



TUGAS AKHIR – SM141501

***TEXT MINING DENGAN METODE NAÏVE
BAYES CLASSIFIER UNTUK
MENGKLASIFIKAN BERITA BERDASARKAN
KONTEN***

AIN NUOR ROIFA
NRP. 0611134000009

Dosen Pembimbing:

1. Prof. Dr. Moh. Isa Irawan, MT
2. Alvida Mustika Rukmi, S.Si, M.Si

DEPARTEMEN MATEMATIKA
Fakultas Matematika, Komputasi dan Sains Data
Institut Teknologi Sepuluh Nopember
Surabaya 2018



FINAL PROJECT – SM141501

***TEXT MINING METHOD FOR CLASSIFYING
NAÏVE BAYES CLASSIFIER BASED ON THE
CONTENT OF NEWS***

AIN NUOR ROIFA
NRP. 0611134000009

Supervisor:

1. Prof. Dr. Moh. Isa Irawan, MT
2. Alvida Mustika Rukmi, S.Si, M.Si

DEPARTMENT MATHEMATICS
Faculty of Mathematics, Computing and Data Science
Institut Teknologi of Sepuluh Nopember
Surabaya 2018

LEMBAR PENGESAHAN

**TEXT MINING DENGAN METODE NAÏVE BAYES
CLASSIFIER UNTUK MENGLASIFIKAN BERITA
BERDASARKAN KONTEN
TEXT MINING METHOD FOR CLASSIFYING NAÏVE
BAYES CLASSIFIER BASED ON THE CONTENT OF NEWS**

Diajukan untuk memenuhi salah satu syarat
Untuk memperoleh gelar Sarjana Sains
Pada bidang studi Ilmu Komputer
Program Studi S-1 Departemen Matematika
Fakultas Matematika, Komputasi dan Sains Data
Institut Teknologi Sepuluh Nopember Surabaya

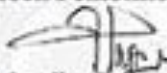
Oleh :

Ain Nuor Roifa

NRP. 0611134000009

Menyetujui,

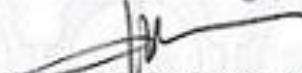
Dosen Pembimbing II,



Alvida Mustika Rukmi, S.Si, M.Si

NIP. 19720715 199802 2 001

Dosen Pembimbing I,



Prof. Dr. Moh. Isa Irawan, MT

NIP. 19631225 198903 1 001

Mengetahui,

Ketua Departemen Matematika

FMKSD ITS



Dr. Imam Mukhlash, S.Si, MT

NIP. 19700831 199403 1 003

Surabaya, Januari 2018



TEXT MINING DENGAN METODE NAÏVE BAYES CLASSIFIER UNTUK MENGLASIFIKAN BERITA BERDASARKAN KONTEN

Nama Mahasiswa : Ain Nuor Roifa
NRP : 0611134000009
Departemen : Matematika-FMKSDITS
Dosen Pembimbing : 1. Prof. Dr. Moh. Isa Irawan, MT
2. Alvida Mustika Rukmi, S.Si, M.Si

Abstrak

Berita merupakan salah satu bentuk informasi yang disampaikan melalui media. Ada berita yang disampaikan melalui media cetak seperti koran, dan ada pula berita yang disiarkan melalui media elektronik seperti televisi. Banyak orang mengetahui setiap berita atau informasi yang disampaikan di televisi. Sebelum berita disampaikan di televisi, berita diklasifikasikan terlebih dahulu. Dalam proses klasifikasi masih dilakukan secara manual oleh jurnalis di perusahaan, sehingga dibutuhkan waktu yang cukup banyak. Oleh karena itu perlu adanya metode yang bisa mengklasifikasi secara tepat dan otomatis. *Text mining* merupakan variasi data mining yang berusaha menemukan pola menarik dari sekumpulan data tekstual dengan jumlah cukup besar. Sedangkan algoritma *naïve bayes classifier* merupakan algoritma metoda untuk melakukan klasifikasi. Dalam penelitian ini data yang digunakan berupa teks berita dari TV9. Berita terdiri dari 4 kategori yaitu agama, ekonomi, politik dan sosial. Setiap kategori memiliki jumlah yang berbeda dengan total semua berita ada 935 teks berita; 500 teks berita digunakan untuk proses *training*, dan 435 teks berita digunakan untuk proses *testing*. Hasil dari penelitian ini adalah program klasifikasi teks berita dengan menggunakan *software* JAVA Neatbeans dapat mengklasifikasi berita kedalam empat kelas dengan tingkat akurasi 94,71265%.

Kata kunci: Program klasifikasi berita, *Text Mining*, dan *Naïve Bayes Classifier*.

TEXT MINING METHOD FOR CLASSIFYING NAÏVE BAYES CLASSIFIER BASED ON THE CONTENT OF NEWS

Student Name : Ain Nuor Roifa
NRP : 0611134000009
Department : Mathematics- FMKSD ITS
Supervisor : 1. Prof. Dr. Moh. Isa Irawan, MT
2. Alvida Mustika Rukmi, S.Si, M.Si

Abstract

News are one form of information submitted through the media. There is news delivered through print media such as newspapers, and there are also news broadcast via electronic media such as television. Many people know every news or information submitted on television. Before news is delivered on television, news is classified first. In the process of classification is still done manually by journalists in the company, so it takes considerable time. Therefore the need for a method that can classify precisely and automatically. Text mining is a variation of data mining that seeks to find interesting patterns from a batch of textual data with large numbers. While the naïve bayes classifier algorithm is a method algorithm for classification. In this study the data used in the form of news text from TV9. News consists of 4 categories: religion, economy, politics and social. Each category has a different amount to the total of all the news there are 935 news texts; 500 news texts are used for the training process, and 435 news texts are used for the testing process. The result of this research is news classification program using JAVA Neatbeans software can classify the news into four classes with accuracy level 94,71265%.

Keywords: News classification program, Text Mining, and Naïve Bayes Classifier.

KATA PENGANTAR

Puji syukur kepada Tuhan Yang Maha Esa atas segala karunia dan rahmat-Nya penulis dapat menyelesaikan tugas akhir yang berjudul “*Text Mining* dengan Metode *Naïve bayes classifier* Untuk Mengklasifikasikan Berita Berdasarkan Konten”.

Melalui lembar ini, penulis ingin menyampaikan ucapan terima kasih dan penghormatan yang sebesar-besarnya kepada:

1. Prof. Dr. Moh. Isa Irawan, MT selaku Dosen Pembimbing I yang telah memberikan bimbingan dan motivasi kepada penulis dalam mengerjakan Tugas Akhir ini sehingga dapat terselesaikan dengan baik.
2. Alvida Mustika Rukmi S.Si., M.Si selaku Dosen Pembimbing II yang telah memberikan bimbingan dan motivasi kepada penulis dalam mengerjakan Tugas Akhir ini sehingga dapat terselesaikan dengan baik.
3. Dr. Imam Mukhlash, S.Si., MT selaku Kepala Departemen Matematika ITS.
4. Dr. Subiono, M.Sc selaku Dosen Wali yang telah memberikan arahan akademik selama penulis menempuh pendidikan di Departemen Matematika ITS sekaligus sebagai dosen penguji yang banyak memberi masukan.
5. Dr. Budi Setiyono, S.Si., MT sebagai dosen penguji yang menambah masukan kepada penulis.
6. Dr. Didik Khusnul Arif, S.Si., M.Si selaku Kepala Program Studi S1 Departemen Matematika ITS.
7. Drs. Iis Herisman, M.Si selaku Sekretaris Program Studi S1 Departemen Matematika ITS.
8. Seluruh jajaran dosen dan staf Jurusan Matematika ITS.
9. Orang tua penulis dan adik penulis yang tiada henti-hentinya mencurahkan kasih sayang, perhatian, semangat dan do'a kepada penulis selama ini.
10. Teman-teman Grup S.Si 16/17 dan Grup S.Math 17/18 yang senantiasa memberikan motivasi dan hiburan pada penulis.

11. Rekan-rekan Matematika ITS terutama angkatan 2013 yang telah menjadi teman penulis hingga sekarang serta segenap Keluarga Besar PMII Sepuluh Nopember yang senantiasa memberi semangat dan menghibur Penulis.
12. Semua pihak yang tak bisa penulis sebutkan satu-persatu, terima kasih telah membantu sampai terselesaikannya Tugas Akhir ini.

Sebagai manusia biasa, penulis menyadari bahwa tugas akhir ini masih jauh dari kesempurnaan dan memiliki banyak kekurangan. Sehingga dengan segala kerendahan hati, penulis mengharapkan saran dan kritik yang membangun dari pembaca. Akhir kata, semoga Tugas Akhir ini dapat bermanfaat bagi semua pihak yang berkepentingan.

Surabaya, Januari 2018

Penulis

DAFTAR ISI

LEMBAR PENGESAHAN.....	Error! Bookmark not defined.
Abstrak	vii
Abstract	ix
KATA PENGANTAR.....	xi
DAFTAR ISI	xiii
DAFTAR GAMBAR	xv
DAFTAR TABEL	xvii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah.....	3
1.3 Batasan Masalah	3
1.4 Tujuan	3
1.5 Manfaat	3
1.6 Sistematika Penulisan Tugas Akhir	3
BAB II TINJAUAN PUSTAKA.....	5
2.1 Penelitian Terdahulu	5
2.2 Proses Pengolahan Berita.....	6
2.3 Pengertian <i>Text Mining</i>	6
2.4 Prosedur <i>Text Mining</i>	8
2.5 Teorema <i>Bayes</i>	8
2.6 Pengertian <i>Naïve Bayes Classifier</i>	12
2.7 Pengertian <i>Naïve Bayes Classifier</i>	13
2.8 Kelebihan <i>Naïve Bayes Classifier</i>	13
2.9 Metode <i>Naïve Bayes Classifier</i>	14
2.10 Mengukur Kinerja Klasifikasi dengan Metode Confusion Matrix.....	16
BAB III METODE PENELITIAN	21
3.1 Pengumpulan Data	21
3.2 Pre-Processing	21
3.2.1 <i>Case Folding</i>	21
3.2.2 Filtering.....	21
3.2.3 Stemming.....	21
3.2.4 Lemmatitation.....	21

3.3	Proses Ekstraksi Kata	22
3.4	Proses Klasifikasi Teks berita	22
3.5	Penarikan Kesimpulan dan Penyusunan Laporan	22
BAB IV PERANCANGAN DAN IMPLEMENTASI SISTEM		25
4.1	Perancangan Data	25
4.1.1	Data Masukan	25
4.1.2	Data Proses	26
4.1.3	Data Keluaran	26
4.2	Perancangan Diagram Alir Sistem	27
4.2.1	Pre-proses Teks	28
4.2.2	Ekstraksi Kata	30
4.2.3	Klasifikasi menggunakan <i>Naïve Bayes Classifier</i>	35
4.3	Implementasi Sistem	36
4.3.1	Implementasi Pre-proses Teks	36
4.3.2	Implementasi ekstraksi kata	38
4.3.3	Implementasi proses klasifikasi menggunakan <i>Naïve Bayes Classifier</i>	41
4.3.4	Implementasi Metode Evaluasi	44
4.4	Implementasi Antarmuka GUI	46
BAB V HASIL DAN PEMBAHASAN		53
5.1	Hasil Pre-proses Data	53
5.2	Ekstraksi Nilai dan Perhitungan <i>Likelihood</i>	58
5.3	Hasil Klasifikasi <i>Naïve Bayes Classifier</i>	62
BAB VI PENUTUP		69
6.1	Kesimpulan	69
6.2	Saran	69
DAFTAR PUSTAKA		71
LAMPIRAN A		73
LAMPIRAN B		75
LAMPIRAN C		77
LAMPIRAN D		79
LAMPIRAN E		81
BIODATA PENULIS		83

DAFTAR GAMBAR

Gambar 2. 1 Proses Pengolahan Berita	6
Gambar 3. 1 Diagram Alir Metodologi Penelitian.....	23
Gambar 4. 1 Diagram Alir Sistem.....	27
Gambar 4. 2 Diagram Alir <i>Case Folding</i> dan <i>Filtering</i>	28
Gambar 4. 3 Gambaran Potongan Teks Berita.....	28
Gambar 4. 4 Contoh Hasil <i>Casefolding</i> dan <i>filtering</i>	29
Gambar 4. 5 Diagram Alir <i>Stemming</i> dan <i>Tokenizing</i>	29
Gambar 4. 6 Sebelum Melalui <i>Stemming</i>	30
Gambar 4. 7 Setelah Melalui <i>Stemming</i> dan <i>Tokenizing</i>	30
Gambar 4. 8 Diagram Alir Ekstraksi Kata	30
Gambar 4. 9 Diagram Alir Klasifikasi NBC	35
Gambar 4. 10 Tampilan GUI tab Proses Awal.....	46
Gambar 4. 11 Hasil Pre-processing	46
Gambar 4. 12 Hasil Ekstraksi Kata	47
Gambar 4. 13 Tampilan GUI tab Perhitungan Kata	47
Gambar 4. 14 Hasil Hitung Probabilitas Kata	48
Gambar 4. 15 Tampilan GUI tab Klasifikasi.....	48
Gambar 4. 16 Proses Perhitungan <i>Naïve Bayes Classifier</i>	50
Gambar 4. 17 Hasil <i>Test</i> Satu Berita	51
Gambar 4. 18 Gambaran Proses <i>Testing</i>	51
Gambar 4. 19 Hasil <i>Testing</i>	52
Gambar 5. 1 Satu Berita Studi Kasus.....	66
Gambar 5. 2 Hasil Pre-processing Satu Teks Berita.....	66

DAFTAR TABEL

Tabel 2. 1 Perhitungan Bayes.....	10
Tabel 2. 2 Tabel <i>Confusion Matrix</i>	18
Tabel 4. 1 Hasil Preprocessing Data <i>Training</i>	31
Tabel 5. 1 Kriteria Klasifikasi	53
Tabel 5. 2 Lima Teks Berita Data <i>Testing</i>	54
Tabel 5. 3 Hasil <i>Pre-Processing</i>	56
Tabel 5. 4 Tabel Perhitungan <i>Naïve Bayes Classifier</i>	62
Tabel 5. 5 Hasil Klasifikasi	63
Tabel 5. 6 Hasil <i>Confusion matrix</i>	64
Tabel 5. 7 Perhitungan Presisi, Recall, dan Akurasi	65
Tabel 5. 8 Likelihood studi kasus.....	66
Tabel 5. 9 Perhitungan <i>Naïve bayes classifier</i> pada studi kasus ..	67

BAB I

PENDAHULUAN

Pada bagian ini dijelaskan mengenai latar belakang, permasalahan, batasan masalah, tujuan, manfaat yang ingin dicapai, dan sistematika penulisan yang digunakan dalam pembuatan tugas akhir ini.

1.1 Latar Belakang

Perkembangan informasi saat ini tumbuh pesat, mulai dari informasi politik, ekonomi, sosial budaya dan kemasyarakatan. Media menjadi konsumsi publik tentang kabar terkini atau teraktual di suatu wilayah tertentu. Pemberitaan tersebut mampu memberitakan informasi supaya diketahui publik sebagai pembaca setianya. Dalam penulisan berita di meja redaksi sangat menentukan informasi yang disampaikan kepada publik sebagai santapan pembaca. Media harus memberitakan dengan sejujur-jujurnya karena sangat mempengaruhi publik pembaca tersebut. Apabila kesalahan menulis dapat memberitakan informasi yang salah kepada masyarakat. Ahli media menulis tentang keuntungan media pendapat ahli media Burrell Headly yang membela pers kapitalis mengatakan bahwa berita adalah komoditas yang dijual dan media massa merupakan usaha menjual informasi [17].

Sebagai media untuk menyampaikan informasi, tidak dipungkiri televisi merupakan alat yang ampuh dalam mempengaruhi bahkan membentuk opini publik termasuk anak-anak. Jadi, selain memiliki dimensi informasi, hiburan, dan sosial, media televisi juga memiliki dimensi lain seperti politik, karena kemampuannya menjadi alat propaganda yang dahsyat. Bahkan media televisi dapat berkembang menjadi kelompok penekan atas suatu gagasan, ide atau bahkan kepentingan kelompok tertentu[18]. Salah satu media televisi di Surabaya yaitu stasiun TV9 yang menyiarkan berita minimal dalam sehari 3 kali yaitu pagi, siang, dan petang. Dalam pengolahan database berita masih dikelompokkan berdasarkan waktu penyampaian berita, tanpa mengelompokkan berita-berita itu termasuk dalam kategori apa. Berdasarkan kunjungan yang

dilakukan penulis ke perusahaan TV9 secara langsung penting adanya suatu pengelompokan berita berdasarkan kategori tertentu. Oleh karena itu, disini penulis akan mengelompokkan atau mengklasifikasikan berita-berita tersebut kedalam beberapa kategori diantaranya agama, ekonomi, politik dan sosial. Sehingga setiap butuh kategori tertentu dapat mencari dengan mudah jika berita hendak ditayangkan ulang atau ada pihak yang membutuhkan. Terkadang beberapa orang atau golongan butuh teks berita untuk dimuat di media massa, dan lain lain. Tidak hanya itu, penataan data berita yang baik terhadap database berita juga akan memudahkan pencarian ketika berita sudah terklasifikasi. Maka klasifikasi berita berdasarkan konten teks berita ini menjadi penting, dan dalam pengklasifikasian berita tersebut dibutuhkan isi dari berita secara keseluruhan agar dapat diklasifikasikan secara detail. Setelah data didapatkan dan berita dari wartawan masuk ke redaktur berita, oleh redaktur berita diolah berdasarkan kesesuaian EYD, dan dilakukan editing konten agar menarik ketika disampaikan. Berita yang telah diedit dan disempurnakan perlu disetujui oleh kepala redaktur sebelum berita disampaikan (pimpinan redaksi TV9- Januari 2017). Setelah teks berita mendapat persetujuan dari kepala redaktur maka teks berita akan diklasifikasi berdasarkan isi keseluruhan teks dalam 4 kategori yaitu agama, ekonomi, politik dan sosial. Proses klasifikasi berita dilakukan dengan *text mining* menggunakan metode *naïve bayes classifier* agar berita terklasifikasi sesuai kategorinya.

Text mining adalah salah satu teknik atau metode yang dapat digunakan untuk melakukan klasifikasi, dimana *text mining* merupakan variasi dari data mining yang berusaha menemukan pola yang menarik dari sekumpulan data tekstual yang berjumlah besar[1]. *Naïve bayes classifier*(NBC) merupakan sebuah metode pengklasifikasi probabilitas sederhana yang mengaplikasikan Teorema Bayes dengan asumsi ketidaktergantungan (independent) yang tinggi[2].

1.2 Rumusan Masalah

Berdasarkan latar belakang diatas, rumusan masalah dalam usulan Tugas Akhir ini adalah bagaimana mengklasifikasi berita berdasarkan konten teks pada berita dengan *text mining* menggunakan metode *naïve bayes classifier*.

1.3 Batasan Masalah

Pembahasan pada Tugas Akhir ini dibatasi pada beberapa hal berikut :

1. Data yang digunakan berupa teks berita yang didapat dari TV9 Surabaya pada periode Juni-Desember 2016.
2. File dataset teks berita dengan format .doc (Microsoft Word Document).
3. Kategori yang diambil hanya 4 yaitu agama, ekonomi, politik dan sosial.
4. Metode yang digunakan untuk klasifikasi yaitu *text mining* menggunakan metode *naïve bayes classifier*.

1.4 Tujuan

Tujuan dari Tugas Akhir ini adalah mendapatkan hasil klasifikasi berita berdasarkan konten teks berita kedalam 4 kategori yaitu agama, ekonomi, politik dan sosial.

1.5 Manfaat

Manfaat dari penulisan Tugas Akhir ini adalah sebagai informasi tambahan mengenai proses klasifikasi berita berdasarkan konten teks pada berita dengan program bahasa Java, tidak lagi dilakukan secara manual. Disamping itu, memudahkan pihak TV9 jika sewaktu waktu butuh data berita untuk kategori tertentu karena data sudah rapi dan terklasifikasi.

1.6 Sistematika Penulisan Tugas Akhir

Sistematika penulisan pada tugas akhir ini disusun sebagai berikut:

1. **BAB I PENDAHULUAN**

Bab ini menjelaskan tentang latar belakang pembuatan tugas akhir, rumusan dan batasan masalah yang dihadapi dalam penelitian tugas akhir, tujuan dan manfaat pembuatan tugas akhir, dan sistematika penulisan Tugas Akhir.

2. **BAB II TINJAUAN PUSTAKA**

Bab ini menjelaskan tentang beberapa teori yang digunakan untuk membantu penyelesaian permasalahan tugas akhir.

3. **BAB III METODOLOGI**

Bab ini menjelaskan metode penelitian yang dilakukan dalam menyelesaikan Tugas Akhir.

4. **BAB IV PERANCANGAN DAN IMPLEMENTASI SISTEM**

Bab ini membahas analisa dan perancangan sistem perangkat lunak yang akan dibangun untuk klasifikasi berita berdasarkan konten teks berita menggunakan *text mining* metode *naïve bayes classifier*.

5. **BAB V HASIL DAN PEMBAHASAN**

Bab ini menjelaskan mengenai hasil pengujian yang telah dilakukan pada perangkat lunak. Hasil pengujian kemudian dibahas untuk mengetahui kerja sistem secara keseluruhan.

6. **BAB VI PENUTUP**

Bab ini merupakan penutup, berisi tentang kesimpulan yang dapat diambil berdasarkan data yang ada dan saran yang selanjutnya dilakukan bila tugas akhir ini dilanjutkan.

BAB II TINJAUAN PUSTAKA

Pada bab ini diuraikan mengenai landasan teori dan materi penunjang yang terkait dalam permasalahan Tugas Akhir antara lain penelitian terdahulu, proses pengolahan berita, pengertian *text mining*, prosedur *text mining*, pengertian metode *naïve bayes classifier*, dan metode *naïve bayes classifier*.

2.1 Penelitian Terdahulu

Proses pengklasifikasian dapat dilakukan dengan beberapa cara, untuk klasifikasi data berupa teks dapat dilakukan dengan *text mining*. Salah satunya yaitu penelitian yang dilakukan oleh S.Brindha, Dr.S.Sukumaran, dan Dr.K.Prabha dalam penelitian yang berjudul *A Survey on Classification Techniques for Text Mining*. Pada penelitian ini dihasilkan beberapa teknik pengklasifikasian, ada beberapa teknik yang digunakan diantaranya *K-Nearest Neighbour*, *Naïve Bayes Algorithm*, *Support Vector Machine*, *Decision Tree Methods*, *Log Regression*[10].

Sebagai bentuk implementasi metode, pada tahun 2016 Dea Delvia Arifin, Shaufiah, dan Moch. Arif Bijaksana dalam penelitian yang berjudul *Enhancing Spam Detection on Mobile Phone Short Message Service (SMS) Performance using FP-Growth and Naïve bayes classifier* membuktikan bahwa metode *Naïve bayes classifier* memiliki kemampuan klasifikasi yang lebih tinggi daripada *FP-Growth* dengan keakuratan data diatas 90%[11].

Penelitian yang lain adalah Klasifikasi Emosi Untuk Teks Bahasa Indonesia Menggunakan Metode *Naïve Bayes* yang dilakukan oleh I.Destuardi. Pada penelitian ini, klasifikasi dilihat dari sub bab pendekatan yaitu pendekatan kata kunci, pendekatan hubungan kata dengan kamus, dan pendekatan statistik pemrosesan Bahasa Alami (NLP). Kemudian alur proses selanjutna yaitu penginputan teks, teks diproses dengan *pre*

processing (tokenizing, filetring), dan selanjutnya teks ditransformasikan menjadi sebuah data vektor[12].

2.2 Proses Pengolahan Berita

Proses pengolahan berita di TV9 dilakukan melalui beberapa tahap beberapa tahap. Proses ini dijelaskan langsung oleh pimpinan redaksi TV9 ketika penulis berkunjung ke kantor pada Januari 2017. Berikut proses pengolahan berita pada Gambar 2.1



Gambar 2.1 Proses Pengolahan Berita

2.3 Pengertian *Text Mining*

Text mining memiliki definisi menambang data yang berupa teks dimana sumber data biasanya didapatkan dari data, dan

tujuannya adalah mencari kata-kata yang dapat mewakili isi dari data sehingga dapat dilakukan analisa keterhubungan antar data.

Sedangkan menurut situs wikipedia, Penambangan teks (bahasa Inggris: *text mining*) adalah proses ekstraksi pola berupa informasi dan pengetahuan yang berguna dari sejumlah besar sumber data teks, seperti data Word, PDF, kutipanteks, dll. Jenis masukan untuk penambangan teks ini disebut data tak terstruktur dan merupakan pembeda utama dengan penambangan data yang menggunakan data terstruktur atau basis data sebagai masukan. Penambangan teks dapat dianggap sebagai proses dua tahap yang diawali dengan penerapan struktur terhadap sumber data teks dan dilanjutkan dengan ekstraksi informasi dan pengetahuan yang relevandari data teks terstruktur ini dengan menggunakan teknik dan alat yang sama dengan penambangan data. Proses yang umum dilakukan oleh penambangan teks di antaranya adalah perangkuman otomatis, kategorisasi data, penggugusan teks, dll.

Tujuan dari *text mining* adalah untuk mendapatkan informasi yang berguna dari sekumpulan data. Jadi, sumber data yang digunakan pada *text mining* adalah kumpulan teks yang memiliki format yang tidak terstruktur atau minimal semi terstruktur. Adapun tugas khusus dari *text mining* antara lain yaitu pengkategorisasian teks (*text categorization*) dan pengelompokan teks (*text clustering*).

Text mining merupakan penerapan konsep dan teknik data mining untuk mencari pola dalam teks, yaitu proses penganalisan teks guna menyarikan informasi yang bermanfaat untuk tujuan tertentu. Berdasarkan ketidakteraturan struktur data teks, maka proses text mining memerlukan beberapa tahap awal yang pada intinya adalah mempersiapkan agar teks dapat diubah menjadi lebih terstruktur[7].

2.4 Prosedur *Text Mining*

Text mining merupakan variasi dari data *mining* yang berusaha menemukan pola yang menarik dari sekumpulan data tekstual yang berjumlah besar [1]. Langkah-langkah yang dilakukan dalam *text mining* adalah sebagai berikut :

1. *Text Preprocessing*

Tindakan yang dilakukan pada tahap ini adalah *toLowerCase*, yaitu mengubah semua karakter huruf menjadi huruf kecil, dan *Tokenizing* yaitu proses penguraian deskripsi yang semula berupa kalimat- kalimat menjadi kata-kata dan menghilangkan tanda baca seperti tanda titik (.), koma (,), spasi dan karakter angka yang ada pada kata tersebut [5].

2. *Feature Selection*

Pada tahap ini tindakan yang dilakukan adalah menghilangkan *stoplist* (*stoplist removal*) dan *stemming* terhadap kata yang berimbuhan [1]. *Stoplist* adalah kosakata yang bukan merupakan ciri (kata unik) dari suatu data [6]. Misalnya “di”, “oleh”, “pada”, “sebuah”, “karena” dan lain sebagainya. *Stemming* adalah proses pemetaan dan penguraian berbagai bentuk (*variants*) dari suatu kata menjadi bentuk kata dasarnya (*stem*). Tujuan dari proses *stemming* adalah menghilangkan imbuhan-imbuhan baik itu berupa prefiks, sufiks, maupun konfiks yang ada pada setiap kata[4].

2.5 Teorema *Bayes*

Dalam teori probabilitas dan statistika, teorema Bayes adalah sebuah teorema dengan dua penafsiran berbeda. Dalam *penafsiran Bayes*, teorema ini menyatakan seberapa jauh derajat kepercayaan subjektif harus berubah secara rasional ketika ada petunjuk baru. Dalam *penafsiran frekuentis* teorema ini menjelaskan representasi invers probabilitas dua kejadian. Teorema ini merupakan dasar dari statistika Bayes dan memiliki penerapan dalam sains, rekayasa, ilmu ekonomi (terutama ilmu ekonomi mikro), teori permainan, kedokteran dan hukum. Penerapan

teorema Bayes untuk memperbarui kepercayaan dinamakan inferens Bayes[19].

Teorema Bayes diambil dari nama Rev. Thomas Bayes, menggambarkan hubungan antara peluang bersyarat dari dua kejadian A dan B sebagai berikut:

$$P(A | B) = \frac{P(B | A) P(A)}{P(B)} \quad (2.1)$$

atau

$$P(A | B) = \frac{P(B | A) P(A)}{P(B | A)P(A) + P(B | A^c)P(A^c)} \quad (2.2)$$

Berikut diberikan contoh di sebuah negara, diketahui bahwa 2% dari penduduknya menderita sebuah penyakit langka. 97% dari hasil tes klinik adalah positif bahwa seseorang menderita penyakit itu. Ketika seseorang yang tidak menderita penyakit itu dites dengan tes yang sama, 9% dari hasil tes memberikan hasil positif yang salah.

Jika sembarang orang dari negara itu mengambil test dan mendapat hasil positif, berapakah peluang bahwa dia benar-benar menderita penyakit langka itu?

Secara sepintas, nampaknya bahwa ada peluang yang besar bahwa orang itu memang benar-benar menderita penyakit langka itu. Karena kita tahu bahwa hasil test klinik yang cukup akurat (97%). Tetapi apakah benar demikian? Marilah kita lihat perhitungan matematikanya.

Marilah kita lambangkan informasi di atas sebagai berikut:

- B = Kejadian tes memberikan hasil positif.
- \bar{B} = Kejadian tes memberikan hasil negatif.
- A = Kejadian seseorang menderita penyakit langka itu.

- A^c = Kejadian seseorang tidak menderita penyakit langka itu.

Kita ketahui juga peluang dari kejadian-kejadian berikut:

- $P(A) = 2\%$
- $P(A^c) = 98\%$
- $P(B|A) = 97\%$
- $P(B|A^c) = 9\%$

Dengan menggunakan rumus untuk peluang bersyarat, dapat kita simpulkan peluang dari kejadian-kejadian yang mungkin terjadi dalam tabel di bawah ini:

Tabel 2. 1 Perhitungan Bayes

	A (2%)	A^c (98%)
B	Positif yang benar $P(B \cap A) = P(A) \times P(B A)$ $= 2\% \times 97\% = 0,0194$	Positif yang salah $P(B \cap A^c) = P(A^c) \times P(B A^c)$ $= 98\% \times 9\% = 0,0882$
\bar{B}	Negatif yang salah $P(\bar{B} \cap A) = P(A) \times P(\bar{B} A)$ $= 2\% \times 3\% = 0,0006$	Negatif yang benar $P(\bar{B} \cap A^c) = P(A^c) \times P(\bar{B} A^c)$ $= 98\% \times 91\% = 0,8918$

Misalnya seseorang menjalani tes klinik tersebut dan mendapatkan hasil positif, berapakah peluang bahwa ia benar-benar menderita penyakit langka tersebut? Dengan kata lain, kita mencoba untuk mencari peluang dari A , dimana B atau $P(A|B)$.

Dari tabel di atas, dapat kita lihat bahwa $P(A|B)$ adalah peluang dari positif yang benar dibagi dengan peluang positif (benar maupun salah), yaitu $0,0194 / (0,0194 + 0,0882) = 0,1803$. Kita dapat juga mendapatkan hasil yang sama dengan menggunakan rumus teorema Bayes di atas:

$$\begin{aligned}
 P(A | B) &= \frac{P(B \cap A)}{P(B)} \\
 &= \frac{P(B | A) \times P(A)}{P(B | A)P(A) + P(B | A^c)P(A^c)} \\
 &= \frac{97\% \times 2\%}{(97\% \times 2\%) + (9\% \times 98\%)} \\
 &= \frac{0.0194}{0.0194 + 0.0882} \\
 &= \frac{0.0194}{0.1076} \\
 P(A | B) &= 0.1803
 \end{aligned}$$

Hasil perhitungan ini sangat berbeda dengan intuisi kita di atas. Peluang bahwa orang yang mendapat hasil tes positif itu benar-benar menderita penyakit langka tidak sebesar yang kita

bayangkan. Cuma ada sekitar 18% kemungkinan bahwa dia benar-benar menderita penyakit itu. Mengapakah demikian?

Ketika mengira-ngira peluangnya, seringkali kita lupa bahwa dari seluruh populasi negara itu, hanya 2% yang benar-benar menderita penyakit langka itu. Jadi, walaupun hasil tes adalah positif, peluang bahwa seseorang menderita penyakit langka itu tidaklah sebesar yang kita bayangkan.

Kita bisa juga meninjau situasi di atas sebagai berikut. Misalnya populasi negara tersebut adalah 1000 orang. Hanya 20 orang yang menderita penyakit langka itu (2%). 19 orang dari antaranya akan mendapat hasil tes yang positif (97% hasil positif yang benar). Dari 980 orang yang tidak menderita penyakit itu, sekitar 88 orang juga akan mendapat hasil tes positif (9% hasil positif yang salah).

Jadi, 1000 orang di negara itu dapat kita kelompokkan sebagai berikut:

- 19 orang mendapat hasil tes positif yang benar
- 1 orang mendapat hasil tes negatif yang salah
- 88 orang mendapat hasil tes positif yang salah
- 892 orang mendapat hasil tes negatif yang benar

Bisa kita lihat dari informasi di atas, bahwa ada $(88 + 19) = 107$ orang yang akan mendapatkan hasil tes positif (tidak peduli bahwa dia benar-benar menderita penyakit langka itu atau tidak). Dari 107 orang ini, berapakah yang benar-benar menderita penyakit? Hanya 19 orang dari 107, atau sekitar 18% [20].

2.6 Pengertian *Naïve Bayes Classifier*

Naïve Bayes Classifier (NBC) merupakan sebuah pengklasifikasi probabilitas sederhana yang mengaplikasikan Teorema Bayes dengan asumsi ketidaktergantungan (independent) yang tinggi. Keuntungan penggunaan NBC adalah bahwa metoda ini hanya membutuhkan jumlah data pelatihan (training data) yang kecil untuk menentukan estimasi parameter yang diperlukan dalam proses pengklasifikasian. Karena yang diasumsikan sebagai variable independent, maka hanya varians dari suatu variable

dalam sebuah kelas yang dibutuhkan untuk menentukan klasifikasi, bukan keseluruhan dari matriks kovarians [2].

2.7 Pengertian *Naïve Bayes Classifier*

Naïve Bayes adalah teknik sederhana untuk membangun pengklasifikasi: model yang menetapkan label kelas untuk contoh masalah, direpresentasikan sebagai vektor dari nilai-nilai fitur, di mana label kelas yang diambil dari beberapa himpunan berhingga. Ini bukan algoritma tunggal untuk melatih pengklasifikasi seperti itu, tapi keluarga algoritma berdasarkan pada prinsip umum: semua pengklasifikasi *Naïve Bayes* menganggap bahwa nilai fitur tertentu tidak tergantung pada nilai dari setiap fitur lainnya, mengingat variabel kelas. Misalnya, buah dapat dianggap apel jika berwarna merah, bulat, dan berdiameter sekitar 10 cm. Sebuah *Naïve bayes classifier* menganggap masing-masing fitur ini untuk berkontribusi mandiri untuk probabilitas bahwa buah ini adalah apel, terlepas dari adanya korelasi yang mungkin antara fitur warna, kebulatan, dan diameter.

Untuk beberapa jenis model probabilitas, pengklasifikasi *Naïve Bayes* dapat dilatih dengan sangat efisien dalam pengaturan pembelajaran yang diawasi. Dalam banyak aplikasi praktis, estimasi parameter untuk model *Naïve Bayes* menggunakan metode kemungkinan maksimum; dengan kata lain, seseorang dapat bekerja dengan model *Naïve Bayes* tanpa menerima probabilitas *Bayesian* atau menggunakan metode *Bayesian*[8].

2.8 Kelebihan *Naïve Bayes Classifier*

Naïve bayes classifier bekerja sangat baik dibanding dengan model classifier lainnya. Hal ini dibuktikan oleh Xhemali, Hinde Stone dalam jurnalnya "*Naïve Bayes Vs. Decision Trees vs. Neural Networks in the classification of Training Web Pages*" mengatikan bahwa "*Naïve bayes classifier* memiliki tingkat akurasi yang lebih baik dibanding model lainnya".

Keuntungan penggunaan Metode *Naïve bayes classifier* adalah bahwa metoda ini hanya membutuhkan jumlah data pelatihan

(training data) yang kecil untuk menentukan estimasi parameter yang diperlukan dalam proses pengklasifikasian. Karena yang diasumsikan sebagai *variable independent*, maka hanya varians dari suatu *variable* dalam sebuah kelas yang dibutuhkan untuk menentukan klasifikasi, bukan keseluruhan dari matriks kovarians[9].

2.9 Metode *Naïve Bayes Classifier*

Secara garis besar model NBC (*Naïve Bayes Classifier*) adalah sebagai berikut:

$$p(C|F_1, \dots, F_n) = \frac{p(C) p(F_1, \dots, F_n|C)}{p(F_1, \dots, F_n)} \quad (2.3)$$

Atau dengan kata lain persamaan diatas dapat digambarkan sebagai:

$$\text{Posterior} = \frac{\text{Prior} * \text{Likelihood}}{\text{Evidence}} \quad (2.4)$$

Dalam penggunaannya dalam proses pengklasifikasian data, maka model NBC dapat digambarkan sebagai berikut:

Misalkan terdapat data terkait berita informasi, apakah berita tersebut termasuk kedalam kategori politik atau tidak.

Untuk itu, kita bayangkan bahwa sebuah data-data diambil dari suatu kelas data yang dapat dimodelkan sebagai himpunan kata-kata, dimana probabilitas (*independen*) bahwa suatu kata ke-*i* dalam sebuah data berasal dari kelas *C*. Hal tersebut digambarkan dengan $p(w_i|C)$. Untuk memudahkan proses klasifikasi, kita asumsikan bahwa probabilitas suatu kata dalam suatu data adalah *independen* terhadap ukuran suatu data, atau dengan kata lain semua data diasumsikan berukuran sama. Selanjutnya probabilitas bahwa sebuah data *D* terhadap kelas *C* adalah:

$$p(D|C) = \prod_i p(w_i|C) \quad (2.5)$$

Kemudian muncul pertanyaan, “Berapa probabilitas suatu data D merupakan milik suatu kelas C ?” dengan kata lain adalah berapa nilai probabilitas $p(C|D)$

Berdasarkan aksioma probabilitas:

$$p(D|C) = \frac{p(D \cap C)}{p(C)} \quad (2.6)$$

Dan

$$p(C|D) = \frac{p(D \cap C)}{p(D)} \quad (2.7)$$

Selanjutnya Teorema Bayes dipakai untuk memanipulasi pernyataan probabilitas tersebut diatas menjadi sebuah terminologi likelihood/ kemiripan.

$$p(C|D) = \frac{p(C)}{p(D)} p(D|C) \quad (2.8)$$

Untuk menjawab permasalahan diatas, maka kita asumsikan bahwa hanya ada dua kelas, yaitu kelas Spam (S) dan kelas bukan Spam ($\neg S$). Dengan demikian model dapat digambar menjadi:

$$p(D|S) = \prod_i p(w_i|S) \quad (2.9)$$

Dan

$$p(D|\neg S) = \prod_i p(w_i|\neg S) \quad (2.10)$$

Dari teorema bayes tersebut diatas, dapat kita tuliskan menjadi:

$$p(S|D) = \frac{p(S)}{p(D)} \prod_i p(w_i|S) \quad (2.11)$$

Dan

$$p(\neg S|D) = \frac{p(\neg S)}{p(D)} \prod_i p(w_i|\neg S) \quad (2.12)$$

Dengan membagi satu dengan yang lainnya dapat kita gambarkan menjadi:

$$\frac{p(S|D)}{p(\neg S|D)} = \frac{p(S) \prod_i p(w_i|S)}{p(\neg S) \prod_i p(w_i|\neg S)} \quad (2.13)$$

Model diatas dapat di-refactor kan menjadi:

$$\frac{p(S|D)}{p(\neg S|D)} = \frac{p(S)}{p(\neg S)} \prod_i \frac{p(w_i|S)}{p(w_i|\neg S)} \quad (2.14)$$

Akhirnya, rasio probabilitas dari $\frac{p(S|D)}{p(\neg S|D)}$ dapat diekspresikan dalam suatu terminologi series of likelihood-ratio/ rasio kemiripan beruntun.

Selanjutnya, probabilitas aktual dari $p(S|D)$ dapat dengan mudah dihitung melalui $\log\left(\frac{p(S|D)}{p(\neg S|D)}\right)$, berdasarkan pernyataan bahwa $p(S|D) + p(\neg S|D) = 1$.

Dengan mengambil logaritma dari keseluruhan rasio tersebut, maka kita dapatkan:

$$\ln \frac{p(S|D)}{p(\neg S|D)} = \ln \frac{p(S)}{p(\neg S)} + \sum_i \ln \frac{p(w_i|S)}{p(w_i|\neg S)} \quad (2.15)$$

(teknik menggunakan log likelihood/ kemiripan logaritma ini sangat umum dipakai kan dalam statistika).

Data tersebut merupakan Spam apabila:

$$\ln \frac{p(S|D)}{p(\neg S|D)} > 0 \quad (2.16)$$

Dan sebaliknya apabila < 0 , maka data tersebut Spam.

Maksud dari Spam disini yaitu tidak terklasifikasi kalau data itu mengandung isi atau konten politik[9].

2.10 Mengukur Kinerja Klasifikasi dengan Metode Confusion Matrix

Pengukuran terhadap kinerja suatu sistem klasifikasi merupakan hal yang penting. Kinerja sistem klasifikasi menggambarkan seberapa baik sistem dalam mengklasifikasikan data. *Confusion matrix* merupakan salah satu metode yang dapat digunakan untuk mengukur kinerja suatu metode klasifikasi. Pada dasarnya *confusion matrix* mengandung informasi yang

membandingkan hasil klasifikasi yang dilakukan oleh sistem dengan hasil klasifikasi yang seharusnya.

Berdasarkan jumlah keluaran kelasnya, sistem klasifikasi dapat dibagi menjadi 4 jenis yaitu klasifikasi *binary*, *multi-class*, *multi-label* dan *hierarchical*. Pada klasifikasi *binary*, data masukan dikelompokkan ke dalam salah satu dari dua kelas. Jenis klasifikasi ini merupakan bentuk klasifikasi yang paling sederhana dan banyak digunakan. Contoh penggunaannya antara lain dalam sistem yang melakukan deteksi orang atau bukan, sistem deteksi kendaraan atau bukan, dan sistem deteksi pergerakan atau bukan.

Sementara itu, pada bentuk klasifikasi *multi-class*, data masukan diklasifikasikan menjadi beberapa kelas. Sebagai contoh sistem yang dapat mengklasifikasikan jenis kendaraan seperti sepeda, sepeda motor, mobil, bus, truk, dan sebagainya. Bentuk klasifikasi *multi-label* pada dasarnya sama dengan *multi-class* dimana data dikelompokkan menjadi beberapa kelas, namun pada klasifikasi *multi-label*, data dapat dimasukkan dalam beberapa kelas sekaligus. Bentuk klasifikasi yang terakhir adalah *hierarchical*. Data masukan dikelompokkan menjadi beberapa kelas, namun kelas tersebut dapat dikelompokkan kembali menjadi kelas-kelas yang lebih sederhana secara hirarkis. Contohnya dalam penelitian ini, arah pergerakan dikelompokkan menjadi 12 arah pergerakan yang tentunya dapat disederhanakan menjadi 4 arah.

Pada pengukuran kinerja menggunakan *confusion matrix*, terdapat 4 istilah sebagai representasi hasil proses klasifikasi. Keempat istilah tersebut adalah *True Positive* (TP), *True Negative* (TN), *False Positive* (FP) dan *False Negative* (FN). Nilai *True Negative* (TN) merupakan jumlah data negatif yang terdeteksi dengan benar, sedangkan *False Positive* (FP) merupakan data negatif namun terdeteksi sebagai data positif. Sementara itu, *True Positive* (TP) merupakan data positif yang terdeteksi benar. *False Negative* (FN) merupakan kebalikan dari *True Positive*, sehingga data positif, namun terdeteksi sebagai data negatif.

Pada jenis klasifikasi *binary* yang hanya memiliki 2 keluaran kelas, *confusion matrix* dapat disajikan seperti pada Tabel 2.2.

Tabel 2. 2 Tabel *Confusion Matrix*

Kelas	Terklasifikasi Positif	Terklasifikasi Negatif
Positif	TP (True Positive)	FN (False Negative)
Negatif	FP (False Positive)	TN (True Negative)

Berdasarkan nilai *True Negative* (TN), *False Positive* (FP), *False Negative* (FN), dan *True Positive* (TP) dapat diperoleh nilai akurasi, presisi dan recall. Nilai akurasi menggambarkan seberapa akurat sistem dapat mengklasifikasikan data secara benar. Dengan kata lain, nilai akurasi merupakan perbandingan antara data yang terklasifikasi benar dengan keseluruhan data. Nilai akurasi dapat diperoleh dengan Persamaan 2.15. Nilai presisi menggambarkan jumlah data kategori positif yang diklasifikasikan secara benar dibagi dengan total data yang diklasifikasi positif. Presisi dapat diperoleh dengan Persamaan 2.16. Sementara itu, recall menunjukkan berapa persen data kategori positif yang terklasifikasikan dengan benar oleh sistem. Nilai recall diperoleh dengan Persamaan 2.17.

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \quad (2.17)$$

$$Presisi = \frac{TP}{TP + FP} \times 100\% \quad (2.18)$$

$$Recall = \frac{TP}{TP + FN} \times 100\% \quad (2.19)$$

dimana:

- TP adalah *True Positive*, yaitu jumlah data positif yang terklasifikasi dengan benar oleh sistem.
- TN adalah *True Negative*, yaitu jumlah data negatif yang terklasifikasi dengan benar oleh sistem.
- FN adalah *False Negative*, yaitu jumlah data negatif namun terklasifikasi salah oleh sistem.
- FP adalah *False Positive*, yaitu jumlah data positif namun terklasifikasi salah oleh sistem

Sementara itu, pada klasifikasi dengan jumlah keluaran kelas yang lebih dari dua (*multi-class*), cara menghitung akurasi, presisi dan recall dapat dilakukan dengan menghitung rata-rata dari nilai akurasi, presisi dan recall pada setiap kelas. Persamaan 2.18, 2.19, dan 2.20 merupakan formula untuk menghitung nilai akurasi, presisi dan recall dari sistem klasifikasi multi-class [16].

$$Akurasi = \frac{\sum_{i=1}^l TP_i + TN_i}{\sum_{i=1}^l TP_i + TN_i + FP_i + FN_i} \times 100\% \quad (2.20)$$

$$Presisi = \frac{\sum_{i=1}^l TP_i}{\sum_{i=1}^l (TP_i + FP_i)} \times 100\% \quad (2.21)$$

$$Recall = \frac{\sum_{i=1}^l TP_i}{\sum_{i=1}^l (TP_i + FN_i)} \times 100\% \quad (2.22)$$

dimana:

- TP_i adalah *True Positive*, yaitu jumlah data positif yang terklasifikasi dengan benar oleh sistem untuk kelas ke-i.
- TN_i adalah *True Negative*, yaitu jumlah data negatif yang terklasifikasi dengan benar oleh sistem untuk kelas ke-i.
- FN_i adalah *False Negative*, yaitu jumlah data negatif namun terklasifikasi salah oleh sistem untuk kelas ke-i.
- FP_i adalah *False Positive*, yaitu jumlah data positif namun terklasifikasi salah oleh sistem untuk kelas ke-i.

BAB III

METODE PENELITIAN

Dalam bab ini diuraikan langkah-langkah sistematis yang dilakukan dalam proses pengerjaan Tugas Akhir. Metode penelitian terdiri atas empat tahap, antara lain:

3.1 Pengumpulan Data

Pada tahap ini dilakukan pengumpulan data berita yang diperoleh dari siaran TV9. Data yang digunakan merupakan data harian berita yang disiarkan pada periode Juni – Desember 2016.

3.2 Pre-Processing

Pada tahap ini dilakukan pemrosesan teks dengan beberapa langkah, mulai dari *case folding*, *filtering*, *stemming*, *lemmatitation*, membangun vektor untuk prediksi.

3.2.1 Case Folding

Proses menghilangkan tanda baca kata seperti titik(.), koma(,), kurung buka “(”, kurung tutup “)”, *slash* “/”, *Back slash* “\”.

3.2.2 Filtering

Proses pemilihan kata atau penyaringan kata. Sehingga dipilih lah kata yang memiliki makna dalam kalimat tertentu.

3.2.3 Stemming

Proses pemotongan akhir atau awalan kata. Kata lebih disederhanakan lagi. Seperti kata “pertanggungjawaban”, kata “per” diawal dan “an” diakhir dihapuskan, jadilah kata tanggung jawab.

3.2.4 Lemmatitation

Proses pembentukan kata dasar, sehingga kata yang memiliki kesamaan makna akan dapat dijadikan satu kata dasar yang sama. Seperti kata “perubahan” dan kata “berubah” sama

sama memiliki kata dasar “ubah” dan memiliki makna yang sama.

3.3 Proses Ekstraksi Kata

Proses ekstraksi kata ini merupakan pemotongan kata pada kalimat teks berita berdasarkan spasi. Setelah kata didapatkan dihitung nilai kata di setiap kata berdasarkan kemunculannya. Kemudian dihitung kembali nilai kata pada setiap kelas dengan rumus *likelihood* pada persamaan 4.1 sehingga kata hasil ekstraksi ini yang menjadi acuan dalam proses klasifikasi nantinya.

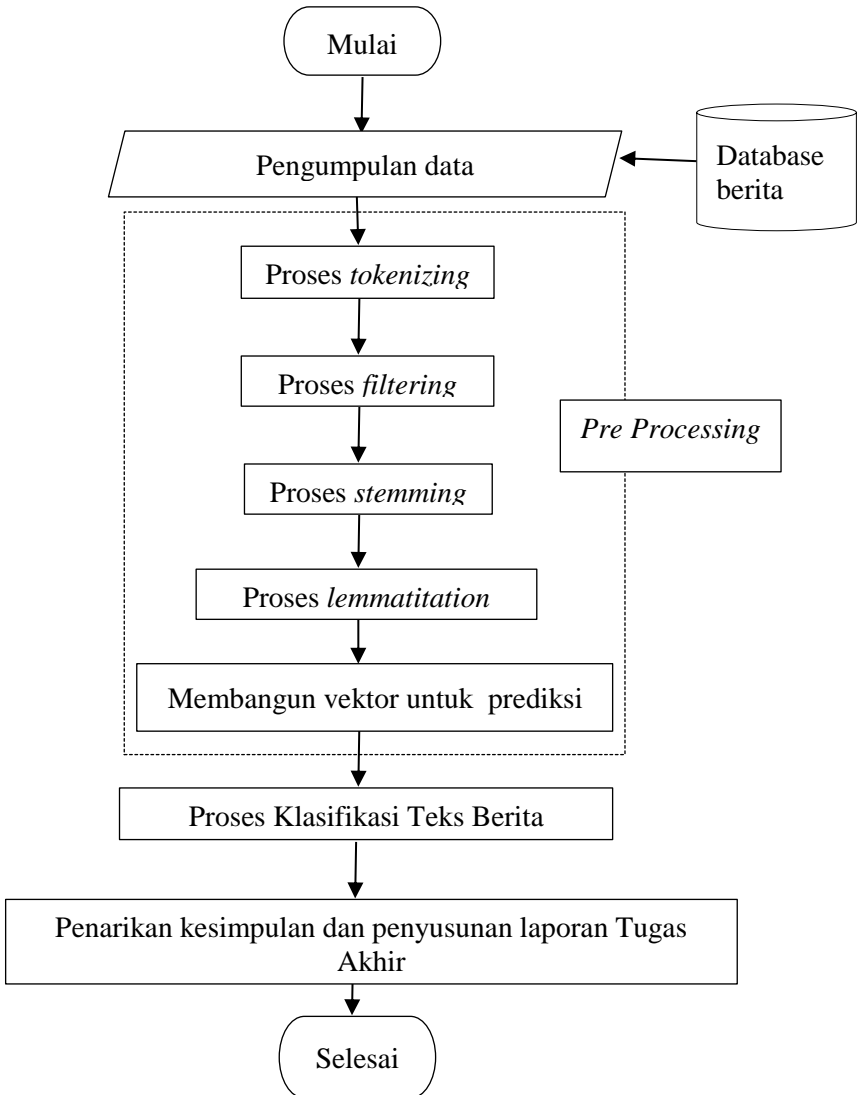
3.4 Proses Klasifikasi Teks berita

Proses pengklasifikasian berita dilakukan dengan *text mining* menggunakan metode *naïve bayes classifier* dari hasil *pre-processing* yang sudah didapat. Kemudian dihasilkan klasifikasi berita kedalam beberapa kategori diantaranya agama, ekonomi, politik, dan sosial. Kemudian dilakukan analisis dengan aplikasi JAVA Neatbeans dalam membuat program klasifikasi yang tepat dan otomatis.

3.5 Penarikan Kesimpulan dan Penyusunan Laporan

Dalam tahap akhir penelitian ini dilakukan penarikan kesimpulan dan penyusunan laporan Tugas Akhir dari hasil analisis dan pembahasan yang telah dilakukan mengenai pengklasifikasian berita dengan *text mining* metode *naïve bays classifier*.

Untuk mempermudah memahami langkah-langkah pengerjaan dalam tugas akhir ini, perhatikan diagram alir pada ditunjukkan pada Gambar 3.1:



Gambar 3. 1 Diagram Alir Metodologi Penelitian

BAB IV PERANCANGAN DAN IMPLEMENTASI SISTEM

Bab ini menjelaskan tentang perancangan data, perancangan diagram alir sistem, dan implementasi sistem. Selain itu juga akan dijelaskan mengenai tampilan antarmuka program.

4.1 Perancangan Data

Data-data yang digunakan dalam program ini dapat dibedakan menjadi tiga jenis, yaitu data masukan, data proses dan data keluaran.

4.1.1 Data Masukan

Data masukan yang digunakan adalah data yang disimpan dalam tabel dari suatu basis data, dalam hal ini peneliti menggunakan Mysql. Data masukan merupakan data berita TV9 yang berjumlah 935 teks berita dengan 500 berita sebagai data *training* dan 435 berita sebagai data *testing*. Untuk data *training* dilakukan klasifikasi manual oleh penulis dan kemudian divalidasi oleh pihak TV9 agar lebih akurat. Data *training* sebanyak 500 teks berita terdiri dari 186 kelas agama; 90 kelas ekonomi; 84 kelas politik; dan 140 kelas sosial, sedang data *testing* sebanyak 435 masih belum diklasifikasi agar nanti program yang akan menghasilkan klasifikasi teks berita kedalam 4 kelas yang telah ditentukan. Untuk setiap teks berita memiliki jumlah kata yang berbeda, ada teks berita yang hanya berisi 50 kata dan adapula teks berita yang berisi sampai 250 kata. Semua data yang dibuat disimpan dalam database MySQL, untuk menghubungkan data pada MySQL dengan bahasa pemrograman Java penulis menggunakan bantuan *library* `mysql-connector-java-5.1.45-bin.jar` yang disediakan *w3school open sources*.

Data masukan ini yang kemudian diolah oleh program melalui tahap-tahap tertentu sehingga menghasilkan keluaran yang diinginkan. Data masukan yang digunakan adalah:

1. Data berita TV9 berupa teks berita yang telah disimpan dalam *database*.
2. Data kumpulan *stoplist* yang telah tersimpan dalam *database*.

Database berita memiliki dua atribut yaitu:

1. Id merupakan nomor berita
2. Teks merupakan konten berita

4.1.2 Data Proses

Data Proses adalah data yang digunakan oleh sistem selama proses berlangsung. Proses yang dimaksud adalah *preprocessing*, *extraction*, *training* dan *testing*. Data proses yang digunakan dalam penelitian ini adalah konten dari setiap teks berita dan 1 kelas (target). Teks berita diproses melalui proses awal yaitu *preprocessing* yang melakukan beberapa tahap diantaranya *case folding* (mengubah huruf besar menjadi huruf kecil), *filtering* (menyaring kata yang tidak penting atau tidak memiliki makna), *stemming* (mengubah kata menjadi kata dasarnya dengan menghilangkan awalan dan akhiran, serta berdasarkan *database stoplist*, dan *tokenizing* (memisahkan kata berdasarkan spasi).

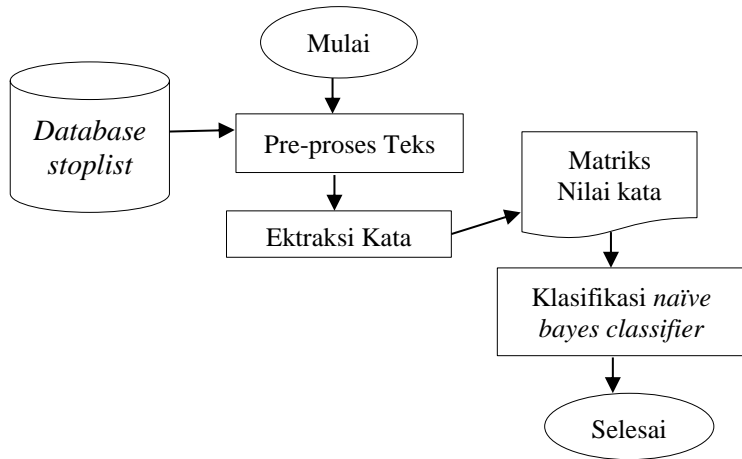
Data kelas (target) dibagi menjadi 4 kelas sebagai berikut:

- a. Kelas agama
- b. Kelas ekonomi
- c. Kelas politik
- d. Kelas sosial

4.1.3 Data Keluaran

Data keluaran merupakan data yang dihasilkan oleh program setelah proses-proses tertentu selesai dilakukan. Terdapat 2 bagian data keluaran, yang pertama data yang dihasilkan oleh proses *training* dan kemudian *testing* yang menghasilkan *output* berupa kelas (target) berita setelah dimasukkan satu *sample* konten teks berita. Sedangkan yang kedua adalah data *testing* yang telah disimpan pada database yang akan dilakukan pengklasifikasian berita kedalam 4 kelas dan didapatkan nilai presisi, recall, dan akurasi.

4.2 Perancangan Diagram Alir Sistem



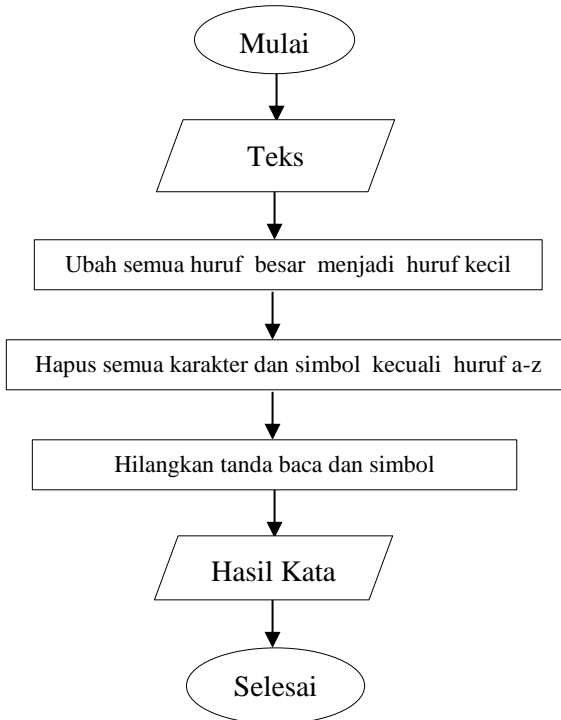
Gambar 4. 1 Diagram Alir Sistem

Diagram alir pada Gambar 4.1 menjelaskan tentang tahapan-tahapan yang akan dilakukan untuk mencapai tujuan tugas akhir. Dalam tugas akhir ini *tools* untuk komputasi menggunakan bahasa pemrograman java dan *software* basis data MySQL.

Tahapan pre-proses data terdiri dari *case folding*, *filtering*, *stemming*, *tokenizing*. Tahapan selanjutnya yaitu ekstraksi kata untuk mengelompokkan berita kedalam beberapa kelas (target) yang sudah ditentukan sebelumnya. Kata dihitung nilainya dengan menggunakan rumus *likelihood* agar tidak terdapat kata yang bernilai nol, dan kata dihitung nilainya untuk setiap kelas. Kelas yang memiliki nilai tertinggi maka kata tertentu masuk di kelas tersebut, dan secara tidak langsung kata yang memiliki nilai tertinggi di suatu kelas maka kata itu merupakan kriteria dari kelas tersebut. Setelah itu, dilakukan proses klasifikasi menggunakan metode *naïve bayes classifier*(NBC) pada tahap klasifikasi.

4.2.1 Pre-proses Teks

4.2.1.1 *Case Folding* dan *Filtering*



Gambar 4.24 Diagram Alir *Case Folding* dan *Filtering*

Pada tahap ini akan dilakukan proses menghilangkan karakter selain huruf, menghilangkan tanda dan simbol serta mengubah huruf besar menjadi huruf kecil dan juga menghapus kata yang tidak memiliki makna. Berikut contoh hasil data *case folding* dan *filtering*:

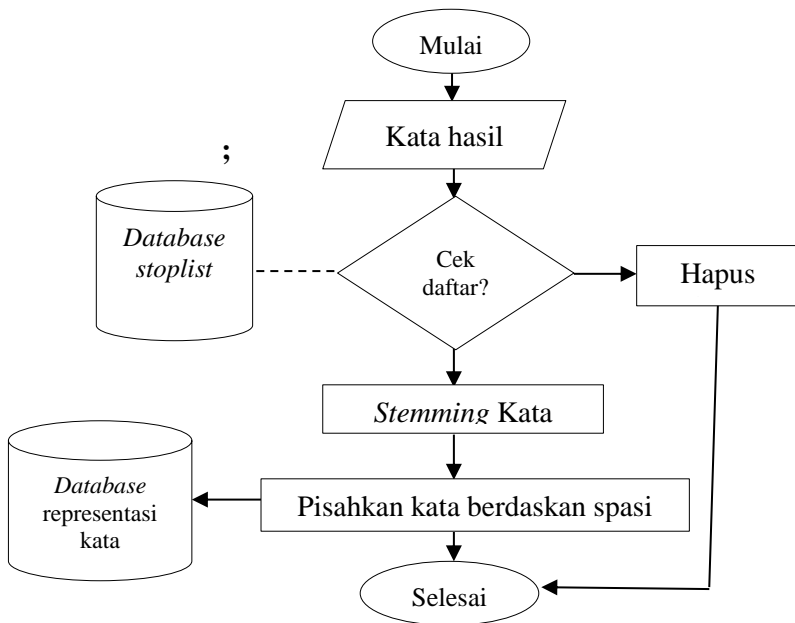
MENYAMBUT HARI JADI PROVINSI JAWA TIMUR KE-71 SANGGAR MERAH PUTIH KEMBALI MENGGELAR PASAR SENI LUKIS SELAMA 10 HARI YANG DI MULAI 10 HINGGA 16 OKTOBER MENDATANG DI JATIM EXPO INTERNASIONAL SURABAYA YANG RESMI DIBUKA OLEH GUBERNUR JAWA TIMUR SOEKARWO

Gambar 4. 40 Gambaran Potongan Teks Berita

menyambut hari provinsi jawa timur sanggar merah putih menggelar pasar seni lukis selama hari mulai hingga oktober mendatang jatim expo internasional surabaya resmi dibuka gubernur jawa timur soekarno

Gambar 4. 49 Contoh Hasil *Casefolding* dan *filtering*

4.2.1.2 *Stemming* dan *Tokenizing*



Gambar 4.72 Diagram Alir *Stemming* dan *Tokenizing*

Setelah data melalui proses *case folding* dan *filtering* selanjutnya data akan di-*stemming* dan di-*tokenizing*. *Stemming* merupakan perubahan kata menjadi kata dasar, dan *tokenizing* memisahkan kata berdasarkan spasi. Berikut contoh hasil data yang melalui proses *stemming* dan *tokenizing* :

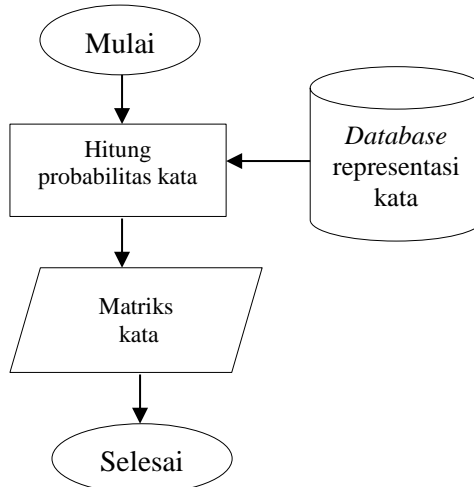
menyambut hari provinsi jawa timur sanggar merah putih menggelar pasar seni lukis selama hari mulai hingga oktober mendatang jatim expo internasional surabaya resmi dibuka gubernur jawa timur soekarwo

Gambar 4. 73 Sebelum Melalui *Stemming*

sambut hari jadi provinsi jawa timur sanggar merah putih gelar pasar seni lukis hari hingga oktober datang jatim expo internasional surabaya resmi gubernur jawa timur soekarwo

Gambar 4. 94 Setelah Melalui *Stemming* dan *Tokenizing*

4.2.2 Ekstraksi Kata



Gambar 4.117 Diagram Alir Ekstraksi Kata

Tahap *pre-processing* teks pada program bahasa Java yang dibuat penulis dilakukan dengan satu tombol preprocessing. Tabel 4.1 berisi 5 berita data training beserta hasil pre-processingnya pada bagian kolom normalisasi.

Tabel 4. 1 Hasil Preprocessing Data *Training*

ID_Berita	Konten Berita	normalisasi	Kelas
487	<p>INTRUKSI PENGURUS BESAR NAHDLATUL ULAMA TERKAIT SATU MILIAR SHOLAWAT NARIYAH DALAM RANGKA HARI SANTRI NUSANTARA INI SUDAH DI LAKSANAKAN OLEH SEJUMLAH ULAMA DI JAWA TIMUR KHUSUSNYA DI SIDOARJO PERSIAPAN SUDAH DILAKUKAN OLEH PANITIA PENYELENGGARA SHOWALAT NARIYAH SATU MILIAR YANG AKAN DI GELAR DI ALUN-ALUN SIDOARJO PADA MALAM NANTI RENCANANYA YANG AKAN HADIR DI SATU MILIAR SHOLAWAT NARIYAH INI AKAN DI HADIRI RIBUAN WARGA NAHDLIYIN DIANTARANYA MULAIPCNU BANOM NUPELAJARMAHASISWASANTRI DAN LEMBAGA NU SELAIN ITUAKAN AHADIR FORUM KOMUNIKASI DAERAH KABUPATEN SIDOARJO DIANTARANYA BUPATI WAKIL BUPATI SIDOARJODPRDPOLRITNI DAN SELURUH ORGANISASI KEPEMUDAAN DI SELURUH KABUPATEN SIDOARJO DALAM PELAKSANAAN SATU MILIAR SHOLAWAT NARIYAH NANTI MALAM INI JUGA AKAN DI AMANKAN RATUSAN ANGGOTA BANSER NUGP ANSOR MENYEBARKAN ANGGOATANYA DI SELURUH SUDUT ALUN-ALUN SIDOARJO UNTUK MENGAMANKAN JAM'YIAH SHOLAWAT NARIYAH MENURUT FATKUL ANAM SELAKU KETUA PANITIA SHOLAWAT NARIYAH BAHWA PELAKSANAAN SATU MILIAR SHOLAWAT NARIYAH AKAN DI HADIRI SEJUMLAH TOKO ULAMA NU RENCANANYA KETUA PWNu JATIM KYAI MUTAWAKIL ALALLAH AKAN MEMIMPIN SHOLAWAT NARIYAH BERSAMA BUPATI DAN PCNU SIDOARJO SELAIN RANGKAIN SATU MILIAR SHOLAWAT NARIYAH INIPANITIA HARI SANTRI NASIONAL JUGA MELAKUKAN KEGIATAN SEPerti</p>	<p>intruksi urus nahdlatul ulama kait miliar sholawat nariyah rangka santri nusantara laksana ulama jawa timur sidoarjo siap panitia selenggara showalat nariyah miliar gelar alunalun sidoarjo malam rencana hadir miliar sholawat nariyah hadir ribu warga nahdliyin mulaipcnu banom nupelajarmahasiswasantri lembaga nu ituakan ahadir forum komunikasi daerah kabupaten sidoarjo bupati wakil bupati sidoarjodprdpolritni organisasi pemuda kabupaten sidoarjo laksana miliar sholawat nariyah malam aman ratus anggota banser nugg ansor sebar angoatanya sudut alunalun sidoarjo aman jamiyah sholawat nariyah fatkul anam ketua panitia sholawat nariyah laksana miliar sholawat nariyah hadir toko ulama nu rencana ketua pwnu jatim kyai mutawakil alallah pimpin sholawat nariyah bupati pcnu sidoarjo rangkain miliar sholawat nariyah inipanitia santri nasional giat pamer produk santripagelaran ketoprak santri jalan sehat pawai santri kabupaten sidoarjo</p>	Agama

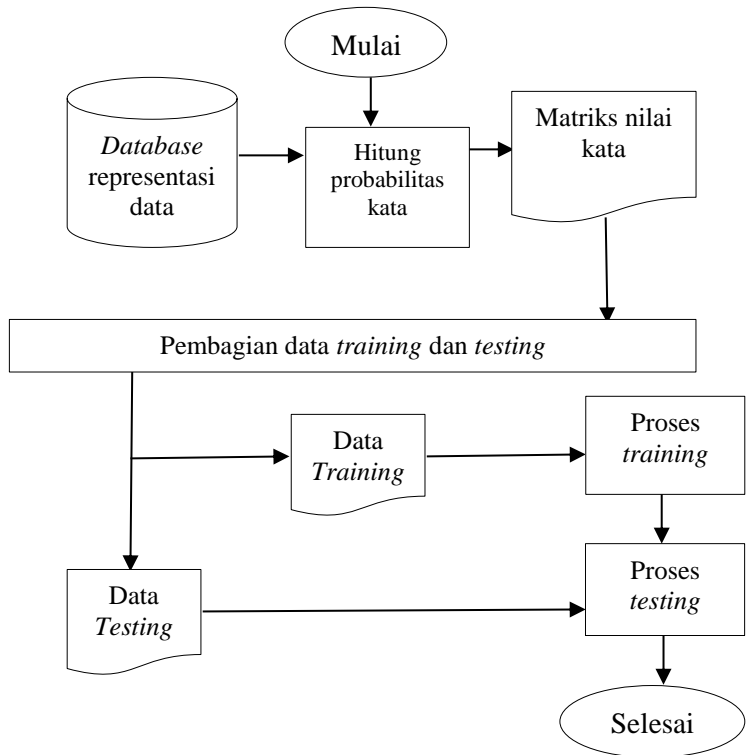
	PAMERAN PRODUK SANTRIPAGELARAN KETOPRAK SANTRI JALAN SEHAT DAN PAWAI BERSAMA SANTRI SE KABUPATEN SIDOARJO		
499	<p>APEL KHUSUS PENDANDATANGAN ANTI PUNGLI DIGELAR DI HALAMAN MAPOLRES JEMBER MENURUT KEPALA KEPOLISIAN RESORT JEMBER AJUN KOMISARIS BESAR POLISI ATAU AKBP SABILUL ALIF POLRES JEMBER SAAT INI SEDANG MENJALANKAN INTRUKSI PRESIDEN JOKO WIDODO UNTUK MENJADI CONTOH SEBAGAI INSTANSI KEPOLISIAN ANTI PUNGLI DIJAWATIMUR BAHKAN POLRES JEMBER ADALAH POLRES PERTAMA YANG MELAKUKAN PENANDTANGAN TERSEBUT KAPOLRES AKBP SABILUL ALIF MENGATAKAN HAL INI MEMANG SANGAT BERAT DILAKUKAN OLEH ANGGOTA KARENA MEREKA TIDAK BOLEH MENERIMA SUAP TERKAIT PELAYANAN APAPUN MULAI DARI S-K-C-KSIMS-T-N-K DAN SELURUH PELAYANAN PUBLIK LAINNYA NAMUN DENGAN SEMAKIN KETATNYA PELAYANAN ANTI PUNGLI TENTU AKAN BERDAMPAK POSITIF KEPADA MASYARAKAT SEKALIGUS AKAN SEMAKIN MEMBUAT ANGGOTA KEPOLISIAN DAN KARYAWAN SIPIL POLRES JEMBER LEBIH PROFESIONAL</p>	<p>apel khusus pendandatangan anti pungli gelar halaman mapolres jember kepala polisi resort jember ajun komisaris polisi akbp sabilul alif polres jember jalan intruksi presiden joko widodo contoh instansi polisi anti pungli dijawatimur polres jember polres penandatangan kapolres akbp sabilul alif berat anggota terima suap kait layanan apa skcksimstnk layanan publik ketat layanan anti pungli dampak positif masyarakat anggota polisi karyawan sipil polres jember profesional</p>	Politik
435	<p>AKTIFITAS DI KEBUN CENGKEH DESA PUSPO KECAMATAN PUSPO KABUPATEN PASURUAN TERLIHAT SIBUK PARA KULI PETIK MENGAMBIL HASIL PANEN CENGKEH DENGAN CEKATAN PANEN KALI INI HASIL YANG DIDAPAT CUKUP MEMUASKAN NAMUN TIDAK MENDUKUNG TINGGINYA OMZET YANG DIDAPAT MENURUT KETERANGAN SALAH SATU PETANI CENGKEH BIAYA YANG DIKELUARKAN UNTUK PEMUPUKAN DAN PANEN LEBIH</p>	<p>aktifitas kebun cengkeh desa puspo camat puspo kabupaten pasuruan sibuk kuli petik ambil hasil panen cengkeh cekat panen kali hasil muas dukung tinggi omzet terang salah tani cengkeh biaya keluar pupu panen musim hujan biaya sewa kuli petik sewa jemur jemur cengkeh wilayah curah hujan jemur</p>	Ekonomi

	<p>BANYAK APALAGI MUSIM HUJAN SEPERTI SAAT INI SELAIN BIAYA UNTUK MENYEWAKAN KULI PETIK JUGA DIPERLUKAN SEWA TEMPAT UNTUK PENJEMURAN KARENA TIDAK MEMUNGKINKAN UNTUK MENJEMUR CENGKEH DI WILAYAH SETEMPAT YANG CURAH HUJANNYA TINGGI SEHINGGA PENJEMURAN CENGKEH HARUS DILAKUKAN DI DAERAH LAIN SELAIN ITU HARGA JUAL DARI PETANI KE PENGEPUL MENURUN HINGGA DUA PULUH LIMA PERSEN YAKNI DELAPAN PULUH LIMA RIBU RUPIAH PER KILOGRAM</p>	<p>cengkeh daerah harga jual tani kepul turun puluh persen delapan puluh ribu rupiah kilogram</p>	
417	<p>POLEMIS PERNYATAAN GUBERNUR DKI JAKARTA BASUKI TJAHAJA PURNAMA HARUS DIJADIKAN PELAJARAN AGAR PENGHINAAN TERHADAP AGAMA ISLAM TIDAK TERJADI LAGI KETUA ALIANSI PENGASUH MUDA PONDOK PESANTREN APMPP JATIM MOCHAMAD FAWAID MENEGASKAN SANGAT TIDAK PENTING MEMBAWABAWA NAMA AGAMA KE RANAH POLITIK MAKA APMPP MEMINTA M-U-I UNTUK LEBIH TEGAS LAGI UNTUK MENYIKAPINYA AGAR TIDAK MELEBAR KE PILKADA LAIN KHUSUSNYA PILKADA JATIM YANG AKAN MELAKSANAKAN PEMILIHAN PADA 2018 MUI SEHARUSNYA MENGELUARKAN FATWA YANG TEGAS JIKA AGAMA DIGUNAKAN DALAM PILKADA SEBAGAI LEMBAGA YANG MEMBAWAI ULAMA-ULAMA M-U-I HARUS DAPAT BERSIKAP TEGAS MENYAMPAIKAN TAFSIR AL QUR'AN APAKAH BOLEH MEMILIH CALON PEMIMPIN NON MUSLIM ATAU TIDAK SEHINGGA TIDAK DIMAINKAN OLEH CALON PEMIMPIN NON MUSLIM</p>	<p>polemik nyata gubernur dki jakarta basuki tjahaja purnama jadi ajar hina agama islam ketua aliansi asuh muda pondok pesantren apmpp jatim mochamad fawaid membawabawa nama agama ranah politik apmpp mui sikap lebar pilkada pilkada jatim laksana pilih mui keluar fatwa agama pilkadasebagai lembaga bawa ulamaulama mui sikap tafsir al quran pilih calon pimpin non muslim main calon pimpin non muslim</p>	<p>politik</p>
107	<p>BEKERJA SAMA DENGAN PT. GRAND ELEPHANT BATA RINGAN SEMEN MORTAR DAN PT. SEMEN INDOGREEN SENTOSA YAYASAN HAJI MUHAMMAD CHENG HOO INDONESIA KEMBALI MENYELENGGARAKAN</p>	<p>pt grand elephant bata ringan semen mortar pt semen indogreen sentosa yayasan haji muhammad cheng hoo indonesia selenggara bagi sembako</p>	<p>Sosial</p>

	PEMBAGIAN SEMBAKO KEPADA WARGA TIDAK MAMPU DI RUMAH SUSUN TANAH MERAH SURABAYA BAKTI SOSIAL YANG RUTIN DI LAKUKAN OLEH WARGA TIONGHOA INI MERUPAKAN UPAYA MENINGKATKAN KEBERSAMAAN ANTAR UMAT BERAGAMA DAN SALING MENGHORMATI DALAM SETIAP PERBEDAAN SERTA MENJUNJUNG TINGGI NILAI-NILAI PANCASILA SEBAGAI DASAR NEGARA BANTUAN SEMBAKO YANG DIBERIKAN WARGA KETURUNAN TIONGHOA INI SANGAT BERMANFAAT DAN TEPAT SASARAN KARENA KONDISI WARGA KORBAN GUSURAN SAAT INI SEDANG MENGEMBALIKAN KONDISI PEREKONOMIAN KELUARGA PENGHUNI RUSUN TANAH MERAH SENDIRI TERDIRI DARI 111 ANAK YATIM PIATU 100 JANDA DAN 239 KELUARGA TIDAK MAMPU	warga rumah susun tanah merah surabaya bakti sosial rutin laku warga tionghoa upaya meningkatkan sama umat agama hormat beda junjung nilainilai pancasila dasar negara bantu sembako warga turun tionghoa manfaat sasaran kondisi warga korban gusur kembali kondisi ekonomi keluarga huni rusun tanah merah anak yatim piatu janda keluarga	
--	---	--	--

Setelah pre-processing telah dilakukan selanjutnya yaitu ekstraksi kata dari semua teks berita yang *ditraining*. Proses ekstraksi kata ini merupakan langkah lanjutan dari proses tokenisasi yang memisahkan kata berdasarkan kalimat, dan kata itulah yang menjadi hasil dari ekstraksi kata. Tokenisasi dilakukan menggunakan software bahasa Java Netbeans IDE 8.1 serta MySQL Navicat Premium. Kata yang dihasilkan dari ekstraksi kata disimpan dalam database term_list pada MySQL.

4.2.3 Klasifikasi menggunakan *Naïve Bayes Classifier*



Gambar 4.133 Diagram Alir Klasifikasi NBC

Sebelum memasuki metode klasifikasi, proses inputan akan dibagi menjadi 2 yaitu proses *training* dan *testing*. Dengan data 500 untuk *training* dan 435 untuk *testing*. Pembagian ini dilakukan secara acak. Dengan adanya *training* akan didapat basis kata agar suatu berita terklasifikasi pada suatu kelas. Sehingga akan mempermudah proses *testing* dalam mengklasifikasi suatu teks berita.

Setelah model prediksi telah dibangun pada data *training*, selanjutnya mengklasifikasi data yang belum diketahui label kelasnya.

Langkah-langkah pada tahap ini adalah:

1. Penentuan peluang bersyarat $P(X_i = x_i | Y = y)$ sesuai dengan hasil *training* dimana x_i sesuai dengan *input user*.
2. Perhitungan $\prod_{i=1}^d P(X_i = x_i | Y = y)$
3. Perhitungan peluang posterior $P(Y|X) = \frac{P(Y) \prod_{i=1}^d P(X_i|Y)}{P(X)}$
4. Untuk mengklasifikasi berita, ada beberapa kondisi yang harus diperhatikan.
 - a. Jika $P(\text{agama}|X) > P(\text{ekonomi}|X), P(\text{agama}|X) > P(\text{politik}|X), P(\text{agama}|X) > P(\text{sosial}|X)$, maka berita diklasifikasikan masuk kelas agama.
 - b. Jika $P(\text{ekonomi}|X) > P(\text{agama}|X), P(\text{ekonomi}|X) > P(\text{politik}|X), P(\text{ekonomi}|X) > P(\text{sosial}|X)$, maka berita akan diklasifikasikan masuk kelas ekonomi.
 - c. Jika $P(\text{politik}|X) > P(\text{agama}|X), P(\text{politik}|X) > P(\text{ekonomi}|X), P(\text{politik}|X) > P(\text{sosial}|X)$, maka berita akan diklasifikasikan masuk kelas politik.
 - d. Jika $P(\text{sosial}|X) > P(\text{agama}|X), P(\text{sosial}|X) > P(\text{ekonomi}|X), P(\text{sosial}|X) > P(\text{politik}|X)$, maka ia tergolong dalam kelas sosial.

4.3 Implementasi Sistem

Penelitian ini menggunakan bahasa pemrograman JAVA pada platform Development Kit 7 dan Netbeans IDE 8.1 dengan *library*. Dan *software* basis data yang digunakan adalah MySQL.

4.3.1 Implementasi Pre-proses Teks

Pre-proses teks merupakan proses awal dari penelitian ini. Data yang sebelumnya berisi angka, huruf kapital dan kata-kata yang tidak penting atau tidak memiliki makna akan diubah menjadi data yang hanya terdiri dari huruf bukan kapital dan diambil kata-kata yang penting saja.

Proses penyeragaman huruf menjadi huruf bukan kapital, penghilangan tanda baca dan angka dilakukan pada proses *case folding*. Setelah proses *case folding*, dilakukan penyaringan atau penghapusan tanda baca dan simbol simbol serta kata yang tidak memiliki makna disebut proses *filtering*.

Setelah didapat data yang sudah di *case folding* dan *filtering*. Selanjutnya akan dilakukan *stemming* dan *tokenizing*. Apabila kata yang memiliki awalan dan akhiran akan diubah menjadi kata dasarnya yang dinamakan proses *stemming*, dan setelah ketiga proses dilakukan akan dilakukan *tokenizing* yaitu pemisahan kata berdasarkan spasi. Jika kata terdapat dalam *stoplist* maka kata tersebut akan dihapus. Kemudian kata yang sudah didapat akan disimpan kedalam *database*. *Stemming* yang dibuat oleh penulis menggunakan bantuan *library* lib jsastrawi-all-0.1.jar yang sudah disediakan github *open sources*. Begitupun untuk tekonesasi juga menggunakan bantuan *library* lucene-analyzers-common-4.2.1-sources.jar yang sudah disediakan w3school *open sources*. Berikut adalah implementasi dari *preprocessing* :

```
// implementasi pre-processing teks
private void
PreprocessingActionPerformed(java.awt.event.Action
Event evt) {
    database koneksi = new database();
    ResultSet rset;
    ResultSet rset2;
    try {
        preprocessing pr = new
preprocessing();
        koneksi.connectFirst();
        rset = koneksi.executeSelect("SELECT
no, isi FROM `databerita_ain`");
        while (rset.next()) {
            String no = rset.getString("no");
            System.out.print("precessing id ");
            System.out.println(no);
            System.out.println("steaming");
            String kalimat =
rset.getString("isi").replaceAll("[^a-z A-Z]",
"");
            kalimat = pr.process(kalimat);
            System.out.println("steaming2");
```

```

koneksi.executeUpdate("UPDATE `databerita_ain` SET
`normalisasi` = '" + kalimat.replace(" ", " ") + "'
WHERE `databerita_ain`.`no` = " + no + "");
    }
    koneksi.destroyConnection();
    } catch (IOException | SQLException ex) {
Logger.getLogger(GUI.class.getName()).log(Level.SEVERE,
null, ex);
    }
    UpdateTabel();
}

```

4.3.2 Implementasi Ekstraksi Kata

Kata-kata penting yang telah disimpan dalam *database* akan dilakukan pembobotan dengan menghitung nilai probabilitas kemunculan kata pada tiap data. Tahap ini dimulai dengan membentuk matriks kemunculan setiap kata penting pada tiap data dengan menggunakan rumus teknik *Laplacian Smoothing*. Berikut merupakan implementasi dari ekstraksi kata:

```

// implementasi ekstraksi kata
private void
fiturExtractionActionPerformed(java.awt.event.ActionEvent evt) {
    database koneksi = new database();
    ResultSet rset;
    try {
        koneksi.connectFirst();
        koneksi.executeUpdate("DELETE FROM
term_list");
        rset = koneksi.executeQuery("SELECT normalisasi
FROM `databerita_ain`");
        int cek = 0;
        while (rset.next()) {
            String kalimat =
rset.getString("normalisasi");
            for (String kata : kalimat.split(" ")) {
                if (!kalimat.equals(" ")) {
                    koneksi.executeUpdate("insert into term_list
(term,agama,ekonomi,politik,sosial) values ('" +
kata + "', 0, 0, 0, 0);");
                }
            }
        }
    }
}

```



```

        System.out.println("check doc num " + (++cek));
    }
    koneksi.destroyConnection();
} catch (SQLException ex) {

Logger.getLogger(GUI.class.getName()).log(Level.SEVERE, null, ex);
    }
    UpdateTabel1();
}

```

Setelah ekstraksi kata dihasilkan, terdapat beberapa nilai kemunculan kata yang nol sehingga tidak dapat diklasifikasikan menggunakan *Naïve Bayes Classifier*. Oleh karena itu, penulis menggunakan metode Teknik *Laplacian Smoothing* untuk mengatasi permasalahan tersebut.

Teknik *laplacian smoothing* digunakan untuk mengatasi nilai kata pada suatu kelas yang bernilai 0 sehingga kondisi pada *Naïve Bayes Classifier* dapat bernilai 0. Cara yang digunakan pada teknik ini adalah dengan menambahkan angka 1 pada perhitungan kata untuk setiap kelas sehingga didapatkan nilai yang tidak lagi 0 melainkan nilai kata menjadi desimal antara 0 sampai 1. Persamaan di bawah ini menunjukkan perhitungan nilai kata dengan Teknik *Laplacian Smoothing*[14].

$$P(X_i|Y) = \frac{1 + n(X_i, Y)}{|W| + n(Y)} \quad (4.1)$$

dimana,

$n(X_i, Y)$ = Jumlah *term* X_i yang ditemukan di seluruh data latih dengan kategori Y .

$n(Y)$ = Jumlah *term* di seluruh data latih dengan kategori Y .

$|W|$ = Jumlah seluruh *term* X_i dari seluruh data latih

Untuk mendapatkan nilai dengan perhitungan persamaan 4.1 berikut implementasi perhitungan probabilitas kata :

```

// implementasi hitung nilai kata
private void
setNilaiKataActionPerformed(java.awt.event.ActionEvent
evt) {

    database koneksi = new database();
    ResultSet rset;
    ResultSet rset2;
    try {
        preprocessing pr = new preprocessing();
        koneksi.connectFirst();
        rset = koneksi.executeSelect("SELECT term
FROM `term_list`");
        rset2 = koneksi.executeSelect("SELECT
normalisasi, class FROM `databerita_ain`");
        while (rset.next()) {
            String term = rset.getString("term");
            int[] jml = {0, 0, 0, 0};
            rset2.beforeFirst();
            while (rset2.next()) {
                String kalimat =
rset2.getString("normalisasi");
                String c = rset2.getString("class");
                if (kalimat.contains(term)) {
                    if (c.equals("agama")) { jml[0]++;
                    } else if (c.equals("ekonomi"))
{
                        jml[1]++;
                    } else if (c.equals("sosial")) {
                        jml[2]++;
                    } else if (c.equals("politik"))
{
                        jml[3]++;
                    }
                }
            }
            int total = 0;
            for (int i = 0; i < jml.length; i++) {
                total += jml[i];
            }
            String upd = "UPDATE term_list SET agama=" + jml[0] +
", ekonomi=" + jml[1] + ", sosial=" + jml[2] + ",
politik=" + jml[3] + ", total=" + total + " where term
='" + term + "';";

```

```

koneksi.executeInsert (upd);
System.out.println (term);
    }
    koneksi.destroyConnection ();
} catch (SQLException ex) {

Logger.getLogger (GUI.class.getName ()) .log (Level.SEVERE, null, ex);
    }
    UpdateTabell ();
}

```

4.3.3 Implementasi Proses Klasifikasi Menggunakan *Naïve Bayes Classifier*

Berikut merupakan implementasi dari proses klasifikasi. Data hasil pembobotan pada ekstraksi kata akan diinputkan kedalam data masukkan *Naïve Bayes Classifier*. Setelah itu akan melalui proses *training*. Berikut merupakan implementasi proses *training* :

```

// implementasi proses training
private void
TrainingActionPerformed (java.awt.event.ActionEvent
evt) {

    database koneksi = new database ();
    ResultSet rset, jumTerm;
    ResultSet rset2;
    try {
        preprocessing pr = new preprocessing ();
        koneksi.connectFirst ();
        rset = koneksi.executeSelect ("SELECT *
FROM `term_list` ");
        rset2 = koneksi.executeSelect ("SELECT
normalisasi, class FROM `databerita_ain`");
        //jumTerm =
koneksi.executeSelect ("SELECT count (term) as x FROM
`term_list`");
        String [] namaClass = {"agama",
"ekonomi", "politik", "sosial"};
        int [] ny = {0, 0, 0, 0};
        for (int i = 0; i < namaClass.length;
i++) {

```

```

String sql = "SELECT COUNT(*) AS con FROM
`term_list` WHERE "
                + namaClass[i] + "!=0";
        jumTerm =
koneksi.executeSelect(sql);
        jumTerm.next();
        ny[i] = jumTerm.getInt("con");
    }
while (rset.next()) {
    String term =
rset.getString("term");
    for (int i = 0; i < namaClass.length; i++) {
        int nxy =
rset.getInt(namaClass[i]);
        int w = rset.getInt("total");
        System.out.println(ny);
        double p = (double) (1 + nxy) /
(double) (w + ny[i]);
        System.out.println(namaClass[i]
+ "-" + term + " = " + p);
        String upd = "UPDATE term_list
SET " + namaClass[i] + "=" + p + " where term =' " +
term + "';";
        koneksi.executeInsert(upd);
    }
    System.out.println("-----
");
    }
    koneksi.destroyConnection();
} catch (SQLException ex) {

Logger.getLogger(GUI.class.getName()).log(Level.SEVERE, null, ex);
}
}

```

Disamping algoritma diatas, untuk proses *training* penulis membuat sebuah kelas yang diberi nama *naïveBayes.java*. Berikut implementasi sistem pada kelas *naïveBayes.java* :

```

// proses perhitungan Naïve Bayes Classifier
public class naïveBayes {
    public String process(String kalimat /* sudah
normal */, String[] namaClass, int[] jmlDataClass)
throws SQLException{
        int jmlData = 0;
        float max = -1;
        String result = null;
        for(int jml: jmlDataClass){
            jmlData += jml;
        }
        for(int i = 0; i < namaClass.length; i++){
            float globalPTC = 1;
            for(String term : kalimat.split(" ")){
                globalPTC *=
searchPTC(term.replace("\\'", "'\\'"),
namaClass[i]);
            }
            float PC = (float) 0.25;
            float PCT = PC * globalPTC;
            if(PCT > max){
                max = PCT;
                result = namaClass[i];
            }
        }
        return result;
    }
    public float searchPTC /* P(T / C) =>
probability term in class */
(String term, String classType) throws
SQLException{
        float ptc = 1;
        database koneksi = new database();
        ResultSet rset;
        koneksi.connectFirst();
        rset = koneksi.executeSelect("SELECT * FROM
`term_list` where term = '"+term+"'");
        if(rset.next()){
            ptc = rset.getFloat(classType)*1000;
        }
        else ptc = 1;
        koneksi.destroyConnection();
        return ptc;
    }
}

```

Keluaran dari proses *training* adalah model prediksi. Selanjutnya akan dilakukan proses *testing* untuk menguji model prediksi tersebut. Berikut merupakan implementasi proses *testing* terhadap satu berita :

```
// testing berita yang dimasukkan
private void
TestingActionPerformed(java.awt.event.ActionEvent evt)
{
    String kalimat = InputBerita.getText();
    preprocessing pr = new preprocessing();
    naiveBayes nb = new naiveBayes();
    try {
        String hasil =
nb.process(pr.process(kalimat));
        paneHasil.setText(hasil);
    }

catch (SQLException ex) {
Logger.getLogger(GUI.class.getName()).log(Level.SEVERE
, null, ex);
    } catch (IOException ex) {
Logger.getLogger(GUI.class.getName()).log(Level.SEVERE
, null, ex);
    }
}
```

4.3.4 Implementasi Metode Evaluasi

Pada tugas akhir ini penulis menggunakan evaluasi *confusion matrix* untuk menguji data apakah sudah baik atau belum dengan melihat nilai akurasi. Pada tahap ini merupakan evaluasi keseluruhan data berita apakah hasil prediksi yang dilakukan program sudah benar atau belum dengan kelas. Berikut merupakan implementasi metode evaluasi *confusion matrix* :

```
// implementasi evaluasi
private void
ujiDataActionPerformed(java.awt.event.ActionEvent evt)
{
    ResultSet rset;
    database koneksi = new database();
    preprocessing pr = new preprocessing();
    naiveBayes nb = new naiveBayes();
```

```

String[] namaClass = {"agama", "ekonomi",
"politik", "sosial"};
    int[] TP = {0, 0, 0, 0};
    int[] TN = {0, 0, 0, 0};
    int[] FP = {0, 0, 0, 0};
    int[] FN = {0, 0, 0, 0};

float jmlTP, jmlTN, jmlFP, jmlFN;
float precision;
float recall;
float accuracy;

try {
        koneksi.connectFirst();
        rset = koneksi.executeSelect("SELECT isi,
class FROM `testing`");
while (rset.next()) {
        String kalimat =
rset.getString("isi").replaceAll("[^a-z A-Z]", "");
        kalimat = pr.process(kalimat);
        String hasil =
nb.process(pr.process(kalimat));

System.out.print(rset.getString("class") + " -> ");
        System.out.println(hasil);
        for (int i = 0; i < 4; i++) {
                if
(rset.getString("class").equalsIgnoreCase(namaClass[i]
) && hasil.equalsIgnoreCase(namaClass[i])) {
                        TP[i]++;
                } else if
(rset.getString("class").equalsIgnoreCase(namaClass[i]
) && !hasil.equalsIgnoreCase(namaClass[i])) {
                        FN[i]++;
                } else if
(!rset.getString("class").equalsIgnoreCase(namaClass[i]
) && hasil.equalsIgnoreCase(namaClass[i])) {
                        FP[i]++;
                } else {
                        TN[i]++;
                } } }
        jmlTP = TP[0] + TP[1] + TP[2] + TP[3];
        jmlTN = TN[0] + TN[1] + TN[2] + TN[3];
        jmlFP = FP[0] + FP[1] + FP[2] + FP[3];
        jmlFN = FN[0] + FN[1] + FN[2] + FN[3];

```

```

koneksi.executeInsert("INSERT INTO `data_uji`
VALUES (" + jmlTP + ", " + jmlTN + ", " + jmlFP
+ ", " + jmlFN + ")");

        precision = jmlTP / (jmlTP + jmlFP);
        recall = jmlTP / (jmlTP + jmlFN);
        accuracy = ((jmlTP + jmlTN) / (jmlTP
+ jmlTN + jmlFP + jmlFN))*100;

System.out.println(precision);
        System.out.println(recall);
        System.out.println((accuracy)+"%");

paneAkurasi.setText(String.valueOf(accuracy)+"%");
);

panePresisi.setText(String.valueOf(precision));

paneRecall.setText(String.valueOf(recall));
        koneksi.destroyConnection();
    } catch (SQLException ex) {

Logger.getLogger(GUI.class.getName()).log(Level.
SEVERE, null, ex);
        System.out.println(ex);
    } catch (IOException ex) {

Logger.getLogger(GUI.class.getName()).log(Level.
SEVERE, null, ex);
        System.out.println(ex);
    }
    UpdateTabel2();
}

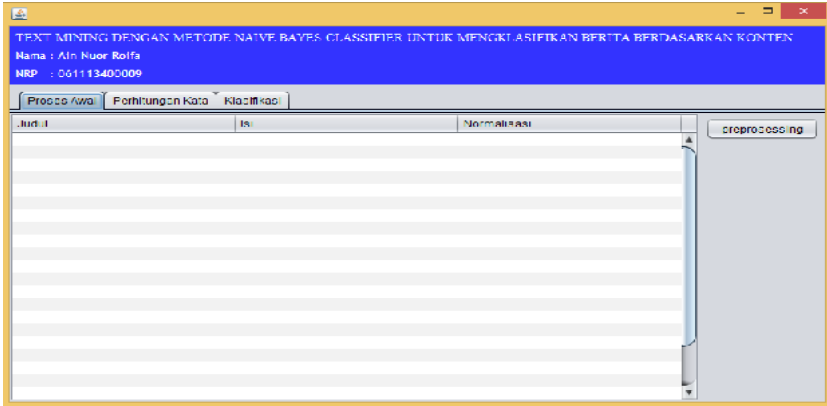
```

4.4 Implementasi Antarmuka GUI

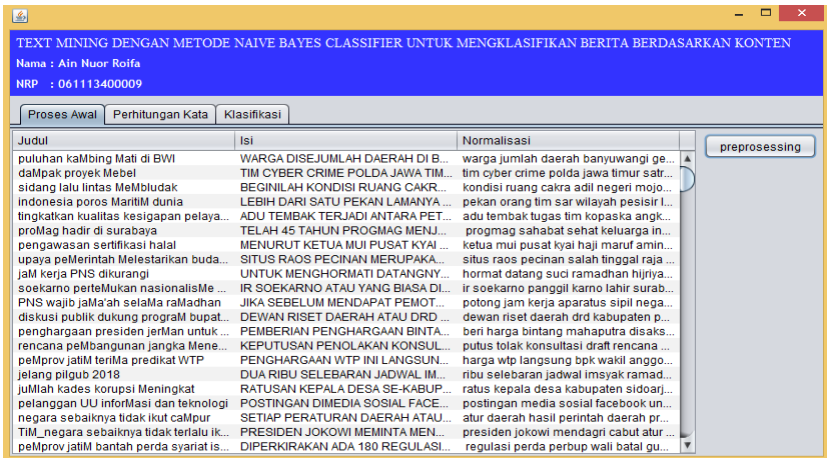
Setelah perancangan selesai, tahap selanjutnya adalah implementasi. Tahap ini bertujuan agar user dapat menggunakan program yang telah dirancang. Antarmuka program ini dibagi menjadi 3 tab, yaitu tab “Proses Awal”, “Perhitungan Kata” dan

“Klasifikasi”. Pada bagian proses awal penulis menampilkan berita TV pada periode Juni-Desember 2016.

Langkah awal adalah melakukan pre-proses teks dengan tombol preprocessing. Kemudian muncul data yang sudah dilakukan *pre-processing*.

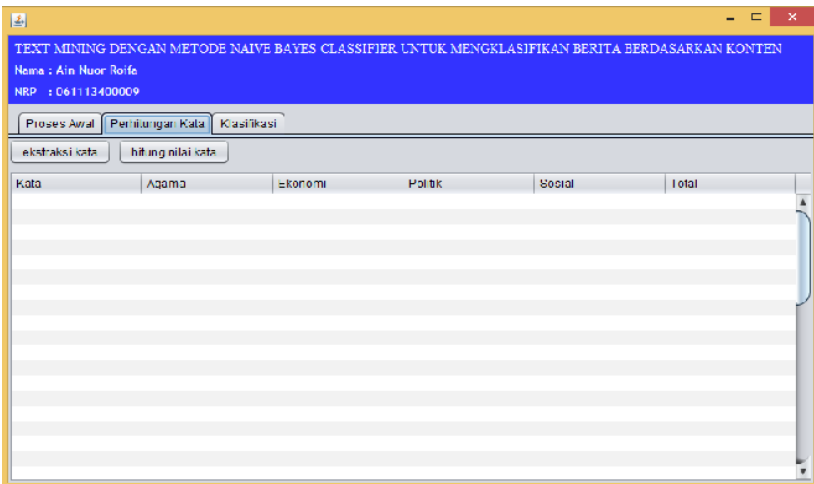


Gambar 4. 10 Tampilan GUI tab Proses Awal

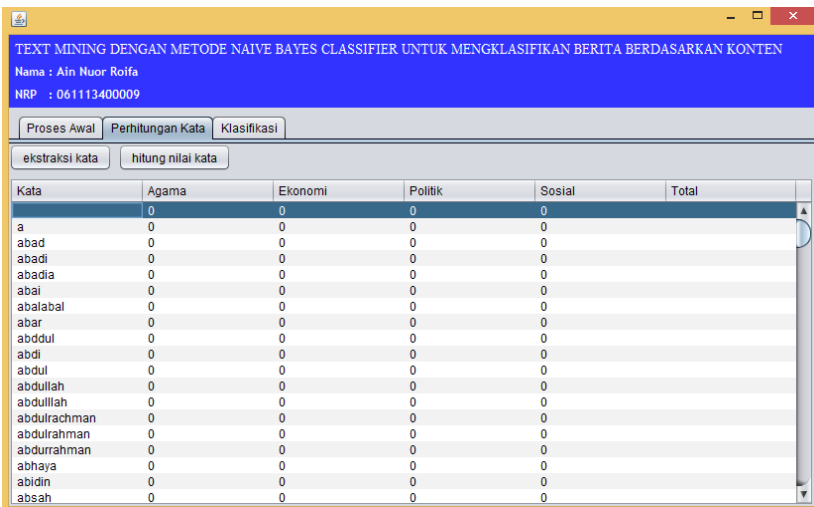


Gambar 4.11 Hasil Pre-processing

Kemudian pada tab selanjutnya yaitu perhitungan kata.



Gambar 4.12 Tampilan GUI tab Perhitungan Kata



Gambar 4.13 Hasil Ekstraksi Kata

term	ptc	agama	ekonomi	politik	sosial	total	
abad	(Null)	0,00103013	0,000753864	0,000878735	0,00121448	10	
abadi	(Null)	0,000515996	0,000755858	0,000293513	0,000486618	3	
abadia	(Null)	0,000258131	0,00075643	0,000293686	0,000243427	1	
abai	(Null)	0,00103093	0,00113208	0,000596338	0,000486145	7	
abalabal	(Null)	0,000258131	0,000378215	0,000293686	0,000486855	1	
abar	(Null)	0,00103119	0,000377501	0,000879765	0,000486263	6	
abbas	(Null)	0,000516262	0,000378215	0,000293686	0,000243427	1	
abd	(Null)	0,00461184	0,00112233	0,00232964	0,00120861	30	
abddul	(Null)	0,000516262	0,000378215	0,000293686	0,000243427	1	
abdi	(Null)	0,00103146	0,000377644	0,000586682	0,000486381	5	
abdul	(Null)	0,00179995	0,00112824	0,00175439	0,000970167	16	
abdullah	(Null)	0,000516262	0,000378215	0,000293686	0,000243427	1	
abdullah	(Null)	0,000516262	0,000378215	0,000293686	0,000243427	1	
abdulrehman	(Null)	0,000258131	0,000378215	0,000293686	0,000486855	1	
abdulrehman	(Null)	0,000258131	0,000378215	0,000293686	0,000486855	1	
abdur	(Null)	0,00128932	0,000377644	0,000586682	0,000243191	5	
abdurrahman	(Null)	0,00128966	0,000377786	0,000293427	0,00024325	4	
abdusshomad	(Null)	0,000516262	0,000378215	0,000293686	0,000243427	1	
abhaya	(Null)	0,000258131	0,000378215	0,000293686	0,000486855	1	
abidin	(Null)	0,000258131	0,00075643	0,000293686	0,000243427	1	
absah	(Null)	0,000774194	0,000378072	0,0002936	0,000243368	2	
absen	(Null)	0,000258065	0,000756144	0,0002936	0,000486736	2	
absensi	(Null)	0,000258131	0,00075643	0,000293686	0,000243427	1	

Gambar 4. 14 Hasil Hitung Probabilitas Kata

Kemudian setelah nilai probabilitas kata didapatkan dan disimpan dalam matriks atau array [] selanjutnya yaitu tahap klasifikasi dengan proses *training* (memperlakukan data latih sebagai proses klasifikasi *Naive Bayes Classifier*). Baru kemudian coba memasukkan satu berita pada kolom input yang terdapat di tab klasifikasi dan ditombol *testing* maka akan keluar hasil klasifikasi (prediksi) yang dilakukan oleh program berdasarkan data latih yang telah di proses sebelumnya.

TEXT MINING DENGAN METODE NAIVE BAYES CLASSIFIER UNTUK MENGLASIFIKASI BERITA BERDASARKAN KONTEN

Nama : Air Nuor Roifa

NRP : 06111340200009

Proses Awa
Perhitungan Kata
Klasifikasi

Input Berita :

Hasil :

Kelas	ID Berita

Presisi :

Recall :

Akurasi :

Gambar 4. 15 Tampilan GUI tab Klasifikasi

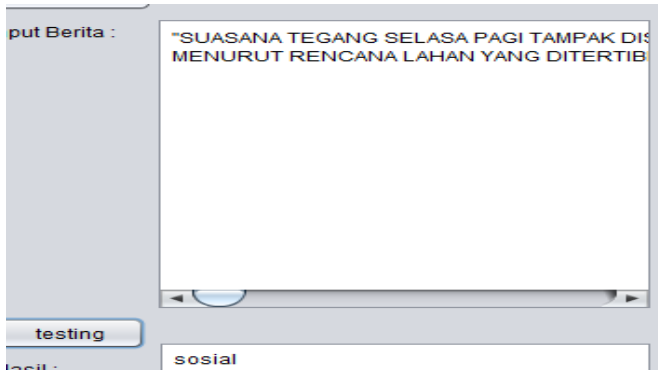
```

-----
[I@369578e5
agama-spek = 0.001544004117344313
[I@369578e5
ekonomi-spek = 0.0011295180722891566
[I@369578e5
politik-spek = 0.001755926251097454
[I@369578e5
sosial-spek = 4.854368932038835E-4
-----
[I@369578e5
agama-spesial = 5.158627805003869E-4
[I@369578e5
ekonomi-spesial = 7.555723460521345E-4
[I@369578e5
politik-spesial = 2.9342723004694836E-4
[I@369578e5
sosial-spesial = 7.297494526879105E-4
-----
[I@369578e5
agama-spesialis = 2.5799793601651185E-4
[I@369578e5
ekonomi-spesialis = 7.558578987150416E-4
[I@369578e5
politik-spesialis = 2.93513354857646E-4
[I@369578e5
sosial-spesialis = 7.2992700729927E-4
-----

```

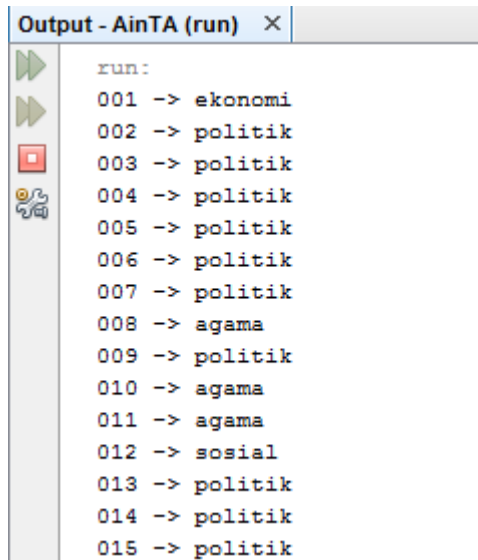
Gambar 4. 16 Proses Perhitungan *Naïve Bayes Classifier*

Setelah proses *training* selesai dilakukan, selanjutnya yaitu coba *testing* terhadap satu berita dimasukkan dan didapat hasil prediksinya. Maka untuk mengetahui keakuratan atau mengukur tingkat kedekatan antara nilai prediksi dengan nilai aktual. Maka penulis menggunakan menggunakan metode *confusion matrix* untuk mengevaluasi program dan mendapatkan nilai akurasi.



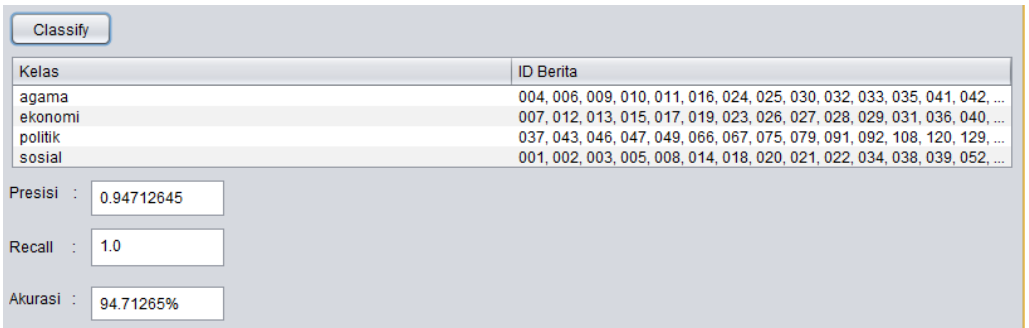
Gambar 4.17 Hasil Test Satu Berita

Untuk mengetahui apakah prediksi sudah baik atau belum, kita lakukan uji data (*testing*) dengan data teks berita sebanyak 435.



Gambar 4.18 Gambaran Proses Testing

Kemudian hasil nilai dari proses *testing* akan dimunculkan pada GUI bagian tab klasifikasi tombol *classify* . Berikut hasil klasifikasi dari 435 teks berita yang disimpan pada database *testing* serta nilai presisi, recall dan akurasi.



Kelas	ID Berita
agama	004, 006, 009, 010, 011, 016, 024, 025, 030, 032, 033, 035, 041, 042, ...
ekonomi	007, 012, 013, 015, 017, 019, 023, 026, 027, 028, 029, 031, 036, 040, ...
politik	037, 043, 046, 047, 049, 066, 067, 075, 079, 091, 092, 108, 120, 129, ...
sosial	001, 002, 003, 005, 008, 014, 018, 020, 021, 022, 034, 038, 039, 052, ...

Presisi : 0.94712645

Recall : 1.0

Akurasi : 94.71265%

Gambar 4.19 Hasil *Testing*

BAB V HASIL DAN PEMBAHASAN

Pada bab ini dijelaskan pembahasan mengenai pengujian yang telah dilakukan terhadap perangkat lunak. Hasil pengujian kemudian dibahas untuk mengetahui kerja sistem secara keseluruhan.

5.1 Hasil Pre-proses Data

Pada penelitian ini data yang digunakan adalah data berita TV yang ditayangkan pada rentang waktu Juni-Desember 2016. Data tersebut didapat dari perusahaan TV9 Surabaya dan berjumlah 935 data teks berita. Data awal dari perusahaan masih sangat primer, masih dalam bentuk *microsoft word* sehingga penulis membuat database manual dengan membuka setiap file *word* yang ada. Kemudian data teks berita yang sudah disimpan pada excel diklasifikasikan manual untuk data *testing* yang berjumlah 500. Setelah itu agar lebih akurat hasil klasifikasi manual penulis dikonsultasikan kepada pihak TV9, baru kemudian dilakukan proses selanjutnya. Saat mengklasifikasikan teks berita penulis memiliki beberapa kriteria yang melihat kata kunci atau kata penting pada sebuah teks berita. Berikut tabel sebagian kata kunci yang menjadi landasan penulis dalam mengklasifikasi berita, untuk lengkapnya dapat dilihat pada Lampiran E.

Tabel 5. 1 Kriteria Klasifikasi

Kelas	Kata Kunci
Agama	islam, santri, ulama, pondok, pesantren, ibadah, haji, idul, fitri, adha, raamdhan, puasa, zakat, fitrah.
Ekonomi	uang, pajak, industri, tax, amnesty, gaji, ekonomi, turun, naik, ekspor, impor, rupiah, bank, makmur.
Politik	pemerintah, demokrasi, presiden, rakyat, dpr, dpd, dprd, politik, hukum, sidang, jaksa, korupsi, mpr.

Sosial	warga, masyarakat, sosial, gotong, royong, desa, bantu, peduli, dinas, desa, kerja, keras, ikhlas.
--------	--

Setelah diklasifikasi, baru memasuki tahap selanjutnya yaitu *running program* yang diawali dari proses *pre-processing*. Berikut merupakan beberapa data berita :

Tabel 5. 2 Lima Teks Berita Data Testing

ID_Berita	Konten Berita
085	MENYAMBUT HARI JADI PROVINSI JAWA TIMUR KE-71SANGGAR MERAH PUTIH KEMBALI MENGGELAR PASAR SENI LUKIS SELAMA 10 HARI YANG DI MULAI 10 HINGGA 16 OKTOBER MENDATANG DI JATIM EXPO INTERNASIONAL SURABAYA YANG RESMI DIBUKA OLEH GUBERNUR JAWA TIMUR SOEKARWO PENYELENGGARAAN PASAR SENI LUKIS 2016 INIBERBEDA DARI TAHUN SEBELUMNYA PADA PASAR LUKIS TAHUN INIBERBAGAI RANGKAIAN AKAN BERLANGSUNG SERTA PARA PELUKIS JUGA AKAN MENGUNJUNGI PT PALDAN MEREKA NANTINYA AKAN MELUKIS DI ATS KAPAL PASAR SENI YANG DIKUTI SEBANYAK 210 PELUKIS DARI SELURUH INDONESIA INI OMSET YANG DITARGETKAN SEBANYAK 2 MILIYAR RUPIAHKARENA PASAR SENI LUKIS INDONESIA 2016 INI MENJADI AJANG TRANSAKSI DAN APRESIASI DARI PEMERINTAH PROVINSI ANIS BERHARAP MELALUI PASAR SENI INI BISA MENJADI PERTIMBANGAN DARI HIRUK-PIKUK POLITIK YANG ADA DI INDONESIAAN MASYARAKAT DI INDONESIA KEMBALI MENCINTAI KESENIAN GUBERNUR JAWA TIMUR SOEKARWO MENYAMBUT BAIK PASAR SENI LUKIS MENURUT PAK DE KARWO ACARA INI TIDAK HANYA BERMANFAAT UNTUK MENGEMBANGKAN DAN MELESTARIKAN SENI BUDAYA JAWA TIMURAKAN TETAPI JUGA SEBAGAI UPAYA UNTUK MELESTARIKANPARA SENIMAN LUKIS SENI MERUPAKAN BAGIAN DARI KEBUDAYAAN DAN KEKAYAAN YANG DIMILIKI OLEH INDONESIAserta MEMBERIKAN KEADABAN SESEORANGSENI JUGA MAMPU MENYENTUH TERHADAP PERASAAN KEINDAHAN DAN JUGA SUARA HATI MANUSIA OLEH KARENA ITU PAK DE KARWO BERJANJI AKAN MEMBERIKAN FASILITAS KEPADA SENIMAN-SENIMAN UNTUK MEMAMERKAN KARYA-KARYANYA DI JAWA TIMUR AGAR PENYELENGGARAAN PASAR SENI LUKIS
077	KEDATANGAN KOMISI TIGA DPR RI YANG DIPIMPIM DESMON J MAHESA INI LANGUSNG DISAMBUT KAPOLDA JAWA TIMUR IRJEND POL. ANTON SETIADJI DI RUANG RUPATAMA MAPOLDA JATIM DALAM KUNJUNGANNYA KOMISI III DPR RI INI MEMINTA PROSES PENYIDIKAN KASUS PEMBUNUHAN DAN DUGAAN PENIPUAN OLEH TERSANGKA TAAT PRIBADI ALIAS DIMAS KANJENG BERLANGSUNG TRANSPARAN DAN TIDAK DITUTUP TUTUPI TERMASUK SOAL RAIBNYA BARANG BUKTI ASET DAN UANG DARI PADEPOKAN DIMAS KANJENG TAAT PRIBADI KOMISI TIGA DPR RI INI MEMPERTANYAKAN KINERJA POLISI YANG GAGAL MENYITA DAN MENEMUKAN UANG DIMAS KANJENG DARI PADEPOKAN YANG KONON

	<p>JUMLAHNYA MENCAPAI TRILYUNAN RUPIAH UNTUK ITU POLISI HARUS SERIUS MENGUAK RAIBNYA UANG DIMAS INI AGAR TIDAK TIMBUL SPEKULASI DAN FITNAH DI LUAR PROSES HUKUM JANGAN-JANGAN ADA ORANG TERTENTU YANG SENGAJA MENGHILNGANKAN DAN MENGUAPKAN BARANG BUKTI INI SELAIN MEMPERTANYAKAN UANG RAIBNYA UANG DIMAS KANJENG KEDATANGAN PANJA PENEGAKAN HUKUM KOMISI TIGA DPR RI INI KE POLDA JATIM INI JUGA UNTUK MENGAWASI PROSES PENYIDIKAN KASUS DIMAS KANJENG APAKAH SUDAH SESUAI PROSEDUR ATAU TIDAK KOMISI III INGIN MENGETAHUI SECARA JELAS PROSES PENYIDIKANNYA SEKALIGUS MENGETAHUI PASTI APA SAJA YANG DIAMANKAN PENYIDIK SEBAGAI BARANG BUKTI KARENA ITU PENYIDIK HARUS MENGUAK INFORMASI APAPUN TERKAIT ASET DAN UANG YANG DISIMPAN DIMAS KANJENG TERMASUK SOAL ADANYA NAMA PENYIMPAN UANG TRILIUNAN DIMAS KANJENG, DODI WAHYUDI DAN ABAH DHOFIR</p>
101	<p>NAMUN MESKI DEMIKIAN BUPATI SIDOARJO DAN WAKIL BUPATI SIDOARJO SERTA JAJARAN PEJABAT LAINNYA BELUM JUGA MELAKUKAN SIDAK UNTUK MELIHAT LANGSUNG KONDISI BANJIR YANG MELUMPUHKAN KOTA SIDOARJO MEREKA JUSTRU MELAKUKAN SIDAK UNTUK MELIHAT PEMBANGUNAN GEDUNG SERBA GUNA YANG MENGHABISKAN ANGGARAN PULUHAN MILIAR RUPIAH BUPATI SIDOARJO HANYA MENJELASKAN BANJIR DIKOTA SIDOARJO INI RUTIN TERJADI KARENA PADA TANGGAL TERTENTU AIR LAUT PASANG SEHINGGA BISA MEMERPARAH KONDISI BANJIR YANG MELANDA KOTA SIDOARJO UPAYA PEMBANGUNAN DAN PERBAIKAN GORONG-GORONG DENGAN BOX CULVERT SUDAH DILAKUKAN DISEJUMLAH TITIK RAWAN BANJIR NAMUN BANJIR TAHUN INI JUSTRU SEMAKIN PARAH</p>
087	<p>ERA MASYARAKAT EKONOMI ASEAN SAAT INI SUDAH MULAI DIRASAKAN OLEH PENDUDUK INDONESIA TAK TERKECUALI MADURA OLEH KARENA ITU KITA HARUS BERBENAH MEMPERSIAPKAN DIRI BERSAING DENGAN PASAR INTERNASIONAL AGAR KITA TIDAK MENJADI OBJEK DI NEGERI SENDIRI MELIHAT POTENSI MADURA YANG SANGAT KAYA DIREKTUR KERJASAMA FUNGSIONAL ASEAN KEMENTERIAN LUAR NEGERI RI J.S. GEORGE LANTU SAAT MENGISI SEMINAR INTERNASIONAL ASEAN DI KAMPUS UNIVERSITAS WIRARAJA SUMENEP OPTIMIS DENGAN POTENSI YANG ADA DI MADURA KEKAYAAN ALAM MADURA KHUSUSNYA SUMENEP SUNGGUH SANGAT LUAR BIASA MULAI DARI KEINDAHAN PANTAI HINGGA KANDUNGAN ALAMNYA YANG BISA MENGUNDANG DAYA TARIK INTERNASIONAL KHUSUS SUMENEP J.S. GEORGE LANTU SANGAT OPTIMIS DENGAN KEKAYAAN SEJARAHNYA YANG SANGAT LUAR BIASA DI SUMENEP PERNAH ADA KERATON SUMENEP YANG MENJADI DESTINASI PERDAGANGAN INTERNASIONAL SALAH SATU BUKTI DULUNYA KERATON SUMENEP MENGGUNAKAN UANG KOIN MARIA TEREZA DIMANA INI MERUPAKAN ALAT TUKAR INTERNASIONAL LEBIH LANJUT GEORGE MENEGASKAN MADURA MEMPUNYAI KEKAYAAN BUDAYA YANG UNIK YANG MEMPUNYAI DAYA TARIK INTERNASIONAL DENGAN KULTUR KEAGAMAAN YANG SANGAT KENTAL SELAIN ITU MAKANAN YANG ADA DI MADURA ENAK-ENAK DAN JIKA DIBAWA KE</p>

	FORUM INTERNASIONAL PASTI LAKU TINGGAL BAGAIMANA KITA MENGEMAS DAN METODE PEMASARANNYA
057	<p>BEBERAPA HARI TERKAHIR WILAYAH KABUPATEN SUMENEP MADURA JAWA TIMUR SERING DIGUYUR HUJAN HINGGA PETANI TEMBAKAU DI SUMENEP MENGELUS DADA KARENA BANYAK TANAMAN TEMBAKAU GENANG AIR SELAIN ITU PETANI JUGA MERASA KKHAWATIR TERHADAP NASIB TEMBAKAU MEREKA AKIBAT ADANYA PERNYATAAN PIHAK GUDANG YANG AKAN MENUTUP GUDANG JIKA TURUN HUJAN PADAHAL HINGGA KINI TEMBAKAU YANG PENEN BARU 50 PERSEN MENANGGAPI PERNYATAAN PIHAK GUDANG YANG AKAN TUTUP JIKA HUJAN TURUN DINAS KEHUTANAN DAN PERKEBUNAN KABUPATEN SUMENEP MENGIMBAU PIHAK GUDANG TIDAK SEWENANG-WENANG MELAKUKAN PENUTUPAN GUDANG KARENA HAL TERSEBUT AKAN MEMUNCULKAN MASALAH DI TINGKATAN PETANI HUTBUN MINTA JIKA MEMANG GUDANG MAU TUTUP HARUS MELAKUKAN PEMBERITAHUAN SEBELUMNYA MINIMAL 5 HARI SEHINGGA PETANI DAN PEDAGANG DIBAWAH BISA MENGANTISIPASI KEPALA BIDANG KEBUN DISHUTBUN SUMENEP JOKO SUWARNO MENGANGGAP PERNYATAAN GUDANG TERSEBUT HANYA MENAKUT-NAKUTI PETANI AGAR SECEPATNYA PANEN DAN BISA MEMBELI TEMBAKAU PETANI DENGAN HARGA MURAH UNTUK MENGECEK SERAPAN TEMBAKAU PIHAKNYA DALAM WAKTU DEKAT AKAN TURUN KEGUDANG GUNA MEMINTA KEBIJAKAN KEPADA PIHAK GUDANG TERKAIT TATA NIAGA TEMBAKAU DI SUMENEP JOKO MENAMBAHKAN SAAT INI HARGA TEMBAKAU DI GUDANG BERKISAR ANTARA 25 RIBU HINGGA 52 RIBU RUPIAH PERKILOGRAM SEMENTARA YANG LAKUKAN PEMBELIAN HANYA 2 GUDANG YAKNI GUDANG GARAM DI DESA PATEAN KECAMATAN BATUAN DAN GUDANG GARAM DI KECAMATAN GULUK-GULUK PADAHAL TAHUN-TAHUN SEBELUMNYA YANG PEMLAKUKAN PEMBELIAN TEMBAKAU ADA 4 HINGGA 5 GUDANG.</p>

Data tersebut selanjutnya dilakukan pre-proses teks yang terdiri dari *case folding*, *tokenizing*, *filtering*, dan *stemming*. Berikut hasil 5 data *testing* berita setelah dilakukan pre-proses teks :

Tabel 5. 3 Hasil Pre-Processing

ID_Berita	Konten Berita
085	<p>sambut hari jadi provinsi jawa timur ke sanggar merah putih kembali gelar pasar seni lukis lama hari mulai hingga oktober datang di jatim expo internasional surabaya resmi buka gubernur jawa timur soekarwo selenggara pasar seni lukis beda tahun sebelum pasar lukis tahun bagi rangkai langsung para pelukis kunjungi pt paldan mereka nanti lukis atas kapal pasar seni ikut banyak pelukis seluruh indonesia omset target banyak miliar rupiah pasar seni lukis indonesia menjadi ajang transaksi apresiasi pemerintah provinsi anis harap pasar seni bisa jadi timbang hiruk-pikuk politik ada indonesia masyarakat indonesia kembali cinta seni gubernur jawa timur soekarwo sambut baik pasar seni lukis turut pak de karwo acara tidak hanya manfaat kembang lestari seni</p>

	budaya jawa timur upaya lestari para seniman lukis seni bagian budaya kaya milik indonesia beri adab orang seni mampu sentuh hadap rasa indah suara hati manusia pak de karwo janji berik fasilitas pada seniman-seniman pamer karya-karya jawa timur selenggara pasar seni lukis
077	datang komisi tiga dpr ri pimpim desmon j mahesa sambut kapolda jawa timur irjend pol. anton setiadji ruang rupertama mapolda jatim kunjungan komisi iii dpr ri minta proses sidik kasus bunuh dugaan tipu sangka taat pribadi alias dimas kanjeng transparan tutup tutupi soal raib barang bukti aset uang dari padepokan dimas kanjeng taat pribadi komisi tiga dpr ri ini tanya kinerja polisi gagal sita temu uang dimas kanjeng padepokan yang konon jumlahnya mencapai trilyunan rupiah untuk itu polisi harus serius menguak raib uang dimas timbul spekulasi fitnah luar proses hukum jangan-jangan orang tentu sengaja menghihgankan uap barang bukti tanya uang raib uang dimas kanjeng datang panja penegakan hukum komisi tiga dpr rike polda jatim awas proses sidik kasus dimas kanjeng apakah sudah sesuai prosedur komisi iii ingin mengetahui secara jelas proses sidikan sekaligus mengetahui pasti aman sidik sebagai barang bukti karena sidik harus nguak informasi apapun terkait aset uang simpan dimas kanjeng soal ada nama simpan uang triliunan dimas kanjeng, dodi wahyudi abah dhofir
101	meski bupati sidoarjo wakil bupati sidoarjo jajaran pejabat lain belum laku sidak lihat langsung kondisi banjir lumpuh kota sidoarjo mereka justru laku sidak lihat bangun gedung serba guna habis anggar puluh miliar rupiah bupati sidoarjo hanya jelas banjir kota sidoarjo rutin jadi pada tanggal tertentu air laut pasang hingga bisa parah kondisi banjir landa kota sidoarjo upaya bangun baik gorong-gorong box culvert sudah laku jumlah titik rawan banjir banjir tahun justru makin parah
087	era masyarakat ekonomi asean mulai rasa duduk indonesia tak kecuai madura kita harus benah siap diri saing pasar internasional kita tidak jadi objek negeri sendiri lihat potensi madura sangat kaya direktur kerjasama fungsional asean menteri luar negeri ri j.s.georgelantu isi seminar internasional asean kampus universitas wiraraja sumenep optimis potensi madura kaya alam madura khusus sumenep sungguh sangat luar biasa mulai indah pantai hingga kandung alam bisa undang daya tarik internasional khusus sumenep j.s.georgelantu optimis kaya sejarah luar biasa sumenep pernah keraton sumenep jadi destinasi dagang internasional salah satu bukti dulu keraton sumenep guna uang koin maria tereza mana rupa alat tukar internasional lebih lanjut george tegas madura punya kaya budaya unik punya daya tarik internasional kultur agama kental lain makan madura enak-enak bawa forum internasional pasti laku tinggal bagaimana kita kemas metode
057	hari terkahir wilayah kabupaten sumenep madura jawa timur sering guyur hujan hingga petani tembakau sumenep elus dada karena banyak tanaman tembakau genang air petani merasa kkhawatir nasib tembakau mereka akibat adanya pernyataan pihak gudang tutup gudang jika turun hujan padahal hingga kini tembakau penen baru 50 persen tanggapi pernyataan pihak gudang akan tutup hujan turun dinas hutan kebun kabupaten sumenep imbau pihak gudang sewenang-wenang melakukan penutupan gudang karena hal tersebut muncul masalah tingkat petani hutbun minta memang gudang mau tutup harus melakukan pemberitahuan sebelumnya minimal 5 hari hingga petani pedagang dibawah bisa antisipasi kepala bidang kebun dishutbun sumenep joko suwarno anggap pernyataan gudang hanya menakut-nakuti petani agar secepatnya panen dan bisa membeli tembakau petani dengan harga murah untuk mengecek serapan tembakau pihaknya dalam waktu dekat akan turun kegudang guna meminta kebijakan kepada pihak gudang terkait tata niaga tembakau sumenep joko tambah saat harga tembakau

	gudang kisar antara 25 ribu hingga 52 ribu rupiah perkilogram sementara lakukan beli hanya 2 gudang yakni gudang garam desa patean kecamatan batuan gudang garam kecamatan guluk-guluk padahal tahun-tahun sebelum lakukan beli tembakau ada 4 hingga 5 gudang.
--	---

5.2 Ekstraksi Nilai dan Perhitungan *Likelihood*

Setelah data melalui pre-processing teks, selanjutnya akan di ekstraksi kata dari data teks berita. Kata diambil dari semua teks yang di proses sebelumnya. Kemudian kata dihitung nilai probabilitasnya pada setiap kelas dengan menggunakan *likelihood* agar nilai kata tidak ada yang nol. Karena jika ada yang nol nanti hasil dari proses klasifikasi tidak ditemukan. Rumus *likelihood* ini sangat membantu perhitungan proses klasifikasi, dengan menambah angka 1 pada proses perhitungannya sehingga nilainya tidak lagi nol. Awal perhitungan ekstraksi kata yaitu menghitung kemunculan dari setiap kata pada semua teks berita. Berikut hasil nilai ekstraksi kata pada beberapa yang diambil dari database :

Tabel 5. 3 Hasil Nilai Kata berdasarkan Kemunculan

No	Kata	Agama	Ekonomi	Politik	Sosial	W
1	wafat	2	0	0	0	2
2	warga	23	26	25	63	137
3	Jual	4	25	5	6	40
4	Turun	8	23	10	13	54
5	Jalan	43	12	27	49	131
6	Vonis	0	0	1	0	1
7	bupati	12	3	19	14	48
8	Ketua	35	10	21	26	92
	N(Y)	3872	2643	3404	4107	14026

Hasil nilai dari tiap kata pada setiap kelas menjadi sebuah kata kunci untuk kelas yang memiliki nilai kata paling besar, seperti pada Tabel 5.3 diatas diketahui kata wafat dan ketua masuk kelas agama, kata jual dan ketua masuk kelas ekonomi, kata negeri dan

bupati masuk kelas politik, kata warga dan jalan masuk kelas sosial.

Dikarenakan terdapat nilai nol pada kata wafat dan vonis untuk beberapa kelas. Maka untuk menghindari itu semua nilai ekstraksi kata di-*update* menjadi seperti tabel 5.4 dibawah ini. Perhitungan dilakukan dengan menggunakan persamaan teknik *Laplacian Smoothing* berikut[14]:

$$P(X_i|Y) = \frac{1 + n(X_i, Y)}{|W| + n(Y)} \quad (5.1)$$

dimana,

$n(X_i, Y)$ = Jumlah *term* X_i yang ditemukan di seluruh data latih dengan kategori Y .

$n(Y)$ = Jumlah *term* di seluruh data latih dengan kategori Y .

$|W|$ = Jumlah seluruh *term* X_i dari seluruh data latih.

Data 1-8 merupakan sebagian kata yang didapatkan dari proses ekstraksi kata, kemudian nilai $n(Y)$ dibagian bawah merupakan jumlah dari semua kata yang ditemukan pada setiap kelas. Total semua kata yang berhasil di ekstraksi dari 500 data *training* sebanyak 14.026 kata. Dan nilai $|W|$ yaitu total kata dari semua kelas. Kata kunci inilah yang menjadi dasar acuan untuk perhitungan klasifikasi pada data *testing* nantinya. Ambil 2 kata dari tabel 5.4 yaitu kata wafat dan vonis. Perhitungan *likelihood* kata wafat pada kelas agama :

$$P(\text{wafat}|\text{agama}) = \frac{1 + 2}{2 + 3782}$$

$$P(\text{wafat}|\text{agama}) = \frac{3}{3784}$$

$$P(\text{wafat}|\text{agama}) = 0,0007741$$

kata wafat pada kelas ekonomi :

$$P(\text{wafat}|\text{ekonomi}) = \frac{1 + 0}{2 + 2643}$$

$$P(\text{wafat}|\text{ekonomi}) = \frac{1}{2645}$$

$$P(\text{wafat}|\text{ekonomi}) = 0,0003780$$

kata wafat pada kelas politik :

$$P(\text{wafat}|\text{politik}) = \frac{1 + 0}{2 + 3404}$$

$$P(\text{wafat}|\text{politik}) = \frac{1}{3406}$$

$$P(\text{wafat}|\text{politik}) = 0,0002936$$

kata wafat pada kelas sosial :

$$P(\text{wafat}|\text{sosial}) = \frac{1 + 0}{2 + 4107}$$

$$P(\text{wafat}|\text{sosial}) = \frac{1}{4109}$$

$$P(\text{wafat}|\text{sosial}) = 0,0002433$$

Perhitungan *likelihood* kata vonis pada kelas agama :

$$P(\text{vonis}|\text{agama}) = \frac{1 + 0}{1 + 3782}$$

$$P(\text{wafat}|\text{agama}) = \frac{1}{3783}$$

$$P(\text{wafat}|\text{agama}) = 0,0002581$$

kata vonis pada kelas ekonomi :

$$P(\text{wafat}|\text{ekonomi}) = \frac{1 + 0}{1 + 2643}$$

$$P(\text{wafat}|\text{ekonomi}) = \frac{1}{2644}$$

$$P(\text{wafat}|\text{ekonomi}) = 0,0003782$$

kata vonis pada kelas politik :

$$P(\text{wafat}|\text{politik}) = \frac{1 + 1}{1 + 3404}$$

$$P(\text{wafat}|\text{politik}) = \frac{2}{3405}$$

$$P(\text{wafat}|\text{politik}) = 0,0005873$$

kata wafat pada kelas sosial :

$$P(\text{wafat}|\text{sosial}) = \frac{1 + 0}{1 + 4107}$$

$$P(\text{wafat}|\text{sosial}) = \frac{1}{4108}$$

$$P(\text{wafat}|\text{sosial}) = 0,0002434$$

Berikut hasil perhitungan dengan *likelihood*:

Tabel 5. 4 Hasil Perhitungan dengan *Likelihood*

No	Kata	Agama	Ekonomi	Politik	Sosial
1	wafat	0,0007741	0,0003780	0,0002936	0,0002433
2	warga	0,005985	0,009712	0,007343	0,01508
3	jual	0,001278	0,009691	0,001742	0,001688
4	turun	0,002292	0,008899	0,003181	0,003365
5	jalan	0,010989	0,004686	0,007921	0,011798
6	vonis	0,0002581	0,0003782	0,0005873	0,0002434
7	Bupati	0,003315	0,001486	0,005794	0,00361
8	Ketua	0,009079	0,004022	0,006293	0,00643

Kemudian masuk proses klasifikasi dengan menggunakan *naïve bayes classifier*. Pada perhitungan *naïve bayes* berdasarkan rumus

$$P(Y|X) = \frac{P(Y) \prod_{i=1}^d P(X_i|Y)}{P(X)} \quad (5.2)$$

untuk nilai $P(Y)$ diasumsikan sama agar setiap kelas memiliki nilai peluang yang sama yaitu 0,25 atau 25% setiap kelas. Sehingga pada perhitungannya dapat diabaikan. Dan juga untuk nilai $P(X)$ tidak perlu dihitung karena semua sama yaitu bernilai 1(satu), didapat dari jumlah data testing dibagi dengan semua data secara keseluruhan $\frac{435}{435} = 1$.

Secara tidak langsung perhitungan klasifikasi menggunakan metode *naïve bayes classifier* yaitu $P(Y|X) = \prod_{i=1}^d P(X_i|Y)$ yang merupakan perkalian dari semua kata yang ada di setiap kelas, dapat dilihat perbandingan nilai peluang yang paling besar dan kelas tersebut yang dipilih menjadi hasil keluaran suatu klasifikasi. Seperti yang ditunjukkan Gambar 4.13 menghasilkan klasifikasi

kelas sosial dari data berita yang dimasukkan manual pada GUI Java. Karena berdasarkan konten berita yang ada data tersebut memiliki nilai peluang paling besar pada kelas sosial.

Berdasar pada hasil perhitungan *likelihood* Tabel 5.4 dapat dicontohkan misal terdapat berita A1 yang mengandung kata warga, wafat, jalan turun. Maka berita A1 perhitungan *naïve bayes* sebagai berikut :

Tabel 5. 4 Tabel Perhitungan *Naïve Bayes Classifier*

$P(\text{agama} A1) = 0,0007741 \times 0,005985 \times 0,010989$ $\times 0,002292$ $P(\text{agama} A1) = 1,167 \times 10^{-10}$
$P(\text{ekonomi} A1)$ $= 0,0003780 \times 0,009712 \times 0,004686$ $\times 0,003181$ $P(\text{ekonomi} A1) = 5,472 \times 10^{-11}$
$P(\text{politik} A1) = 0,0002936 \times 0,007343 \times 0,007921$ $\times 0,003181$ $P(\text{politik} A1) = 5,432 \times 10^{-11}$
$P(\text{sosial} A1) = 0,0002433 \times 0,01508 \times 0,011798$ $\times 0,003365$ $P(\text{sosial} A1) = 1,456 \times 10^{-10}$

Dari keempat hasil kemungkinan berita A1 masuk kedalam kelas sosial karena nilai pada kelas sosial yang paling besar diantara yang lain yaitu sebesar $1,456 \times 10^{-10}$.

5.3 Hasil Klasifikasi *Naïve Bayes Classifier*

Klasifikasi dilakukan untuk membuat suatu model prediksi dari data hasil pre-proses dan ekstraksi serta perhitungan nilai kata. Data berita yang berjumlah 935 tersebut dipisahkan menjadi data *training* dan data *testing*, yaitu sebanyak 500 untuk *training* dan 435 untuk *testing*.

Model prediksi yang telah terbentuk selanjutnya akan diuji menggunakan data *testing*. Hasil pengujian menunjukkan laporan yang benar diprediksi sebanyak 412 dari 435 data *testing*. Nilai 412 diambil dari nilai TP yaitu data yang benar terklasifikasi. Hasil klasifikasi dari data *testing* sebagai berikut :

Tabel 5. 5 Hasil Klasifikasi

Kelas	ID_Berita
Agama	004, 006, 009, 010, 011, 016, 024, 025, 030, 032, 033, 035, 041, 042, 044, 048, 050, 051, 058, 059, 068, 077, 081, 083, 084, 085, 088, 090, 095, 098, 099, 102, 103, 104, 106, 114, 115, 116, 117, 119, 122, 123, 124, 125, 126, 127, 128, 130, 134, 136, 138, 141, 142, 147, 152, 153, 154, 156, 157, 158, 159, 160, 161, 162, 163, 164, 166, 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177, 178, 179, 181, 182, 183, 184, 185, 188, 190, 191, 192, 193, 195, 196, 197, 198, 199, 200, 201, 202, 203, 206, 207, 209, 210, 211, 213, 221, 222, 223, 225, 226, 230, 231, 232, 233, 234, 235, 239, 240, 241, 242, 243, 247, 249, 253, 254, 255, 258, 259, 260, 262, 265, 266, 269, 271, 272, 273, 274, 276, 277, 281, 282, 283, 284, 285, 287, 288, 289, 292, 296, 299, 300, 301, 303, 305, 306, 307, 308, 310, 311, 312, 313, 314, 315, 316, 318, 319, 320, 321, 329, 330, 331, 332, 334, 335, 337, 338, 339, 340, 348, 349, 350, 351, 352, 358, 360, 362, 365, 367, 369, 374, 381, 385, 388, 389, 390, 391, 394, 395, 399, 400, 401, 402, 403, 406, 407, 412, 413, 414, 415, 416, 419, 422, 426, 427, 431, 432, 433, 434, 435.
Ekonomi	007, 012, 013, 015, 017, 019, 023, 026, 027, 028, 029, 031, 036, 040, 045, 054, 056, 057, 061, 065, 072, 073, 078, 080, 087, 094, 107, 113, 135, 137, 140, 146, 150, 155, 165, 186, 208, 212, 218, 224, 227, 236, 237, 244, 250, 251, 256, 257, 280, 286,

	293, 304, 309, 326, 327, 345, 346, 354, 363, 366, 368, 370, 378, 384, 387, 392, 408, 409, 417, 420, 423.
Politik	037, 043, 046, 047, 049, 066, 067, 075, 079, 091, 092, 108, 120, 129, 143, 144, 148, 149, 187, 189, 194, 216, 219, 248, 252, 267, 268, 290, 295, 317, 322, 323, 324, 325, 336, 341, 342, 343, 344, 353, 355, 356, 361, 373, 375, 376, 380, 393, 405, 411, 424, 425, 428, 429, 430.
Sosial	001, 002, 003, 005, 008, 014, 018, 020, 021, 022, 034, 038, 039, 052, 053, 055, 060, 062, 063, 064, 069, 070, 071, 074, 076, 082, 086, 089, 093, 096, 097, 100, 101, 105, 109, 110, 111, 112, 118, 121, 131, 132, 133, 139, 145, 151, 180, 204, 205, 214, 215, 217, 220, 228, 229, 238, 245, 246, 261, 263, 264, 270, 275, 278, 279, 291, 294, 297, 298, 302, 328, 333, 347, 357, 359, 364, 371, 372, 377, 379, 382, 383, 386, 396, 397, 398, 404, 410, 418, 421.

Keterangan : Pada kolom Hasil Klasifikasi merupakan ID Berita pada data *testing* sebanyak 435.

Proses klasifikasi dapat mengetahui jumlah data yang benar terprediksi terhadap jumlah data *testing* atau data ujinya. Sehingga bisa didapatkan kesimpulan bahwa model prediksi yang telah dibentuk sudah baik atau masih belum berdasarkan hasil akurasi. Untuk menentukan besar nilai akurasi digunakan metode *confusion matrix* yang melihat nilai *True and False prediction*.

Tabel 5. 6 Hasil *Confusion matrix*

TP	TN	FP	FN
412	0	23	0

Setelah diketahui hasil *confusion matrix* dari proses *testing* yang nilainya tersebut tersimpan dalam database data_uji MySQL. Perhitungan nilai presisi, recall, dan akurasi sebagai berikut :

Tabel 5. 7 Perhitungan Presisi, Recall, dan Akurasi

$\text{Presisi} = \frac{TP}{TP + FP} = \frac{412}{412 + 23} = \frac{412}{435} = 0,9471264368$	(5.3)
$\text{Recall} = \frac{TP}{TP + FN} = \frac{412}{412 + 0} = \frac{412}{412} = 1,0$	(5.4)
$\begin{aligned} \text{Akurasi} &= \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \\ &= \frac{412 + 0}{412 + 0 + 23 + 0} \times 100\% \\ &= \frac{412}{435} \times 100\% \\ &= 0,9471264368 \times 100\% \\ \text{Akurasi} &= 94,71264368 \% \end{aligned}$	(5.5)

Berdasarkan persamaan 5.3, 5.4, dan 5.5 didapatkan hasil dari proses *testing* program ini yaitu klasifikasi berita seperti pada Tabel 5.6, serta nilai presisi, recall, dan akurasi pada Tabel 5.8.

Study Case :

Apabila terdapat kata lebih dari satu kelas maka diambil nilai tertinggi saat proses perhitungan *likelihood*. Dan ketika perhitungan *naïve bayes* secara otomatis juga akan semakin besar. Karena *likelihood* berbanding lurus dengan rumus *naïve bayes classifier*. Berikut satu berita yang akan dibahas apakah berita ini masuk kelas agama atau politik?

PULUHAN PELAJAR MAUPUN SANTRI DARI GRESIK DAN LAMONGAN MENGIKUTI SEKOLAH PARLEMEN SANTRI SEKOLAH PARLEMEN MERUPAKAN WACANA PROGRAM DPR RI SEBAGAI UPAYA PENINGKATAN KUALITAS ANGGOTA DEWAN SERTA MEMBERIKAN WAWASAN MENGENAI TUGAS DAN FUNGSI DPR BAGI KALANGAN NON ANGGOTA DEWAN MESKI PROGRAM TERSEBUT BELUM DISAHKAN DPR RI NAMUN FRAKSI PKB MENGINISIASI PROGRAM ITU DENGAN MENGGELAR SEKOLAH PARLEMEN YANG DITUJUKAN BAGI KALANGAN SANTRI MAUPUN PELAJAR SEKOLAH PARLEMEN SANTRI INI MEMBERI PEMAHAMAN KEPADA KAUM SANTRI TENTANG POLITIK DAN KERJA-KERJA KEPARLEMENAN AGAR DAPAT MEMBUKA WAWASAN POLITIK KAUM SANTRI SUPAYA TIDAK SALAH MENILAI PEKERJAAN ANGGOTA DEWAN TAK SEDIKIT KALANGAN POLITISI DI PARLEMENYANG DULUNYA BERLATAR BELAKANG SEBAGAI SANTRI OLEH KARENA ITU PENDIDIKAN POLITIK DIRASA SANGAT PERLU TERUTAMA DALAM MEMERSIAPKAN KADER-KADER DARI PONDOK PESANTREN SEBAB SEORANG SANTRI YANG TELAH DITEMPA AKHLAK DAN MENTALNYA SEJAK DI PONDOK

PESANTREN MEMILIKI BEKAL ILMU YANG DAPAT DIGUNAKAN UNTUK MEMBANGUN BANGSA MELALUI JALUR POLITIK

Gambar 5. 3 Satu Berita Studi Kasus

Pertama dilakukan preprocessing terlebih dahulu, maka berita diatas mengalami normalisasi sebagai berikut :

puluh pelajar santri gresik lamongan ikut sekolah parlemen santri sekolah parlemen wacana program dpr ri upaya tingkat kualitas anggota dewan beri wawasan kena tugas fungsi dpr kalangan non anggota dewan program belum sah dpr ri namun fraksi pkb inisiasi program dengan menggelar sekolah parlemen tuju kalangan santri pelajar sekolah parlemen santri beri paham kaum santri politik kerja parlemenan buka wawasan politik kaum santri nilai kerja anggota dewan kalangan politisi parlemenyang dulu latar belakang santri didik politik rasa utama persiapan kader pondok pesantren orang santri telah tempa akhlak mental pondok pesantren miliki bekal ilmu guna bangun bangsa jalur politik

Gambar 5. 4 Hasil Pre-processing Satu Teks Berita

Kemudian didapatkan kata kunci dari berita yang sudah dinormalisasi pada Gambar 5.4, berikut diantaranya :

Tabel 5. 8 Likelihood studi kasus

Kata	Agama	Ekonomi	Politik	Sosial
pelajar	0,000516	0,0003782	0,000293	0,000243
santri	0,010935	0,0014803	0,002310	0,001920
parlemen	1,0	1,0	1,0	1,0
dpr	0,001530	0,0044609	0,005215	0,003610
sekolah	0,005085	0,0018498	0,002598	0,007199
politik	0,002569	0,0007513	0,002629	0,000484
kaum	0,003335	0,0014998	0,001166	0,001694
parlemenyang	1,0	1,0	1,0	1,0

Nilai pada Tabel 5.8 diambil dari database term_list dalam MySQL yang mana nilai tersebut dihasilkan dari proses data *training* sebelumnya. Untuk data yang bernilai 1,0 yaitu kata yang tidak ditemukan pada term_list dianggap 1,0 agar perhitungan *naïve bayes classifier* tetap didapatkan hasil.

Kemudian dihitung setiap kelas dan dipilihlah nilai terbesar dari keempat kelas, itulah kelas dari berita yang sedang diproses. Berikut perhitungan setiap kelas :

Tabel 5. 9 Perhitungan Naïve bayes classifier pada studi kasus

$P(\text{agama} A1) = 0,000516 \times 0,010935 \times 0,001530$ $\times 0,005085 \times 0,002569 \times 0,003335$ $P(\text{agama} A1) = 3,761 \times 10^{-15}$
$P(\text{ekonomi} A1)$ $= 0,0003782 \times 0,0014803 \times 0,0044609$ $\times 0,0018498 \times 0,0007513 \times 0,0014998$ $P(\text{ekonomi} A1) = 5,205 \times 10^{-18}$
$P(\text{politik} A1) = 0,000293 \times 0,002310 \times 0,005215$ $\times 0,002598 \times 0,002629 \times 0,001166$ $P(\text{politik} A1) = 2,811 \times 10^{-17}$
$P(\text{sosial} A1) = 0,000243 \times 0,001920 \times 0,003610$ $\times 0,007199 \times 0,000484 \times 0,001694$ $P(\text{sosial} A1) = 9,941 \times 10^{-18}$

Dari perhitungan tabel 5.10 diketahui nilai 1,0 untuk kata DPR dan kata parlemen yang tidak dihitung, karena tidak mempengaruhi perhitungan ketika dikalikan dengan satu. Kata parlemen menunjukkan bahwa pada teks berita terdapat kesalahan penulisan sehingga dikenali sebuah kata baru. Dan disimpulkan bahwa berita tersebut nilai terbesar berada pada kelas agama yaitu sebesar $3,761 \times 10^{-15}$. Maka berita pada studi kasus ini masuk ke dalam kelas agama. Karena orientasi kata pendukung lainnya mengarah ke berita agama, walaupun terdapat kata politik tidak selalu berita ini masuk ke dalam kelas politik.

BAB VI PENUTUP

Bab ini berisi tentang beberapa kesimpulan yang dihasilkan berdasarkan penelitian yang telah dilaksanakan. Di samping itu, pada bab ini juga diberikan saran yang dapat digunakan jika penelitian ini akan dikembangkan.

6.1 Kesimpulan

Berdasarkan hasil pengujian program, maka dapat diambil beberapa kesimpulan sebagai berikut :

1. Didapatkan klasifikasi terhadap 435 teks berita yang termasuk dalam data *testing* terdapat 212 berita agama, 79 berita ekonomi, 54 berita politik, dan 90 berita sosial.
2. Didapatkan akurasi rata-rata sebesar 94.71265% dengan presisi 0.94712645 dan recall 1,0 dari data testing sebanyak 435 data.
3. Terkadang dikaji jika kata-kata berita termuat dalam lebih dari satu kelas seperti yang dicontohkan dalam studi kasus pada bab V, harus tetap dihitung kata penting yang muncul. Barulah diketahui berita tersebut masuk kedalam kelas apa.

6.2 Saran

Dalam penelitian selanjutnya sebaiknya dipastikan semua redaksi tidak ada yang salah, karena sedikit berbeda saja akan tidak dikenali sehingga muncul sebagai kata baru yang diterima oleh program. Serta untuk kedepannya dapat dikembangkan dengan jumlah data *training* yang lebih banyak.

DAFTAR PUSTAKA

- [1] Feldman, R & Sanger, J. 2007. *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*. Cambridge University Press: New York.
- [2] Ulfa, Maria. 2007. **Pengertian Naïve Bayes Classifier**. <https://jagoanana.wordpress.com/2007/12/13/naive-bayes-classifier-2/> . Diakses tgl 14 Februari 2017.
- [3] Adriani, M., Asian, J., Nazief, B., Tahaghoghi, S.M.M., Williams, H.E. 2007. Stemming Indonesian: *A Confix-Stripping Approach*. *Transaction on Asian Lantage Information Processing*. Vol. 6, No. 4, Artikel 13. Association for Computing Machinery : New York .
- [4] Tala, Fadillah Z. 2003. *A Study of Stemming Effects on Information Retrieval in Bahasa Indonesia*. *Institute for Logic, Language and ComputationUniversiteit van Amsterdam The Netherlands*. <http://www.illc.uva.nl/Research/Reports/MoL-2003-02.text.pdf>. Diakses tanggal 2 Februari 2017.
- [5] Weiss, S.M., Indurkha, N., Zhang, T., Damerou, F.J. 2005. *Text Mining : Predictive Methods fo Analyzing Unstructured Information*. Springer : New York.
- [6] Dragut, E., Fang, F., Sistla, P., Yu, S. & Meng, W. 2009. *Stop Word and Related Problems in Web Interface Integration*. <http://www.vldb.org/pvldb/2/vldb09-384.pdf>. Diakses tanggal 8 Februari 2017.
- [7] Maria, Ratna. 2003. **Pengertian Data Mining, Text Mining dan Web Mining**. <http://analisis-proses-bisnis-koperasi.blogspot.co.id/2013/04/pengertian-data-mining-text-mining-dan.html> . Diakses tanggal 14 Februari 2017.
- [8] wikipedia. 2004. *Naïve Bayes classifier*. https://en.wikipedia.org/wiki/Naïve_Bayes_classifier. Diakses pada 3 Maret 2017.
- [9] Media Algoritma. 2013. **Perhitungan Metode Naïve bayes classifier**.*Java*. <http://www.metode-algoritma.com/2013/06/conoth-perhitungan-naive-bayes.html>. Diakses pada 3 Maret 2017.

- [10] Brindha,S., Dr.Sukumaran, S., Dr.Prabha, K. 2016. *A Survey on Classification Techniques for Text Mining*. IEEE: India.
- [11] Arifin, Dea Delvia., Shaufiah., Bijaksana, Moch.Arif. 2016. *Enhancing Spam Detection on Mobile Phone Short Message Service (SMS) Performance using FP-Growth and Naïve Bayes Classifier*. IEEE: Indonesia.
- [12] Destuardi, I. 2009. **Klasifikasi Emosi untuk Teks Bahasa Indonesia Menggunakan Metode Naïve Bayes**. ITS: Surabaya.
- [13] Ja Lee, D.L., (1997), *Document Ranking and the Vector-Space Model*. Hong Kong University of Science and Technology. Hongkong.
- [14] Saptano R., Wiranto, Suryono W. D. (2016). **Sistem Klasifikasi Keluhan Di UPT TIK UNS Menggunakan Algoritma Naïve Bayes Classifier**. Seminar Nasional Teknologi Informasi dan Komunikasi. Yogyakarta.
- [15] Kusumadewi, S., 2009. **Klasifikasi Status Gizi Menggunakan Naïve Bayes Classification**. in Communication and Information Technology Journal., Vol: 3, 6-11.
- [16] M. Sokolova dan G. Lapalme. *A systematic analysis of performance measures for classification tasks*. nf. Process. Manag., vol. 45, no. 4, hal. 427–437, 2009.
- [17] Kristian, Papua. 2015. **Media Sebagai Sumber Informasi**. https://www.kompasiana.com/ernard/media-sebagai-sumber-informasi_55106750a33311513dba7d40. Diakses tanggal 24 Januari 2018.
- [18] Drs. Surbakti, EB, MA. 2008. **Awas Tayangan Televisi-Tayangan Misteri dan Kekerasan Mengancam Anak Anda**. Gramedia. Jakarta. Hal 49.
- [19] wikipedia. 2017. **Teorema Bayes**. https://id.wikipedia.org/wiki/Teorema_Bayes. Diakses pada 24 Januari 2018.
- [20] Sie, Jimmy. 2017. **Teorema Bayes**. <https://www.idomaths.com/id/peluang5.php>. Diakses pada 24 Januari 2018.

LAMPIRAN A
Tabel Dataset Training

No	Konten Berita	Normalisasi	Kelas
1	<p>DAHLAN ISKAN MANTAN MENTERI BUMN RABU SORE AKHIRNYA TURUN DARI RUANG PENYIDIKAN LANTAI 5 TINDAK PIDANA KHUSUS KEJAKSAAN TINGGI JAWA TIMUR PEMERIKSAAN HARI KE-3 DAHLAN SEBAGAI SAKSI KASUS DUGAAN KORUPSI 2 ASET PT PWU DENGAN TERSANGKA WISNU WARDHANA BELUM JUGA TUNTAS MESKI SUDAH 3 KALI DIPERIKSA DAN HARUS DILANJUTKAN PADA SENIN MINGGU DEPAN USAI DIPERIKSA DAHLAN AKHIRNYA MAU ANGKAT BICARA DAN MEMBERIKAN PENJELASAN SECARA UMUM TERKAIT PEMERIKSAAN YANG IA HADAPI DIDEPAN PENYIDIK KEJAKSAAN SEMENTARA TERKAIT PEMERIKSAAN UNTUK HARI INI KEPALA KEJAKSAAN TINGGI JAWA TIMUR MENGAKU JIKA TIM PENYIDIK MELONTARKAN SEBANYAK 19 PERTANYAAN TERHADAP DAHLAN ISKAN TOTAL SELAMA 3 HARI INI ADA SEBANYAK 82 PERTANYAAN DARI HASIL PEMERIKSAAN HARI INI SUDAH MULAI MENGKERUCUT TATANAN PELAKSANAAN PENJUALAN ASET YANG DILAKUKAN OLEH DAHLAN ISKAN NAMUN PINGERUCUTAN INI MASIH PADA 1 ASET YANG DI KEDIRI</p>	<p>dahlan iskan mantan menteri bumrn rabu sore turun ruang sisi lantai tindak pidana khusus jaksa jawa timur periksa dahlan saksi duga korupsi aset pt pwu sangka wisnu wardhana tuntas kali periksa lanjut senin minggu periksa dahlan angkat bicara jelas kait periksa hadap depan sidik jaksa kait periksa kepala jaksa jawa timur aku tim sidik lontar dahlan iskan total hasil periksa kerucut tatanan laksana jual aset dahlan iskan kerucut aset diri</p>	Politik
2	<p>MOMENTUM HARI SANTRI NASIONAL INI SANGAT TEPAT UNTUK MEREFLAKSI KEMBALI PERJUANGAN SANTRI DAN KYAI DALAM UPAYA MEMPERTAHANKAN KEMERDEKAAN REPUBLIK INDONESIA 22 OKTOBER ADALAH TANGGAL DIKELUARKANNYA FATWA RESOLUSI JIHAD FATWA TERSEBUT MERUPAKAN AWAL KEBANGKITAN SEMANGAT PERLAWANAN PADA PENJAJAH KETUA LP MAARIF JATIM PROFESOR ABD HARIS MENEGASKAN BAHWA HARI SANTRI HARUS DIPERINGATI OLEH SEMUA MASYARAKAT TIDAK TERBATAS OLEH ORANG ISLAM MELAINKAN MASYARAKAT SECARA KESELURUHAN</p>	<p>momentum santri nasional refleksi juang santri kyai upaya tahan merdeka republik indonesia oktober tanggal keluar fatwa resolusi jihad fatwa bangkit semangat lawan jajah ketua lp maarif jatim profesor abd haris santri ingat masyarakat batas orang islam masyarakat</p>	agama

3	<p>POLRES JOMBANG Mendukung Program Presiden Jokowi yang berupaya memberantas praktek pungli Polres mempermudah pelayanan masyarakat melalui sistem online berbasis hp android fasilitas pelayanan terdiri dari pelayanan perpanjangan sim informasi kecelakaan informasi BPKB serta kritik dan saran sistem online tersebut menurut Kapolres Jombang AKBP Agung Marlianto merupakan salah satu upaya untuk menghindari praktek pungli praktek pungli bisa dihindari sebab fasilitas tersebut bisa memotong panjangnya birokrasi ditambahkan jika masyarakat menemukan praktek pungli bisa melaporkan secara online atau sms sementara itu dengan adanya fasilitas tersebut masyarakat merasa terbantu sebab tidak perlu antri</p>	<p>polres jombang dukung program presiden jokowi upaya berantas praktek pungli polres mudah layan masyarakat sistem online bas hp android fasilitas layan layan panjang sim informasi celaka informasi bpkb kritik saran sistem online kapolres jombang akbp agung marlianto salah upaya hindar praktek pungli praktek pungli hindar fasilitas potong panjang birokrasi masyarakat temu praktek pungli lapor online sms fasilitas masyarakat bantu antri</p>	Sosial
4	<p>Aktifitas di kebun cengkeh desa Puspo Kecamatan Puspo Kabupaten Pasuruan terlihat sibuk para kuli petik mengambil hasil panen cengkeh dengan cekatan panen kali ini hasil yang didapat cukup memuaskan namun tidak mendukung tingginya omzet yang didapat menurut keterangan salah satu petani cengkeh biaya yang dikeluarkan untuk pemupukan dan panen lebih banyak apalagi musim hujan seperti saat ini selain biaya untuk menyewa kuli petik juga diperlukan sewa tempat untuk penjemuran karena tidak memungkinkan untuk menjemur cengkeh di wilayah setempat yang curah hujannya tinggi sehingga penjemuran cengkeh harus dilakukan di daerah lain selain itu harga jual dari petani ke pengepul menurun hingga dua puluh lima persen yakni delapan puluh lima ribu rupiah per kilogram</p>	<p>aktifitas kebun cengkeh desa puspo camat puspo kabupaten pasuruan sibuk kuli petik ambil hasil panen cengkeh ceekat panen kali hasil muas dukung tinggi omzet terang salah tani cengkeh biaya keluar pupu panen musim hujan biaya sewa kuli petik sewa jemur jemur cengkeh wilayah curah hujan jemur cengkeh daerah harga jual tani kepul turun puluh persen delapan puluh ribu rupiah kilogram</p>	ekonomi

LAMPIRAN B
Tabel Dataset Testing

No	Konten Berita
1	<p>PERARUTAN BUPATI SUMENEP SUMENEP TENTANG PERLINDUNGAN TANAH WARGA YANG DIAJUKAN KE PEMERINTAH PROVINSI JAWA TIMUR SUDAH DIPASTIKAN DITOLAK PADAHAL PERBUP TERSEBUT DIMAKSUDKAN BISA MENJADI PENGONTROL AGAR WARGA SUMENEP TIDAK MUDAH MENJUAL BELIKAN TANAH TANPA MEMPERTIMBANGKAN PERUNTUKAN DAN PERTIMBANGAN MATANG DENGAN DITOLAKNYA PERBUP TANAH KARENA PEMKAB DIANGGAP TIDAK PUNYA HAK UNTUK MENGATUR PRIVAT WARGA NAMPAKNYA TIDAK MENGURANGI KOMITMEN PEMKAB UNTUK MEMBIARKAN PEMBALAKAN TANAH BESAR-BESARAN DI SUMENEP PEMKAB BERJANJI AKAN LEBIH MEMPERKETAT PROSES IZIN MENDIRIKAN BANGUNAN JIKA MAKSUD DAN TUJUANNYA TIDAK JELAS SELAMA INI DIKABUPATEN SUMENEP ADA BEBERAPA PENGEMBANG DENGAN DALIH INVESTASI MENYEPELEKAN IZIN MENDIRIKAN BANGUNAN MEREKA SERINGKALI MENGURUS IZIN SETELAH SELESAI MEMBANGUN HAL INI SEBAGAI MODUS DESAKAN AGAR PEMERINTAH LEBIH GAMPANG UNTUK MEMBERIKAN IZIN SEKRETARIS DAERAH KABUPATEN SUMENEP HADI SOETARTO BERJANJI PIHAKNYA TIDAK AKAN MENTOLERIR JIKA ADA OKNUM YANG MENYEPELEKAN IZIN DAN BERANI MENDIRIKAN BANGUNAN TANPA IZIN JIKA ITU TERJADI PIHAKNYA TIDAK AKAN SEGAN-SEGAN BERKOORDINASI DENGAN PIHAK TERKAIT UNTUK MENERTIBKAN SOETARTO MENAMBAHKAN PIHAKNYA AKAN MEMROSES IZIN JIKA PEMOHON YANG TERTERA DISERTIFIKAT DATANG SENDIRI KE PEMKAB JIKA YANG DATANG BUKAN NAMA YANG TERTERA MAKA DIPASTIKAN PERMOHONANNYA AKAN DITOLAK KECUALI MEMBALIK NAMA SERTIFIKAT ATAU MEMBAWA SURAT KUASA SELAIN ITU SEBELUM DITERBITKAN IZIN PIHAKNYA AKAN KOORDINASIKAN DENGAN PIHAK TERKAIT DAN SELANJUTNYA AKAN MELAKUKAN SURVEI LOKASI</p>
2	<p>UNTUK PERTAMA KALINYA DI KABUPATEN SIDOARJO JAWA TIMUR PENGURUS CABANG IPNU IPPNU DILANTIK OLEH KETUA UMUM IPNU DAN IPPNU PUSAT DALAM PELANTIKAN INI JUGA DIHADIRI KETUA PENGURUS WILAYAH IPNU IPPNU JAWA TIMUR WAKIL BUPATI SIDOARJO NUR ACHMAD SYAIFUDDIN PENGURUS CABANG NAHDLATUL ULAMA SIDOARJO FORKOPIMDA DAN BEBERAPA ORAGANISASI KEPEMUDAAN PELANTIKAN PENGURUS CABANG IPNU IPPNU SIDOARJO PERIODE DUA RIBU ENAM BELAS-DUA RIBU DELAPAN BELAS INI BERTAJUK "DENGAN KERJA DAN KARYA NYATA KITA WUJUDKAN PELAJAR SANTRI YANG MAJU MANDIRI DAN BERPERADABAN FAHRIZAL MU'AFI KETUA PC IPNU SIDOARJO MENGATAKAN IPNU-IPPNU SIDOARJO AKAN BEKERJASAMA DENGAN BEBERAPA PIHAK UNTUK MENCEGAH PEREDARAN NARKOBA DAN KENAKALAN REMAJA DI LINGKUNGAN SEKOLAH</p>
3	<p>PULUHAN PETUGAS KEPOLISIAN DARI SATLANTAS POLRES SUMENEP MADURA JAWA TIMUR MEMBERHENTIKAN SEMUA PENGENDARA YANG MELINTAS DI JALAN TRUNOJOYO KOTA SUMENEP BAIK PENGENDARA RODA DUA MAUPUN RODA EMPAT SEMUA PENGENDARA DIPERIKSA KELENGKAPAN SURAT-SURAT KENDARAAN MAUPUN SURAT IZIN MENGEMUDI DALAM OPERASI INI POLISI BERHASIL MENILANG PULUHAN PENGENDARA YANG TIDAK MELENGKAPI SURAT-SURATNYA OPERASI INI DILAKSANAKAN DALAM RANGKA MENCIPTAKAN KONDISI AMAN DAN KONDISIF DI KABUPATEN SUMENEP MENJELANG HARI RAYA IDUL ADHA KARENA BIASANYA SETIAP AKAN DATANG PERAYAAN HARI BESAR TINDAKAN KRIMINALITAS SELALU MENINGKAT NAMUN PADA OPERASI KALI INI POLISI BELUM MENEMUKAN KENDARAAN BODONG ATAU HASIL</p>

	<p>TINDAKAN KRIMINAL SALAH SATU PENGENDARA MOTOR YANG DITILANG PETUGAS H. HIDAYATULLAH MENGAKU IA DITILANG KARENA LUPA MEMBAWA STNK MOTORNYA SEHINGGA IA HARUS MENGURUSNYA DI PENGADILAN PADA TANGGAL 16 SEPTEMBER YANG AKAN DATANG SEMENTARA KANIT TURJAWALI SATLANTAS POLRES SUMENEP IPTU RIZAL NUGRAWIJAYA MENGATAKAN OPERASI KALI INI DILAKUKAN DIBEBERAPA TITIK DI WILAYAH SUMENEP SEMENTARA UNTUK PELANGGAR DIDOMINASI OLEH PENGENDARA YANG TIDAK LENGKAP SURAT-SURATNYA SELAIN ITU MASIH BANYAK PENGENDARA TIDAK MENGHIDUPKAN LAMPU DI SIANG HARI</p>
4	<p>BEGINILAH KESEHARIAN SUKMINI LIMA PULUH TAHUN WARGA DUSUN JEJERAN DESA GUNUNGGANGSIR KECAMATAN BEJI KABUPATEN PASURUAN YANG NAMPAK SIBUK MEMASAK KEPERLUAN WARUNGNYA GUNA MELAYANI MAKAN SIANG PARA PELANGGANNYA YANG KEBANYAKAN BURUH PABRIK SETEMPAT PEREMPUAN PARUH BAYA INI ADALAH KORBAN DARI DODI WAHYUDI YANG MERUPAKAN PENGIKUT DARI DIMAS KANJENG TAAT PRIBADI DI PROBOLINGGO KORBAN ADALAH SALAH SATU TETANGGANYA SENDIRI MENURUT KETERANGAN SUKMINI PENIPUAN TERSEBUT BERMULA SEKITAR DELAPAN BULAN YANG LALU DODI SEDANG MEMBELI KOPI DI WARUNGNYA DODI BERPENAMPILAN MEYAKINKAN DAN DI KAWAL OLEH BODYGUARD SETELAH BERCERITA BANYAK HAL DODI MENAWARKAN UNTUK MEMBELIKAN MOBIL BARU DENGAN HARGA MURAH KEPADA KORBAN SEHINGGA KORBAN PUN TERTARIK DENGAN TAWARAN TERSEBUT NAMUN DENGAN SYARAT KORBAN HARUS MEMBAYAR UANG TUNAI SEBESAR SERATUS DUA PULUH JUTA RUPIAH PADA DODI UNTUK PEMBAYARAN MOBIL SETELAH DIRINYA MELAKUKAN PEMBAYARAN MOBIL YANG DIJANJIKAN TIDAK KUNJUNG DATANG BAHKAN PELAKU TERKESAN MENGHINDAR PADA SAAT DI TAGIH JANJI TIDAK PUTUS ASA KORBAN TERUS MENAGIH JANJI KEPADA PELAKU NAMUN HINGGA SAAT INI TIDAK ADA KEJELASAN DARI PELAKU KORBAN BERHARAP AGAR PELAKU MENGEMBALIKAN UANG MILIKNYA</p>

LAMPIRAN C
Tabel Stoplist

NO	STOPLIST	NO	STOPLIST
1	ada	58	banyak
2	adakala	59	bapak
3	adakalanya	60	baru
4	adalah	61	barusan
5	adanya	62	bawa
6	adapun	63	bawah
7	aduh	64	beberapa
8	agak	65	begini
9	agaknya	66	beginian
10	agar	67	banyak
11	akan	68	bapak
12	akankah	69	baru
13	akhir	70	barusan
14	akhiri	71	bawa
15	akhirnya	72	bawah
16	akibat	73	beberapa
17	aku	74	begini
18	akulah	75	beginian
19	alih	58	banyak
20	amat	59	bapak
21	amatlah	60	baru
22	anda	61	barusan
23	andalah	62	bawa
24	antar	63	bawah
25	antara	64	beberapa
26	antaranya	65	begini
27	apa	66	Beginian
28	apaan	67	beginikah
29	apabila	68	beginilah

30	apakah	69	begitu
31	apalagi	70	begitukah
32	apatah	71	begitulah
33	arti	72	begitupun
34	artinya	73	bekerja
35	asal	74	belakang
36	asalkan	75	belakangan
37	atas	76	belas
38	atau	77	belum
39	ataukah	78	belumlah
40	ataupun	79	benar
41	awal	80	benarkah
42	awalnya	81	benarlah
43	awas	82	berada
44	bagai	83	berakhir
45	bagaikan	84	berakhirilah
46	bagaimana	85	berakhirnya
47	bagaimanakah	86	berapa
48	bagaimanapun	87	berapakah
49	bagi	88	berapalah
50	bagian	89	berapapun
51	bahkan	90	berarti
52	bahwa	91	berawal
53	bahwasanya	92	berbagai
54	baik	93	berdatangan
55	bakal	94	beri
56	bakalan	95	berikan
57	balik	96	berikut

Keterangan: sebagian data stoplist yang diambil dari sumber <https://github.com/masdevi/id-stopwords> pada tanggal 1 Desember 2017.

LAMPIRAN D
Tabel Kata Ekstrasi dan Nilainya

Kata	Agama	Ekonomi	Politik	Sosial	Total
abad	0,00103	0,000754	0,000879	0,001214	10
abadi	0,000516	0,000756	0,000294	0,000487	3
abadia	0,000258	0,000756	0,000294	0,000243	1
abai	0,001031	0,001132	0,000586	0,000486	7
abalabal	0,000258	0,000378	0,000294	0,000487	1
abar	0,001031	0,000378	0,00088	0,000486	6
abbas	0,000516	0,000378	0,000294	0,000243	1
abd	0,004612	0,001122	0,00233	0,001209	30
abddul	0,000516	0,000378	0,000294	0,000243	1
abdi	0,001031	0,000378	0,000587	0,000486	5
abdul	0,0018	0,001128	0,001754	0,00097	16
abdullah	0,000516	0,000378	0,000294	0,000243	1
abdulllah	0,000516	0,000378	0,000294	0,000243	1
abdulrachman	0,000258	0,000378	0,000294	0,000487	1
abdulrahman	0,000258	0,000378	0,000294	0,000487	1
abdur	0,001289	0,000378	0,000587	0,000243	5
abdurrahman	0,00129	0,000378	0,000293	0,000243	4
abdusshomad	0,000516	0,000378	0,000294	0,000243	1
abhaya	0,000258	0,000378	0,000294	0,000487	1
abidin	0,000258	0,000756	0,000294	0,000243	1
absah	0,000774	0,000378	0,000294	0,000243	2
absen	0,000258	0,000756	0,000294	0,000487	2
absensi	0,000258	0,000756	0,000294	0,000243	1
abu	0,009852	0,013781	0,014481	0,013973	187
abuya	0,000516	0,000378	0,001173	0,000486	5
ac	0,015913	0,005372	0,008725	0,010103	149
acak	0,000258	0,000378	0,000881	0,000487	3
acaman	0,000516	0,000378	0,000294	0,000243	1
acang	0,000258	0,000756	0,000294	0,000243	1
acara	0,006366	0,000742	0,00347	0,004566	54

aceh	0,000258	0,000378	0,000587	0,000243	1
achamd	0,000774	0,000378	0,000294	0,000487	3
achmad	0,00206	0,000754	0,000878	0,000486	11
achyar	0,000516	0,000378	0,000294	0,000243	1
acmad	0,000258	0,000378	0,000587	0,000243	1
acu	0,000516	0,001133	0,001173	0,000243	6
acuh	0,000258	0,000756	0,000294	0,000243	1
ada	0,018919	0,008451	0,009442	0,015335	197
adad	0,000774	0,000378	0,000587	0,000243	3
adakan	0,000516	0,000378	0,000587	0,000243	2
adalah	0,000258	0,000378	0,000294	0,000487	1
adanaya	0,000258	0,000378	0,000587	0,000243	1
adat	0,001545	0,000377	0,000293	0,001457	10
ade	0,0023	0,001864	0,004065	0,003858	40
adelia	0,000258	0,000378	0,000294	0,000487	1
adha	0,006368	0,006306	0,001736	0,002163	53
adhim	0,000258	0,000378	0,000587	0,000243	1
adi	0,01592	0,010866	0,015495	0,014362	210
adik	0,003853	0,000376	0,001168	0,000969	20
adil	0,001028	0,001127	0,003799	0,000485	18
administrasi	0,000773	0,000755	0,00088	0,000729	7
adu	0,005067	0,006993	0,004025	0,005979	74
aduk	0,000516	0,000756	0,000294	0,000487	3
advokasi	0,000258	0,000378	0,000294	0,000487	1
adz	0,001288	0,000754	0,000586	0,000729	8
adzan	0,000258	0,000756	0,000294	0,00073	3
adzkia	0,000516	0,000378	0,000294	0,000243	1

Keterangan: Nilai total didapatkan dari perhitungan jumlah kata pada semua kelas berdasarkan banyaknya kemunculan kata. Angka tersebut diperoleh sebelum nilai *likelihood* dihasilkan.

LAMPIRAN E
Tabel Kata Kunci suatu Kelas

Agama	Ekonomi	Politik	Sosial
nahdlatul	jual	negeri	warga
ulama	turun	bupati	jalan
nu	ribu	partai	masyarakat
islam	bank	hukum	dinas
pcnu	usaha	ketua	data
pbnu	wilayah	pimpin	asal
agama	tingkat	pkb	anti
ranting	rupiah	polisi	air
mwc	makmur	politik	awas
pondok	bawa	dprd	temu
pesantren	sejahtera	korup	tinggal
umat	ton	korupsi	arah
ibadah	gram	polisi	nilai
haji	operasi	saksi	ikut
nkri	pasar	tangkap	perhati
waliyulloh	ternak	terima	ulang
idul	tanam	calo	masuk
fitri	ekonomi	kantor	awet
adha	upah	gubernur	rumah
ramadhan	gaji	calon	program
kurban	kas	pidana	korban
doa	kilo	jaksa	aman
santri	pangsa	tindak	proses
ngaji	mes	lapor	jadi
kitab	dapat	calon	para
kiai	wiraswasta	gedung	kondisi
masjid	saham	pemkot	peduli
muslim	capai	pemda	sara
makam	butuh	pemerintah	kuat
puasa	kilo	presiden	ajang

sedekah	naik	hak	hasil
amal	uang	wajib	ras
jariah	pokok	sidang	harap
allah	untung	putus	salah
swt	rugi	tanda	sosial
maulid	banding	tangan	desa
kanjeng	harga	dewan	daerah
nabi	pajak	wakil	kerja
muhammad	amnesty	rakyat	ikhlas
alquran	tax	media	orang
musholla	barang	pers	luar
takbir	produk	negara	kota
sholat	biaya	tanah air	nasi
zakat	tinggi	dpr	laku
ketupat	rendah	dpd	air
fakir	target	dprd	bak
miskin	mudah	mpr	nilai
amil	juta	mahkamah	awan
muallaf	tingkat	agung	cuaca
aliran	kurang	konstitusi	sekolah
golongan	buka	pilih	siswa
nusantara	penuh	pemilu	hidup
kua	nganggur	pilkada	hadap
nikah	belanja	kepala	sabar
gus	anggar	golkar	anggap
sah	dana	misi	pemuda
mui	defisit	buruh	selenggara
halal	impor	demo	asing
habib	ekspor	anggota	dasar
syekh	beli	pdip	proses
ziarah	persen	saksi	lomba
silaturahmi	nego	pilkades	kasih

BIODATA PENULIS



Penulis bernama lengkap Ain Nuor Roifa, lahir di Gresik, 22 Desember 1994. Penulis merupakan anak pertama dari pasangan Nyoto dan Karni, serta memiliki satu adik laki-laki yaitu Ilham Arie Hanafi. Penulis telah menempuh pendidikan formal yaitu TK Charma wanita Kedayang, MI Al-Falah Ma'arif Kedayang, MTs Ma'arif Sidomukti Kebomas dan SMAN 1 Gresik. Setelah lulus SMA pada tahun 2013, Penulis melanjutkan pendidikan S1 di Jurusan Matematika melalui jalur SNMPTN dan mengambil rumpun mata kuliah pada semester 5 yaitu Ilmu Komputer.

Pengalaman organisasi yang sudah dilakukan yaitu menjadi anggota UKM LPM (Lembaga Pers Mahasiswa) 1.0 ITS pada tahun 2013/2014. Pada tahun 2014/2015 aktif menjadi Staf Departemen RT (Rumah Tangga) di LMB (Lembaga Minat Bakat) ITS. Kemudian pada tahun 2015/2016 menjadi BPH UKM Penalaran ITS dan penulis juga aktif dalam kepanitiaan acara tingkat Nasional yaitu Olimpiade Matematika ITS sebagai *Crew of Accomodation and Transportation* di dalam kampus. Pada tahun 2016/2017 penulis aktif di organisasi ekstra kampus dan diamanahi sebagai wakil ketua satu. Selain itu, penulis juga pernah melakukan Kerja Praktek di PT. Petrokimia Kayaku Gresik dengan *project* pembuatan sistem prediksi penjualan produk dari perusahaan setiap bulannya agar semakin meningkat. Adapun mengenai informasi lebih lanjut atau ingin berdiskusi mengenai Tugas Akhir ini dapat ditujukan ke email penulis ainnuorroifa@gmail.com.

