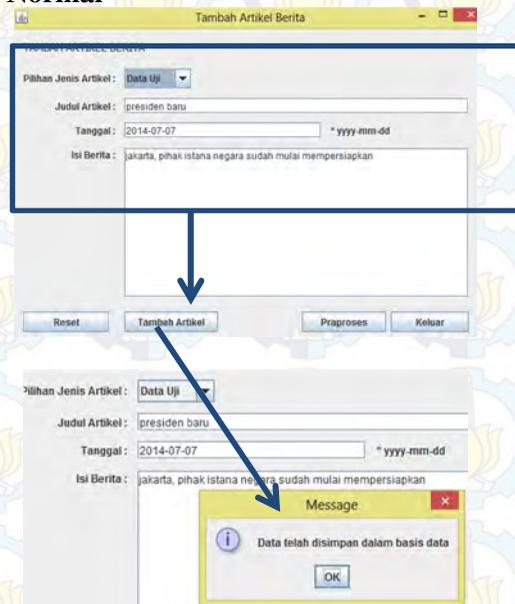
## B.2.2 Alur Alternatif



Lampiran B - 5 Menambahkan Data Artikel Untuk Data Latih Alur Alternatif

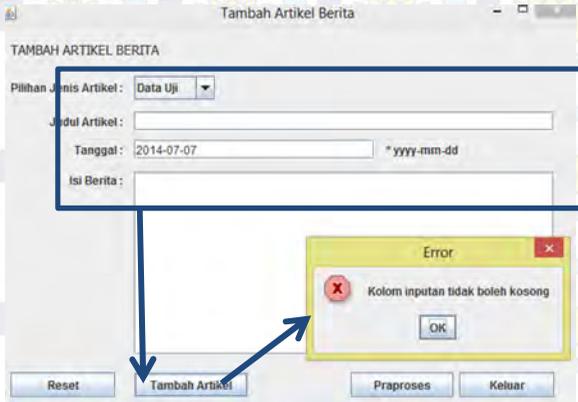
## B.3. Menambahkan Data Artikel Untuk Data Uji

### B.3.1 Alur Normal



Lampiran B - 6 Menambahkan Data Artikel Untuk Data Uji Alur Normal

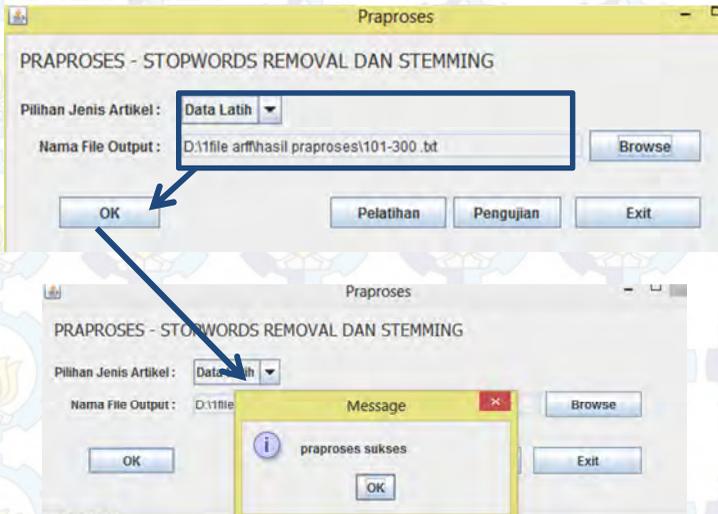
### B.3.2 Alur Alternatif



Lampiran B - 7 Menambahkan Data Artikel Untuk Data Uji Alur Alternatif

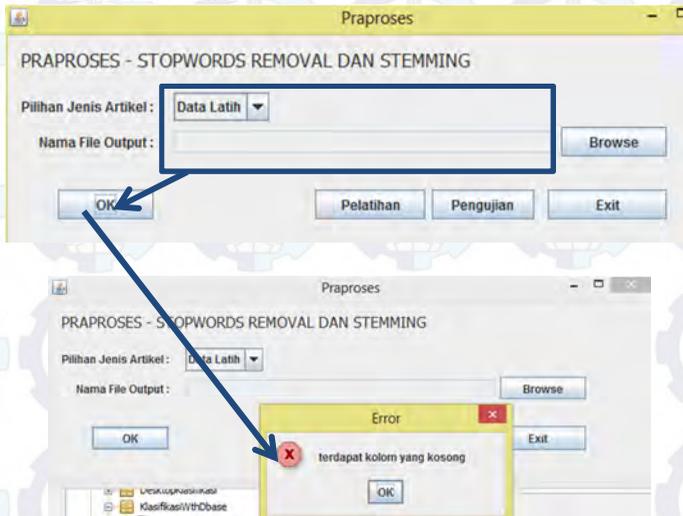
## B.4 Melakukan Praproses Data Latih

### B.4.1 Alur Normal



Lampiran B - 8 Melakukan Praproses Data Latih Alur Normal

## B.4.2 Alur Alternatif



Lampiran B - 9 Melakukan Praproses Data Latih Alur Alternatif

## B.5 Melakukan Praproses Data Uji

### B.5.1 Alur Normal

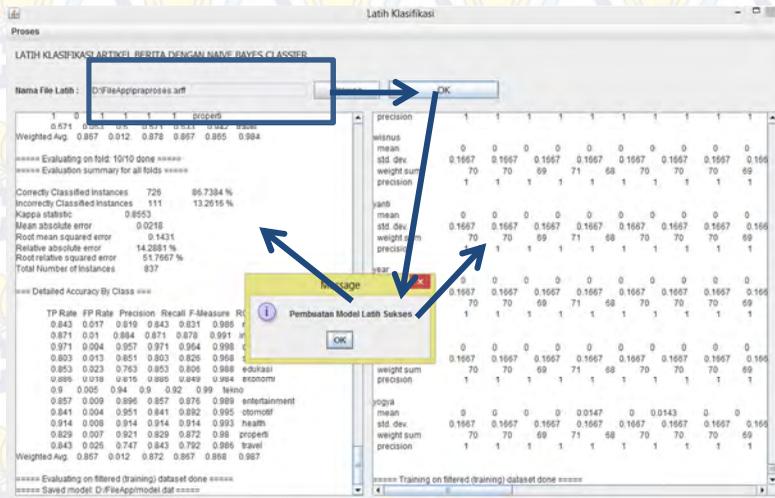


Lampiran B - 10 Melakukan Praproses Data Uji Alur Normal

B-6

## B.6 Melakukan Latih Klasifikasi

### B.6.1 Alur Normal



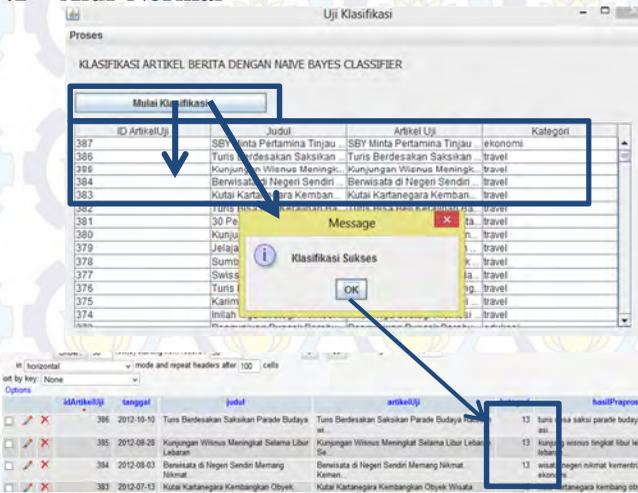
Lampiran B - 11 Melakukan Latih Klasifikasi Alur Normal

### B.6.2 Alur Alternatif



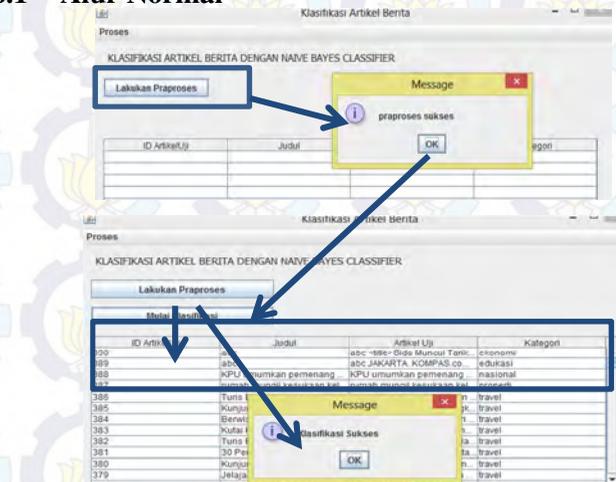
Lampiran B - 12 Melakukan Latih Klasifikasi Alur Alternatif

**B.7 Melakukan Uji Klasifikasi**  
**B.7.1 Alur Normal**



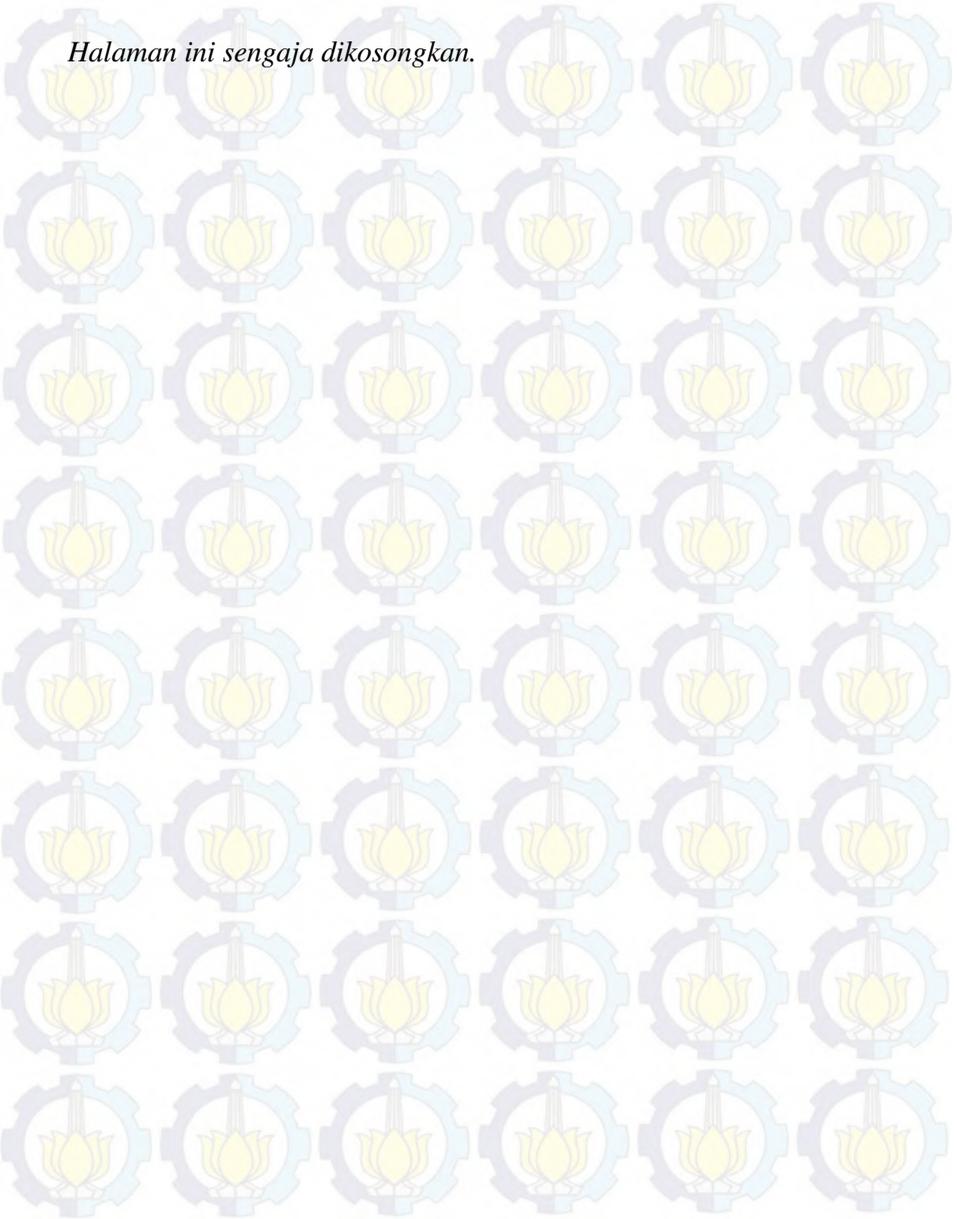
Lampiran B - 13 Melakukan Uji Klasifikasi Alur Normal

**B.8 Mengklasifikasikan Artikel Berita**  
**B.8.1 Alur Normal**



Lampiran B - 14 Mengklasifikasikan Artikel Berita Alur Normal

*Halaman ini sengaja dikosongkan.*



## LAMPIRAN C

Pada lampiran ini dicantumkan 100 urutan kata/fitur dari masing-masing kategori. Berikut ini daftar kata/fitur dan rata-rata kemunculannya dari seluruh data latih. Kata yang dicetak tebal merupakan kata yang juga muncul pada kategori lain.

Lampiran C - 1 Daftar Fitur Bagian 1

no	nasional		internasional		olahraga		sains		edukasi		ekonomi	
	kata	mean	kata	mean	kata	mean	kata	mean	kata	mean	kata	mean
1	<b>jakarta</b>	0.786	<b>negara</b>	0.529	juara	0.638	<b>temu</b>	0.479	pendidikan	0.735	<b>indonesia</b>	0.586
2	komisi	0.414	<b>as</b>	0.414	<b>dunia</b>	0.536	<b>hasil</b>	0.437	guru	0.471	<b>persen</b>	0.586
3	<b>ketua</b>	0.414	<b>presiden</b>	0.357	<b>hasil</b>	0.464	<b>teliti</b>	0.437	ajar	0.456	<b>jakarta</b>	0.529
4	partai	0.371	<b>orang</b>	0.357	<b>tim</b>	0.406	<b>milik</b>	0.352	<b>indonesia</b>	0.412	<b>usaha</b>	0.457
5	korupsi	0.343	<b>negeri</b>	0.357	raih	0.406	ilmuwan	0.310	<b>nasional</b>	0.412	<b>rp</b>	0.443
6	<b>indonesia</b>	0.343	<b>menteri</b>	0.343	<b>kali</b>	0.391	<b>bentuk</b>	0.282	sekolah	0.412	<b>tingkat</b>	0.429
7	<b>politik</b>	0.314	<b>pemerintah</b>	0.314	<b>nomor</b>	0.377	<b>hidup</b>	0.268	<b>jakarta</b>	0.397	<b>harga</b>	0.414
8	<b>anggota</b>	0.314	<b>aman</b>	0.300	unggul	0.377	<b>nama</b>	0.254	siswa	0.382	<b>pemerintah</b>	0.400
9	kpk	0.300	<b>warga</b>	0.271	<b>buka</b>	0.362	<b>alam</b>	0.239	<b>menteri</b>	0.368	<b>capai</b>	0.371
10	<b>kait</b>	0.300	<b>salah</b>	0.257	menang	0.362	<b>kawasan</b>	0.239	<b>tingkat</b>	0.338	<b>ekonomi</b>	0.329

no	nasional		internasional		olahraga		sains		edukasi		ekonomi	
	kata	mean	kata	mean	kata	mean	kata	mean	kata	mean	kata	mean
11	<b>hukum</b>	0.286	<b>temu</b>	0.257	<b>indonesia</b>	0.348	<b>indonesia</b>	0.225	mahasiswa	0.338	<b>pasar</b>	0.329
12	<b>wakil</b>	0.271	<b>bicara</b>	0.257	tanding	0.333	<b>salah</b>	0.225	<b>pemerintah</b>	0.324	triliun	0.314
13	<b>presiden</b>	0.271	<b>wilayah</b>	0.257	pebalap	0.319	<b>capai</b>	0.211	<b>program</b>	0.294	<b>negara</b>	0.300
14	<b>dpr</b>	0.271	senjata	0.257	<b>tiga</b>	0.304	Bumi	0.211	<b>universitas</b>	0.279	<b>turun</b>	0.300
15	<b>salah</b>	0.257	<b>lapor</b>	0.229	<b>gelar</b>	0.290	<b>Juta</b>	0.197	<b>dasar</b>	0.265	<b>pt</b>	0.286
16	<b>orang</b>	0.257	<b>bantu</b>	0.229	hadap	0.290	<b>Tim</b>	0.183	<b>budaya</b>	0.265	<b>milik</b>	0.271
17	<b>hasil</b>	0.257	kelompok	0.229	lawan	0.290	<b>dasar</b>	0.183	<b>milik</b>	0.250	<b>bank</b>	0.257
18	berantas	0.257	<b>serang</b>	0.229	balap	0.290	<b>Air</b>	0.183	<b>salah</b>	0.235	<b>jumat</b>	0.243
19	<b>proses</b>	0.243	tewas	0.229	<b>catat</b>	0.275	<b>manusia</b>	0.183	<b>negeri</b>	0.221	<b>juta</b>	0.243
20	<b>selasa</b>	0.229	<b>internasional</b>	0.229	<b>tampil</b>	0.275	<b>bahan</b>	0.183	<b>terima</b>	0.221	<b>nilai</b>	0.243
21	<b>sabtu</b>	0.229	<b>amerika</b>	0.229	peringkat	0.275	jurnal	0.183	<b>ketua</b>	0.221	<b>tumbuh</b>	0.243
22	putus	0.214	<b>selasa</b>	0.214	final	0.275	<b>Kali</b>	0.169	<b>selenggara</b>	0.221	<b>miliar</b>	0.243
23	<b>masyarakat</b>	0.214	<b>senin</b>	0.214	<b>minggu</b>	0.261	<b>akibat</b>	0.169	<b>orang</b>	0.206	kenaikan	0.243
24	duga	0.214	<b>kamis</b>	0.214	empat	0.261	<b>kepala</b>	0.169	<b>bidang</b>	0.206	<b>menteri</b>	0.229
25	<b>dasar</b>	0.214	nyata	0.214	musim	0.261	<b>badan</b>	0.169	uji	0.206	<b>selasa</b>	0.229
26	undang	0.200	pasu	0.214	tumamen	0.261	Jarak	0.169	<b>kepala</b>	0.191	<b>dunia</b>	0.229

no	nasional		internasional		olahraga		sains		edukasi		ekonomi	
	kata	mean	kata	mean	kata	mean	kata	mean	kata	mean	kata	mean
27	tindak	0.200	militer	0.214	gp	0.261	buat	0.169	daerah	0.191	jual	0.229
28	temu	0.200	dukung	0.200	milik	0.246	sumber	0.169	jawa	0.191	harap	0.214
29	tangan	0.200	juru	0.200	babak	0.246	tanah	0.155	kamis	0.191	uang	0.214
30	hubung	0.200	timur	0.200	sabtu	0.232	daerah	0.155	ikut	0.191	industri	0.214
31	atur	0.200	kota	0.200	pekan	0.232	habis	0.155	bangun	0.191	minyak	0.214
32	menilai	0.186	sabtu	0.186	ikut	0.232	baru	0.155	anak	0.191	hasil	0.200
33	jalan	0.186	serikat	0.186	kalah	0.232	universitas	0.155	studi	0.191	presiden	0.200
34	dana	0.186	angkat	0.186	china	0.217	hutan	0.155	sma	0.191	banding	0.200
35	bentuk	0.186	hasil	0.171	gagal	0.217	university	0.155	kelas	0.191	produksi	0.200
36	bambang	0.186	tingkat	0.171	baik	0.217	konservasi	0.155	sd	0.191	target	0.200
37	pidana	0.171	polisi	0.171	posisi	0.217	harap	0.141	kali	0.177	sektor	0.200
38	pemerintah	0.171	pejabat	0.171	seri	0.217	orang	0.141	harap	0.177	masyarakat	0.186
39	menteri	0.171	obama	0.171	menit	0.217	masuk	0.141	jumat	0.177	masuk	0.186
40	mantan	0.171	asing	0.171	harap	0.203	negara	0.141	hadir	0.177	senin	0.186
41	ilu	0.171	anggota	0.157	mantan	0.203	kondisi	0.141	rp	0.177	tambah	0.186
42	calon	0.171	tahan	0.157	cepat	0.203	as	0.141	smp	0.177	produk	0.186
43	adil	0.171	rencana	0.157	bagus	0.203	utara	0.141	hasil	0.162	negeri	0.171

no	nasional		internasional		olahraga		sains		edukasi		ekonomi	
	kata	mean	kata	mean	kata	mean	kata	mean	kata	mean	kata	mean
44	<b>usaha</b>	0.157	<b>rabu</b>	0.157	petenis	0.203	<b>ambil</b>	0.141	<b>capai</b>	0.162	<b>kerja</b>	0.171
45	<b>uang</b>	0.157	<b>juta</b>	0.157	<b>tahan</b>	0.188	<b>data</b>	0.141	<b>proses</b>	0.162	<b>direktur</b>	0.171
46	<b>selesai</b>	0.157	bebas	0.157	<b>keras</b>	0.188	<b>pakai</b>	0.141	<b>masyarakat</b>	0.162	<b>gabung</b>	0.171
47	rapat	0.157	<b>pekan</b>	0.157	enam	0.188	<b>wilayah</b>	0.141	<b>selasa</b>	0.162	<b>batas</b>	0.171
48	pimpin	0.157	resmi	0.157	<b>rumah</b>	0.188	tanam	0.141	<b>giat</b>	0.162	dagang	0.171
49	<b>pilih</b>	0.157	<b>umum</b>	0.157	<b>berat</b>	0.188	matahari	0.141	<b>kerja</b>	0.162	<b>investasi</b>	0.171
50	<b>nilai</b>	0.157	<b>selatan</b>	0.157	latih	0.188	spesies	0.141	<b>sesuai</b>	0.162	impor	0.171
51	<b>negara</b>	0.157	korban	0.157	putra	0.188	Nasa	0.141	<b>masuk</b>	0.147	persero	0.171
52	lembaga	0.157	<b>asia</b>	0.157	poin	0.188	antariksa	0.141	<b>negara</b>	0.147	<b>salah</b>	0.157
53	<b>langsung</b>	0.157	barack	0.157	singkir	0.188	<b>cepat</b>	0.127	<b>rabu</b>	0.147	<b>ketua</b>	0.157
54	laksana	0.157	<b>indonesia</b>	0.143	<b>prestasi</b>	0.188	<b>Jalan</b>	0.127	<b>butuh</b>	0.147	<b>kamis</b>	0.157
55	kpu	0.157	<b>ambil</b>	0.143	grand	0.188	<b>usia</b>	0.127	<b>kait</b>	0.147	<b>utama</b>	0.157
56	<b>kerja</b>	0.157	<b>minggu</b>	0.143	<b>inggris</b>	0.174	<b>rabu</b>	0.127	<b>buka</b>	0.147	<b>dana</b>	0.157
57	<b>harap</b>	0.157	<b>capai</b>	0.143	<b>jalan</b>	0.174	<b>alami</b>	0.127	<b>bantu</b>	0.147	<b>dukung</b>	0.157
58	dewan	0.157	bahas	0.143	tujuh	0.174	<b>nasional</b>	0.127	ilmu	0.147	<b>dpr</b>	0.157
59	<b>bukti</b>	0.157	<b>akibat</b>	0.143	<b>total</b>	0.174	<b>selatan</b>	0.127	beasiswa	0.147	dorong	0.157

no	nasional		internasional		olahraga		sains		edukasi		ekonomi	
	kata	mean	kata	mean	kata	mean	kata	mean	kata	mean	kata	mean
60	<b>bicara</b>	0.157	<b>keras</b>	0.143	putri	0.174	<b>beda</b>	0.127	<b>biaya</b>	0.132	<b>rencana</b>	0.157
61	<b>beda</b>	0.157	<b>provinsi</b>	0.143	set	0.174	<b>ubah</b>	0.127	<b>langsung</b>	0.132	<b>as</b>	0.157
62	terang	0.143	<b>kawasan</b>	0.143	<b>usaha</b>	0.159	<b>pusat</b>	0.127	peserta	0.132	dollar	0.157
63	suap	0.143	<b>informasi</b>	0.143	<b>jakarta</b>	0.159	<b>kembang</b>	0.127	<b>wakil</b>	0.132	saham	0.157
64	<b>senin</b>	0.143	henti	0.143	jelang	0.159	<b>daya</b>	0.127	<b>acara</b>	0.132	ekspor	0.157
65	<b>satu</b>	0.143	<b>kunjung</b>	0.143	australia	0.159	meter	0.127	<b>kompas</b>	0.132	<b>bidang</b>	0.143
66	<b>rp</b>	0.143	<b>utara</b>	0.143	kuat	0.159	<b>proses</b>	0.127	<b>dinas</b>	0.132	<b>butuh</b>	0.143
67	ri	0.143	pbb	0.143	<b>gabung</b>	0.159	panas	0.127	didik	0.132	<b>kait</b>	0.143
68	polri	0.143	<b>inggris</b>	0.143	serah	0.159	<b>ukur</b>	0.127	nuh	0.132	<b>biaya</b>	0.143
69	pemilihan	0.143	afganistan	0.143	cedera	0.159	amat	0.127	<b>timur</b>	0.118	<b>upaya</b>	0.143
70	<b>milik</b>	0.143	<b>politik</b>	0.129	ajang	0.159	darat	0.127	<b>upaya</b>	0.118	<b>catat</b>	0.143
71	<b>mengaku</b>	0.143	<b>kait</b>	0.129	games	0.159	energi	0.127	<b>direktur</b>	0.118	<b>ubah</b>	0.143
72	libat	0.143	<b>wakil</b>	0.129	urut	0.159	<b>riset</b>	0.127	<b>khusus</b>	0.118	<b>atur</b>	0.143
73	<b>kompas</b>	0.143	<b>masyarakat</b>	0.129	tunggal	0.159	hewan	0.127	<b>kualitas</b>	0.118	<b>alami</b>	0.143
74	<b>khusus</b>	0.143	<b>jalan</b>	0.129	semifinal	0.159	<b>dunia</b>	0.113	<b>prestasi</b>	0.118	<b>tahan</b>	0.143
75	<b>kantor</b>	0.143	<b>usaha</b>	0.129	medali	0.159	<b>minggu</b>	0.113	<b>selesai</b>	0.118	<b>dampak</b>	0.143
76	<b>badan</b>	0.143	<b>kantor</b>	0.129	formula	0.159	<b>senin</b>	0.113	<b>internasional</b>	0.118	tutup	0.143

no	nasional		internasional		olahraga		sains		edukasi		ekonomi	
	kata	mean	kata	mean	kata	mean	kata	mean	kata	mean	kata	mean
77	<b>ambil</b>	0.143	<b>terima</b>	0.129	<b>detik</b>	0.159	kilometer	0.113	daftar	0.118	subsidi	0.143
78	<b>tingkat</b>	0.129	<b>Jumat</b>	0.129	<b>salah</b>	0.145	<b>bukti</b>	0.113	jenderal	0.118	global	0.143
79	<b>tahan</b>	0.129	<b>bangun</b>	0.129	<b>selasa</b>	0.145	<b>dekat</b>	0.113	kampus	0.118	krisis	0.143
80	susilo	0.129	<b>dekat</b>	0.129	<b>utama</b>	0.145	terbang	0.113	<b>com</b>	0.118	jangka	0.143
81	surat	0.129	tindakan	0.129	<b>jepang</b>	0.145	<b>mati</b>	0.113	<b>presiden</b>	0.118	investor	0.143
82	suara	0.129	te	0.129	<b>mobil</b>	0.145	<b>laut</b>	0.113	pelaksanaan	0.118	<b>bangun</b>	0.129
83	<b>pusat</b>	0.129	jadwal	0.129	juang	0.145	<b>gambar</b>	0.113	bahasa	0.118	<b>sesuai</b>	0.129
84	pasal	0.129	al	0.129	tarung	0.145	<b>luas</b>	0.113	un	0.118	<b>wakil</b>	0.129
85	<b>nomor</b>	0.129	sipil	0.129	<b>usia</b>	0.145	<b>jenis</b>	0.113	lulus	0.118	<b>akibat</b>	0.129
86	<b>nasional</b>	0.129	regional	0.129	maju	0.145	listrik	0.113	<b>alam</b>	0.103	<b>daya</b>	0.129
87	<b>nama</b>	0.129	pesawat	0.129	<b>main</b>	0.145	rekam	0.113	<b>juta</b>	0.103	dapat	0.129
88	mahkamah	0.129	<b>perdana</b>	0.129	malaysia	0.145	<b>teknologi</b>	0.113	<b>akibat</b>	0.103	angka	0.129
89	<b>lapor</b>	0.129	perang	0.129	ganda	0.145	publikasi	0.113	<b>sumber</b>	0.103	<b>bahan</b>	0.129
90	<b>kamis</b>	0.129	duta	0.129	lolos	0.145	identifikasi	0.113	<b>jalan</b>	0.103	<b>asia</b>	0.129
91	gedung	0.129	damai	0.129	spanyol	0.145	cahaya	0.113	<b>usia</b>	0.103	<b>total</b>	0.129
92	<b>dukung</b>	0.129	<b>aksi</b>	0.129	perak	0.145	<b>jakarta</b>	0.099	<b>beda</b>	0.103	minta	0.129

no	nasional		internasional		olahraga		sains		edukasi		ekonomi	
	kata	mean	kata	mean	kata	mean	kata	mean	kata	mean	kata	mean
93	demokrat	0.129	israel	0.129	emas	0.145	defik	0.099	kembang	0.103	bi	0.129
94	cari	0.129	china	0.129	senin	0.130	utama	0.099	teknologi	0.103	infrastruktur	0.129
95	yudhoyono	0.114	hukum	0.114	rabu	0.130	jumat	0.099	catat	0.103	konsumsi	0.129
96	wartawan	0.114	harap	0.114	tambah	0.130	sulit	0.099	asal	0.103	bursa	0.129
97	upaya	0.114	milik	0.114	langsung	0.130	jawa	0.099	sabtu	0.103	rabu	0.114
98	tetap	0.114	nama	0.114	tinggal	0.130	timur	0.099	mampu	0.103	sumber	0.114
99	terima	0.114	upaya	0.114	olahraga	0.130	butuh	0.099	malam	0.103	badan	0.114
100	sidang	0.114	anggap	0.114	mundur	0.130	muka	0.099	nilai	0.103	hitung	0.114

### Lampiran C - 2 Daftar Fitur Bagian 2

no	teknologi		entertainment		otomotif		health		properti		Travel	
	kata	mean	kata	mean	kata	mean	kata	mean	kata	mean	kata	Mean
1	sedia	0.471	jakarta	0.543	motor	0.551	sehat	0.571	rumah	0.671	wisata	0.729
2	guna	0.471	film	0.429	mobil	0.493	sakit	0.500	bangun	0.457	pariwisata	0.457
3	milik	0.429	lagu	0.357	indonesia	0.449	orang	0.400	ruang	0.386	indonesia	0.429
4	layan	0.329	musik	0.329	mesin	0.406	sebab	0.400	milik	0.371	daerah	0.414
5	sistem	0.329	anak	0.314	kendaraan	0.391	derita	0.357	indonesia	0.343	kunjung	0.400
6	nama	0.329	nama	0.300	model	0.348	dokter	0.329	harga	0.343	kota	0.386
7	perangkat	0.329	malam	0.300	jual	0.333	indonesia	0.314	jakarta	0.314	jalan	0.357
8	harga	0.314	orang	0.286	unit	0.333	tubuh	0.314	salah	0.286	budaya	0.343

no	teknologi		entertainment		otomotif		health		properti		Travel	
	kata	mean	kata	mean	kata	mean	kata	mean	kata	mean	kata	Mean
9	dunia	0.314	the	0.271	produk	0.319	makan	0.314	unit	0.286	milik	0.314
10	produk	0.314	temu	0.271	harga	0.304	salah	0.300	desain	0.286	orang	0.314
11	operasi	0.314	bintang	0.271	pt	0.304	tingkat	0.300	persen	0.271	pemerintah	0.314
12	indonesia	0.300	tampil	0.257	honda	0.304	teliti	0.300	rp	0.271	jakarta	0.300
13	internet	0.300	mengaku	0.257	juta	0.290	alami	0.286	huni	0.271	negara	0.300
14	pasar	0.286	salah	0.243	salah	0.275	hidup	0.286	jual	0.257	salah	0.286
15	hadir	0.286	tonton	0.243	jakarta	0.261	jalan	0.271	beli	0.243	masyarakat	0.286
16	tarik	0.271	cerita	0.243	baru	0.261	milik	0.271	luas	0.243	tarik	0.286
17	usaha	0.257	album	0.243	pasar	0.261	cegah	0.271	bunga	0.243	bangun	0.271
18	jual	0.257	indonesia	0.229	rp	0.261	dr	0.271	pengembang	0.229	ujung	0.271
19	salah	0.257	jalan	0.229	tingkat	0.261	jakarta	0.257	hasil	0.214	kawasan	0.257
20	teknologi	0.257	selatan	0.229	tampil	0.232	persen	0.257	fungsi	0.214	air	0.257
21	akses	0.257	hasil	0.214	pakai	0.232	kurang	0.257	capai	0.214	indah	0.257
22	google	0.257	baru	0.214	lengkap	0.232	akibat	0.257	tinggal	0.214	ekonomi	0.257
23	layar	0.257	main	0.214	tambah	0.232	obat	0.257	buka	0.214	potensi	0.257
24	fitur	0.257	tinggal	0.214	konsumen	0.232	hasil	0.243	warna	0.214	kabupaten	0.257
25	orang	0.243	penyanyi	0.214	bahan	0.232	temu	0.243	tingkat	0.200	rabu	0.243
26	dukung	0.229	as	0.200	hasil	0.217	jantung	0.243	guna	0.200	dunia	0.243
27	langsung	0.229	peran	0.200	bentuk	0.217	serang	0.243	pasar	0.200	tuju	0.243
28	ponsel	0.229	cinta	0.200	luncur	0.217	butuh	0.229	juta	0.200	kepala	0.243
29	lengkap	0.229	artis	0.200	jalan	0.203	kondisi	0.229	lahan	0.200	tingkat	0.229
30	aplikasi	0.229	buat	0.186	sesuai	0.203	jenis	0.229	orang	0.186	makan	0.229

no	teknologi		entertainment		otomotif		health		properti		Travel	
	kata	mean	kata	mean	kata	mean	kata	mean	kata	mean	kata	Mean
31	hasil	0.214	acara	0.186	beda	0.203	ahli	0.229	masyarakat	0.186	promosi	0.229
32	tambah	0.214	vokalis	0.186	khusus	0.203	risiko	0.229	tarik	0.186	libur	0.229
33	baru	0.214	dunia	0.171	direktur	0.203	turun	0.214	investasi	0.186	harap	0.214
34	luncur	0.214	harap	0.171	milik	0.188	usia	0.214	bentuk	0.171	provinsi	0.214
35	komputer	0.214	jumat	0.171	kali	0.188	bantu	0.200	perhati	0.171	obyek	0.214
36	juta	0.200	kawasan	0.171	ganti	0.188	ganggu	0.200	beda	0.171	masuk	0.200
37	tawar	0.200	usia	0.171	desain	0.188	ubah	0.186	negara	0.171	giat	0.200
38	saing	0.200	band	0.171	tenaga	0.188	konsumsi	0.186	tempat	0.171	jawa	0.200
39	fungsi	0.200	sutradara	0.171	pilih	0.174	berat	0.186	fasilitas	0.171	pengembangan	0.200
40	pakai	0.200	milik	0.157	produksi	0.174	andung	0.186	dinding	0.171	taman	0.186
41	buat	0.200	layar	0.157	langsung	0.174	badan	0.186	konsep	0.171	pusat	0.186
42	rabu	0.186	kenal	0.157	ubah	0.174	darah	0.186	bank	0.171	kembang	0.186
43	cepat	0.186	masuk	0.157	produsen	0.174	cepat	0.171	hidup	0.157	industri	0.186
44	mudah	0.186	garap	0.157	jepang	0.174	rumah	0.171	turun	0.157	laut	0.186
45	web	0.186	judul	0.157	standar	0.174	baik	0.171	bahan	0.157	buka	0.171
46	windows	0.186	senang	0.157	otomotif	0.174	anak	0.171	tambah	0.157	juta	0.171
47	software	0.186	pop	0.157	dunia	0.159	faktor	0.171	lengkap	0.157	tambah	0.171
48	microsoft	0.186	kali	0.143	butuh	0.159	fisik	0.171	pilih	0.157	usaha	0.171
49	bisnis	0.171	terima	0.143	presiden	0.159	buruk	0.171	usaha	0.157	daya	0.171
50	kembang	0.171	senin	0.143	banding	0.159	dasar	0.157	pt	0.157	nasional	0.171
51	buka	0.171	inggris	0.143	amerika	0.159	hubung	0.157	penuh	0.157	negeri	0.171
52	jaring	0.171	bawa	0.143	komponen	0.159	masyarakat	0.157	batas	0.157	malam	0.171

no	teknologi		entertainment		otomotif		health		properti		Travel	
	kata	mean	kata	mean	kata	mean	kata	mean	kata	mean	kata	Mean
53	basis	0.171	sang	0.143	varian	0.159	jaga	0.157	<b>tampil</b>	0.157	pulau	0.171
54	situs	0.171	hibur	0.143	bakar	0.159	<b>mati</b>	0.157	<b>miliar</b>	0.157	pantai	0.171
55	apple	0.171	perempuan	0.143	bensin	0.159	timbul	0.157	<b>temu</b>	0.143	<b>dinas</b>	0.171
56	<b>jakarta</b>	0.157	banget	0.143	<b>hadir</b>	0.145	<b>dampak</b>	0.157	<b>butuh</b>	0.143	<b>luas</b>	0.157
57	<b>harap</b>	0.157	akting	0.143	<b>sedia</b>	0.145	urang	0.143	<b>dasar</b>	0.143	<b>temu</b>	0.157
58	<b>as</b>	0.157	<b>harga</b>	0.129	<b>dasar</b>	0.145	<b>atur</b>	0.143	<b>mudah</b>	0.143	<b>lokasi</b>	0.157
59	<b>butuh</b>	0.157	<b>usaha</b>	0.129	<b>tahan</b>	0.145	<b>rendah</b>	0.143	<b>utama</b>	0.143	<b>anak</b>	0.157
60	<b>dasar</b>	0.157	<b>pilih</b>	0.129	<b>persen</b>	0.145	tekan	0.143	gaya	0.143	barat	0.157
61	<b>jalan</b>	0.157	<b>rilis</b>	0.129	<b>target</b>	0.145	of	0.143	<b>satu</b>	0.143	<b>warga</b>	0.157
62	<b>beda</b>	0.157	teman	0.129	<b>capai</b>	0.145	efek	0.143	<b>masuk</b>	0.143	<b>kamis</b>	0.157
63	<b>data</b>	0.157	<b>berat</b>	0.129	gerak	0.145	<b>manusia</b>	0.143	<b>direktur</b>	0.143	<b>langsung</b>	0.157
64	<b>jadi</b>	0.157	<b>produksi</b>	0.129	<b>sistem</b>	0.145	<b>studi</b>	0.143	<b>tawar</b>	0.143	<b>timur</b>	0.157
65	<b>besar</b>	0.157	<b>minggu</b>	0.129	lampu	0.145	biasa	0.143	<b>pemerintah</b>	0.143	singapura	0.157
66	<b>tampil</b>	0.157	<b>kamis</b>	0.129	<b>cepat</b>	0.145	<b>universitas</b>	0.143	<b>sedia</b>	0.143	<b>nilai</b>	0.157
67	operator	0.157	percaya	0.129	nyaman	0.145	<b>olahraga</b>	0.143	<b>menteri</b>	0.143	<b>kenal</b>	0.157
68	<b>bentuk</b>	0.157	<b>hati</b>	0.129	astra	0.145	waspada	0.143	kayu	0.143	<b>bidang</b>	0.157
69	<b>pilih</b>	0.157	<b>hidup</b>	0.129	roda	0.145	gula	0.143	letak	0.143	<b>gelar</b>	0.157
70	smartphone	0.157	<b>aksi</b>	0.129	nissan	0.145	<b>baru</b>	0.129	<b>kawasan</b>	0.143	<b>rp</b>	0.143
71	platform	0.157	wawancara	0.129	<b>masuk</b>	0.130	<b>bahan</b>	0.129	<b>taman</b>	0.143	<b>jual</b>	0.143
72	<b>persen</b>	0.143	lahir	0.129	<b>muncul</b>	0.130	<b>kali</b>	0.129	kamar	0.143	<b>pasar</b>	0.143
73	<b>rp</b>	0.143	kasih	0.129	<b>besar</b>	0.130	kaki	0.129	<b>alami</b>	0.129	<b>pilih</b>	0.143
74	<b>negara</b>	0.143	personel	0.129	<b>konsep</b>	0.130	papar	0.129	<b>jenis</b>	0.129	<b>tawar</b>	0.143

no	tekno		entertainment		otomotif		health		properti		Travel	
	kata	mean	kata	mean	kata	mean	kata	mean	kata	mean	kata	Mean
75	<b>kali</b>	0.143	vokal	0.129	<b>beli</b>	0.130	pria	0.129	<b>khusus</b>	0.129	duduk	0.143
76	<b>kompas</b>	0.143	pentas	0.129	<b>industri</b>	0.130	<b>kait</b>	0.129	<b>pusat</b>	0.129	<b>kerja</b>	0.143
77	<b>terima</b>	0.143	gitar	0.129	pasang	0.130	rusak	0.129	<b>ukur</b>	0.129	<b>alam</b>	0.143
78	<b>desain</b>	0.143	aja	0.129	andal	0.130	<b>tangan</b>	0.129	and	0.129	<b>wilayah</b>	0.143
79	<b>aman</b>	0.143	<b>langsung</b>	0.114	show	0.130	<b>fungsi</b>	0.129	<b>tumbuh</b>	0.129	menikmati	0.143
80	online	0.143	<b>gambar</b>	0.114	<b>ukur</b>	0.130	<b>kembang</b>	0.129	<b>besar</b>	0.129	<b>program</b>	0.143
81	<b>kenal</b>	0.143	<b>sesuai</b>	0.114	<b>merek</b>	0.130	wanita	0.129	sederhana	0.129	<b>Jumat</b>	0.143
82	komunikasi	0.143	<b>ikut</b>	0.114	<b>kualitas</b>	0.130	buah	0.129	<b>kota</b>	0.129	<b>rumah</b>	0.129
83	<b>seri</b>	0.143	<b>air</b>	0.114	samping	0.130	minum	0.129	sisi	0.129	<b>harga</b>	0.129
84	<b>rilis</b>	0.143	<b>selasa</b>	0.114	sport	0.130	<b>riset</b>	0.129	untung	0.129	<b>capai</b>	0.129
85	<b>informasi</b>	0.143	bilang	0.114	cc	0.130	bahaya	0.129	murah	0.129	<b>tinggal</b>	0.129
86	video	0.143	karya	0.114	motors	0.130	aktivitas	0.129	udara	0.129	<b>bentuk</b>	0.129
87	media	0.143	<b>perdana</b>	0.114	dealer	0.130	departemen	0.129	<b>lokasi</b>	0.129	<b>penuh</b>	0.129
88	prosesor	0.143	dengar	0.114	atpm	0.130	otak	0.129	kesan	0.129	<b>menteri</b>	0.129
89	mobile	0.143	suka	0.114	<b>tinggal</b>	0.116	<b>banding</b>	0.114	suku	0.129	<b>kondisi</b>	0.129
90	<b>merek</b>	0.143	<b>gelar</b>	0.114	<b>usaha</b>	0.116	<b>sulit</b>	0.114	area	0.129	<b>banding</b>	0.129
91	android	0.143	kisah	0.114	<b>hati</b>	0.116	<b>mudah</b>	0.114	kredit	0.129	<b>ketua</b>	0.129
92	<b>pt</b>	0.129	<b>jadi</b>	0.100	<b>ikut</b>	0.116	<b>guna</b>	0.114	<b>sehat</b>	0.114	<b>target</b>	0.129
93	<b>target</b>	0.129	<b>kompas</b>	0.100	<b>warna</b>	0.116	positif	0.114	<b>jalan</b>	0.114	<b>dukung</b>	0.129
94	<b>kerja</b>	0.129	<b>penuh</b>	0.100	new	0.116	<b>data</b>	0.114	<b>anak</b>	0.114	<b>acara</b>	0.129
95	<b>ubah</b>	0.129	<b>com</b>	0.100	<b>alami</b>	0.116	sadar	0.114	<b>rendah</b>	0.114	<b>badan</b>	0.129
96	<b>mampu</b>	0.129	<b>muncul</b>	0.100	versi	0.116	hindar	0.114	<b>kait</b>	0.114	yogyakarta	0.129

no	tekno		entertainment		otomotif		health		properti		Travel	
	kata	mean	kata	mean	kata	mean	kata	mean	kata	mean	kata	Mean
97	<b>produsen</b>	0.129	isi	0.100	kemarin	0.116	<b>lapor</b>	0.114	<b>sulit</b>	0.114	<b>selenggara</b>	0.129
98	<b>cari</b>	0.129	<b>tiga</b>	0.100	<b>umum</b>	0.116	sel	0.114	<b>kualitas</b>	0.114	paket	0.129
99	jam	0.129	cipta	0.100	<b>negara</b>	0.116	virus	0.114	<b>air</b>	0.114	<b>hasil</b>	0.114
100	mengembangkan	0.129	<b>coba</b>	0.100	<b>teknologi</b>	0.116	<b>tambah</b>	0.100	<b>coba</b>	0.114	<b>utama</b>	0.114

## LAMPIRAN D

Lampiran ini berisi tabel nilai prediksi dari masing-masing artikel berita yang digunakan sebagai data uji. Nilai prediksi ini berdasarkan hasil perhitungan seluruh fitur dalam satu dokumen berdasarkan model klasifikasi yang dibuat sebelumnya. Penentuan kelas ditentukan dari nilai prediksi paling tinggi. Tanda \* menunjukkan nilai prediksi tertinggi.

**Lampiran D - 1**

No	Judul Artikel Uji	Kelas Asli	Kelas Prediksi												
			nasional	internasi onal	olahraga	sains	edukasi	ekonomi	tekno	<i>entertai nment</i>	otomotif	<i>health</i>	properti	travel	
1	Mendagri Akan Konfirmasi Bupati Singkil	nasional	*1	0	0	0	0	0	0	0	0	0	0	0	0
2	Perampasan Aset Century di Hongkong Butuh Proses	nasional	*1	0	0	0	0	0	0	0	0	0	0	0	
3	Kontras: Priyo Tak Mengerti HAM	nasional	*1	0	0	0	0	0	0	0	0	0	0	0	
4	Kemenag akan Umumkan Pelaku Korupsi Alquran	nasional	*1	0	0	0	0	0	0	0	0	0	0	0	
5	Polri Tanyakan Barang Bukti kepada Pengacara Sukotjo	nasional	*1	0	0	0	0	0	0	0	0	0	0	0	
6	Polri Tunggu Realisasi Pemanggilan Djoko Susilo	nasional	*1	0	0	0	0	0	0	0	0	0	0	0	
7	Tiga Partai Keluarga Cendana Gugur di Tangan KPU	nasional	*1	0	0	0	0	0	0	0	0	0	0	0	
8	Dipo Laporkan Tiga Kementerian, Ada Persaingan	nasional	*1	0	0	0	0	0	0	0	0	0	0	0	

No	Judul Artikel Uji	Kelas Asli	Kelas Prediksi											
			nasional	internasi onal	olahraga	sains	edukasi	ekonomi	tekno	<i>entertai nment</i>	otomotif	<i>health</i>	properti	travel
9	KPK Datangi Korlantas Terkait Barang Bukti Djoko ...	nasional	*1	0	0	0	0	0	0	0	0	0	0	0
10	PB HMI: TNI Telah Menghina Kebebasan Pers	nasional	*1	0	0	0	0	0	0	0	0	0	0	0
11	Mentan Siap Diperiksa KPK Terkait Skandal Impor Daging	nasional	*1	0	0	0	0	0	0	0	0	0	0	0
12	Gerindra: Pilkada Serentak Efisienkan Anggaran	nasional	*1	0	0	0	0	0	0	0	0	0	0	0
13	Demokrat Tantang Anas Laporkan Nama Baru Kasus Century ke KPK	nasional	*1	0	0	0	0	0	0	0	0	0	0	0
14	Mendagri Perlu Lebih Proaktif	nasional	*1	0	0	0	0	0	0	0	0	0	0	0
15	Dimilai Kutu Loncat, Bacaleg Ganda Tak Perlu Diproses ...	nasional	*1	0	0	0	0	0	0	0	0	0	0	0
16	Ada Rekaman Fathanah-Luthfi soal 40 Ribu Dollar AS	nasional	*1	0	0	0	0	0	0	0	0	0	0	0
17	Priyo: Tak Masuk Akal kalau Saya Cuma Dapat 1 Persen	nasional	*1	0	0	0	0	0	0	0	0	0	0	0
18	Presiden: Atasi Kebakaran di Riau	nasional	0.155	0	0	0	0	*0.843	0	0	0	0	0.002	0
19	Persinyalan KRL Buruk, Dahlan Iskan Peringatkan Dirut PT LEN	nasional	0	0	0	0	0	*1	0	0	0	0	0	0
20	Berduka, Megawati Tak Gelar "Open House" Lebaran	nasional	0.115	0.001	0	0	0.006	0	0	*0.877	0	0	0	0

No	Judul Artikel Uji	Kelas Asli	Kelas Prediksi											
			nasional	internasional	olahraga	sains	edukasi	ekonomi	tekno	entertainment	otomotif	health	properti	travel
21	Tampar Legimo, Djoko Tegaskan Tak Terkait Uang	nasional	*1	0	0	0	0	0	0	0	0	0	0	0
22	Kak Seto: Tes Keperawanan Tak Boleh Dilakukan	nasional	0	0	0	0	*1	0	0	0	0	0	0	0
23	ICW Apresiasi Sistem Baru Seleksi CPNS	nasional	0.005	0	0	0	*0.995	0	0	0	0	0	0	0
24	KPK Akan Jemput Paksa Haris Andi jika Mangkir Lagi	nasional	*1	0	0	0	0	0	0	0	0	0	0	0
25	Telat Lomba Lari, Agus Yudhoyono Akhirnya Minta Maaf	nasional	*0.858	0	0	0	0.127	0	0	0.015	0	0	0	0
26	Terpilih Jadi Kapolri, Sutarman Mau Buat Densus Anti-Korupsi	nasional	*1	0	0	0	0	0	0	0	0	0	0	0
27	Agung Laksono: Sahabat Gamawan Bukan Hanya FPL...	nasional	*1	0	0	0	0	0	0	0	0	0	0	0
28	Istri Anas Urbaningrum Siap Diperiksa KPK	nasional	*1	0	0	0	0	0	0	0	0	0	0	0
29	Adrianus: Waktu Dirawat, Tubuh Mulyana Kurus Sekali	nasional	*1	0	0	0	0	0	0	0	0	0	0	0
30	Ratu Atut, Contohnya Andi Mallarangeng...	nasional	*1	0	0	0	0	0	0	0	0	0	0	0
31	Jumlah Penduduk Eropa Tembus 500 Juta	internasional	0	*0.994	0	0.006	0	0	0	0	0	0	0	0
32	Australia Tunggu Ratu Elizabeth II Lengser	internasional	0	0.002	0	0	0	0	0	*0.998	0	0	0	0



No	Judul Artikel Uji	Kelas Asli	Kelas Prediksi											
			nasional	internasional	olahraga	sains	edukasi	ekonomi	tekno	entertainment	otomotif	health	properti	travel
46	Aksi Sabotase Putuskan Pipa Gas Sinai	internasional	0	*1	0	0	0	0	0	0	0	0	0	0
47	Pasukan Keamanan Afganistan Jadi Andalan	internasional	0	*1	0	0	0	0	0	0	0	0	0	0
48	Peringatan Hari Al Quds Dukung Palestina	internasional	0	*1	0	0	0	0	0	0	0	0	0	0
49	NATO Lanjutkan Operasi di Libya	internasional	0.002	*0.998	0	0	0	0	0	0	0	0	0	0
50	Bom Mobil Tewaskan Setidaknya 55 Orang	internasional	0	*1	0	0	0	0	0	0	0	0	0	0
51	Utusan PBB: Hati-hati Libatkan Diri pada Libya	internasional	0	*1	0	0	0	0	0	0	0	0	0	0
52	Obama dan Kepala Negara ASEAN Bertemu	internasional	0	*1	0	0	0	0	0	0	0	0	0	0
53	Ledakan Bom Tewaskan Warga Sipil Termasuk Wanita dan Anak	internasional	0	*1	0	0	0	0	0	0	0	0	0	0
54	Presiden Argentina Siap Jalani Operasi Kanker	internasional	0	*1	0	0	0	0	0	0	0	0	0	0
55	AS Cuma Merotasi Pasukan di Asia Pasifik	internasional	0	*1	0	0	0	0	0	0	0	0	0	0
56	Inggris Perancis Perkuat Kerja Sama Industri Militer	internasional	0	*0.998	0	0	0	0.002	0	0	0	0	0	0
57	Perempuan Aceh Berbicara di DK PBB	internasional	0	*1	0	0	0	0	0	0	0	0	0	0
58	Bom Meledak Saat PM Somalia Berpidato	internasional	0	*1	0	0	0	0	0	0	0	0	0	0



No	Judul Artikel Uji	Kelas Asli	Kelas Prediksi											
			nasional	internasi onal	olahraga	sains	edukasi	ekonomi	tekno	entertai nment	otomotif	health	properti	travel
72	Button: MP4-28 Mobil Terbaik yang Pernah Dibikin McLaren	olahraga	0	0	*1	0	0	0	0	0	0	0	0	0
73	Sananta Sempat Kewalahan Hadapi Juara Bertahan	olahraga	0	0	*1	0	0	0	0	0	0	0	0	0
74	Tontowi/Liliyana Tak Mau Meremehkan Siapa Saja Lawannya	olahraga	0	0	*1	0	0	0	0	0	0	0	0	0
75	Ferrari Tak Boleh Kehilangan Poin Lagi di China	olahraga	0	0.002	0.001	*0.913	0.01	0	0.07	0.004	0	0	0	0
76	Marquez Tak Mau Hanyut dalam Euforia	olahraga	0	0	*1	0	0	0	0	0	0	0	0	0
77	Halep Tantang Serena di Semifinal	olahraga	0	0	*1	0	0	0	0	0	0	0	0	0
78	Sharapova Tembus Babak Keempat French Open	olahraga	0	0	*1	0	0	0	0	0	0	0	0	0
79	Pedrosa, Marquez, dan Bautista Berbagi Ilmu dengan Para Remaja	olahraga	0	0	*1	0	0	0	0	0	0	0	0	0
80	Perempat Final Pertama Del Potro di Wimbledon	olahraga	0	0	*1	0	0	0	0	0	0	0	0	0
81	Daud Yordan Konsentrasi Pulihkan Cedera	olahraga	0	0	*1	0	0	0	0	0	0	0	0	0
82	Julio Bria Ingin Pertahankan Emas SEA Games	olahraga	0	0	*1	0	0	0	0	0	0	0	0	0
83	Indonesia Tambah Satu Medali Perak	olahraga	0	0	*1	0	0	0	0	0	0	0	0	0



No	Judul Artikel Uji	Kelas Asli	Kelas Prediksi											
			nasional	internasional	olahraga	sains	edukasi	ekonomi	teknologi	entertainment	otomotif	health	properti	travel
97	Inilah Planet Pertama di Luar Bima Sakti	sains	0	0	0	*1	0	0	0	0	0	0	0	0
98	Melesat dengan 7 Kali Kecepatan Suara	sains	0	0	0	*1	0	0	0	0	0	0	0	0
99	Satelit SOHO Berhasil Temukan 2000 Komet	sains	0	0	0	*1	0	0	0	0	0	0	0	0
100	Sulitkah Menemukan Planet Baru?	sains	0	0	0	*1	0	0	0	0	0	0	0	0
101	Kelahiran Planet Akhirnya Berhasil Diabadikan	sains	0	0	0	*1	0	0	0	0	0	0	0	0
102	Gempa Jepang, Akibat Kedekatan Bumi-Bulan?	sains	0	0	0	*1	0	0	0	0	0	0	0	0
103	Fosil Berusia 1,5 Juta Tahun Ditemukan	sains	0	0	0	*1	0	0	0	0	0	0	0	0
104	"Styrofoam" Ini Ramah Lingkungan	sains	0	0	0	0.122	0	0	0	0	0	*0.878	0	0
105	Jejak Astronomis di Borobudur	sains	0	0	0	*1	0	0	0	0	0	0	0	0
106	Inilah Kelompok Supernova Paling Terang	sains	0	0	0	*1	0	0	0	0	0	0	0	0
107	Misi Berawak ke Antariksa Tetap Lanjut	sains	0	0	0	*1	0	0	0	0	0	0	0	0
108	Mikroba Bumi Lebih Ganas di Luar Angkasa	sains	0	0	0	0.001	0	0	0	0	0	*0.999	0	0
109	Bulan Ternyata 200 Juta Tahun Lebih Muda	sains	0	0	0	*1	0	0	0	0	0	0	0	0

No	Judul Artikel Uji	Kelas Asli	Kelas Prediksi											
			nasional	internasi onal	olahraga	sains	edukasi	ekonomi	tekno	entertai nment	otomotif	health	properti	travel
110	Jubah Tembus Pandang Ubah Tank Jadi Sapi	sains	0	0	0	*1	0	0	0	0	0	0	0	0
111	Waspadai Penyebab Tagihan Listrik Membengkak	sains	0	0	0	0.005	0	0	0	0	0	0.233	*0.762	0
112	Purnama Jupiter, Jakarta Bisa Melihat Jelas	sains	0	0	0	*1	0	0	0	0	0	0	0	0
113	Mendesak, Penelitian Bulus Raksasa Ciliwung	sains	0	0	0	0.093	*0.907	0	0	0	0	0	0	0
114	Tukirin dan Ironi Krakatau	sains	0	0	0	*1	0	0	0	0	0	0	0	0
115	Badai Matahari Melenyapkan Bulan?	sains	0	0	0	*1	0	0	0	0	0	0	0	0
116	Kue Cinta dari Kerabat Bunga Bangkai	sains	0	0	0	*1	0	0	0	0	0	0	0	0
117	Dulu, Matahari Berukuran Lebih Besar	sains	0	0	0	*1	0	0	0	0	0	0	0	0
118	Dinosaur Mini Sebesar Merpati	sains	0	0	0	*1	0	0	0	0	0	0	0	0
119	LIPI Bisa Memprotes Publikasi "Raja Tawon" ...	sains	0	0	0	*1	0	0	0	0	0	0	0	0
120	Mahasiswa Indonesia Raih Prestasi Matematika ...	edukasi	0	0	0	0	*1	0	0	0	0	0	0	0
121	1.670 Mahasiswa Unnes Terima Beasiswa Bidikmisi	edukasi	0	0	0	0	*1	0	0	0	0	0	0	0
122	Sekolah "Nakal" Diancam Sanksi	edukasi	0	0	0	0	*1	0	0	0	0	0	0	0

No	Judul Artikel Uji	Kelas Asli	Kelas Prediksi											
			nasional	internasional	olahraga	sains	edukasi	ekonomi	tekno	entertainment	otomotif	health	properti	travel
123	Kurikulum Baru Pangkas Jumlah Mata Pelajaran	edukasi	0	0	0	0	*1	0	0	0	0	0	0	0
124	Peraih Nobel Asal Australia Inspirasi Pelajar Surabaya	edukasi	0	0	0	0	*1	0	0	0	0	0	0	0
125	Kemitraan Pemerintah-Swasta Dukung Kemajuan Pendidikan	edukasi	0	0	0	0	*1	0	0	0	0	0	0	0
126	Politeknik Negeri Madura 2014 Akan Selesai	edukasi	0	0	0	0	*1	0	0	0	0	0	0	0
127	Uji Publik Kurikulum 2013 Digelar Tiga Hari	edukasi	0	0	0	0	*1	0	0	0	0	0	0	0
128	Kurikulum 2013 Bagus untuk Asah Kreativitas Anak	edukasi	0	0	0	0	*1	0	0	0	0	0	0	0
129	Siswa Tanpa NPSN Tak Bisa Ikuti SNMPTN	edukasi	0	0	0	0	*1	0	0	0	0	0	0	0
130	Kurikulum 2013 Belum Diperlukan?	edukasi	0	0	0	0	*1	0	0	0	0	0	0	0
131	Sapu Angin UNY Juara Lomba Desain Kincir	edukasi	0	0	0	*0.882	0.118	0	0	0	0	0	0	0
132	Sekolah Swasta Sambut Baik Pembubaran RSBI	edukasi	0	0	0	0	*1	0	0	0	0	0	0	0
133	1.651 Sekolah Belum Isi PDSS	edukasi	0	0	0	0	*1	0	0	0	0	0	0	0
134	Mahasiswa Lajang Diutamakan Dapat Beasiswa...	edukasi	0	0	0	0	*1	0	0	0	0	0	0	0
135	Cegah Pungutan, Dewan Pendidikan Awasi Eks RSBI	edukasi	0	0	0	0	*1	0	0	0	0	0	0	0
136	Kemendikbud Segera Laksanakan Gerakan ...	edukasi	0	0	0	0	*1	0	0	0	0	0	0	0

No	Judul Artikel Uji	Kelas Asli	Kelas Prediksi											
			nasional	internasi onal	olahraga	sains	edukasi	ekonomi	tekno	<i>entertai nment</i>	otomotif	<i>health</i>	properti	travel
137	8.910 Siswa di Mukomuko Siap Ujian Nasional	edukasi	0	0	0	0	*1	0	0	0	0	0	0	0
138	Sekolah/Madrasah Aman Terus Dikembangkan	edukasi	0	0	0	0	*1	0	0	0	0	0	0	0
139	Tak Aktif Dua Tahun, Mahasiswi Ini Ikut Ujian ...	edukasi	0	0	0	0	*1	0	0	0	0	0	0	0
140	Aktivitas Perkuliahan STAN Sumbawa Dimulai April 2013	edukasi	0	0	0	0	*1	0	0	0	0	0	0	0
141	Unsyiah Siap Menampung Mahasiswa Tidak Mampu	edukasi	0	0	0	0	*1	0	0	0	0	0	0	0
142	Citilink Creative Academy Dukung Kreativitas Anak ...	edukasi	0	0	0	0	0	0	0	0	0	0	0	*1
143	Jaktim Targetkan SMA dan SMK Lulus 99,9 Persen	edukasi	0	0	0	0	*1	0	0	0	0	0	0	0
144	Sudah Cair, Anggaran Bidik Misi Segera Disalurkan	edukasi	0	0	0	0	*1	0	0	0	0	0	0	0
145	Kemendikbud: Kami Sangat Terbuka pada KPK	edukasi	*0.701	0	0	0	0,299	0	0	0	0	0	0	0
146	50 Peserta UN SMP di Semarang Absen	edukasi	0	0	0	0	*1	0	0	0	0	0	0	0
147	Pendaftaran SBMPTN 2013 Dibuka 13 Mei	edukasi	0	0	0	0	*1	0	0	0	0	0	0	0
148	Universitas Paramadina Gelar Seminar Pendidikan	edukasi	0	0	0	0	*1	0	0	0	0	0	0	0
149	Penundaan Kenaikan Harga Gas Rugikan Negara	ekonomi	0	0	0	0	0	*1	0	0	0	0	0	0

No	Judul Artikel Uji	Kelas Asli	Kelas Prediksi											
			nasional	internasional	olahraga	sains	edukasi	ekonomi	tekno	entertainment	otomotif	health	properti	travel
150	Bebas Bea Impor Kedelai Hanya Untungkan Importir	ekonomi	0	0	0	0	0	*1	0	0	0	0	0	0
151	2013, Penerimaan Negara Bukan Pajak Turun Tipis	ekonomi	0	0	0	0	0	*1	0	0	0	0	0	0
152	Bank Mandiri Dukung Forum Investasi	ekonomi	0	0	0	0	0	*1	0	0	0	0	0	0
153	Ini Dia Somasi Nasabah Gadaai Emas ke BRI Syariah	ekonomi	0	0	0	0	0	*1	0	0	0	0	0	0
154	Kenaikan Impor China Kerek Harga Kedelai	ekonomi	0	0	0	0	0	*1	0	0	0	0	0	0
155	Bursa Berjangka Indonesia Harus Jadi Acuan	ekonomi	0	0	0	0	0	*1	0	0	0	0	0	0
156	Perbanas: Pangkas Subsidi BBM, Alihkan Untuk ...	ekonomi	0	0	0	0	0	*1	0	0	0	0	0	0
157	Izin Maskapai Kingfisher Airlines Dicabut	ekonomi	0	*1	0	0	0	0	0	0	0	0	0	0
158	Ganti Rugi Tiket Penumpang Batavia Air Menggantung	ekonomi	*0,999	0	0	0	0,001	0	0	0	0	0	0	0
159	Lima Berita Bisnis Keuangan Terpopuler Pekan Ini	ekonomi	0	0	0	0	0	*1	0	0	0	0	0	0
160	Gabah Dengan Kadar Air Tinggi Kurang Diminati ...	ekonomi	0	0	0	0,348	0	0,001	0	0	0	*0,651	0	0
161	AirAsia Tawarkan 2 Juta Kursi Promo	ekonomi	0	0	0	0	0	0	0	0	0	0	0	*1
162	Taufiq Kiemas: Kompensasi BLT Tak Fair	ekonomi	*1	0	0	0	0	0	0	0	0	0	0	0

No	Judul Artikel Uji	Kelas Asli	Kelas Prediksi												
			nasional	internasi onal	olahraga	sains	edukasi	ekonomi	teknologi	entertai nment	otomotif	health	properti	travel	
163	Bank Indonesia Borong Emas 933 Kg	ekonomi	0	0	0	0	0	0	*1	0	0	0	0	0	0
164	Menkeu: BBM Naik, Rupiah Bakal Menguat	ekonomi	0	0	0	0	0	0	*1	0	0	0	0	0	0
165	Data China dan The Fed Menekan Wall Street di Sesi ..	ekonomi	0	0	0	0	0	0	*1	0	0	0	0	0	0
166	Sentimen Positif dari Asia Angkat Wall Street	ekonomi	0	0	0	0	0	0	*1	0	0	0	0	0	0
167	Kredit Bank Mandiri Tumbuh 22,3 persen	ekonomi	0	0	0	0	0	0	*1	0	0	0	0	0	0
168	Data Ekonomi Bagus, Wall Street Langsung Melaju	ekonomi	0	0	0	0	0	0	*1	0	0	0	0	0	0
169	Rupiah Susah Kembali ke Bawah Level 10.000	ekonomi	0	0	0	0	0	0	*1	0	0	0	0	0	0
170	"Current Account" Negatif Tidak Selalu Tabu	ekonomi	0	0	0	0	0	0	*1	0	0	0	0	0	0
171	Dibayangi Keraguan atas The Fed, Wall Street Dibuka ...	ekonomi	0	0	0	0	0	0	*1	0	0	0	0	0	0
172	Hatta: 100.000 Lahan Baru Akan Ditanam Kedelai	ekonomi	0	0	0	0	0	0	*1	0	0	0	0	0	0
173	Kinerja Emiten di Kuartal III Gelincirkan Wall Street	ekonomi	0	0	0	0	0	0	*1	0	0	0	0	0	0
174	Data China Beri "Tenaga" Bagi Wall Street	ekonomi	0	0	0	0	0	0	*1	0	0	0	0	0	0
175	Emirates NBD Resmikan Kantor Perwakilan di Jakarta	ekonomi	0	0	0	0	0	0	*0.997	0	0	0	0	0.003	0

No	Judul Artikel Uji	Kelas Asli	Kelas Prediksi											
			nasional	internasi onal	olahraga	sains	edukasi	ekonomi	tekno	entertai nment	otomotif	health	properti	travel
176	DPR Tak Restui Penambahan BBM Subsidi akibat Mobil ...	ekonomi	0	0	0	0	0	0	*1	0	0	0	0	0
177	Bank Mutiara Bisa Dibanderol Rp 8,2 Triliun	ekonomi	0	0	0	0	0	0	*1	0	0	0	0	0
178	Slap-siap, Pajak Reksa Dana Mulai Naik pada 2014	ekonomi	0	0	0	0	0	0	*1	0	0	0	0	0
179	Bulan Depan, Sharp Mulai Kirim Layar iPhone 5	tekno	0	0	0	0	0	0	0	*1	0	0	0	0
180	Samsung Siapkan Cara Hindari Pencekalan Apple	tekno	0	0	0	0	0	0	0	*1	0	0	0	0
181	November, Tablet Windows 8 Samsung Masuk Indonesia	tekno	0	0	0	0	0	0	0	*1	0	0	0	0
182	Telkomsel Obral Ponsel Pintar dan Tablet di Indocomtech	tekno	0	0	0	0	0	0	0	*1	0	0	0	0
183	Karyawan Facebook Diminta Lirik Android	tekno	0	0	0	0	0	0	0	*1	0	0	0	0
184	ZTE Hadirkan Ponsel Jelly Bean Murah di Indonesia	tekno	0	0	0	0	0	0	0	*1	0	0	0	0
185	iPhone Murah Dibanderol 1,5 Juta?	tekno	0	0	0	0	0	0	0	*1	0	0	0	0
186	Peluncuran PlayStation 4 Tinggal Menghitung Hari?	tekno	0	0	0	0	0	0	0	*1	0	0	0	0
187	Dibanderol Rp 190.000, Nokia 105 Bisa "Melek" 1 Bulan	tekno	0	0	0	0	0	0	0	*1	0	0	0	0
188	Jelajahi Puncak-puncak Tertinggi Dunia dengan ...	tekno	0	0	0	*0.952	0	0	0	0	0	0	0.048	0

No	Judul Artikel Uji	Kelas Asli	Kelas Prediksi												
			nasional	internasi onal	olahraga	sains	edukasi	ekonomi	tekno	entertai nment	otomotif	health	properti	travel	
189	Konsumen AS Tak Sadar Ada BlackBerry 10	tekno	0	0	0	0	0	0	0	0	*1	0	0	0	0
190	Bagaimana Rasanya Kerja di Facebook?	tekno	0	0	0	0	0	0	0	0	*1	0	0	0	0
191	Galaxy S4 Dibongkar, Profit Samsung Terungkap	tekno	0	0	0	0	0	0	0	0	*1	0	0	0	0
192	Remaja 18 Tahun Ciptakan Baterai "30 Detik"	tekno	0	0	0	*1	0	0	0	0	0	0	0	0	0
193	Demi Apple WWDC, Rela Antre Tengah Malam	tekno	0	0	0	0	*0.988	0	0	0	0	0	0	0	0.012
194	10 Aplikasi Android untuk Teman Berpuasa	tekno	0	0	0	0	0.005	0	0	*0.994	0	0	0	0	0
195	Internet di Indonesia Bakal Disurvei	tekno	0	0	0	0	*1	0	0	0	0	0	0	0	0
196	9 Skandal Seks Petinggi Perusahaan Teknologi	tekno	*1	0	0	0	0	0	0	0	0	0	0	0	0
197	Parade Foto iPhone 5S Warna Emas	tekno	0	0	0	0	0	0	0	0	*1	0	0	0	0
198	Inikah Spesifikasi Tablet Android Nvidia?	tekno	0	0	0	0	0	0	0	0	*1	0	0	0	0
199	Menanti Jawaban di Pesta Developer BlackBerry	tekno	0	0	0	0	0	0	0	0	*1	0	0	0	0
200	BIA 2013 Mendorong Lahirnya Inovator dan Inventor Generasi Muda	tekno	0	0	0	0	0	*1	0	0	0	0	0	0	0

No	Judul Artikel Uji	Kelas Asli	Kelas Prediksi												
			nasional	internasional	olahraga	sains	edukasi	ekonomi	tekno	entertainment	otomotif	health	properti	travel	
201	BBM Bisa Dipakai di iPad Tertentu	tekno	0	0	0	0	0	0	0	0	*1	0	0	0	0
202	Pekerja Pabrik Apple Terjerat Hutang	tekno	0	0	0	0	0	0	0	0	*1	0	0	0	0
203	Intel Mau Android Saingi Windows	tekno	0	0	0	0	0	0	0	0	*1	0	0	0	0
204	Frekuensi Rendah Disebut Lebih Baik untuk 4G LTE	tekno	0	0	0	0	0	0	0	0	*1	0	0	0	0
205	Nokia "Iklankan" Kelemahan iPad	tekno	0	0	0	0	0	0	0	0	*1	0	0	0	0
206	Ponsel Android Nokia Bakal Hadir Warna-warni	tekno	0	0	0	0	0	0	0	0	*1	0	0	0	0
207	Aplikasi Peta Nokia Hilang dari Toko Apple	tekno	0	0	0	0	0	0	0	0	*1	0	0	0	0
208	Google Disiapkan Tak Cuma Jadi Mesin Pencari	tekno	0	0	0	0	0	0	0	0	*1	0	0	0	0
209	"Eat, Pray, Love" Harus Puas di Posisi Kedua	entertainment	0	0	0	0	0	0	0	0	*1	0	0	0	0
210	Dicari, Pemain untuk "Bumi Manusia"	entertainment	0	0	0	0	0	0	0	0	*1	0	0	0	0
211	Jadi Produser, Anjasmara Diwanti-wanti Istri	entertainment	0	0	0	0	0	0	0	0	*1	0	0	0	0
212	Boyong Penghargaan, ST12 Persembahkan untuk Anak Pepeng	entertainment	0	0	0	0	0	0	0	0	*1	0	0	0	0
213	Piyu: PADI Libur, Bukan Bubar	entertainment	0	0	0	0	0	0	0	0	*1	0	0	0	0

No	Judul Artikel Uji	Kelas Asli	Kelas Prediksi												
			nasional	internasional	olahraga	sains	edukasi	ekonomi	tekno	entertainment	otomotif	health	properti	travel	
214	Si Kembar Olsen: Yang Penting Karya Kami	entertainment	0	0	0	0	0	0	0	0	0.322	0	0	*0.678	0
215	Indra Herlambang Keringat Dingin Baca Al Quran	entertainment	0	0	0	0	0	0	0	0	*1	0	0	0	0
216	Walau Sudah Pernah, Afgan Tetap Kaget	entertainment	0	0	0	0	0	0	0	0	*1	0	0	0	0
217	Rekor iTunes untuk Lady Gaga	entertainment	0	0	0	0	0	0	0	0	*1	0	0	0	0
218	Flu Tak Halangi Once Hibur Pekanbaru	entertainment	0	0	0	0	0	0	0	0	*1	0	0	0	0
219	Tak Kebagian Tiket, Fans Bruno Mars Gigit Jari	entertainment	0	0	0	0	0	0	0	0	*1	0	0	0	0
220	Yuk, Sikat Gigi Sambil Nikmati Lagu Bieber	entertainment	0	0	0	0	0	0	0	0	*0.874	0.126	0	0	0
221	Eva Herzigova Pamer Kaki Jenjang di Cannes	entertainment	0	0	0	0	0	0	0	0	*1	0	0	0	0
222	Konser Batal karena Cuaca, Taylor Swift Kecewa Berat	entertainment	0	0	0	0	0	0	0	0	*1	0	0	0	0
223	Christopher Chong: Mencipta Musik Layaknya Meracik Parfum	entertainment	0	0	0	0	0	0	0	0	*1	0	0	0	0
224	Tompri: Saya Selalu Terlambat	entertainment	0	0	0	0	0	0	0	0	*1	0	0	0	0
225	Paramore: Indonesia Merdeka!	entertainment	0	0	0	0	0	0	0	0	*1	0	0	0	0

No	Judul Artikel Uji	Kelas Asli	Kelas Prediksi												
			nasional	internasional	olahraga	sains	edukasi	ekonomi	tekno	entertainment	otomotif	health	properti	travel	
226	Sarah Jessica Parker, Perempuan Kesuksesan Oscar de la Renta	entertainment	0	0	0	0	0	0	0	0	*1	0	0	0	0
227	Konser Tiga Kota Maher Zain	entertainment	0	0	0	0	0	0	0	0	*1	0	0	0	0
228	Pelajar Wakatobi: Baru Kali Ini Pegang Kamera Ini	entertainment	0	0	0	0	0	0	0	0	*1	0	0	0	0
229	Konser Lagu Nostalgia	entertainment	0	0	0	0	0	0	0	0	*1	0	0	0	0
230	Evan Sanders Rancang Sendiri Kostumnya	entertainment	0	0	0	0	0	0	0	0	*1	0	0	0	0
231	Evanescence Tolak Istora Senayan	entertainment	0	0	0	0	0	0	0	0	*1	0	0	0	0
232	Nikita Mirzani: Soal Pakaian, Tak Ada yang Harus Diubah	entertainment	0	0	0	0	0	0	0	0	*1	0	0	0	0
233	Nonton Kembali "Titanic", Kate Winslet Merasa Aneh	entertainment	0	0	0	0	0	0	0	0	*1	0	0	0	0
234	Promotor Yakin Tiket Lady Gaga Terjual dalam Sehari	entertainment	0	0	0	0	0	0	0.393	0.038	0	0	0.001	*0.568	
235	Menyelamatkan Zeus dari Neraka	entertainment	0	0	0	0.004	0	0	0	*0.996	0	0	0	0	
236	Sempat Ada Gangguan "Sound System" di Konser SuJu	entertainment	0	0	0	0	0	0	0	*1	0	0	0	0	
237	Andien Keluar dari Zona Aman	entertainment	0	0	0	0	0	0	0	*1	0	0	0	0	
238	Manik: La Luna Malah Dapat Tambahan Penggemar	entertainment	0	0	0	0	0	0	0	*1	0	0	0	0	

No	Judul Artikel Uji	Kelas Asli	Kelas Prediksi												
			nasional	internasi onal	olahraga	sains	edukasi	ekonomi	tekno	<i>entertai nment</i>	otomotif	<i>health</i>	properti	travel	
268	Terminal Konektor Khusus untuk Mobil Listrik	otomotif	0	0	0	0	0	0	0	0	0	*1	0	0	0
269	Paket Velg dan Ban Murah serta Panel buat Freed	otomotif	0	0	0	0	0	0	0	0	0	*1	0	0	0
270	Dua Model Volvo "Made in" China	otomotif	0	0	0	0	0	0	0	0	0	*1	0	0	0
271	Uji Coba Keseharian Mobil Listrik dengan Audi A1 e-tron	otomotif	0	0	0	0	0	0	0	0	0	*1	0	0	0
272	Lucu, Mobil Komuter Listrik dari Oslo	otomotif	0	0	0	0	0	0	0	0	0	*1	0	0	0
273	Peminat Scoopy Antre, CBR 250R Bisa Daftar	otomotif	0	0	0	0	0	0	0	0	0	*1	0	0	0
274	Renault Siapkan Baterai 240 Km untuk Zoe	otomotif	0	0	0	0	0	0	0	0	0	*1	0	0	0
275	Tarian" Honda Blade di Pakuwon Supermall	otomotif	0	0	0	0	*0.970	0	0	0	0.029	0	0	0	0
276	Honda Perkenalkan Civic Gen	otomotif	0	0	0	0	0	0	0	0	0	*1	0	0	0
277	Toyota Yaris dan Dyna Melejit di Awal Tahun Kelinci	otomotif	0	0	0	0	0	0	0	0	0	*1	0	0	0
278	Menaklukkan Sentul dengan Honda CBR 250R	otomotif	0	0	0	0	0	0	0	0	0	*1	0	0	0
279	Dua Grand Cherokee Terbaru Ramaikan Pasar SUV Indonesia	otomotif	0	0	0	0	0	0	0	0	0	*1	0	0	0
280	Perancis Batasi Usia	otomotif	0	0	0	0	0	0	0.013	0	0	*0.987	0	0	0

No	Judul Artikel Uji	Kelas Asli	Kelas Prediksi												
			nasional	internasi onal	olahraga	sains	edukasi	ekonomi	tekno	entertai nment	otomotif	health	properti	travel	
	Kendaraan														
281	Jualan Turun, Daihatsu Tetap Bangun Pabrik	otomotif	0	0	0	0	0	0.063	0	0	*0.937	0	0	0	
282	Honda Motor Kebut Pembangunan Pabrik di India	otomotif	0	0	0	0	0	0	0	0	*1	0	0	0	
283	Honda Serious Pasarkan Motor CBU	otomotif	0	0	0	0	0	0	0	0	*1	0	0	0	
284	Beli New Toyota Hilux Inden Dua Bulan	otomotif	0	0	0	0	0	0	0	0	*1	0	0	0	
285	Masuk Citos, New Blade Diserbu Konsumen	otomotif	0	0	0	0	0	0	0	0	*1	0	0	0	
286	Yamaha Lepas 1.000 Pemudik Asik	otomotif	0	0	0	0	0	0	0	0	0	0	0	*1	
287	Peluncuran All New Avanza 2012 Terancam Mundur?	otomotif	0	0	0	0	0	0	0	0	*1	0	0	0	
288	Kawasaki Ninja Berkulit Serat Karbon	otomotif	0	0	0	0	0	0	0	0	*1	0	0	0	
289	Toyota Perkenalkan 86 Versi Produksi	otomotif	0	0	0	0	0	0	0	0	*1	0	0	0	
290	Nissan Luncurkan Kuro dan Shiro	otomotif	0	0	0	0	0	0	0	0	*1	0	0	0	
291	Jangan Salah Pilih Pelek	otomotif	0	0	0	0	0	0	0	0	*1	0	0	0	
292	VW Beetle Mesin Diesel untuk Amerika	otomotif	0	0	0	0	0	0	0	0	*1	0	0	0	
293	Suzuki Tambah Dealer, Dua Sekaligus	otomotif	0	0	0	0	0	0	0	0	*0.967	0	0.033	0	

No	Judul Artikel Uji	Kelas Asli	Kelas Prediksi												
			nasional	internasi onal	olahraga	sains	edukasi	ekonomi	tekno	<i>entertai nment</i>	otomotif	<i>health</i>	properti	travel	
294	Gema Tokyo Motor Show Mulai Membahana	otomotif	0	0	0	0	0	0	0	0	0	*1	0	0	0
295	Ford Gunakan Serat Karbon di Mobil Hibrida dan Listrik	otomotif	0	0	0	0	0	0	0	0	0	*1	0	0	0
296	Gunawan Tancap Gas 200 kpj Bersama Ducati	otomotif	0	0	0	0	0	0	0	0	0.115	*0.885	0	0	0
297	Halal-Haram Vaksin Meningitis	health	*1	0	0	0	0	0	0	0	0	0	0	0	0
298	Inilah Bukti Vaksin Meningitis Sumbernya Sama	health	0	0	0	*1	0	0	0	0	0	0	0	0	0
299	Inilah Penyebab Nyeri Leher	health	0	0	0	0	0	0	0	0	0	0	*1	0	0
300	Tambah 5 Menit untuk Pembakaran Lemak	health	0	0	0	0	0	0	0	0	0	0	*1	0	0
301	Mikrochip Gantikan Fungsi Retina	health	0	0	0	0	0	0	0	0	0	0	*1	0	0
302	Lindungi Otak dengan Kopi	health	0	0	0	0	0	0	0	0	0	0	*1	0	0
303	Kuatkan Tubuh, Jauhi Flu	health	0	0	0	0	0	0	0	0	0	0	*1	0	0
304	Hindari Minum Obat dengan Susu	health	0	0	0	0	0	0	0	0	0	0	*1	0	0
305	Menderita Kanker Tak Perlu Ditutupi	health	0	0	0	0	0	0	0	0	0	0	*1	0	0
306	Bahaya Cacar Air pada Ibu Hamil	health	0	0	0	0	0	0	0	0	0	0	*1	0	0
307	Pasien Terduga Flu Burung Bertambah Lagi	health	0	0	0	0	0	0	0	0	0	0	*1	0	0

No	Judul Artikel Uji	Kelas Asli	Kelas Prediksi												
			nasional	internasi onal	olahraga	sains	edukasi	ekonomi	tekno	entertai nment	otomotif	health	properti	travel	
308	Pasien dan Dokter Sama-sama Picu TB Resisten	health	0	0	0	0	0	0	0	0	0	0	*1	0	0
309	Susah Makan, Perlukah Anak Minum Vitamin?	health	0	0	0	0	0	0	0	0	0	0	*1	0	0
310	Merawat Kulit dengan Isi Kulkas	health	0	0	0	0	0	0	0	0	0	0	*1	0	0
311	Pria Penyebab Utama Gangguan Kesuburan?	health	0	0	0	0	0	0	0	0	0	0	*1	0	0
312	8 Manfaat Kedelai Bagi Kesehatan	health	0	0	0	0	0	0	0	0	0	0	*1	0	0
313	Rumahan saja Rakyat, Tak Perlu Tarik-Menarik!	health	0	0	0	0	0	0	0	0	0	0	0	*1	0
314	Vaksin Hepatitis C Semakin Dekat	health	0	0	0	0	0	0	0	0	0	0	*1	0	0
315	10 Alasan Anda Harus Mengayuh Sepeda	health	0	0	0	0	0	0	0	0	0	0	*1	0	0
316	Kulit Kentang Kleci Berpotensi Jadi Obat Kanker	health	0	0	0	0	0	0	0	0	0	0	*1	0	0
317	Infeksi Cacing Cambuk Bisa Memicu Anemia	health	0	0	0	0	0	0	0	0	0	0	*1	0	0
318	8 Alasan Menunda Vaksinasi pada Anak	health	0	0	0	0	0	0	0	0	0	0	*1	0	0
319	Kiat Menurunkan Kolesterol Tanpa Obat	health	0	0	0	0	0	0	0	0	0	0	*1	0	0
320	Mengompol karena Stres	health	0	0	0	0	0	0	0	0	0	0	*1	0	0
321	Gigi Tak Rata Bukan Sekadar Soal Estetika	health	0	0	0	0	0	0	0	0	0	0	*1	0	0

No	Judul Artikel Uji	Kelas Asli	Kelas Prediksi												
			nasional	internasional	olahraga	sains	edukasi	ekonomi	tekno	entertainment	otomotif	health	properti	travel	
322	Perubahan Gaya Hidup 30 Persen Cegah Kanker	health	0	0	0	0	0	0	0	0	0	0	*1	0	0
323	Waspadai Kadar Gula dalam Susu Pertumbuhan	health	0	0	0	0	0	0	0	0	0	0	*1	0	0
324	Mengenal Penyakit Batu Empedu	health	0	0	0	0	0	0	0	0	0	0	*1	0	0
325	9 Makanan Ini Memicu "Heartburn"	health	0	0	0	0	0	0	0	0	0	0	*1	0	0
326	Kemandirian Bahan Baku Obat Terus Ditingkatkan	health	0	0	0	0	0	*1	0	0	0	0	0	0	0
327	BTN Salurkan KPR Dengan Fasilitas Likuiditas	properti	0	0	0	0	0	0	0	0	0	0	0	*1	0
328	Harga Tanah di Jakarta Pusat dan Jakarta Utara Termahal	properti	0	0	0	0	0	0	0	0	0	0	0	*1	0
329	Perumnas Bangun 20.000 Rumah Pada Tahun 2011	properti	0	0	0	0	0	0	0	0	0	0	0	*1	0
330	Dapur Dengan Meja Makan Mungil di Sudutnya	properti	0	0	0	0	0	0	0	0	0	0	0	*1	0
331	Sofa dan Meja Pojok Berkaki Tembok	properti	0	0	0	0	0	0	0	0	0	0	0	*1	0
332	Pemkot Tangsel Kembangkan Perumahan Swadaya	properti	0	0	0	0	0	0	0	0	0	0	0	*0.524	0.476
333	Menjelajahi Ujung Genteng	properti	0	0	0	0	0	0	0	0	0	0	0	0	*1
334	Proyek Apartemen Baru Agung Podomoro di Pos Pengumben	properti	0	0	0	0	0	0.009	0	0	0	0	0	*0.991	0

No	Judul Artikel Uji	Kelas Asli	Kelas Prediksi												
			nasional	internasional	olahraga	sains	edukasi	ekonomi	tekno	entertainment	otomotif	health	properti	travel	
335	Pengembang Garap Pinggiran Jakarta Barat	properti	0	0	0	0	0	0	0	0	0	0	0	*1	0
336	Arvin: Gapuraprima Tetap Sediakan Hunian Menengah	properti	0	0	0	0	0	0	0	0	0	0	0	*1	0
337	Transformasi Unik Si Fosil Kayu	properti	0	0	0	0	0	0	0	0	0	0	0	*1	0
338	REI Gandeng Pemkab Tabanan Bangun Permukiman Terpadu	properti	0	0	0	0	0	0	0	0	0	0	0	*1	0
339	Kemenpera Bangun Rusunawa Baru di Palembang	properti	0	0	0	0	0	0	0	0	0	0	0	*1	0
340	Pemkot Ambon Benahi Obyek Wisata	properti	0	0	0	0	0	0	0	0	0	0	0	0	*1
341	Rusunawa Gunungsari untuk Korban Penggusuran	properti	0	0	0	0	0	0	0	0	0	0	0	*1	0
342	Kalau Harga Tanah Gila-gilaan, Siapa Mau Beli?	properti	0	0	0	0	0	0	0	0	0	0	0	*1	0
343	"Rumah Keong" dan Ketenangan Batin Dokter ...	properti	0	0	0	0	0	0	0	0	*1	0	0	0	0
344	Hijau Muda-Putih, Perpaduan yang Cocok di Kamar Anak	properti	0	0	0	0	0	0	0	0	0	0	0	*1	0
345	"Menyobek" Dinding Demi Memasukkan Cahaya	properti	0	0	0	0	0	0	0	0	0	0	0	*1	0
346	RUU Rusun Memberatkan Pengembangan Kecil	properti	0	0	0	0	0	0	0	0	0	0	0	*1	0
347	Waduh... Kabar Buruk Bagi yang Butuh Rumah Murah!	properti	0	0	0	0	0	0	0	0	0	0	0	*1	0



No	Judul Artikel Uji	Kelas Asli	Kelas Prediksi												
			nasional	internasi onal	olahraga	sains	edukasi	ekonomi	tekno	entertai nment	otomotif	health	properti	travel	
361	Museum Proklamasi Segera Terwujud	travel	0	0	0	0	0	0	0	0	0	0	0	0	*1
362	Festival Dongdang Ini yang Pertama Kali	travel	0	0	0	0	0	0	0	0	0	0	0	0	*1
363	Danau Batur Hapus Citra Buruk...	travel	0	0	0	0	0	0	0	0	0	0	0	0	*1
364	Menggapung di Jumbo Kingdom	travel	0	0	0	0	0	0	0	0	0	0	0	0	*1
365	Batal ke Jepang, Wisman Pilih Bromo	travel	0	0	0	0	0	0	0	0	0	0	0	0	*1
366	Wisata Religi Seribu Kuil	travel	0	0	0	0	0	0	0	0	0	0	0	0	*1
367	Bandung Bakal Punya Museum Layang-layang	travel	0	0	0	0	0	0	0	0	0	0	0	0	*1
368	Tips Berwisata di Taman Wisata	travel	0	0	0	0	0	0	0	0	0	0	0	0	*1
369	Rekreasi Pedesaan di Taman Wisata Matahari	travel	0	0	0	0	0	0	0	0	0	0	0	0	*1
370	Teluk Kendari Jadi Wisata Andalan Kendari	travel	0	0	0	0	0	0	0	0	0	0	0	0	*1
371	Forum Pariwisata Antarkota di Asia Digelar	travel	0	0	0	0	0	0	0	0	0	0	0	0	*1
372	"Traveling" Kok Malah Stres? Simak Tips Ini...	travel	0	0	0	0	0	0	0	0	0	0	0	0	*1
373	Pengunjung Puncak Borobudur Akan Dibatasi	travel	0	0	0	0	0	*1	0	0	0	0	0	0	0
374	Inilah Tiga Strategi Promosi Wisata 2012	travel	0	0	0	0	0	0	0	0	0	0	0	0	*1

