



ITS
Institut
Teknologi
Sepuluh Nopember

TUGAS AKHIR - KS141501

**KLASIFIKASI LAPORAN KRIMINALITAS PADA
SITUS LAPOR! DENGAN NAÏVE BAYES
CLASSIFIER (NBC) MENGGUNAKAN BAHASA
PEMROGRAMAN R**

RIZKA AMALIA PUTRI
NRP 5212 100 072

Dosen Pembimbing
Renny Pradina Kusumawardani, S.T, M.T
Irmasari Hafidz, S.Kom, M.Sc

JURUSAN SISTEM INFORMASI
Fakultas Teknologi Informasi
Institut Teknologi Sepuluh Nopember
Surabaya 2016

FINAL PROJECT - KS141501

***CLASSIFICATION OF CRIME REPORT ON THE
SITE OF LAPOR! WITH NAÏVE BAYES
CLASSIFIER (NBC) USING R PROGRAMMING***

**RIZKA AMALIA PUTRI
NRP 5212 100 072**

**Supervisor
Renny Pradina Kusumawardani, S.T, M.T
Irmasari Hafidz, S.Kom, M.Sc**

**DEPARTMENT OF INFORMATION SYSTEM
Faculty of Information Technology
Institute of Technology Sepuluh Nopember
Surabaya 2016**

LEMBAR PENGESAHAN

**KLASIFIKASI LAPORAN KRIMINALITAS PADA
SITUS LAPOR! DENGAN *NAÏVE BAYES CLASSIFIER*
(NBC) MENGGUNAKAN BAHASA PEMROGRAMAN R**

TUGAS AKHIR

Disusun Untuk Memenuhi Salah Satu Syarat
Memperoleh Gelar Sarjana Komputer
pada

Jurusan Sistem Informasi
Fakultas Teknologi Informasi
Institut Teknologi Sepuluh Nopember

Oleh:

RIZKA AMALIA PUTRI

NRP. 5210 100 072

Surabaya, Juni 2016

**KETUA
JURUSAN SISTEM INFORMASI**

Dr. Ir. Ari S. Tjahyanto, M. Kom.

NIP. 19650318 199102 001



**KLASIFIKASI LAPORAN KRIMINALITAS PADA
SITUS LAPOR! DENGAN *NAÏVE BAYES CLASSIFIER*
(NBC) MENGGUNAKAN BAHASA PEMROGRAMAN R**

Nama Mahasiswa : Rizka Amalia Putri
NRP : 5212 100 072
Jurusan : Sistem Informasi FTIF – ITS
Pembimbing I : Renny Pradina K., S.Kom, M.Kom
Pembimbing II : Irmasari Hafidz, S.Kom, M.Sc

ABSTRAK

Data LAPOR! yang tergolong sebagai laporan kriminalitas merupakan laporan yang diarsipkan dan tidak memiliki prosedur tindak lanjut. Laporan ini berasal dari pengguna yang menjadi korban atau saksi dari kejadian kriminalitas dan mereka laporkan melalui situs LAPOR!. Pada penelitian ini, dibangun model klasifikasi dengan Naïve Bayes Classifier (NBC) untuk mengetahui kategori kriminalitas yang tepat diterapkan pada data LAPOR!. Kategori kriminalitas yang digunakan bersumber dari Reserse Kriminal Umum POLDA Jatim. Hasil penelitian menunjukkan bahwa kategori 5 subdit lebih baik digunakan pada data penelitian ini daripada kategori 30 jenis kejahatan. Semua model pada percobaan subdit dapat menjadi pengklasifikasi walaupun dengan performa yang tidak semuanya baik. Sedangkan untuk model dengan 30 kategori, tidak semua dapat menjadi pengklasifikasi karena keterbatasan data. Model klasifikasi terbaik yang dihasilkan mampu bekerja akurat pada kelas penyuaipan dengan nilai f – measure 100%. Disusul dengan kelas perjudian, penipuan, pengrusakan, korupsi, dan narkoba dengan nilai f – measure lebih dari 80%. Sedangkan untuk kelas pencurian ringan, lantasi, perkosaan,

penganiayaan, penghinaan, curas, curat, jabatan, penggelapan, pemerasan, curanmor, dan implikasi kontijensi memiliki nilai f – measure kurang dari 80%. Sementara untuk 12 kelas lainnya tidak mampu untuk melakukan proses klasifikasi. Sedangkan performa yang dihasilkan pada kelas subdit yaitu subdit II dan II mampu melakukan kasifikasi dengan nilai f – measure masing – masing 85.5% dan 85.4%. Disusul dengan model subdit V sebesar 77.7% dan subdit I sebesar 66.6%. Sedangkan model IV hanya mampu mengklasifikasi model dengan f – measure 14.2%. Pada penelitian ini juga dibentuk visualisasi berupa grafik alir dan wordcloud yang ditampilkan pada aplikasi R berbasis web yang interaktif yaitu Shiny.

Kata Kunci — Kriminalitas, Klasifikasi, Text Mining, Naïve Bayes Classifier (NBC), R, Shiny, LAPOR!

***CLASSIFICATION OF CRIME REPORT ON THE SITE
OF LAPOR! WITH NAÏVE BAYES CLASSIFIER (NBC)
USING R PROGRAMMING***

Name : Rizka Amalia Putri
NRP : 5212 100 072
Department : Sistem Informasi FTIF – ITS
Supervisor 1 : Renny Pradina K, S.Kom, M.Kom
Supervisor 2 : Irmasari Hafidz, S.Kom, M.Sc

ABSTRACT

LAPOR! dataset which classified as a crime report are the archived reports and there are no follow-up procedures. This report comes from users who are victims or witness of crime then they reported through the site of LAPOR!. In this research, we deploy classification model with Naive Bayes Classifier (NBC) to determine the categories of crime are proper to the data of LAPOR!. We obtain categories from the General Criminal Investigation Regional Police of East Java. The results of this research showed that the category of 5 Sub-Directorate better used in this research data than 30 categories of crimes. All models in the Sub-Directorate scenarios can be classifying although with a performance that is not good. However, the model with 30 categories, not all can be a classifier because of data limitations. Best classification model able to work accurately on bribery class with f - measure of 100%. Followed by classes with f - measure more than 80% are gambling, fraud, vandalism, corruption, and drugs. While classes that have an f - measure less than 80% are lightweight theft, traffic crimes, rape, torture, humiliation, theft with violence, theft by weighting, position, embezzlement, extortion, theft of motor vehicles, and

the implications of contingency. The number of instances of the other 12 classes are too small to train adequately accurate classifiers, resulting in f - measures of 0 %. In the Sub-Directorate scenario, resulting in f-measure of 85.5% and 85.4% for class II and II. Followed by f-measure of 77.7% and 66.6% for class V and I. While class IV is only able to classify model in f - measure of 14.2%. In this research, visualizations of the data are also created in the form of stream graph and word clouds. These are then made available to the public using R's web-based publishing library, Shiny.

Keywords — Crime, Classification, Text Mining, Naïve Bayes classifier (NBC), R, Shiny, LAPOR!

DAFTAR ISI

LEMBAR PENGESAHAN.....	iii
LEMBAR PERSETUJUAN.....	iv
ABSTRAK.....	v
ABSTRACT.....	vii
KATA PENGANTAR	ix
DAFTAR ISI.....	xi
DAFTAR GAMBAR	xvii
DAFTAR TABEL.....	xix
DAFTAR KODE.....	xxi
BAB I PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	4
1.3. Batasan Tugas Akhir	4
1.4. Tujuan Tugas Akhir	4
1.5. Manfaat Tugas Akhir	5
1.6. Sistematika Penulisan.....	5
1.7. Relevansi.....	6
BAB II TINJAUAN PUSTAKA.....	7
2.1. Studi Sebelumnya.....	7
2.2. Dasar Teori.....	11
2.2.1. Kriminalitas.....	11
2.2.2. LAPOR!	21

2.2.3.	<i>Text Mining</i>	25
2.2.4.	Algoritma Pengklasifikasi	29
2.2.5.	Evaluasi Performa Pengakurasion	31
BAB III	METODE PENELITIAN	35
3.1.	Diagram Alir Pengerjaan Tugas Akhir	35
3.2.	Tahapan Pengerjaan Tugas Akhir.....	37
BAB IV	PERANCANGAN	39
4.1.	Data.....	39
4.2.	Pengumpulan Data.....	40
4.3.	Penggabungan Data	40
4.4.	Pelabelan Data	41
4.4.1.	Jenis Kejahatan.....	41
4.4.2.	Subdit.....	42
4.5.	Metode Penelitian	43
BAB V	IMPLEMENTASI	49
5.1.	Data Implementasi.....	49
5.2.	Lingkungan Implementasi	49
5.3.	Klasifikasi.....	50
5.3.1.	Pengerjaan Algoritma Klasifikasi.....	50
5.3.2.	Percobaan Klasifikasi	59
5.4.	Visualisasi.....	79
5.4.1.	<i>Streamgraph</i>	80
5.4.2.	<i>Wordcloud</i>	82
5.4.3.	<i>Shiny</i>	83
BAB VI	HASIL DAN PEMBAHASAN.....	88

6.1	Analisis Model Klasifikasi	88
6.1.1	Hasil Percobaan 1	89
6.1.2	Hasil Percobaan 2	89
6.1.3	Hasil Percobaan 3	90
6.1.4	Hasil Percobaan 4	91
6.1.5	Hasil Percobaan 5	92
6.1.6	Hasil Percobaan 6	92
6.1.7	Hasil Percobaan 7	93
6.1.8	Hasil Percobaan Subdit	95
6.2	Analisis Performa Model Klasifikasi	96
6.3	Pembahasan Kategori Klasifikasi	99
6.4	Analisis Hasil Visualisasi	100
BAB VII KESIMPULAN DAN SARAN		108
7.1.	Kesimpulan	108
7.2.	Saran	109
DAFTAR PUSTAKA		112
DAFTAR LAMPIRAN		116
LAMPIRAN A		1
A.1	Kode Program Percobaan	1
A.1.1	Kode Program Percobaan 1	1
A.1.2	Kode Program Percobaan 2	2
A.1.3	Kode Program Percobaan 3	4
A.1.4	Kode Program Percobaan 4	6
A.1.5	Kode Program Percobaan 5	9
A.2	Shiny	13

A.2.1	Server.R.....	13
A.2.2	Ui.R	14
A.2.3	Global.R.....	15
LAMPIRAN B.....		1
B.1	Daftar <i>Stopwords</i> Indonesia	1
B.2	Analisis Hasil 20 Kata Teratas	5
B.3	Proses Penentuan Daftar Kata untuk <i>Dictionary</i>	11
LAMPIRAN C.....		1
C.1	Akibat Orang Luka	1
C.2	Curanmor.....	1
C.3	Curas.....	2
C.4	Curat	2
C.5	Himbauan	3
C.6	Hutang Piutang	3
C.7	Implikasi Kontijensi	4
C.8	Jabatan	4
C.9	Kebakaran.....	5
C.10	Kesusilaan.....	5
C.11	Korupsi	6
C.12	Lain-Lain	6
C.13	Lantas	7
C.14	Narkoba	7
C.15	Pemalsuan Surat	8
C.16	Pembakaran	8
C.17	Pembunuhan	9

C.18	Pemerasan	9
C.19	Penadahan	10
C.20	Penculikan	10
C.21	Pencurian Ringan	11
C.22	Penganiayaan.....	11
C.23	Pengerusakan.....	12
C.24	Penggelapan	12
C.25	Penghinaan	13
C.26	Penipuan.....	13
C.27	Penyuapan	14
C.28	Perdagangan Manusia	14
C.29	Perjudian	15
C.30	Perkosaan	15
LAMPIRAN D	1
D.1	Kebakaran	1
D.2	Pembakaran	1
D.3	Jabatan.....	1
D.4	Pemalsuan Surat.....	2
D.5	Penipuan.....	2
D.6	Penggelapan	2
D.7	Sumpah Palsu	2
D.8	Utang Piutang.....	3
D.9	Jabatan.....	3
D.10	Curat.....	4
D.11	Curas	4

D.12	Curanmor.....	5
D.13	Pembunuhan	5
D.14	Pemerasan/ Ancaman	6
D.15	Perjudian.....	6
D.16	Merusak Barang.....	7
D.17	Penculikan	7
D.18	Penghinaan	7
D.19	Aniaya Ringan.....	8
D.20	Akibat Orang Luka	8
D.21	Pencurian Ringan.....	9
D.22	Penadahan.....	9
D.23	Menerima Suap.....	9
D.24	Penganiayaan.....	10
D.25	Perkosaan.....	10
D.26	Kesusilaan.....	10
LAMPIRAN E.....		1
E.1	2014 - Triwulan 1 (Januari – Maret 2014).....	1
E.2	2014 - Triwulan 2 (April – Juni 2014)	2
E.3	2014 - Triwulan 3 (Juli – September 2014).....	3
E.4	2014 - Triwulan 4 (Oktober – Desember 2014)	4
E.5	2015 - Triwulan 1 (Januari – Maret 2015).....	5
E.6	2015 - Triwulan 2 (April – Juni 2015)	6
E.7	2015 - Triwulan 3 (Juli – September 2015).....	7
BIODATA PENULIS.....		113
UCAPAN TERIMA KASIH		115

DAFTAR GAMBAR

Gambar 2.1 Halaman utama situs LAPOR!	22
Gambar 2.2 Alur kerja LAPOR!	23
Gambar 2.3 Kerangka Proses Analisis <i>Text Mining</i>	26
Gambar 2.4 <i>Confusion Matrix</i>	31
Gambar 3.1 Diagram Alir Pengerjaan Tugas Akhir.....	36
Gambar 4.1 Dataset dengan Atribut Terpilih	41
Gambar 4.2 Metode Penelitian Tahap 3, 4, dan 5	44
Gambar 4.3 Metode Penelitian Tahap 3	45
Gambar 4.4 Metode Penelitian Tahap 4.....	45
Gambar 4.5 Metode Penelitian Tahap 5	46
Gambar 5.1 Proses Penemuan Percobaan Klasifikasi Terbaik	60
Gambar 5.2 Algoritma Percobaan 1	62
Gambar 5.3 Algoritma Percobaan 2	63
Gambar 5.4 Statistik Jumlah Laporan	64
Gambar 5.5 Algoritma Percobaan 3	65
Gambar 5.6 Contoh <i>Wordcloud</i> Tanpa Eliminasi Kata.....	66
Gambar 5.7 Algoritma Percobaan 4.....	68
Gambar 5.8 Algoritma Percobaan 5	70
Gambar 5.9 Lanjutan Algoritma Percobaan 5.....	71
Gambar 5.10 Algoritma Percobaan 6.....	73
Gambar 5.11 Lanjutan Algoritma Percobaan 6.....	73
Gambar 5.12 Algoritma Percobaan 7	76
Gambar 5.13 Lanjutan Algoritma Percobaan 7	76
Gambar 5.14 Distribusi Laporan Subdit	77
Gambar 5.15 Skenario Algoritma Subdit 1	78
Gambar 5.16 Algoritma Percobaan Subdit 2.....	79
Gambar 5.17 Lanjutan Algoritma Percobaan Subdit 2	79
Gambar 5.18 Hasil Visualisasi <i>Streamgraph</i> Pada Triwulan 4	81
Gambar 5.19 Hasil Visualisasi <i>Wordcloud</i> Pada Kelas Korupsi	83

Gambar 5.20 Tab <i>Shiny</i> yang Menampilkan Visualisasi <i>Wordcloud</i>	84
Gambar 5.21 Menu Pilihan Jenis Kejahatan Pada <i>Shiny</i>	84
Gambar 5.22 Menu Konfigurasi <i>Minimum Frequency</i> dan <i>Maximum Number of Words</i> Pada <i>Shiny</i>	85
Gambar 5.23 Tab <i>Shiny</i> yang Menampilkan Visualisasi <i>Streamgraph</i>	85
Gambar 5.24 Menu Pilihan Periode	86
Gambar 5.25 Menu Pilihan Jenis Kejahatan	86
Gambar 6.1 <i>Timeline Stram Graph</i> 21 Bulan.....	101
Gambar 6.2 Pola Laporan pada Periode 2014 – Triwulan 1.	101
Gambar 6.3 Pola Laporan pada Periode 2015 – Triwulan 1.	102
Gambar 6.4 Pola Laporan pada Periode 2014 – Triwulan 2.	103
Gambar 6.5 Pola Laporan pada Periode 2015 – Triwulan 2.	103
Gambar 6.6 Pola Laporan pada Periode 2014 – Triwulan 3.	104
Gambar 6.7 Pola Laporan pada Periode 2015 – Triwulan 3.	104
Gambar 6.8 Pola Laporan pada Periode 2014 – Triwulan 4.	105
Gambar E.1 <i>Streamgraph</i> 2014 – Triwulan 1	1
Gambar E.2 <i>Streamgraph</i> 2014 – Triwulan 2	2
Gambar E.3 <i>Streamgraph</i> 2014 – Triwulan 3	3
Gambar E.4 <i>Streamgraph</i> 2014 – Triwulan 4	4
Gambar E.5 <i>Streamgraph</i> 2015 – Triwulan 1	5
Gambar E.6 <i>Streamgraph</i> 2015 – Triwulan 2	6
Gambar E.7 <i>Streamgraph</i> 2015 – Triwulan 3	7

DAFTAR TABEL

Tabel 2.1 Jenis Kejahatan Konvensional	15
Tabel 2.2 Jenis Kejahatan Implikasi Kontijensi	20
Tabel 2.3 Jenis Kejahatan Transnasional	21
Tabel 4.1 Contoh Data Laporan Beserta Atribut.....	39
Tabel 4.2 Label Jenis Kejahatan	41
Tabel 4.3 Label Subdit	43
Tabel 5.1 Lingkungan Uji Coba Perangkat Keras.....	49
Tabel 5.2 Lingkungan Uji Coba Perangkat Lunak.....	50
Tabel 5.3 Package dan Fungsi R Studio dalam Pengerjaan Algoritma Klasifikasi	50
Tabel 5.4 Perbandingan Laporan Sebelum dan Sesudah Melewati Proses <i>Case Folding</i>	52
Tabel 5.5 Perbandingan Laporan Sebelum dan Sesudah Melewati Proses <i>Filtering</i>	54
Tabel 5.6 Perbandingan Laporan Sebelum dan Sesudah Melewati Proses <i>Stopwords</i>	55
Tabel 5.7 Kutipan Perbandingan 20 Kata Teratas.....	68
Tabel 5.8 Daftar Kelas Setiap Model Pada Percobaan 6.....	72
Tabel 5.9 Daftar Kelas Setiap Model Pada Percobaan 7.....	74
Tabel 5.10 Daftar Kelas Setiap Model Pada Percobaan Subdit	78
Tabel 6.1 Hasil Akurasi Percobaan 2	89
Tabel 6.2 Hasil Akurasi Percobaan 3	90
Tabel 6.3 Hasil Akurasi Percobaan 4	91
Tabel 6.4 Hasil Akurasi Percobaan 6	93
Tabel 6.5 Hasil Akurasi Percobaan 7	94
Tabel 6.6 Hasil Akurasi Percobaan Subdit.....	95
Tabel 6.7 Nilai <i>Precision</i> , <i>Recall</i> , dan <i>F – Measure</i> dari Percobaan Subdit.....	95
Tabel 6.8 <i>Precision</i> , <i>Recall</i> , dan <i>F – Measure</i> Percobaan 7 ..	96
Tabel 6.9 Jumlah Dataset Pada Kelas 19 hingga 29.....	98
Tabel 6.10 Hasil Visualisasi <i>Wordcloud</i>	106

Tabel B.1 Daftar Stopwords Indonesia.....	1
Tabel B.2 Analisis 20 Kata Teratas Kelas Penipuan, Pemerasan, Narkoba, Perjudian, Curat.....	5
Tabel B.3 Analisis 20 Kata Teratas Kelas Implikasi Kontijensi, Himbauan, Korupsi, Curas, Penggelapan	6
Tabel B.4 Analisis 20 Kata Teratas Kelas Kesusilaan, Penghinaan, Pencurian Ringan, Pembunuhan, Jabatan	8
Tabel B.5 Analisis 20 Kata Teratas Kelas Penyuapan, Pemalsuan Surat, Curanmor, Penganiayaan	10
Tabel B.6 Uji Coba <i>Dictionary</i> 1.....	11
Tabel B.7 Hasil Uji Coba <i>Dictionary</i> 1	14
Tabel B.8 Uji Coba <i>Dictionary</i> 2.....	14
Tabel B.9 Hasil Uji Coba <i>Dictionary</i> 2	15

DAFTAR KODE

Kode 5.1 Membaca Data.....	51
Kode 5.2 Menjadikan Data sebagai Korpus.....	52
Kode 5.3 Mengkonversi Data ke Dalam Huruf Kecil.....	52
Kode 5.4 Menghapus, Nomor, Simbol, dan Mengelemiasi Kelebihan Spasi.....	53
Kode 5.5 Menghapus Kata.....	55
Kode 5.6 Tokenisasi Korpus.....	57
Kode 5.7 Mengidentifikasi Frequent Words dan Membuat DTM.....	57
Kode 5.8 Mengkonversi Informasi ke Dalam Yes & No dan Mengaplikasikan ke DTM.....	58
Kode 5.9 Membagi Data Latih dan Uji.....	58
Kode 5.10 Membuat Pengklasifikasi.....	58
Kode 5.11 Mengevaluasi Performa Pada Data Tes.....	59
Kode 5.12 Menghitung Akurasi dalam Persen dan Membuat <i>Confusion Matrix</i>	59
Kode 5.13 Kode Program untuk Membuat <i>Wordcloud</i>	66
Kode 5.14 Menghapus Kata dengan <i>Stopwords</i> Manual.....	67
Kode 5.15 Menampilkan 20 Kata Teratas.....	68
Kode 5.16 Menggunakan <i>Dictionary</i> Manual Pada Pembobotan.....	70
Kode 5.17 Menseleksi Data Mode.....	71
Kode 5.18 Menghitung Nilai Akurasi Total.....	72
Kode 5.19 Visualisasi <i>Streamgraph</i>	82

BAB I

PENDAHULUAN

Pada bab pendahuluan ini akan membahas mengenai latar belakang, perumusan masalah, batasan masalah, tujuan dan manfaat dari tugas akhir, sistematika penulisan, serta relevansi yang digunakan dalam pengerjaan tugas akhir.

1.1. Latar Belakang

Kebutuhan masyarakat yang tidak terpenuhi secara merata berpotensi menimbulkan gangguan terhadap hubungan sosial. Indonesia, dengan kondisi masyarakat yang heterogen membuat banyak sekali dijumpai permasalahan sosial, seperti tindak kejahatan atau kriminalitas. Hal ini ditandai dengan angka kriminalitas yang cenderung meningkat dari jumlah dan jenisnya. Pemberitaan di media massa belakangan ini menunjukkan angka terjadinya tindak kriminal secara kuantitatif meningkat, baik kejahatan kekerasan (*violence*) maupun kejahatan tanpa kekerasan (*non violence*).

Namun demikian, tidak semua informasi kriminalitas tersebut dapat tercatat dalam statistik kriminal, terbatas pada peristiwa kriminalitas yang diketahui saja oleh kepolisian. Keterbatasan ini membuat penegak hukum seringkali mengalami kesulitan dalam melakukan kajian dan analisis tindak kriminalitas.

Seiring berkembangnya teknologi informasi dan komunikasi yang semakin pesat, cara manusia berkomunikasi tidak harus selalu bertatap muka. Berbagai jenis media sosial banyak digunakan sebagai sarana berbagi informasi. Salah satu media tersebut yaitu LAPOR!, sistem *e-government* di bawah pengelolaan Kantor Staf Presiden yang dibuat pada tahun 2012.

Laporan Aspirasi dan Pengaduan Online Rakyat (LAPOR!) merupakan sebuah aplikasi media sosial yang dibangun dan dikelola oleh Unit Kerja Presiden Bidang Pengawasan dan Pengendalian Pembangunan (UKP4) untuk melibatkan partisipasi publik dan meningkatkan interaksi dua arah antara masyarakat dan pemerintah dalam pengawasan program – program pembangunan [1]. Masyarakat dapat menyampaikan laporan pengaduan dengan mengakses LAPOR! melalui situs LAPOR!, SMS, ataupun aplikasi LAPOR! yang dapat dipasang pada perangkat seluler. Hingga saat ini tercatat bahwa jumlah laporan yang diterima LAPOR! rata-rata 10000-14000 laporan diterima setiap bulannya. Besarnya antusiasme masyarakat ini menunjukkan bahwa LAPOR! merupakan salah satu media penyampaian aspirasi dan pengaduan masyarakat yang mudah dan terpadu secara nasional.

Data menunjukkan bahwa hingga Maret 2015 LAPOR! telah menerima lebih dari 610.000 laporan dengan rata-rata lebih dari 900 laporan setiap harinya [2]. Dari keseluruhan data tersebut, hanya 75.870 laporan atau sekitar 12,5% yang langsung disetujui, 6600 laporan atau 1% yang dipending, dan 527530 atau sekitar 86% diarsipkan [2]. Hal ini mengindikasikan bahwa hanya sekitar 13-14% laporan yang diproses dan diketahui perihalnya, sedangkan sekitar 86% laporan tidak diketahui perihalnya.

Berdasarkan hasil wawancara dengan Bapak Ferdy, Kedeputan I Bidang Pengendalian Pembangunan Program Prioritas Nasional, mengatakan bahwa ada sangat banyak laporan masuk yang tidak dapat didisposisikan kepada lembaga terkait karena bukan merupakan aspirasi atau pengaduan terkait dengan layanan pemerintahan. Tidak ada perihal laporan yang sesuai dengan kategori yang ada dalam sistem LAPOR! sehingga laporan tersebut diarsipkan dan tidak dapat diproses lebih lanjut. Setelah ditelusuri, dari 86% laporan yang masuk mayoritas merupakan data pengaduan

mengenai tindak kriminalitas yang terjadi di masyarakat. Tentu sangat disayangkan jika laporan-laporan ini tidak diproses dan didisposisikan kepada pihak yang berwenang untuk ditindaklanjuti.

Namun begitu, LAPOR! belum memiliki metode efektif untuk menganalisis banyaknya laporan tersebut. Jumlah laporan yang begitu besar ini dapat disebut sebagai *Big Data*. *Big Data* merupakan data yang memiliki jumlah dan variasi yang sangat besar serta bergerak sangat cepat, sehingga mampu melebihi kapasitas pengolahan sistem database konvensional [3]. Untuk mengolah *Big Data*, *Data Mining* merupakan metode yang dapat mengotomatisasi proses pengolahan data untuk mengekstraksi informasi yang tidak dapat diamati hanya dengan melihat data karena terlalu kompleks atau multidimensi. Pada studi kasus LAPOR! yang merupakan data bertipe tekstual, sehingga jenis metode yang tepat digunakan adalah metode *Text Mining*.

Penentu kebijakan dan pengambil keputusan, yang dalam hal ini adalah pemerintah dan kepolisian, perlu untuk mengetahui segala bentuk aduan masyarakat mengenai tindak kriminalitas. Dengan bantuan teknik *Text Mining*, analisis laporan dapat lebih lebih cepat dan efisien untuk dilakukan, sehingga nantinya dapat diketahui pola permasalahan kriminalitas yang terjadi di masyarakat. Kategori kriminalitas yang tepat juga perlu diketahui agar didapatkan hasil klasifikasi yang bisa merepresentasikan laporan kriminalitas pada data LAPOR! serta memudahkan kepolisian dalam melakukan tindak lanjut. Dalam penelitian ini menggunakan *Naïve Bayes Classifier (NBC)* sebagai metode pengklasifikasi. Diharapkan, hasil penelitian ini dapat menjadi masukan bagi pemerintah maupun kepolisian dalam mengambil tindakan dan kebijakan yang tepat kedepannya.

1.2. Rumusan Masalah

Berdasarkan latar belakang yang telah dipaparkan sebelumnya, berikut merupakan rumusan masalah yang akan diselesaikan dalam penelitian ini.

1. Bagaimana menentukan kategori kriminalitas yang tepat digunakan dalam klasifikasi pada data LAPOR!?
2. Bagaimana cara melakukan proses klasifikasi laporan kriminalitas pada data LAPOR! menggunakan *Naïve Bayes Classifier (NBC)*?
3. Bagaimana performa *Naïve Bayes Classifier (NBC)* dalam melakukan klasifikasi laporan kriminalitas pada data LAPOR!?

1.3. Batasan Tugas Akhir

Dalam pengerjaan tugas akhir ini, berikut merupakan batasan masalah yang perlu diperhatikan.

1. Data laporan yang digunakan adalah laporan yang disampaikan masyarakat melalui LAPOR!
2. Data yang digunakan adalah data LAPOR! pada bulan Januari 2014 hingga bulan September 2015
3. Data yang digunakan adalah data LAPOR! mengenai kriminalitas
4. Keluaran yang dihasilkan pada tugas akhir ini berupa model klasifikasi laporan kriminalitas

1.4. Tujuan Tugas Akhir

Tujuan dari pengerjaan tugas akhir ini adalah untuk mengetahui kategori kriminalitas yang tepat digunakan dalam mengklasifikasi data kriminalitas LAPOR! menggunakan *Naïve Bayes Classifier (NBC)*. Hasilnya diharapkan dapat digunakan oleh penentu kebijakan dan pengambil keputusan untuk melihat pola permasalahan kriminalitas yang terjadi di masyarakat sebagai acuan dalam mengambil tindakan dan kebijakan yang tepat kedepannya.

1.5. Manfaat Tugas Akhir

Tugas akhir ini diharapkan dapat memberikan manfaat antara lain.

1. Bagi LAPOR!, dapat membantu menentukan kategori kriminalitas yang tepat diterapkan pada data LAPOR!
2. Bagi kepolisian, dapat membantu dalam hal penangkapan pelaku kejahatan, pengurangan gangguan keamanan, dan evaluasi prosedur organisasi penegak hukum untuk didapatkan tindakan dan kebijakan yang tepat.
3. Bagi pemerintah, dapat dijadikan sebagai referensi dalam menentukan kebijakan dan pengambilan keputusan.
4. Bagi masyarakat, dapat mengetahui adanya pola permasalahan kriminalitas yang terjadi di masyarakat sehingga dapat lebih waspada.
5. Hasil penelitian ini dapat dijadikan sebagai rujukan untuk penelitian-penelitian selanjutnya.

1.6. Sistematika Penulisan

Berikut merupakan sistematika penulisan dari buku tugas akhir yang dibagi ke dalam beberapa bab.

BAB I PENDAHULUAN

Pada bab ini akan dibahas mengenai latar belakang, perumusan masalah, batasan, tujuan, manfaat tugas akhir, sistematika penulisan, dan relevansi.

BAB II TINJAUAN PUSTAKA

Pada bab ini akan dibahas mengenai tinjauan pustaka dan teori-teori yang mendukung dalam pengerjaan tugas akhir. Teori-teori tersebut antara lain: Kriminalitas, LAPOR!, Text Mining, Klasifikasi, *Naïve Bayes Classifier* (NBC), Evaluasi Performa dan Keakurasian

BAB III METODOLOGI

Pada bab ini akan dibahas mengenai metode pengerjaan tugas akhir. Metode pengerjaan tugas akhir dimulai dengan melakukan tahap pendahuluan,

studi literatur, pengumpulan dan pengolahan data, analisis hasil dan keakurasian, dan penyusunan laporan tugas akhir.

BAB IV DESAIN DAN IMPLEMENTASI SISTEM

Pada bab ini dijelaskan mengenai perancangan dari hasil pengambilan data dan implementasi metode klasifikasi yaitu *Naïve Bayes Classifier (NBC)*.

BAB V UJI COBA DAN ANALISIS HASIL

Pada bab ini akan dilakukan analisis hasil klasifikasi dan evaluasi performa dan keakurasian dengan menggunakan *confusion matrix*.

BAB VI PENUTUP

Bab penutup ini berisi kesimpulan yang diperoleh dari hasil pengerjaan tugas akhir dan rekomendasi serta saran untuk pengembangan penelitian selanjutnya yang dilakukan dalam tugas akhir ini.

1.7. Relevansi

Topik yang diangkat pada tugas akhir ini mengenai *text mining* dan visualisasi yang berkaitan dengan mata kuliah Sistem Cerdas, Kecerdasan Bisnis serta Penggalian Data dan Analitika Bisnis khususnya pada bagian klasifikasi. Mata kuliah tersebut merupakan bagian dalam bidang keilmuan Akuisisi Data dan Diseminasi Informasi.

BAB II TINJAUAN PUSTAKA

Pada bab tinjauan pustaka ini akan dibahas mengenai studi sebelumnya dan dasar teori yang digunakan sebagai acuan dan mendukung dalam pengerjaan tugas akhir.

2.1. Studi Sebelumnya

Tabel 2.1 Ringkasan Studi Sebelumnya

No	Judul	Metode	Dataset	Hasil
1.	<i>A decision support system: Automated crime report analysis and classification for e-government</i>	<i>Naïve Bayes Classifier</i> dibandingkan dengan <i>binary logistic regression</i> dan <i>human expert's performance</i>	10 set laporan kejahatan yang membahas 2-5 jenis kejahatan serta 40 dan 60 laporan kejahatan yang membahas 16 jenis kejahatan	Akurasi sistem klasifikasi tertinggi mencapai 94,82% sedangkan akurasi klasifikasi dari analisis kejahatan 93,74%
2	Analisis Aspirasi dan Pengaduan di Situs Laporan! dengan Menggunakan <i>Text Mining</i>	Klasifikasi dengan <i>Support Vector Machine (SVM)</i> dan <i>Naïve Bayes Classifier</i> .	Data LAPOR! bulan Oktober 2014 - Maret 2015	Laporan terkait kemiskinan memiliki jumlah terbanyak dengan topic mayoritas

No	Judul	Metode	Dataset	Hasil
		Sedangkan klasterisasi dengan <i>Self-Organizing Map</i> (SOM) dan <i>K-means Clustering</i>		yang dibahas mengenai jenis bantuan sosial seperti KPS dan BLSM yang tidak didistribusikan dengan baik atau tidak tepat sasaran
3.	Klasifikasi Artikel Berita Berbahasa Indonesia Berbasis <i>Naïve Bayes Classifier</i> Menggunakan <i>Confix-Stripping Stemmer</i>	<i>Naïve Bayes Classifier</i>	Artikel berita online kompas.com hingga akhir tahun 2013 dengan kategori sebanyak 12	<i>Precision, recall</i> dan <i>f-measure</i> berturut-turut sebesar 87.675%, 88.21%, 87.65%, dan 87.69%
4.	Klasifikasi Data Forum dengan menggunakan Metode <i>Naïve Bayes Classifier</i>	<i>Naïve Bayes Classifier</i>	15 data forum	Akurasi kebenaran pengklasifikasian sebesar 73%

No	Judul	Metode	Dataset	Hasil
5.	Klasifikasi Teks dengan <i>Naïve Bayes Classifier</i> (NBC) untuk Pengelompokan Teks Berita dan Abstract Akademis	<i>Naïve Bayes Classifier</i> (NBC)	1000 dokumen berita dan 450 dokumen abstrak akademik	Dokumen berita memiliki akurasi maksimal 91% sedangkan dokumen akademik 82%

“*A decision support system: Automated crime report analysis and classification for e-government*” oleh Chih-Hao Ku dan Gondy Leroy dari School of Information Systems & Technology, Claremont Graduate University, 150 E 10th St, Claremont, CA 91711, United States. Paper ini bertujuan untuk mengetahui bagaimana analisis teks dan teknik klasifikasi dapat digunakan untuk meningkatkan *e-government* yaitu dengan mengembangkan DSS yang dapat mengidentifikasi laporan kejahatan yang serupa atau sama dengan sebuah classifier *Naïve Bayes*. Dengan teknik ini diharapkan dapat mendukung analisis kejahatan dan mengklasifikasikan laporan kejahatan dengan membahas kejahatan yang sama dan berbeda. Pada penelitian ini dilakukan dua jenis penelitian dengan dataset kecil dan besar dilakukan untuk membandingkan kinerja sistem dengan kinerja ahli manusia. Penelitian pertama meliputi 10 set laporan kejahatan yang membahas 2-5 jenis kejahatan. Akurasi algoritma tertinggi dengan menggunakan regresi logistik biner (89%) sedangkan pengklasifikasi *Naïve Bayes* hanya sedikit lebih rendah (87%). Ahli dicapai masih kinerja yang lebih baik (96%) ketika diberikan waktu yang cukup. Studi kedua mencakup dua dataset dengan 40 dan 60 laporan kejahatan membahas 16 jenis kejahatan untuk setiap dataset. Hasil penelitian menunjukkan bahwa sistem kami mencapai

akurasi klasifikasi tertinggi (94,82%), sedangkan akurasi klasifikasi analisis kejahatan ini (93,74%) sedikit lebih rendah.

“Analisis Aspirasi dan Pengaduan di Situs Laporan! dengan Menggunakan Text Mining” oleh Chyntia Megawati dari Fakultas Teknik, Program Studi Teknik Industri, Universitas Indonesia. Pada penelitian ini, peneliti menggunakan metode text mining untuk menganalisis data tekstual yang berupa opini atau keluhan dengan mengklasifikasikannya menjadi beberapa kelas dan kemudian data set setiap kelas akan dikelompokkan lagi menjadi beberapa topik khusus (cluster). Hasil penelitian menunjukkan bahwa laporan terkait kemiskinan memiliki jumlah terbanyak dengan topik mayoritas yang dibahas adalah mengenai beberapa jenis bantuan sosial seperti KPS (Kartu Perlindungan Sosial) dan BLSM (Bantuan Langsung Sementara Masyarakat) yang tidak didistribusikan dengan baik atau tidak tepat sasaran.

“Klasifikasi Artikel Berita Berbahasa Indonesia Berbasis Naïve Bayes Classifier Menggunakan Confix-Stripping Stemmer” oleh Amalia Anjani A dari Jurusan Sistem Informasi, Fakultas Teknologi Informasi, Institut Teknologi Sepuluh Nopember (ITS). Tugas akhir ini bertujuan untuk membangun sebuah aplikasi pengklasifikasi artikel berita berbahasa Indonesia berbasis Naïve Bayes classifier dengan menggunakan *confix-stripping stemmer*. Aplikasi ini dibangun dengan menggunakan kata dasar yang diperoleh dari berbagai artikel berita yang mewakili semua kategori berita yang akan dikelola. Hasilnya telah diujicobakan terhadap sejumlah sampel artikel berita yang terdiri dari 12 kategori yang dalam pengujian menunjukkan bahwa aplikasi yang dibangun dalam lingkungan sistem operasi Windows mampu menghasilkan pengklasifikasi dengan akurasi, precision, recall dan f-measure berturut-turut sebesar 87.675%, 88.21%, 87.65%, dan 87.69%

“Klasifikasi Data Forum dengan menggunakan Metode Naïve Bayes Classifier” oleh Aida Indriani dari Program Studi Teknik Informatika, STMIK PPKIA. Paper ini bertujuan untuk

melakukan klasifikasi kelas dalam forum untuk memudahkan dosen dan mahasiswa dalam mencari bahan diskusi. Metode klasifikasi yang digunakan untuk mengklasifikasikan data forum secara otomatis pada penelitian ini yaitu *Naïve Bayes Classifier* (NBC). Dengan NBC, diperoleh hasil akurasi kebenaran pengklasifikasian terhadap 15 data uji sebesar 73%.

“Klasifikasi Teks dengan Naïve Bayes Classifier (NBC) untuk Pengelompokan Teks Berita dan Abstract Akademis” oleh Amir Hamzah dari Jurusan Teknik Informatika, Fakultas Teknologi Industri, Institut Sains dan Teknologi AKPRIND. Pada penelitian ini mengkaji kinerja NBC untuk kategorisasi teks berita dan teks akademis. Penelitian menggunakan data 1000 dokumen berita dan 450 dokumen abstrak akademik. Hasil penelitian menunjukkan pada dokumen berita akurasi maksimal dicapai 91% sedangkan pada dokumen akademik 82%. Seleksi kata dengan minimal muncul pada 4 atau 5 dokumen memberikan akurasi yang paling tinggi.

2.2. Dasar Teori

2.2.1. Kriminalitas

Kriminalitas merupakan segala macam bentuk tindakan dan perbuatan yang merugikan secara ekonomis dan psikologis yang melanggar hukum yang berlaku dalam negara Indonesia serta norma-norma sosial dan agama. Dapat diartikan bahwa, tindak kriminalitas adalah segala sesuatu perbuatan yang melanggar hukum dan melanggar norma-norma sosial, sehingga masyarakat menentangnya [4].

Apabila berbicara tentang kriminalitas, seringkali yang dimaksudkan adalah jenis kriminalitas yang tercantum dalam Kitab Undang-undang Hukum Pidana (KUHP), seperti pembunuhan, penganiayaan, pemerkosaan, perampasan, pencurian, perampokan, penipuan, atau berbagai jenis kejahatan yang disebut sebagai violent offenses (kejahatan yang disertai kekerasan pada orang lain) dan property offenses (kejahatan yang menyangkut hak milik orang lain). Seperti

juga dengan jenis kejahatan yang digunakan Kepolisian Daerah Jawa Timur sebagai berikut.

Tabel 2.2 Jenis Kejahatan Kepolisian Daerah Jawa Timur [5]

TIPE KEJAHATAN	KATEGORI
Konvensional	Curat, curas, curanmor, kebakaran, pembunuhan, pemerasan, pemerkosaan, penjudian, pemalsuan surat, pererusakan, pencurian, penipuan, penggelapan, kenakalan remaja.
Transaksional	People smuggling, perdagangan anak dan wanita, selundup senjata api, perompakan.
Implikasi Kontijensi	Antara pelajar, mahasiswa, warga, suku, agama, pendukung dalam pilkada, oknum TNI dan POLRI, pemogokan kerja, mengganggu ketertiban.
Kejahatan Lain	Pembakaran, sumpah palsu, penghinaan, perzinahan, pencurian, pornografi, penadahan, suap, bahayakan keamanan umum, pekerjaan anak bawah umur, keimigrasian, ekstradisi, sistem peradilan pidana, penyelenggaraan pemilu.

Sementara itu, berbeda dengan klasifikasi yang dianut masyarakat atau penegak hukum, ahli sosiologis Light, Keller, dan Callhoun dalam bukunya yang berjudul *Sociology* (1989) mengklasifikasikan kriminalitas menjadi empat tipe, yaitu *crime without victim*, *organized crime*, *white collar crime*, dan *corporate crime* [6].

a. *White Collar Crime* (Kejahatan Keraf Putih)

Kejahatan ini mengacu pada kejahatan yang dilakukan oleh orang yang terpandang atau berstatus tinggi dalam hal pekerjaannya. Contohnya penghindaran pajak, penggelapan uang perusahaan, manipulasi data keuangan sebuah perusahaan (korupsi), dan lain sebagainya.

b. *Crime Without Victim* (Kejahatan Tanpa Korban)

Kejahatan tidak menimbulkan penderitaan pada korban secara langsung akibat tindak pidana yang dilakukan. Contohnya berjudi, mabuk, dan hubungan seks yang tidak sah tetapi dilakukan secara sukarela.

c. *Organized Crime* (Kejahatan Terorganisir)

Kejahatan ini dilakukan secara terorganisir dan berkesinambungan dengan menggunakan berbagai cara

untuk mendapatkan sesuatu yang diinginkan (biasanya lebih ke materiil) dengan jalan menghindari hukum. Contohnya penyedia jasa pelacuran, penadah barang curian, perdagangan perempuan ke luar negeri untuk komoditas seksual, dan lain sebagainya.

- d. *Corporate Crime* (Kejahatan Korporasi)
Kejahatan ini dilakukan atas nama organisasi formal dengan tujuan menaikkan keuntungan dan menekan kerugian. Lebih lanjut Light, Keller, dan Callhoun membagi tipe kejahatan korporasi ini menjadi empat, yaitu kejahatan terhadap konsumen, kejahatan terhadap publik, kejahatan terhadap pemilik perusahaan, dan kejahatan terhadap karyawan.

2.2.1.1. Kategori Kriminalitas

2.2.1.1.1. Reskrim POLDA JATIM

Reserse Kriminal dan Umum (Reskrim) Kepolisian Daerah Jawa Timur berungsi untuk melaksanakan tugas penyelidikan dan penyidikan tindak pidana umum yang terjadi di wilayah Jawa Timur. Reskrim terdiri dari 4 subdit dengan 5 unit untuk masing-masing subdit [7].

a. Subdit I. Keamanan Negara

Subdit I bertugas untuk menyelenggarakan fungsi penyelidikan dan penyidikan tindak pidana yang berhubungan dengan keamanan Negara, bahan peledak, senjata api, pemilu/ pemilukada, tindak pidana oleh pejabat public atau politisi, serta tindak pidana yang berimplikasi kontigensi yang terjadi di wilayah hukum Kepolisian Daerah Jawa Timur.

- Unit I. Tindak pidana handak (bahan peledak) dan sempu (senjata api)
- Unit II. Tindak pidana pemilu/ pemilukada dan pejabat public/ politik
- Unit III. Tindak pidana berimplikasi kontijensi
- Unit IV. Tindak pidana terorisme

- Unit V. Tindak pidana kebakaran/ pembakaran, penghinaan dan pencemaran nama baik

b. Subdit II. Harta Benda serta Bangunan dan Tanah

Subdit II bertugas untuk menyelenggarakan fungsi penyelidikan dan penyidikan tindak pidana yang berhubungan dengan harta benda, bangunan, tanah, penipuan, penggelapan, dan pemalsuan yang terjadi di wilayah hukum Kepolisian Daerah Jawa Timur.

- Unit I. Tindak pidana harta benda
- Unit II. Tindak pidana bangunan
- Unit III. Tindak pidana tanah
- Unit IV. Tindak pidana penipuan dan penggelapan
- Unit V. Tindak pidana pemalsuan

b. Subdit III. Kejahatan dan Kekerasan

Subdit III bertugas untuk menyelenggarakan fungsi penyelidikan dan penyidikan tindak pidana yang berhubungan dengan pembajakan dan sandera, pembunuhan dan penganiayaan, pencurian kendaraan bermotor, premanisme, pemerasan dan pencurian, dan perjudian yang terjadi di wilayah hukum Kepolisian Daerah Jawa Timur.

- Unit I. Tindak pidana pembajakan dan sandera
- Unit II. Tindak pidana pembunuhan dan penganiayaan
- Unit III. Tindak pidana curanmor
- Unit IV. Tindak pidana premanisme, pemerasan, pencurian
- Unit V. Tindak pidana perjudian

c. Subdit IV. Remaja, Wanita dan Anak

Subdit IV bertugas untuk menyelenggarakan fungsi penyelidikan dan penyidikan tindak pidana yang berhubungan dengan remaja, anak, dan wanita, perdagangan dan penyelundupan manusia, asusila,

KDRT, dan tenaga kerja yang berada di wilayah hukum Kepolisian Daerah Jawa Timur.

- Unit I. Tindak pidana remaja, anak, dan wanita
- Unit II. Tindak pidana perdagangan dan penyelundupan manusia
- Unit III. Tindak pidana asusila
- Unit IV. Tindak pidana KDRT
- Unit V. Tindak pidana tenaga kerja

2.2.1.1.2. Jenis Kejahatan Reskrim POLDA JATIM

Jenis – jenis kejahatan yang ditangani oleh Reserse Kriminal dan Umum POLDA JATIM terdiri dari tiga bagian, yaitu konvensional, implikasi kontijensi, dan transaksional. Setiap bagian memiliki subbagian berupa jenis – jenis kejahatan. Berikut merupakan rincian jenis kejahatan beserta deskripsi berdasarkan Kitab Undang-Undang Hukum Pidana (KUHP).

a. Konvensional

Kejahatan konvensional adalah perilaku ilegal yang kebanyakan orang anggap sebagai kejahatan [8]. Tabel 2.1 merupakan penjelasan apa saja yang tergolong ke dalam jenis kejahatan konvensional. Deskripsi dan sanksi lebih lanjut dapat dilihat pada **LAMPIRAN D**.

Tabel 2.1 Jenis Kejahatan Konvensional [9]

Jenis Kejahatan	Deskripsi Kejahatan
Kebakaran	Barang siapa karena kesalahan (kealpaan) menyebabkan kebakaran, ledakan atau banjir, jika karena perbuatan itu timbul bahaya umum bagi barang, nyawa orang lain, atau mengakibatkan orang mati. (KUHP Pasal 188)
Pembakaran	Barang siapa dengan sengaja menimbulkan kebakaran, ledakan atau banjir, jika karena perbuatan tersebut timbul bahaya umum bagi barang, nyawa orang lain, dan mengakibatkan orang mati.

Jenis Kejahatan	Deskripsi Kejahatan
Jabatan	<p>(KUHP Pasal 187)</p> <p>Seorang pejabat yang dengan sengaja menggelapkan uang atau surat berharga yang disimpan karena jabatannya, atau membiarkan diambil atau digelapkan oleh orang lain, atau menolong sebagai pembantu dalam melakukan perbuatan tersebut.</p> <p>(KUHP Pasal 415)</p>
Pemalsuan surat	<p>Barang siapa membuat surat palsu atau memalsukan surat yang dapat menimbulkan sesuatu hak, perikatan atau pembebasan hutang, atau yang diperuntukkan sebagai bukti daripada sesuatu hal dengan maksud untuk memakai atau menyuruh orang lain memakai surat tersebut seolah-olah isinya benar dan tidak dipalsu.</p> <p>(KUHP Pasal 263)</p>
Penipuan	<p>Barang siapa dengan maksud untuk menguntungkan diri sendiri atau orang lain, dengan memakai nama palsu atau martabat palsu, dengan tipu muslihat, ataupun rangkaian kebohongan, menggerakkan orang lain untuk menyerahkan barang sesuatu kepadanya, atau supaya memberi hutang rnaupun menghapuskan piutang.</p> <p>(KUHP Pasal 378)</p>
Penggelapan	<p>Barang siapa dengan sengaja memiliki barang sesuatu yang seluruhnya atau sebagian adalah kepunyaan orang lain.</p> <p>(KUHP Pasal 372)</p>
Sumpah palsu	<p>Barang siapa dengan sengaja memberi keterangan palsu di atas sumpah, baik dengan lisan atau tulisan, secara pribadi maupun oleh kuasanya yang khusus ditunjuk untuk itu.</p> <p>(KUHP Pasal 242)</p>

Jenis Kejahatan	Deskripsi Kejahatan
Utang piutang	Seorang pengusaha yang dinyatakan dalam keadaan pailit merugikan pemiutang karena pengeluarannya melewati batas, dan lain – lain. (KUHP Pasal 396)
Curat	Pencurian ternak; pencurian pada waktu ada kebakaran, letusan, banjir gempa bumi, atau gempa laut, gunung meletus, kapal karam, kapal terdampar, kecelakaan kereta api, huru-hara, pemberontakan atau bahaya perang; pencurian di waktu malam dalam sebuah rumah atau pekarangan tertutup yang ada rumahnya, yang dilakukan oleh orang yang tidak diketahui atau tidak dikehendaki oleh yang berhak. (KUHP Pasal 363)
Curas	Pencurian yang didahului, disertai atau diikuti dengan kekerasan atau ancaman kekerasan, terhadap orang dengan maksud untuk mempersiapkan atau mempermudah pencurian, atau dalam hal tertangkap tangan, untuk memungkinkan melarikan diri sendiri atau peserta lainnya, atau untuk tetap menguasai barang yang dicuri. (KUHP Pasal 365)
Curanmor	Barang siapa mengambil suatu barang (kendaraan bermotor), yang seluruhnya atau sebagian kepunyaan orang lain, dengan maksud untuk memiliki (kendaraan bermotor) secara melawan hukum. (KUHP Pasal 362)
Pembunuhan	Barang siapa dengan sengaja dan dengan rencana terlebih dahulu merampas nyawa orang lain. (KUHP Pasal 340)
Pemerasan/ Ancaman	Barang siapa memaksa seorang dengan kekerasan atau ancaman kekerasan untuk memberikan barang sesuatu, yang

Jenis Kejahatan	Deskripsi Kejahatan
	seluruhnya atau sebagian adalah kepunyaan orang itu atau orang lain, atau supaya membuat hutang maupun menghapuskan piutang. (KUHP Pasal 368)
Perjudian	Barang siapa dengan sengaja memberi kesempatan untuk bermain judi dan menjadikannya sebagai mata pencarian, atau dengan sengaja turut serta. (KUHP Pasal 303)
Merusak barang	Barangsiapa dengan sengaja menghancurkan, merusak, membuat tak dapat dipakai atau menghilangkan barang sesautu yang seluruhnya atau sebagian miliki orang lain. (KUHP Pasal 406)
Penculikan	Barang siapa membawa pergi seorang dari tempat kediamannya atau tempat tinggalnya sementara dengan maksud untuk menempatkan orang itu secara melawan hukum di bawah kekuasaannya atau kekuasaan orang lain, atau untuk menempatkan dia dalam keadaan sengsara. (KUHP Pasal 328)
Penghinaan	Barang siapa sengaja menyerang kehormatan atau nama baik seseorang dengan menuduhkan sesuatu hal, yang maksudnya terang supaya hal itu diketahui umum, (KUHP Pasal 310)
Aniaya ringan	Penganiayaan yang tidak menimbulkan penyakit atau halangan untuk menjalankan pekerjaan jabatan atau pencarian, diancam, sebagai penganiayaan ringan. (KUHP Pasal 352)
Akibat Orang Luka	Barang siapa karena kesalahannya menyebabkan orang lain mati, mendapat

Jenis Kejahatan	Deskripsi Kejahatan
	luka-luka berat, luka – luka sedemikian rupa sehingga timbul penyakit atau halangan menjalankan pekerjaan jabatan atau pencarian selama waktu tertentu. (KUHP Pasal 360)
Pencurian ringan	Barang siapa mengambil barang sesuatu, yang seluruhnya atau sebagian kepunyaan orang lain, dengan maksud untuk dimiliki secara melawan hukum. (KUHP Pasal 362)
Penadahan	Barang siapa membeli, menyewa, menukar, menerima gadai, menerima hadiah, atau untuk menarik keuntungan, menjual, menyewakan, menukarkan, menggadaikan, mengangkut, menyimpan atau menyembunyikan sesuatu benda, yang diketahui diperoleh dari kejahatan. (KUHP Pasal 480)
Menerima suap	Seorang pejabat yang menerima hadiah atau janji padahal diketahui bahwa hadiah atau janji itu diberikan karena kekuasaan atau kewenangan yang berhubungan dengan jabatannya, atau yang menurut pikiran orang yang memberi hadiah atau janji itu ada hubungan dengan jabatannya. (KUHP Pasal 418)
Penganiayaan	Barang siapa melakukan perbuatan yang mengakibatkan luka-luka berat dan mengakibatkan mati. (KUHP Pasal 351)
Perkosaan	Barang siapa dengan kekerasan atau ancaman kekerasan memaksa seorang wanita bersetubuh dengan dia di luar perkawinan. (KUHP Pasal 285)
Kesusilaan	Barang siapa dengan sengaja dan terbuka melanggar kesusilaan, (KUHP Pasal 281)

b. Implikasi Kontijensi

Kejahatan berimplikasi kontijensi adalah kejahatan yang dapat mengganggu aspek-aspek keamanan, politik, sosial, dan ekonomi serta meresahkan masyarakat yang terjadi secara mendadak dan sulit diprediksi [10]. Tabel 2.2 merupakan daftar jenis kejahatan yang tergolong ke dalam jenis kejahatan implikasi kontijensi.

Tabel 2.2 Jenis Kejahatan Implikasi Kontijensi

No	Jenis Kejahatan
1	Antar pelajar
2	Antar mahasiswa
3	Antar warga/ kampung/ desa
4	Antar suku
5	Antar agama
6	Antar kelompok/ ormas
7	Antar pendukung dalam pilkada
8	Antar oknum TNI dan POLRI
9	Antar oknum TNI dengan masyarakat
10	Antar oknum POLRI dengan masyarakat
11	Oleh orang terhadap barang
12	Oleh orang terhadap orang
13	Pemogokan kerja oleh buruh
14	Mengganggu ketertipan
15	Pemogokan kerja oleh PNS
16	Tokoh adat/ suku/ masyarakat/ agama
17	Kejahatan yang berhubungan dengan orang asing
18	Bidang pertahanan senjata lahan/ tapal

c. Transnasional

Kejahatan transnasional adalah kejahatan terorganisasi yang terjadi pada lintas perbatasan negara dan melibatkan kelompok atau jaringan yang bekerja di lebih dari satu negara untuk

merencanakan dan melaksanakan bisnis ilegal [11]. Tabel 2.3 merupakan jenis kejahatan yang tergolong ke dalam jenis kejahatan transaksional.

Tabel 2.3 Jenis Kejahatan Transnasiona

No	Jenis Kejahatan
1	Selundup senjata api
2	Perompakan
3	<i>People smuggling</i>
4	Perdagangan anak dan wanita

2.2.2. LAPOR!

Laporan Aspirasi dan Pengaduan Online Rakyat (LAPOR!) merupakan sebuah aplikasi media sosial yang dibangun dan dikelola oleh Unit Kerja Presiden Bidang Pengawasan dan Pengendalian Pembangunan (UKP4) untuk melibatkan partisipasi publik dan meningkatkan interaksi dua arah antara masyarakat dan pemerintah dalam pengawasan program-program pembangunan [1]. Pasal 36 dan 37 Undang-Undang Nomor 25 tahun 2009 tentang Pelayanan Publik mengamanatkan pemerintah wajib memberikan akses seluas luasnya kepada masyarakat untuk memberikan masukan atas pemberian layanannya. Sebagai tindak lanjut amanat Undang-Undang Nomor 25 tahun 2009, telah diterbitkan Peraturan Presiden Nomor 76 Tahun 2013 Tentang Pengelolaan Pengaduan Pelayanan Publik, yang mengisyaratkan dibentuknya Sistem Pengelolaan Pengaduan Pelayanan Publik Nasional (SP4N) yang merupakan integrasi pengelolaan pengaduan pelayanan publik secara berjenjang pada setiap penyelenggara dalam kerangka sistem informasi pelayanan publik dan LAPOR! sebagai alat atau sarana pengaduan yang disediakan [2].

Menurut Agung Hardjono, Deputy III UKP4 mengatakan bahwa Sistem LAPOR! dibuat karena pengawasan kinerja dan layanan langsung ke kementerian, lembaga, dan Pemda masih terkendala di teknisnya. Di sisi lain, hasil penelitian

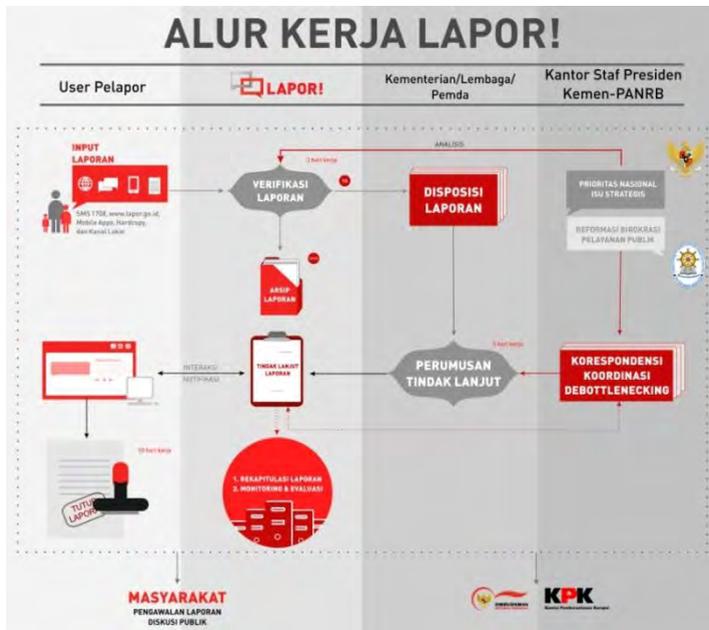
menunjukkan bahwa warga Indonesia merupakan pengguna gadget terbanyak sekitar 10-20 juta orang. Oleh karena itu, untuk memenuhi tagline-nya —Mudah”, LAPOR! menggunakan alat yang dianggap sudah dimiliki oleh kebanyakan penduduk Indonesia [12]. Hingga April 2015, LAPOR! telah digunakan oleh lebih dari 290.000 pengguna dan menerima rata-rata lebih dari 800 laporan masyarakat per harinya. LAPOR! menjadi cikal-bakal sistem aspirasi dan pengaduan masyarakat yang terpadu secara nasional.



Gambar 2.1 Halaman utama situt LAPOR!

2.2.2.1. Sistem Pelaporan LAPOR!

LAPOR! kini terhubung dengan 81 Kementerian/ Lembaga, 5 Pemerintah Daerah, serta 44 BUMN di Indonesia dalam pengelolaan aspirasi dan pengaduan masyarakat. Masyarakat tidak perlu bingung mengenai bagaimana cara menyampaikan aspirasi dan pengaduan, karena LAPOR! akan mendisposisikan laporan masyarakat kepada Kementerian/ Lembaga, Pemerintah Daerah atau BUMN terkait untuk ditindaklanjuti.



Gambar 2.2 Alur kerja LAPOR!

Masyarakat umum dapat mengirimkan laporan pada LAPOR! melalui berbagai media termasuk situs <https://www.lapor.go.id/>, SMS 1708 dan juga aplikasi mobile. Laporan kemudian diverifikasi terlebih dahulu oleh administrator LAPOR! untuk kejelasan dan kelengkapan, dan selanjutnya diteruskan ke instansi K/L/D terkait paling lambat 3 hari kerja setelah pelaporan dilakukan.

Setiap laporan yang diterima diproses menjadi tiga jenis perlakuan yaitu, disetujui, dipending, atau diarsipkan. Laporan yang disetujui adalah laporan yang sudah jelas dan merupakan aspirasi yang bagus, sehingga akan langsung diproses dan diteruskan untuk ditindaklanjuti. Laporan yang dipending adalah laporan yang dirasa bagus tetapi belum jelas, sehingga perlu dilakukan konfirmasi kembali kepada pelapor. Sedangkan laporan yang diarsipkan adalah laporan yang dirasa tidak jelas, laporan yang berulang atau sudah

pernah dilaporkan sebelumnya, atau merupakan saran yang bersifat sangat umum sehingga tidak diproses lebih lanjut. LAPOR! akan mempublikasikan setiap laporan yang sudah diteruskan sekaligus memberikan notifikasi kepada pelapor. Instansi K/L/D diberikan waktu paling lambat 5 hari kerja untuk melakukan koordinasi internal dan perumusan tindak lanjut dari pelaporan yang diberikan oleh masyarakat umum. Apabila sudah ada rumusan tindak lanjut, maka instansi K/L/D memberikan informasi kepada pelapor pada halaman tindak lanjut laporan. Laporan dianggap selesai apabila sudah terdapat tindak lanjut dari instansi K/L/D pada laporan, dan telah berjalan 10 hari kerja setelah tindak lanjut dilakukan tanpa adanya balasan dari pelapor maupun administrator LAPOR! di halaman tindak lanjut.

2.2.2.2. Fitur LAPOR!

LAPOR! mendorong keterbukaan dari instansi yang menggunakan serta menjadi alat reformasi dalam pelaksanaan pembangunan dan upaya peningkatan layanan publik. Semua ini dimungkinkan karena LAPOR! mengedepankan prinsip mudah, terpadu, dan tuntas. Oleh karena itu, berbagai dukungan fitur disediakan oleh LAPOR! untuk memeberikan kemudahan dan kenyamanan masyarakat dalam menggunakan LAPOR!. Berikut merupakan fitur yang ada pada LAPOR! [1].

Tracking ID LAPOR! .Tracking ID LAPOR! merupakan sebuah kode unik yang secara otomatis melengkapi setiap laporan yang dipublikasikan pada situs LAPOR!. Tracking ID dapat digunakan pengguna untuk melakukan penelusuran atas suatu laporan.

Anonim dan Rahasia. Fitur anonim tersedia bagi pelapor untuk merahasiakan identitasnya, sedangkan fitur rahasia dapat digunakan untuk membatasi akses atas laporan hanya bagi pelapor dan instansi terlapor. Kedua fitur ini dapat digunakan untuk pelaporan isu-isu sensitif dan sangat privat.

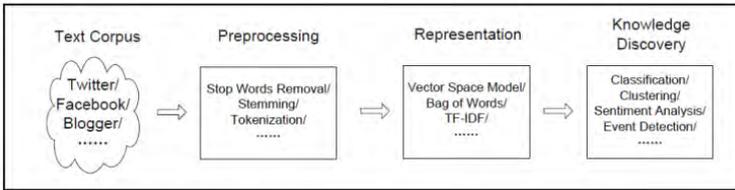
Peta dan Kategorisasi. Setiap laporan dapat dilabeli dengan dengan lokasi geografis, topik, status ketuntasan laporan, dan institusi terkait sehingga pemerintah maupun masyarakat dapat memonitor isu dengan berbagai skala dan sudut pandang. Peta LAPOR! dipergunakan sebagai pusat informasi banjir pada saat bencana banjir besar Jakarta di tahun 2012 dan 2014 sebagai rujukan dalam rangka penyaluran bantuan kepada para korban.

Opini Kebijakan. Fitur ini dapat digunakan oleh instansi pemerintah yang terhubung sebagai sarana jajak pendapat masyarakat. Beberapa jajak pendapat yang telah dilakukan melalui fitur ini diantaranya tentang Badan Penyelenggara Jaminan Sosial Kesehatan dan Rencana Implementasi Kurikulum Baru Pendidikan 2013.

2.2.3. Text Mining

Text mining merupakan variasi dari data mining yang berusaha menemukan pola yang menarik dari sekumpulan data tekstual yang berjumlah besar [13]. Pada dasarnya, text mining merupakan bidang interdisiplin yang mengacu pada perolehan informasi, data mining, pembelajaran mesin, statistik, dan komputasi linguistik [14]. Namun, secara umum konsep keduanya sama, yaitu penggalian prediktif dan penggalian deskriptif. *Text mining* mengekstrak indeks numerik yang bermakna dari teks dan kemudian informasi yang terkandung dalam teks akan diakses dengan menggunakan berbagai algoritma data mining [15].

Text mining dapat menganalisis dokumen dan mengelompokkan dokumen berdasarkan kata-kata yang terkandung di dalamnya, serta menentukan kesamaan di antara dokumen untuk mengetahui bagaimana mereka berhubungan dengan variabel lainnya [16]. Untuk memperoleh tujuan akhirdari text mining, diperlukan beberapa tahapan yang harus dilakukan seperti yang ditunjukkan pada Gambar 2.3.



Gambar 2.3 Kerangka Proses Analisis *Text Mining* [13]

Data yang nantinya terpilih akan dianalisis dimana pertamakali akan melewati tahap *text corpus*, *praprocessing*, *representation*, hingga yang terakhir adalah tahap *knowledge discovery*.

2.2.3.1. Text Corpus

Korpus adalah kumpulan teks yang bisa digunakan untuk proses *training* dan pengembangan data. Istilah *corpus* terutama yang dikenal dalam bidang linguistik pada prinsipnya bermakna koleksi yang memiliki lebih dari satu teks. Menurut McEnery & Wilson (2001), suatu corpus modern memiliki beberapa karakteristik yakni: [17]

- *Sampling & representativeness*
- *Finite size*
- *Machine-readable form*
- *A standard reference*

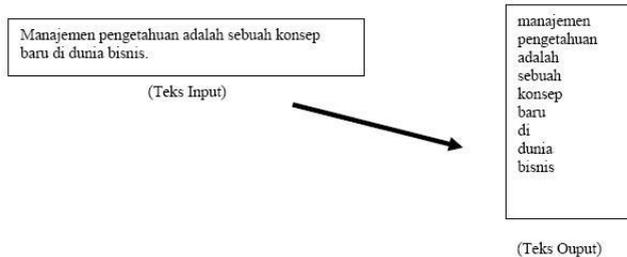
2.2.3.2. Praprocessing

Struktur data yang baik dapat memudahkan proses komputerisasi secara otomatis. Pada *text mining*, informasi yang akan digali merupakan informasi yang tidak terstruktur. Oleh karena itu, perlu dilakukan proses perubahan bentuk menjadi data yang terstruktur yang biasanya akan menjadi nilai-nilai numerik. Proses ini sering disebut *Text Preprocessing*. Setelah data menjadi data terstruktur atau berupa nilai numeric, maka data dapat diolah lebih lanjut. Fitur-fitur tahap praproses adalah sebagai berikut: [18]

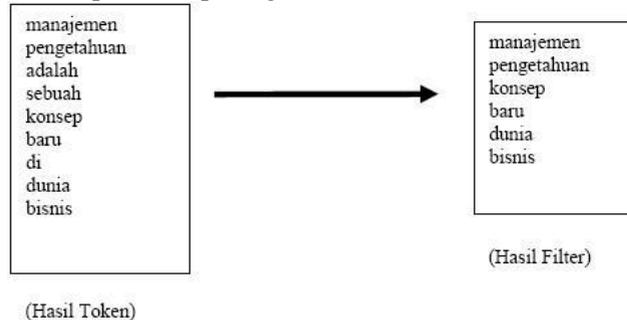
- a. *Casefolding*, merupakan proses perubahan huruf dalam dokumen menjadi satu bentuk, misalnya huruf kapital

menjadi huruf kecil dan sebaliknya. Dengan batasan hanya huruf a sampai dengan z yang diterima. Karakter selain huruf dihilangkan dan dianggap delimiter.

- b. *Tokenization*, merupakan proses pemisahan teks menjadi potongan kalimat dan kata yang disebut token. Contoh dari tahap ini ada pada gambar berikut.



- c. *Filtering*, merupakan proses pengambilan kata-kata penting dari hasil token ataupun membuang kata maupun tanda yang tidak bermakna secara signifikan. Contoh dari tahap ini ada pada gambar berikut.



2.2.3.3. Representation

Proses operasi *learning algorithms* tidak bisa langsung memproses dokumen teks dalam bentuk aslinya [2]. Sehingga, perlu proses untuk mengubah dokumen menjadi suatu representasi yang mudah dikelola lebih lanjut, yang biasanya direpresentasikan dalam bentuk vector. Representasi yang umum digunakan dalam text mining adalah Term Frequency –

Inverse Document Frequency (TF-IDF). [19] Skema persamaan TF-IDF ditunjukkan oleh persamaan berikut.

$$tfidf(w) = tf \times \log \frac{N}{df(w)} \quad [19]$$

Keterangan:

- $tf(w)$ = *Term frequency* (jumlah kemunculan suatu kata dalam suatu dokumen)
- $df(w)$ = *Document frequency* (jumlah dokumen yang mengandung suatu kata)
- N = Jumlah dokumen

Setelah bobot masing-masing dokumen diketahui, selanjutnya dilakukan proses pemeringkatan atau perankingan dokumen berdasarkan besarnya tingkat kesesuaian dokumen terhadap query, dimana semakin besar nilai bobot dokumen terhadap query maka semakin besar tingkat similaritas dokumen tersebut terhadap query yang dicari.

2.2.3.4. Knowledge Discovery

Tahap akhir penggalian informasi pada text mining yaitu ekstraksi ilmu pengetahuan (*knowledge discovery*), dimana terdapat beberapa jenis kategori utama yang bisa dilakukan sebagai berikut [15].

- a. Klasifikasi, merupakan bentuk analisis data yang mengekstrak model untuk menggambarkan kelas data [14].
- b. Klasterisasi, merupakan proses pengelompokan satu set data objek menjadi beberapa kelompok atau klaster sehingga objek dalam sebuah klaster memiliki kemiripan yang tinggi satu sama lain, tetapi sangat berbeda dengan objek dalam kelompok lainnya [20].
- c. Asosiasi, merupakan proses pencarian hubungan antar elemen data. Dalam dunia industri retail, analisis asosiasi biasanya disebut Market Basket Analysis [15].

2.2.4. Algoritma Pengklasifikasi

Klasifikasi merupakan bagian dari algoritma data mining, yang menggunakan data dengan target (class/label) berupa nilai kategorikal/nominal. Algoritma yang biasanya digunakan untuk melakukan pengklasifikasian diantaranya adalah *Naïve Bayes Classifier* (NBC).

2.2.4.1. Naïve Bayes Classifier (NBC)

Naïve Bayes Classifier atau bisa disebut sebagai multinomial Naïve Bayes merupakan model penyederhanaan dari algoritma bayes yang cocok dalam pengklasifikasian teks atau dokumen [21]. Dasar dari teorema Naïve Bayes [22] yang digunakan dalam pemrograman adalah sebagai berikut.

$$P(A|B) = \frac{P(B|A) \times P(A)}{P(B)} \dots (1)$$

Sedangkan persamaan *Naïve Bayes Classifier* (NBC) adalah:

$$V_{MAP} = \arg \max P(V_j | a_1, a_2, \dots a_n) \dots (2)$$

Berdasarkan persamaan (2), maka persamaan (1) dapat ditulis

$$V_{MAP} = \arg \max_{v_j \in V} \frac{P(a_1, a_2, \dots a_n | v_j) P(v_j)}{P(a_1, a_2, \dots a_n)} \dots (3)$$

$P(a_1, a_2, \dots a_n)$ konstan, sehingga dapat dihilangkan menjadi

$$V_{MAP} = \arg \max_{v_j \in V} P(a_1, a_2, \dots a_n | v_j) P(v_j) \dots (4)$$

Karena $P(a_1, a_2, \dots a_n | v_j)$ sulit untuk dihitung, maka akan diasumsikan bahwa setiap kata pada dokumen tidak mempunyai keterkaitan.

$$V_{MAP} = \arg \max_{v_j \in V} P(v_j) \prod_i P(a_i | v_j) \dots (5)$$

Keterangan:

$$P(v_j) = \frac{|docs_j|}{|contoh|} \dots (6)$$

$$P(w_k|v_j) = \frac{n_k + 1}{n + |kosakata|} \dots (7)$$

Dimana untuk :

$P(v_j)$: Probabilitas setiap dokumen terhadap sekumpulan dokumen

$P(w_k|v_j)$: Probabilitas kemunculan kata(w_k) pada suatu dokumen dengan kategori class (v_j)

$|docs_j|$: frekuensi dokumen pada setiap kategori

$|contoh|$: jumlah dokumen

n_k : frekuensi kata ke-k pada setiap kategori

$|kosakata|$: jumlah kata padadokumen test

Metode NBC menempuh dua tahap dalam proses klasifikasi teks, yaitu tahap pelatihan dan tahap klasifikasi [23] [24].

d. Tahap Pelatihan

Pada tahap pelatihan dilakukan proses analisis terhadap sampel dokumen berupa pemilihan *vocabulary*, yaitu kata yang mungkin muncul dalam koleksi dokumen sampel yang sedapat mungkin dapat menjadi representasi dokumen. Selanjutnya adalah penentuan probabilitas prior bagi tiap kategori berdasarkan sampel dokumen.

- Langkah 1 : Bentuk *vocabulary* pada setiap dokumen data latih
- Langkah 2 : Hitung probabilitas pada setiap kategori $P(v_j)$
- Langkah 3 : Tentukan frekuensi setiap kata (w_k) pada setiap kategori $P(w_k|v_j)$

e. Tahap Klasifikasi

Pada tahap klasifikasi ditentukan nilai kategori dari suatu dokumen berdasarkan kata yang muncul dalam dokumen yang diklasifikasi. Sedangkan untuk menentukan klasifikasi pada data uji, digunakan persamaan 5.

- Langkah 1 : Hitung $P(v_j) \prod P(w_k|v_j)$ untuk setiap kategori
- Langkah 2 : Tentukan kategori dengan nilai $P(v_j) \prod P(w_k|v_j)$ maksimal

2.2.5. Evaluasi Performa Pengakurasian

Sebuah model klasifikasi diharapkan dapat melakukan klasifikasi data dengan benar, tetapi tidak dapat dipungkiri bahwa kinerja suatu sistem tidak bisa 100% benar. Sehingga perlu dilakukan pengukuran kinerja dan keakurasian dari hasil klasifikasi. Umumnya, pengukuran kinerja klasifikasi dilakukan dengan matriks konfusi (*confusion matrix*) yang ditunjukkan pada gambar di bawah ini.

	Prediksi c = A	Prediksi c = B	Prediksi c = C	Prediksi c = D	
Aktual c = A	True A	(FP B c = A)	(FP C c = A)	(FP D c = A)	Total Aktual c = A
Aktual c = B	(FP A c = B)	True B	(FP C c = B)	(FP D c = B)	Total Aktual c = B
Aktual c = C	(FP A c = C)	(FP B c = C)	True C	(FP D c = C)	Total Aktual c = C
Aktual c = D	(FP A c = D)	(FP B c = D)	(FP C c = D)	True D	Total Aktual c = D
	Total Prediksi c = A	Total Prediksi c = B	Total Prediksi c = C	Total Prediksi c = D	TOTAL

Gambar 2.4 *Confusion Matrix* [19]

Keterangan:

- *True Positives* (TP), yaitu jumlah dokumen dari kelas yang benar dan diklasifikasikan sebagai kelas positif. (prediksi benar)
- *True Negatives* (TN), yaitu jumlah dokumen dari kelas yang benar diklasifikasikan sebagai kelas negatif. (prediksi benar)
- *False Positives* (FP), yaitu jumlah dokumen dari kelas yang salah diklasifikasikan sebagai kelas positif. (prediksi salah)

- *False Negatives* (FN) yaitu jumlah dokumen dari kelas yang salah diklasifikasikan sebagai kelas negatif. (prediksi salah)

Perhitungan-perhitungan yang dapat dilakukan untuk mengukur kinerja klasifikasi adalah sebagai berikut. [25]

- Akurasi, merupakan nilai accuracy merepresentasikan seberapa banyak keseluruhan dokumen diklasifikasikan dengan benar.

$$Akurasi = \frac{True\ A + True\ B + True\ C + True\ D}{jumlah\ total\ data} \quad [25]$$

- *Precision*, merupakan pengukuran ketepatan prediksi pengklasifikasi pada kelas tertentu

$$Precision(A) = \frac{TP}{TP+FP} \quad [25]$$

- *Recall*, merupakan pengukuran berapa banyak instans dari kelas tertentu yang dapat diprediksikan secara benar

$$Recall(A) = \frac{TP}{TP+FN} \quad [25]$$

- *F-Measure*, merupakan pengukuran untuk mengetahui akurasi secara keseluruhan yang didapatkan dari pengukuran *precision* dan *recall* antara kelas hasil *cluster* dengan kelas sebenarnya. Ukuran yang menampilkan timbal balik antara kedua pengukuran ini disebut sebagai *harmonic mean* dari *precision* dan *recall*.

$$F_{\beta} = (1 + \beta^2) \cdot \frac{precision \cdot recall}{(\beta^2 \cdot precision) + recall} \quad [25]$$

Rumus di atas adalah rumus perhitungan *F-Measure* dimana nilai *precision* dan *recall* pada suatu keadaan dapat memiliki bobot yang berbeda, sehingga nilai β yang dipakai juga dapat berbeda.

- $\beta = 0.5$ berarti pemberat untuk *precision* 2x lipat
- $\beta = 2$ berarti pemberat untuk *recall* 2x lipat
- $\beta = 1$ berarti *precision* dan *recall* bernilai setara

Nilai $\beta = 1$ ini sering ditulis menjadi F_1 , singkatan dari $F_{\beta=1}$. Hasil perhitungannya sering disebut juga sebagai *balanced F-score*.

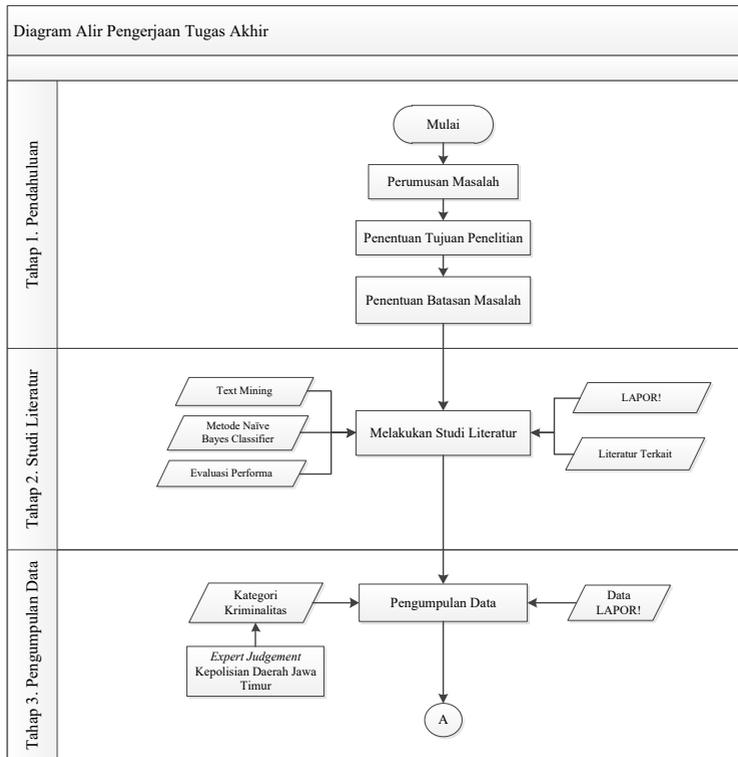
$$F_1 = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} \quad [19]$$

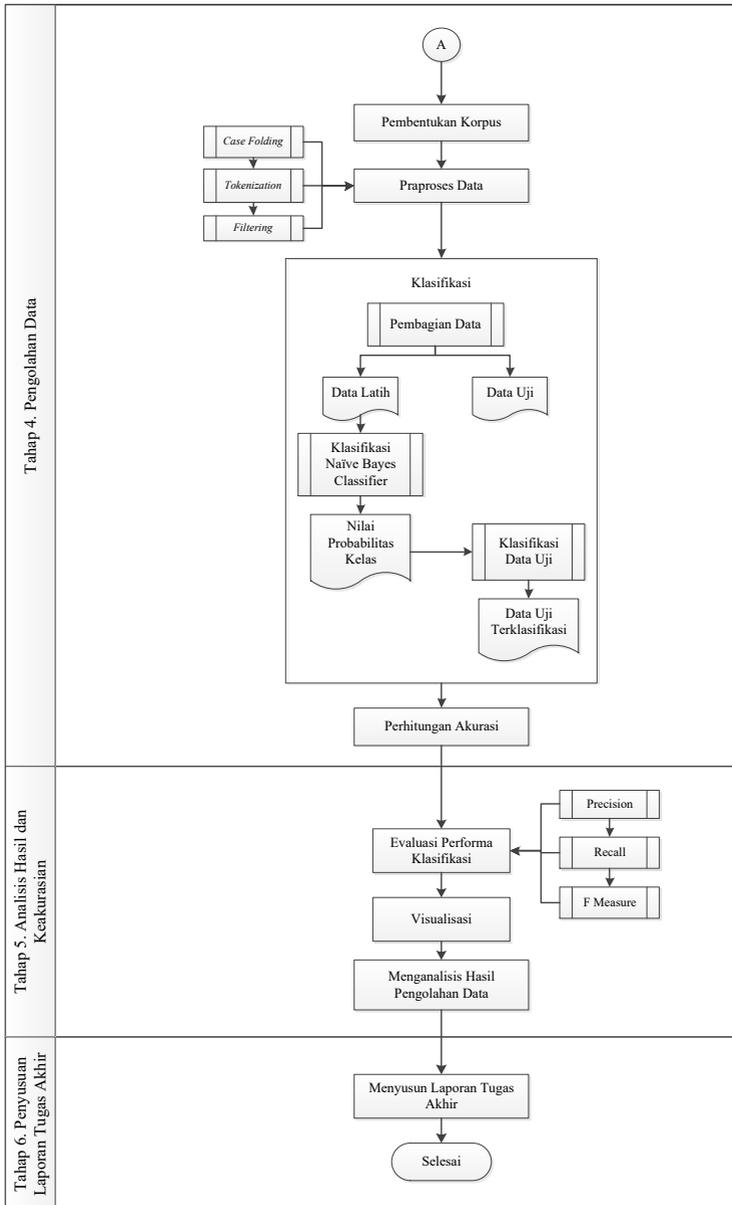
BAB III METODE PENELITIAN

Pada bab metode penelitian ini akan menjelaskan bagaimana tahapan yang harus dilakukan dalam pengerjaan tugas akhir ini. Tahap-tahap tersebut memiliki jangka waktu masing-masing yang telah ditentukan untuk membantu penulis dalam menyelesaikan pengerjaan tugas akhir ini.

3.1. Diagram Alir Pengerjaan Tugas Akhir

Permasalahan pada tugas akhir ini akan diselesaikan dengan metode yang tergambar pada diagram alir berikut.





Gambar 3.1 Diagram Alir Pengerjaan Tugas Akhir

3.2. Tahapan Pengerjaan Tugas Akhir

Berikut merupakan penjelasan mengenai metodologi pada diagram alir diatas.

1. Tahap Pendahuluan
Untuk memulai pengerjaan tugas akhir ini, tahap awal yang dilakukan yaitu merumuskan masalah, menentukan tujuan, dan batasan masalah yang akan dipakai sebagai pedoman dalam pengerjaan pada tahap selanjutnya.
2. Tahap Studi Literatur
Pada tahap ini dilakukan pencarian literatur atau referensi baik dari sumber online maupun offline dengan tujuan untuk menggali informasi sebagai penunjang dari penelitian tugas akhir ini. Studi literatur yang dilakukan adalah yang berkaitan dengan *text mining*, praproses data, *Naive Bayes Classifier*, kriminalitas, LAPOR!, dan evaluasi performa dan keakurasian dari hasil penelitian yang nantinya digunakan dalam menunjang penelitian.
3. Tahap Pengumpulan Data
Pada tahap ini dilakukan pengumpulan data dengan teknik wawancara kepada pihak terkait, yakni LAPOR! dan kepolisian. Data yang digunakan adalah data laporan kriminalitas yang disampaikan melalui situs LAPOR! pada bulan Januari 2014 hingga bulan September 2015. Sedangkan kategori kriminalitas yang akan digunakan adalah *expert judgement* dari Kepolisian Daerah Jawa Timur.
4. Tahap Pengolahan Data
Pada tahap ini dilakukan beberapa langkah pengolahan data. Pertama, melakukan pembentukan korpus dari kategori kriminalitas yang didapat dari kepolisian. Korpus ini nantinya digunakan dalam pelabelan data LAPOR!. Setelah itu melakukan praproses data yang terdiri dari *case folding*, *filtering*, dan *stopwords* untuk kemudian direpresentasikan dalam bentuk *data term matrix*.

5. Tahap Analisis Data dan Keakurasian

Setelah didapatkan hasil olah data, langkah selanjutnya yaitu melakukan analisis data dan keakurasian. Dalam melakukan pengukuran kinerja dan keakurasian dari pengklasifikasi, digunakan *confusion matrix* sebagai acuan pengukuran-pengukurannya.

6. Tahap Penyusunan Laporan Tugas Akhir

Tahapan terkahir ini adalah pembuatan laporan tugas akhir yang dibuat dan disusun sesuai dengan format yang telah ditentukan. Laporan ini mencakup keseluruhan proses pengerjaan tugas akhir dan analisis hasil pengerjaan tugas akhir.

BAB IV PERANCANGAN

Pada bab ini akan dijelaskan bagaimana rancangan dari penelitian tugas akhir yang meliputi penjelasan data, pengumpulan, penggabungan, dan pelabelan dari obyek penelitian dan bagaimana penelitian akan dilakukan.

4.1. Data

Data teks laporan yang digunakan adalah laporan kriminalitas yang disampaikan masyarakat melalui situs LAPOR! pada periode Januari 2014 – September 2015. Jumlah data yang diterima pada periode itu sebanyak 1136 laporan. Laporan tersebut pada sistem LAPOR! tergolong laporan yang diarsipkan karena tidak dapat ditindaklanjuti. Tabel 2.1 merupakan contoh data laporan beserta atribut yang didapatkan.

Tabel 4.1 Contoh Data Laporan Beserta Atribut

No	Atribut	Keterangan Atribut	Contoh Isi Data
1	nid	Nomor id laporan	1253802
2	tanggal	Tanggal laporan dibuat	8/20/2014 6:53:00 PM
3	isiLaporan	Isi laporan	saya mengadu kepada aparat pemerintah kalau benar akan memeriksanya saya korban penipuan oleh aparat desa saya tida bukti akan ada bantuan rumah di desa saya ini udah kena pungutan uang sebesar Rp.20.000

No	Atribut	Keterangan Atribut	Contoh Isi Data
			perrumah dari masakat desa sinarasa kampung ciganas rt.o2 kec.cikakak kab. Sukabumi

Data yang diperoleh pada periode tersebut telah mengalami proses penyaringan. Laporan yang mengandung unsur kata mengenai caci – maki, ancaman, dan tidak relevan pada kinerja pemerintah digolongkan ke dalam laporan kriminalitas. Laporan inilah yang digunakan sebagai data utama pada penelitian yaitu dengan total kata 92.976 dan rata-rata untuk setiap laporannya adalah 81.845 kata.

4.2. Pengumpulan Data

Pengumpulan data merupakan bagian terpenting yang akan menentukan keberhasilan dari penelitian tugas akhir ini. Kualitas data juga sangat mempengaruhi proses klasifikasi yang akan dilakukan. Pada penelitian ini, digunakan metode wawancara kepada pihak narasumber yang dalam hal ini adalah Kepolisian Daerah Jawa Timur serta pihak pengelola LAPOR!. Data yang didapatkan berupa informasi mengenai data kriminalitas LAPOR! yang selama ini belum ditindaklanjuti dan bagaimana pengkategorian laporan kriminalitas yang dilakukan oleh Kepolisian Daerah Jawa Timur.

4.3. Penggabungan Data

Data laporan yang didapat dari LAPOR! berupa file dengan format csv dimana setiap file berisi laporan bulanan dalam periode Januari 2014 – September 2015 dengan jumlah 21 bulan. Sehingga, perlu dilakukan penggabungan data ke dalam satu dataset berformat csv.

4.4. Pelabelan Data

Pelabelan data dilakukan untuk mengkategorikan setiap laporan ke dalam kategori kejahatan tertentu. Dasar dari pelabelan ini menggunakan informasi yang didapat dari POLDA JATIM serta pendefinisian jenis kejahatan berdasarkan Kitab Undang – Undang Hukum Pidana (KUHP). Label data yang digunakan ada dua jenis, yaitu berdasarkan jenis kejahatan dan berdasarkan subdit.

	A	B	C
1	nid	isiLaporan	tag
2	1182391	aGeng motor berawal dr perkelahian sebulan yg lalu di g akibatorangluka	
3	1197418	Anak2 kecil melempari kereta yg jalan sampai ada korbar akibatorangluka	
4	1225550	Kejadian jam 18.00, 29 mei 2014)meski mengenai muka, akibatorangluka	
5	1248777	Slmt siang. Sy mau lpr bhw d blng psr rumput ada tawur: akibatorangluka	
6	1166162	Bahwa Pak Kabag Humasy Tapanuli Tengah , Iwan RM Si akibatorangluka	
7	1372744	Anak Bua Hugua, Menabrak Siswi 5 Orang, 1 Orang Mati akibatorangluka	
8	1372916	Keracunan berakibat fatal. Saya diberikan minuman oleh akibatorangluka	
9	1181596	perampasan motor di kawasan tubagus ismail dengan al curanmor	
10	1200965	Masalah pencurian kendaraan mobil sering terjadi. Ini m curanmor	
11	1270931	Polantas gub dki,mohon adakn razia cctv tambahan di sk curanmor	
12	1295077	Yth Pengurus Laporan Rakyat, Saya dengan nama Bobby curanmor	

Gambar 4.1 Dataset dengan Atribut Terpilih

4.4.1. Jenis Kejahatan

Pada label jenis kejahatan terdapat 30 label dengan 25 merupakan jenis kejahatan pidana kriminal umum, 3 merupakan jenis kejahatan kriminal khusus, serta 2 merupakan label himbauan dan lain-lain.

Tabel 4.2 Label Jenis Kejahatan

No	Jenis Kejahatan	Label
1	Akibat orang luka	akibatorangluka
2	Curanmor	curanmor
3	Curat	curat
4	Curas	curas

No	Jenis Kejahatan	Label
5	Utang piutang	hutangpiutang
6	Implikasi Kontijensi	imlikasikontijensi
7	Jabatan	jabatan
8	Kebakaran	kebakaran
9	Kesusilaan	kesusilaan
10	Pemalsuan Surat	pemalsuansurat
11	Pembakaran	pembakaran
12	Pembunuhan	pembunuhan
13	Pemerasan	pemerasan
14	Penadahan	penadahan
15	Penculikan	penculikan
16	Pencurian ringan	pencurianringan
17	Penganiayaan	penganiayaan
18	Merusak barang	pengerusakan
19	Penggelapan	penggelapan
20	Penghinaan	penghinaan
21	Penipuan	penipuan
22	Menerima suap	penyuapan
23	Perjudian	perjudian
24	Perkosaan	perkosaan
25	Perdagangan anak dan wanita	perdaganganmanusia
26	Korupsi	korupsi
27	Kejahatan lalu lintas	lantas
28	Narkoba	narkoba
29	Bukan merupakan laporan kejahatan	himbauan
30	Tidak termasuk ke dalam jenis kejahatan	lain-lain

4.4.2. Subdit

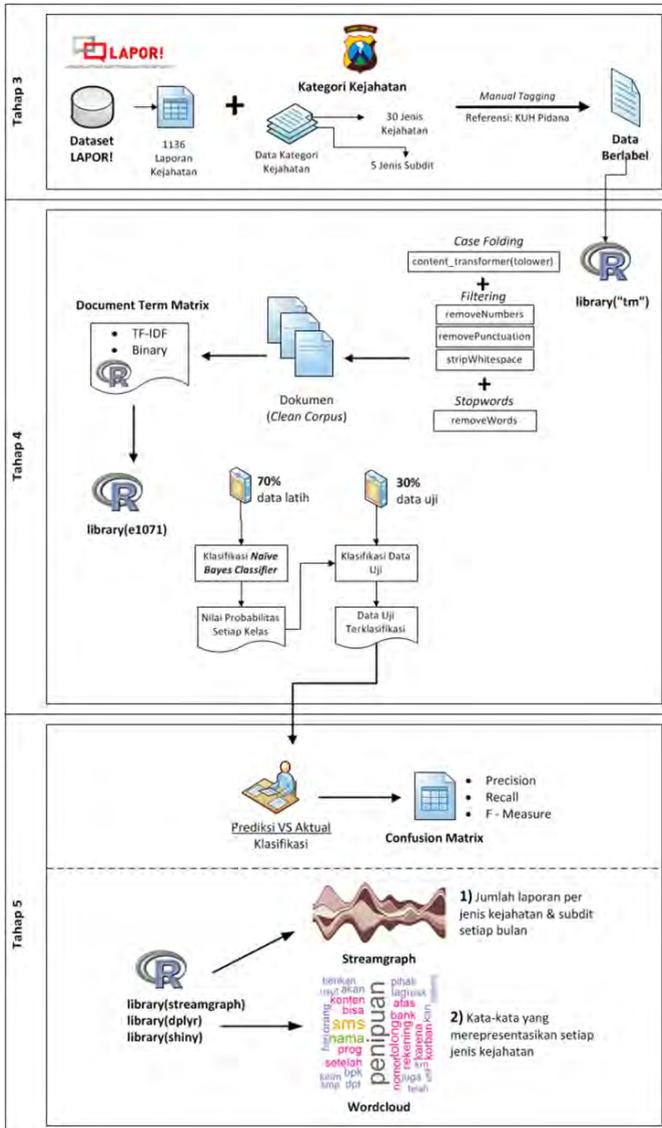
Label Subdit merupakan pengelompokan label jenis kejahatan ke dalam Sub Direktorat berdasar pada fungsi masing-masing bidang penyidikan dan penyelidikan. Terdapat lima label Subdit, dimana empat merupakan jenis Sub Direktorat dan satu merupakan label dari jenis kejahatan yang tidak tergolong ke dalam empat Subdit tersebut.

Tabel 4.3 Label Subdit

Label	Jenis Kejahatan
I	Kebakaran Pembakaran Implikasi Kontijensi
II	Pemalsuan surat Penipuan Penggelapan Utang piutang Jabatan
III	Curat Curas Curanmor Pembunuhan Pemerasan Perjudian Pengerusakan Penculikan Penghinaan Akibat orang luka Pencurian ringan Penadahan Menerima suap Penganiayaan
IV	Perkosaan Kesusilaan Perdagangan manusia
V	Lain – lain Himbauan

4.5. Metode Penelitian

Metode penelitian ini merupakan penjelasan dari metode penelitian pada Gambar 3.1 tahap 3, 4, dan 5 secara lebih detail. Pada metode ini dijelaskan bagaimana implementasi dari model klasifikasi dengan Naïve Bayes Classifier. Adapun metode implementasi yang digunakan ditunjukkan pada Gambar 4.2.



Gambar 4.2 Metode Penelitian Tahap 3, 4, dan 5

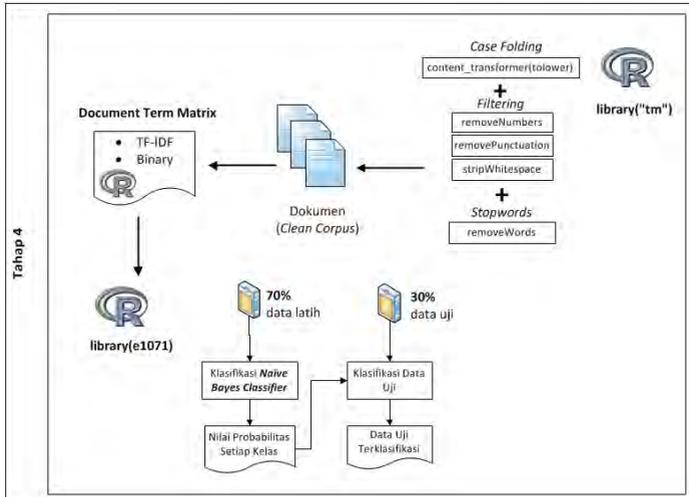
Tahapan yang ada pada metode implementasi ini merupakan bagian dari diagram alir pengerjaan tugas akhir yaitu tahap 3

pengumpulan data, tahap 4 pengolahan data, dan tahap 5 analisis hasil dan keakurasian.



Gambar 4.3 Metode Penelitian Tahap 3

Tahapan pertama pada Gambar 4.3 dimulai dari pengumpulan data LAPOR! yang terdiri dari 1136 laporan kejahatan dan data kategori kejahatan yang terdiri dari dua macam data yaitu 30 jenis kejahatan dan 5 jenis subdit. Kemudian dilakukan *manual tagging* atau pelabelan laporan kejahatan oleh data kategori kejahatan dimana digunakan KUH Pidana sebagai referensi. Pada tahap ini dihasilkan data berlabel yang akan digunakan pada proses selanjutnya.

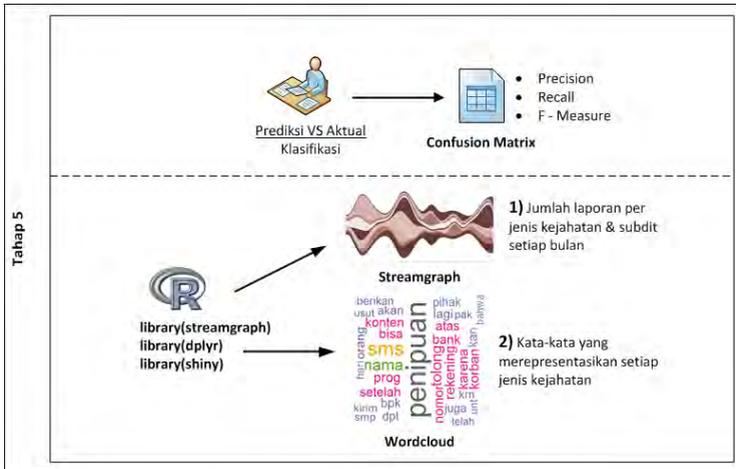


Gambar 4.4 Metode Penelitian Tahap 4

Tahap kedua pada Gambar 4.4 yaitu pengolahan data menggunakan aplikasi R Studio. Aplikasi ini digunakan untuk mempermudah tahapan praproses, yang terdiri dari:

- *Case folding*: merubah huruf dalam dokumen menjadi huruf kecil (*lower case*)
- *Filtering*: menghilangkan angka dan tanda baca yang tidak memiliki arti signifikan
- *Stopwords*: menghilangkan kata atau frasa yang tidak bermakna secara signifikan

Setelah ketiga proses selesai, proses terakhir dari praproses yaitu menghapus kelebihan spasi pada dokumen. Hasil dari tahapan praproses ini yaitu dokumen yang telah bersih (*clean corpus*) dan siap diolah ke dalam bentuk *document term matrix* (dtm). Untuk membentuk *clean corpus* menjadi dtm, digunakan bin atau Tf-Idf sebagai metode pembobotan.



Gambar 4.5 Metode Penelitian Tahap 5

Data yang dihasilkan dari proses pembobotan kemudian dibagi menjadi dua bagian yaitu data latih sebesar 70% dan data uji sebesar 30%. Dari data latih, dilakukan proses klasifikasi menggunakan NBC dan menghasilkan nilai probabilitas setiap kelas. Nilai probabilitas setiap kelas yang dihasilkan

digunakan untuk melakukan klasifikasi data uji sehingga menghasilkan data uji yang terklasifikasi.

Tahap terakhir pada Gambar 4.5 yaitu mengukur presentasi ketepatan (akurasi) dan visualisasi hasil klasifikasi yang dilakukan pada tahapan sebelumnya. Untuk mengukur akurasi klasifikasi ini digunakan *confusion matrix* yang mana akan menghasilkan nilai *precision*, *recall*, dan *F-Measure* dari setiap kelas. Sedangkan untuk mempermudah pembacaan data, dilakukan visualisasi berbasis web menggunakan shiny dengan jenis visualisi *streamgraph* yaitu grafik yang menunjukkan jumlah laporan dari setiap jenis kejahatan berdasarkan waktu dan *wordcloud* yaitu penggambaran kata-kata yang merepresentasikan setiap jenis kejahatan.

Halaman ini sengaja dikosongkan

BAB V IMPLEMENTASI

Pada bab ini akan dijelaskan mengenai proses implementasi *text mining* dengan menggunakan metode klasifikasi *Naïve Bayes Classifier* serta pembahasan visualisasi.

5.1. Data Implementasi

Data laporan yang digunakan merupakan sampel laporan yang diambil dari portal www.lapor.go.id dengan topik laporan yaitu kriminalitas. Total sampel laporan yang diambil berjumlah 1136 laporan. Sampel data tersebut dibagi menjadi dua jenis data, yakni data latih dan data uji yang proporsinya sebesar 70:30.

5.2. Lingkungan Implementasi

Lingkungan implementasi merupakan kriteria perangkat pengujian yang digunakan dalam menguji model yang dibangun pada tugas akhir ini. Lingkungan implementasi terdiri dari perangkat keras dan perangkat lunak. Adapun perangkat keras yang digunakan ditunjukkan pada Tabel 5.1.

Tabel 5.1 Lingkungan Uji Coba Perangkat Keras

Perangkat Keras	Spesifikasi
Jenis	Notebook
Processor	Intel® Pentium® Dual CPU T2370 @ 1.73GHz 1.73GHz
RAM	2.50 GB

Sementara lingkungan perangkat lunak yang digunakan dalam pengujian sistem yang dibangun pada tugas akhir ini yang ditunjukkan pada Tabel 5.2.

Tabel 5.2 Lingkungan Uji Coba Perangkat Lunak

Perangkat Lunak	Spesifikasi
Sistem Operasi	Windows 7
Bahasa Pemrograman	R
Tools	<ul style="list-style-type: none"> • Microsoft Excel • R Studio

5.3. Klasifikasi

Proses klasifikasi dilakukan untuk mengelompokkan data laporan menjadi beberapa kategori kelas yang ditentukan. Dalam penelitian ini digunakan kategori kejahatan berjumlah 30 dengan rincian pada Tabel 4.2

5.3.1. Pengerjaan Algoritma Klasifikasi

Algoritma klasifikasi yang digunakan pada tugas akhir ini adalah Naïve Bayes Classifier. Berdasarkan penjelasan metode penelitian pada Gambar 4.2 tahapan klasifikasi dapat dijabarkan ke dalam empat tahap, yaitu:

1. *Import Data*
2. Praproses Data
3. Representasi Data
4. Klasifikasi Data
5. Evaluasi Data

Pengerjaan algoritma ini dilakukan dengan aplikasi R Studio menggunakan *package* tm [26], e1071 [27], dan caret [28] menggunakan fungsi pada Tabel 5.1 berikut.

Tabel 5.3 Package dan Fungsi R Studio dalam Pengerjaan Algoritma Klasifikasi

<i>Packages</i>	Fungsi	Definisi
Tm	<ol style="list-style-type: none"> 1. Corpus 2. tm_map 3. content_transformer 4. removeNumber 	<ol style="list-style-type: none"> 1. Corpora 2. Transformations on Corpora 3. Content Transformers 4. Remove Numbers from

<i>Packages</i>	Fungsi	Definisi
	5. removePunctuation 6. removeWords 7. stopwords 8. stripWhitespace 9. DocumentTermMatrix	a Text Document 5. Remove Punctuation Marks from a Text Document 6. Remove Words from a Text Document 7. Stopwords 8. Strip Whitespace from a Text Document 9. Term-Document Matrix
e1071	1. naiveBayes 2. predict	1. Naive Bayes Classifier 2. Naive Bayes Classifier
caret	1. confusionMatrix	2. Create a confusion matrix

5.3.1.1. *Import Data*

Data yang akan diolah terlebih dahulu disimpan dalam format .csv kemudian dilakukan *import data* untuk diolah menggunakan R Studio. *Import data* dapat dilakukan dengan menggunakan “import dataset” pada *environment* atau dengan mengetikkan Kode 5.1.

```
# Load the data
tag <- read.csv("E:/MATA KULIAH/Skripsi/bahanta/LAPOR!/20151004/subdit.csv")
```

Kode 5.1 Membaca Data

Setelah data masuk pada *workspace* dan sebelum melakukan pengolahan lebih lanjut, data terlebih dahulu perlu diubah ke dalam bentuk *corpus* dengan menjalankan Kode 5.2.

```
# Load the data as a corpus
crime_corpus <-
Corpus(VectorSource(tag$isiLaporan))
```

Kode 5.2 Menjadikan Data sebagai Korpus

Jika kode program ini telah berhasil dijalankan, maka siap dilakukan pengolahan.

5.3.1.2. Praproses Data

Praproses data merupakan tahap yang dilakukan untuk mentransformasi data tekstual yang bersifat tidak terstruktur menjadi terstruktur. Terdiri dari beberapa tahapan yaitu *case folding*, *filtering* dan *stopwords*. Langkah awal yang harus dilakukan yaitu menjalankan *library* text mining (tm) berikut.

```
>> library(tm)
```

Tahap pertama yaitu *case folding*, merupakan proses perubahan huruf dalam dokumen menjadi satu bentuk, misalnya huruf kapital menjadi huruf kecil dan sebaliknya. Kode program R yang digunakan untuk melakukan *case folding* seperti ditunjukkan pada Kode 5.3.

```
# Convert the text to lower case
clean_corpus <-
tm_map(crime_corpus, content_transformer(tolower))
```

Kode 5.3 Mengkonversi Data ke Dalam Huruf Kecil

Huruf dengan *highlight* berwarna biru pada Tabel 5.4 mengalami proses perubahan huruf menjadi huruf kecil. Sehingga dapat dilihat hasil transformasi proses *case folding* sebagai berikut.

Tabel 5.4 Perbandingan Laporan Sebelum dan Sesudah Melewati Proses *Case Folding*

Contoh Isi Laporan Asli	Isi Laporan Setelah Melewati Proses <i>Case Folding</i>
[1] "Penipuan di ITC Cempaka Mas merugikan banyak kalangan . modulusnya kita di kasih hadiah dari salah satu CV yang lagi	[1] "penipuan di itc cempaka mas merugikan banyak kalangan . modulusnya kita di kasih hadiah dari salah satu cv yang lagi

Contoh Isi Laporan Asli	Isi Laporan Setelah Melewati Proses <i>Case Folding</i>
<p>ulang tahun, disitu kita dikasih hadiah 4 barang, tapi kita disuruh beli 1 barangnya dia yang harganya ga masuk akal . terus dengan paksa dia minta identitas, foto, sama atm kita . nah tanpa kita sadar oknum ini menggesek atm kita di alatnya, uang 3 juta langsung raib dituker sama barang barang yang bisa di bilang ga bermutu . disitu juga katanya kita di iming imingi upah bulanan selama 1 tahun sebesar 1,5 juta mulai bulan juni akan di transfer ke rekening yang kita pakai . tolong dong masalah ini di selesaikan, karna saya salah satu korbannya . dan saya juga baru sadar setelah tiba di rumah . saya searh di blog ternyata kronologisnya sama persis . saya minta tolong tindak pidana mereka yang melakukan kejahatan ini . terima kasih."</p>	<p>ulang tahun, disitu kita dikasih hadiah 4 barang, tapi kita disuruh beli 1 barangnya dia yang harganya ga masuk akal . terus dengan paksa dia minta identitas, foto, sama atm kita . nah tanpa kita sadar oknum ini menggesek atm kita di alatnya, uang 3 juta langsung raib dituker sama barang barang yang bisa di bilang ga bermutu . disitu juga katanya kita di iming imingi upah bulanan selama 1 tahun sebesar 1,5 juta mulai bulan juni akan di transfer ke rekening yang kita pakai . tolong dong masalah ini di selesaikan, karna saya salah satu korbannya . dan saya juga baru sadar setelah tiba di rumah . saya searh di blog ternyata kronologisnya sama persis . saya minta tolong tindak pidana mereka yang melakukan kejahatan ini . terima kasih."</p>

Setelah seluruh kata diubah menjadi satu bentuk, selanjutnya dilakukan *filtering*. Tanda baca yang tidak memiliki arti yang signifikan atau termasuk noise (pengganggu) akan dieliminasi. Langkah eliminasi yang dilakukan ditunjukkan dengan menjalankan Kode 5.4.

```
# Remove numbers
clean_corpus <- tm_map(clean_corpus,
removeNumbers)
# Remove punctuations
clean_corpus <- tm_map(clean_corpus,
removePunctuation)
# Eliminate extra white spaces
```

Kode 5.4 Menghapus, Nomor, Simbol, dan Mengelemiasi Kelebihan Spasi

Angka dan tanda baca dengan *highlight* berwarna kuning pada Tabel 5.5 perlu dieliminasi karena tidak memiliki arti yang signifikan. Sehingga hasil dari laporan setelah melewati proses *filtering* sebagai berikut.

Tabel 5.5 Perbandingan Laporan Sebelum dan Sesudah Melewati Proses *Filtering*

Isi Laporan Setelah Melewati Proses <i>Case Folding</i>	Isi Laporan Setelah Melewati Proses <i>Filtering</i>
<p>[1] "penipuan di itc cempaka mas merugikan banyak kalangan . modusnya kita di kasih hadiah dari salah satu cv yang lagi ulang tahun, disitu kita dikasih hadiah 4 barang, tapi kita disuruh beli 1 barangnya dia yang harganya ga masuk akal , terus dengan paksa dia minta identitas, foto, sama atm kita , nah tanpa kita sadar oknum ini menggesek atm kita di alatnya, uang 3 juta langsung raib dituker sama barang barang yang bisa di bilang ga bermutu . disitu juga katanya kita di iming imingi upah bulanan selama 1 tahun sebesar 1,5 juta mulai bulan juni akan di transfer ke rekening yang kita pakai , tolong dong masalah ini di selesaikan, karna saya salah satu korbannya . dan saya juga baru sadar setelah tiba di rumah , saya searh di blog ternyata kronologisnya sama persis , saya minta tolong tindak pidana mereka yang melakukan kejahatan ini . terima kasih."</p>	<p>[1] "penipuan di itc cempaka mas merugikan banyak kalangan modusnya kita di kasih hadiah dari salah satu cv yang lagi ulang tahun disitu kita dikasih hadiah barang tapi kita disuruh beli barangnya dia yang harganya ga masuk akal terus dengan paksa dia minta identitas foto sama atm kita nah tanpa kita sadar oknum ini menggesek atm kita di alatnya uang juta langsung raib dituker sama barang barang yang bisa di bilang ga bermutu disitu juga katanya kita di iming imingi upah bulanan selama tahun sebesar juta mulai bulan juni akan di transfer ke rekening yang kita pakai tolong dong masalah ini di selesaikan karna saya salah satu korbannya dan saya juga baru sadar setelah tiba di rumah saya searh di blog ternyata kronologisnya sama persis saya minta tolong tindak pidana mereka yang melakukan kejahatan ini terima kasih"</p>

Selain tanda baca, kata atau frasa yang tidak bermakna secara signifikan juga dapat mengurangi performa klasifikasi. Oleh

karena itu, kata atau frase tersebut perlu dieliminasi. Proses stopwords ditunjukkan seperti Kode 5.5.

```
# Remove stopwords english
clean_corpus <- tm_map(clean_corpus, removewords,
stopwords("english"))
# Remove stopwords manual
clean_corpus <- tm_map(clean_corpus, removewords,
c("nama", "hari", "yth", "sdh", "jadi", "org",
"dpt", "krn", "bpk", "lagi", "akan", "sampai",
"kan", "setelah", "pak", "atas", "orang", "saya",
"dgn", "atau", "juga", "bisa", "kepada", "tsb",
"sangat", "mereka", "karena", "dan", "yang", "ini",
"dengan", "ada", "tidak", "untuk", "dari", "nya",
"tdk", "oleh", "kami", "dia", "mohon", "itu",
"tersebut", "pada", "sudah"))
# Remove stopwords indonesian
stopwordID <-
"https://raw.githubusercontent.com/nurandi/nurandi.
net/master/data/stopwords-id.txt"
cstopwordID <- readLines(stopwordID);
clean_corpus <- tm_map(clean_corpus, removewords,
c(cstopwordID))
```

Kode 5.5 Menghapus Kata

Huruf dengan *highlight* berwarna hijau pada Tabel 5.6 merupakan kata atau frasa yang dapat mengganggu proses klasifikasi, sehingga perlu dieliminasi. Hasil akhir dari proses ini ditunjukkan pada berikut.

Tabel 5.6 Perbandingan Laporan Sebelum dan Sesudah Melewati Proses *Stopwords*

Isi Laporan Setelah Melewati Proses <i>Filtering</i>	Isi Laporan Setelah Melewati Proses <i>Stopwords</i>
[1] "penipuan di itc cempaka mas merugikan banyak kalangan modusnya kita di kasih hadiah dari salah satu cv yang lagi ulang tahun disitu kita dikasih hadiah barang tapi kita disuruh beli barangnya dia yang harganya ga masuk akal terus dengan paksa dia minta identitas foto sama atm kita nah tanpa kita sadar oknum	[1] "penipuan itc cempaka mas merugikan kalangan modusnya kasih hadiah salah satu cv ulang tahun disitu dikasih hadiah barang disuruh beli barangnya harganya ga masuk akal terus paksa minta identitas foto atm sadar oknum menggesek atm alatnya uang juta langsung raib dituker barang barang bilang ga

Isi Laporan Setelah Melewati Proses <i>Filtering</i>	Isi Laporan Setelah Melewati Proses <i>Stopwords</i>
<p>ini menggesek atm kita di alatnya uang juta langsung raib dituker sama barang barang yang bisa di bilang ga bermutu disitu juga katanya kita di iming imingi upah bulanan selama tahun sebesar juta mulai bulan juni akan di transfer ke rekening yang kita pakai tolong dong masalah ini di selesaikan karna saya salah satu korbannya dan saya juga baru sadar setelah tiba di rumah saya searh di blog ternyata kronologisnya sama persis saya minta tolong tindak pidana mereka yang melakukan kejahatan ini terima kasih"</p>	<p>bermutu disitu katanya iming imingi upah bulanan tahun sebesar juta mulai bulan juni transfer rekening pakai tolong masalah selesaikan karna salah satu korbannya baru sadar tiba rumah searh blog ternyata kronologisnya persis minta tolong tindak pidana melakukan kejahatan terima kasih"</p>

Setelah melewati ketiga tahap pra-proses tersebut, data sudah bisa dikatakan bersih serta siap olah.

5.3.1.3. Representasi Data

Data yang telah melewati tahapan praproses tidak bisa langsung diolah dalam bentuk dokumen teks aslinya. Perlu proses untuk mengubah dokumen menjadi suatu representasi yang mudah dikelola lebih lanjut. Pada tahapan ini, akan dilakukan proses representasi data menjadi vector atau numeric dalam bentuk matrix yang dapat menggunakan bermacam metode, diantaranya Tf, Tf-Idf, dan binary. Kode program yang digunakan yaitu Kode 5.6.

```
# tokenizing the corpus
dtm <- DocumentTermMatrix(clean_corpus,
control=list(weighting=weightTfIdf))
```

Kode 5.6 Tokenisasi Korpus

Tahap di atas ini merupakan proses tokenisasi corpus dengan metode Tf-Idf yang bertujuan untuk mentransformasi data ke dalam bentuk *Document Term Matrix*.

Setelah dilakukan tokenisasi, *frequent words* digunakan sebagai *dictionary* pada pembobotan yang akan dilakukan selanjutnya. Kata yang digunakan sebagai *dictionary* dapat kita atur berdasarkan jumlah kemunculan kata pada suatu dokumen.

```
#Identify frequently used words
dtm_train <- dtm[train_indices, ]
frequent_words <- findFreqTerms(dtm_train, 5)

#Create document-term matrices using frequent words
crime_train <- DocumentTermMatrix(train_corpus,
control=list(dictionary = frequent_words))
crime_test <-
DocumentTermMatrix(test_corpus,
control=list(dictionary = frequent_words))
```

Kode 5.7 Mengidentifikasi Frequent Words dan Membuat DTM

Selain itu, untuk mempermudah proses pembobotan, dilakukan konversi informasi ke dalam bentuk kata “yes” dan “no” seperti pada Kode 5.8. Setelah konversi diimplementasikan pada *Document Term Matrix*, klasifikasi data siap dilakukan.

```
#Convert count information to "Yes", "No"
convert_count <- function(x) {
  y <- ifelse(x > 0, 1,0)
  y <- factor(y, levels=c(0,1), labels=c("No",
"yes"))
  y
}

#Convert document-term matrices
crime_train <- apply(crime_train, 2,
convert_count)
crime_test <- apply(crime_test, 2,
convert_count)
```

Kode 5.8 Mengkonversi Informasi ke Dalam Yes & No dan Mengaplikasikan ke DTM

5.3.1.4. Klasifikasi Data

Data hasil dari proses pembobotan, dibagi menjadi dua bagian yaitu data latih sebesar 70% dan data uji sebesar 30% dengan menjalankan Kode 5.9.

```
#divide corpus into training and test data
train_size <- round(0.7*nrow(tag))
set.seed(141321)
train_indices <- sample(seq_len(nrow(tag)), size =
train_size)
train_corpus <- clean_corpus[ train_indices]
test_corpus <- clean_corpus[-train_indices]
```

Kode 5.9 Membagi Data Latih dan Uji

Setelah data terbagi, dilakukan proses klasifikasi terhadap data latih menggunakan *Naïve Bayes Classifier* sehingga menghasilkan nilai probabilitas setiap kelas. Nilai probabilitas setiap kelas yang dihasilkan digunakan untuk melakukan klasifikasi data uji sehingga menghasilkan data uji yang terklasifikasi. Proses pada aplikasi R ditunjukkan pada Kode 5.10.

```
#Load the naive bayes library
library(e1071)

#Create a Naive Bayes classifier object
crime_classifier <- naiveBayes(crime_train,
factor(tag$tag[train_indices]))
```

Kode 5.10 Membuat Pengklasifikasi

5.3.1.5. Evaluasi Data

Tahapan terakhir yaitu melakukan evaluasi performa klasifikasi data dengan mengukur presentase ketepatan pada data uji. Nilai probabilitas setiap kelas yang dihasilkan pada data latih digunakan pada tahap ini, yaitu Kode 5.11.

```
#Evaluate the performance on the test data
predicted <- predict(crime_classifier,
crime_test)
actual <- tag$tag[-train_indices]
```

Kode 5.11 Mengevaluasi Performa Pada Data Tes

Hasil data uji yang terklasifikasi selanjutnya akan diukur bagaimana tingkat akurasinya, seperti ditunjukkan pada Kode 5.12. Untuk mengukur akurasi klasifikasi ini digunakan *confusion matrix* yang sekaligus akan menghasilkan nilai *precision*, *recall*, dan *F-Measure* dari setiap kelas.

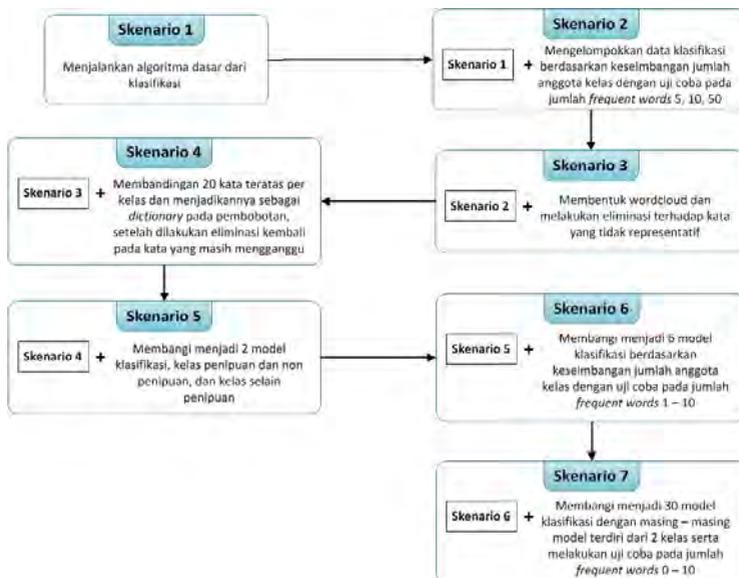
```
# Calculating the accuracy in percents
levels(predicted) <- levels(actual)
100*sum(predicted==actual)/length(predicted)

#Create Confusion Matrix
confusionMatrix(actual, predicted)
```

Kode 5.12 Menghitung Akurasi dalam Persen dan Membuat *Confusion Matrix*

5.3.2. Percobaan Klasifikasi

Untuk mendapatkan hasil klasifikasi terbaik, dibutuhkan konfigurasi yang tepat terhadap algoritma klasifikasi. Setiap percobaan yang nantinya dilakukan akan menghasilkan algoritma – algoritma pengembangan yang lebih baik sehingga didapatkan model klasifikasi yang paling optimal.



Gambar 5.1 Proses Penemuan Percobaan Klasifikasi Terbaik

Pada penelitian ini dilakukan 7 percobaan pengembangan. Dimulai dari percobaan 1 yang menjalankan algoritma dasar dari klasifikasi tanpa penghapusan *stopwords*. Hasil percobaan ini menunjukkan bukan merupakan model optimal. Setelah dianalisis, ditemukan bahwa anggota setiap kelas memiliki jumlah yang tidak berimbang. Oleh karena itu, dilakukan pengelompokan kelas berdasarkan jumlah anggota pada percobaan 2.

Pada percobaan 2 hasil yang didapat menunjukkan bahwa akurasi meningkat tetapi belum optimal. Kemudian pada percobaan 3 dilakukan analisis kata – kata yang mendominasi setiap kelas. Hasil yang ditemukan yaitu adanya kata konjungsi dan frasa yang tidak representatif terhadap kategori kejahatan. Kata – kata tersebut kemudian dieleminasi dan menghasilkan akurasi yang semakin meningkat dengan beberapa juga mengalami penurunan. Oleh karena pada percobaan 3 ditemukan kata yang sama di beberapa kelas berbeda, maka dilakukan percobaan pengembangan 4.

Pada percobaan 4 ini dilakukan perbandingan 20 kata teratas yang muncul pada setiap kelas. Kemudian kata - kata tersebut digabungkan dan digunakan sebagai *dictionary* pada tahap pembobotan. Hasil percobaan ini lebih baik karena ditemukan akurasi optimal.

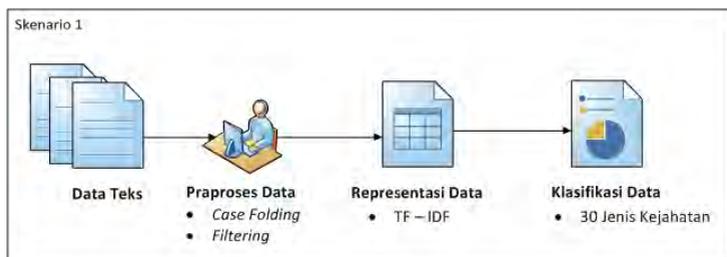
Percobaan 5 melakukan optimalisasi percobaan 4 yaitu dengan membentuk 2 model klasifikasi. Model pertama menggunakan model yang ditemukan pada percobaan 4 yaitu kelas penipuan dan non penipuan. Sedangkan algoritma kedua menggunakan kelas non penipuan pada model 1 dengan melakukan klasifikasi terhadap 29 kelas selain kelas penipuan. Nilai akurasi yang dihasilkan meningkat, tetapi tidak untuk model 2.

Percobaan 6 dibuat untuk menyelesaikan permasalahan pada model 2. Dilakukan pembagian kelas – kelas berdasarkan jumlah anggota kelas sehingga terbentuk 6 model klasifikasi. Secara keseluruhan akurasi meningkat, tetapi semakin menurun seiring berkurangnya jumlah data.

Percobaan 7 dibentuk 29 model klasifikasi dengan masing – masing model hanya mengklasifikasi 2 kelas. Susunan model diawali dari kelas dengan jumlah anggota terbanyak. Hasilnya bisa dikatakan bahwa model ini mampu mempelajari data dengan baik. Sehingga, dapat disimpulkan bahwa pada percobaan 7 ini merupakan model klasifikasi terbaik.

5.3.2.1. Percobaan 1

Pada percobaan pertama ini digunakan algoritma klasifikasi seperti yang dijelaskan pada bab 5.3.1 tanpa dilakukan konfigurasi tambahan. Algoritma yang dijalankan pada percobaan 1 tergambar pada Gambar 5.2 berikut.



Gambar 5.2 Algoritma Percobaan 1

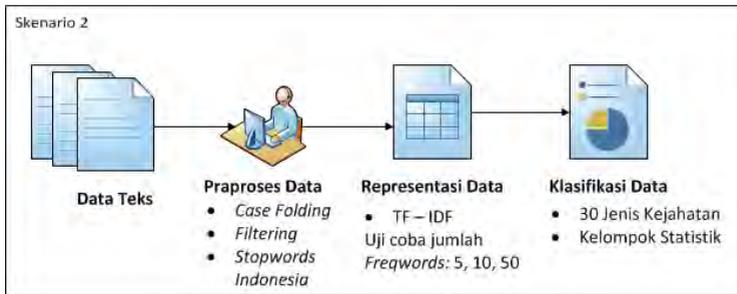
Percobaan di atas menjalankan algoritma dasar dari klasifikasi menggunakan Naïve Bayes Classifier. Tahap awal dilakukan persiapan data dengan mengubah ke dalam bentuk korpus, praproses data, representasi data ke dalam bentuk matriks, klasifikasi data, dan terakhir dilakukan evaluasi terhadap keakurasian hasil klasifikasi.

Hasil yang didapat menunjukkan nilai akurasi kurang dari 10%. Tentu dengan nilai akurasi tersebut belum bisa dikatakan bahwa model sudah optimal.

5.3.2.2. Percobaan 2

Hasil yang didapatkan pada percobaan 1 yaitu nilai akurasi yang sangat minim, kurang dari 10%. Dengan nilai akurasi tersebut tentu belum bisa dikatakan bahwa model sudah

optimal. Setelah dilakukan analisis terhadap hasil klasifikasi, Gambar 5.4 memperlihatkan bahwa jumlah anggota setiap kelas tidak berimbang. Hal ini tentu sangat berpengaruh terhadap hasil klasifikasi. Berikut percobaan pengembangan yang dibuat untuk menyelesaikan permasalahan pada percobaan 1.



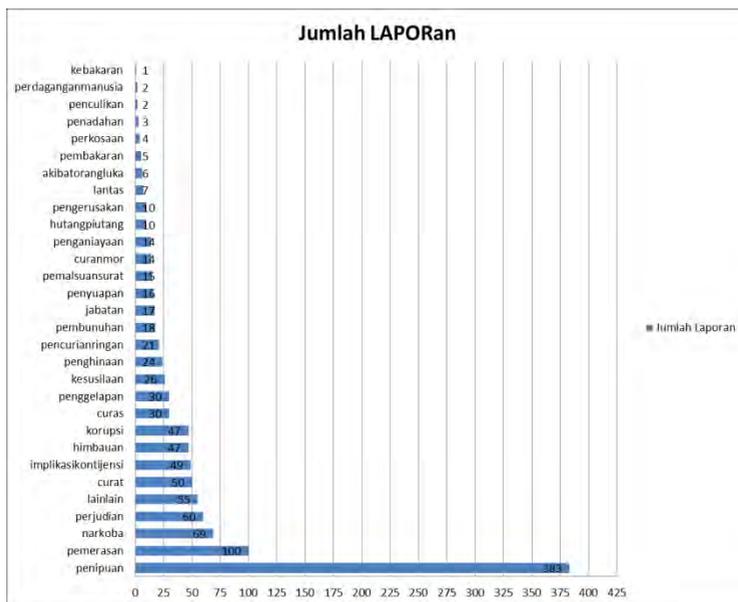
Gambar 5.3 Algoritma Percobaan 2

Grafik di bawah menunjukkan bahwa terdapat ketidakseimbangan jumlah anggota setiap kelas. Ada kelas tertentu yang terlalu mendominasi kelas lainnya. Seperti kelas penipuan dengan anggota terbanyak yang sangat signifikan jumlahnya dibandingkan dengan pemerasan.

Model klasifikasi yang baik menggunakan data dengan jumlah anggota kelas yang berimbang. Hal ini bertujuan agar model dapat belajar dengan optimal guna menghasilkan akurasi yang maksimal. Pada percobaan 2 ini akan dilakukan percobaan dengan menggunakan kelas-kelas dengan jumlah data yang berimbang. Kelas-kelas tersebut jika dikelompokkan berdasarkan jumlah anggota adalah sebagai berikut.

1. Kelas yang beranggotakan lebih dari 10
2. Kelas yang beranggotakan lebih dari 20
3. Kelas yang beranggotakan lebih dari 30
4. Kelas yang beranggotakan lebih dari 30 dan tanpa memakai kelas penipuan.

5. Kelas yang beranggotakan lebih dari 30 dan tanpa memakai kelas lain-lain dan kelas himbauan.



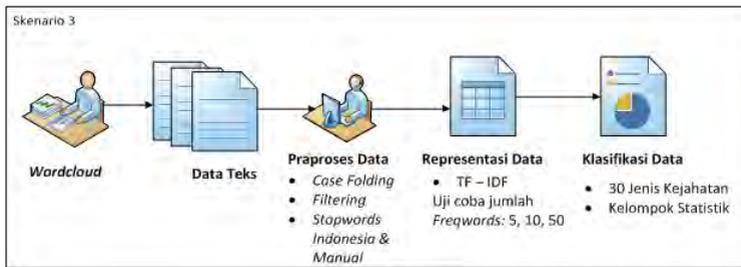
Gambar 5.4 Statistik Jumlah Laporan

Pada representasi data, dilakukan uji coba terhadap jumlah *frequent words* yaitu 5, 10, dan 50. Hal ini dilakukan karena dengan menggunakan jumlah *frequent words* yang beragam akan menghasilkan akurasi yang berbeda pula.

Hasilnya dapat disimpulkan bahwa dengan menggunakan data yang berimbang dan penggunaan *frequent words* yang berbeda, menghasilkan akurasi yang lebih baik. Namun, dengan hasil akurasi yang tidak lebih dari 65% tentu masih perlu dilakukan perbaikan algoritma.

5.3.2.3. Percobaan 3

Percobaan 2 menghasilkan kesimpulan bahwa jumlah anggota kelas yang berimbang akan menghasilkan performa klasifikasi yang lebih baik. Namun, akurasi yang didapatkan tidak lebih dari 65% dan tentu masih perlu dilakukan perbaikan algoritma. Setelah dilakukan analisis pada hasil percobaan 2, ditemukan kata-kata yang mengganggu proses klasifikasi. Maka dari itu perlu ditinjau kata-kata apa yang mendominasi pada setiap kelas dengan cara pembentukan *wordcloud*. Berikut percobaan yang akan dilakukan.



Gambar 5.5 Algoritma Percobaan 3

Pertama kita membentuk *wordcloud* untuk melihat kata-kata apa saja yang merepresentasikan setiap kelas. Untuk membuatnya digunakan *package wordcloud* [29] dengan kode program yang digunakan pada Kode 5.13.

```

# Load library
library("tm")
library("wordcloud")
library("RColorBrewer")

# Load the data as a corpus
docs <- Corpus(vectorSource(lapor))

inspect(docs)

toSpace <- content_transformer(function(x,pattern)
gsub(pattern, " ", x))
docs <- tm_map(docs, toSpace, "/"")
docs <- tm_map(docs, toSpace, "@")
docs <- tm_map(docs, toSpace, "\\|")
  
```

```

# Convert the text to lower case
docs <- tm_map(docs, content_transformer(tolower))
# Remove numbers
docs <- tm_map(docs, removeNumbers)
# Remove punctuations
docs <- tm_map(docs, removePunctuation)
# Specify your stopwords as character vector
stopwordID <-
"https://raw.githubusercontent.com/nurandi/nurandi.
net/master/data/stopwords-id.txt"
cStopwordID <- readLines(stopwordID);
docs <- tm_map(docs, removeWords,
stopwords("english",cStopwordID))
docs <- tm_map(docs, removeWords, c(cStopwordID))

# Eliminate extra white spaces
docs <- tm_map(docs, stripWhitespace)

dtm <- TermDocumentMatrix(docs)
m <- as.matrix(dtm)
v <- sort(rowSums(m),decreasing=TRUE)
d <- data.frame(word = names(v),freq=v)
dx <- head(d, 30)
head(dx, 20)
set.seed(1234)
wordcloud(words = d$word, freq = d$freq, min.freq =
1, max.words=50, random.order=FALSE, rot.per=0.35,
colors=brewer.pal(8, "Dark2"))

```

Kode 5.13 Kode Program untuk Membuat *Wordcloud*

Pada kode program di atas, *wordcloud* yang dihasilkan merupakan representasi dari dataset yang dimasukkan. Sehingga, jika yang diinginkan *wordcloud* dari setiap kelas, berarti dataset yang dimasukkan merupakan dataset dengan label kelas tertentu yang ingin dicari. Berikut merupakan salah satu hasil *wordcloud* pada kelas penipuan.



Gambar 5.6 Contoh Wordcloud Tanpa Eliminasi Kata

Pada *wordcloud* di atas terlihat bahwa ada kata konjungsi yang dominan dan mengganggu hasil dari representasi kelas. Untuk mengatasi hal ini, dilakukan *stopwords* terhadap kata-kata yang tidak representatif. Pada percobaan sebelumnya, sudah dilakukan *stopwords* dengan mengeliminasi kata – kata konjungsi berbahasa Indonesia menggunakan data *stopwords* bahasa Indonesia. Daftar kata – kata lengkapnya dapat dilihat pada Tabel B.1. Namun, masih saja ada kata – kata yang masih mengganggu dan belum tereleminasi. Untuk itu, perlu dilakukan penghapusan terhadap kata – kata tersebut. Berikut adalah Kode 5.14 yang digunakan beserta kata – kata yang perlu dieliminasi.

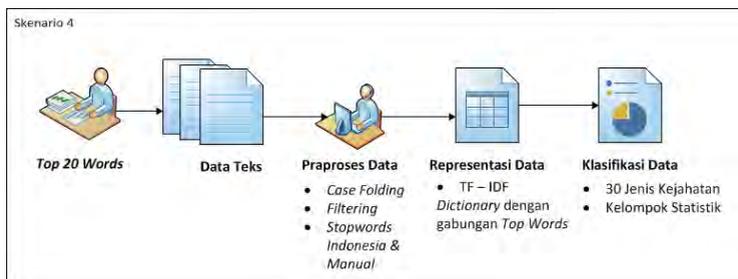
```
# Specify your stopwords as a character vector
docs <- tm_map(docs, removewords, c("unt",
  "nama", "hari", "yth", "sdh", "jadi", "org",
  "dpt", "krn", "bpk", "lagi", "akan", "sampai",
  "kan", "setelah", "pak", "atas", "orang", "saya",
  "dgn", "atau", "juga", "bisa", "kepada", "tsb",
  "sangat", "mereka", "karena", "dan", "yang",
  "ini", "dengan", "ada", "tidak", "untuk", "dari",
  "nya", "tdk", "oleh", "kami", "dia", "mohon",
  "itu", "tersebut", "pada", "sudah"))
```

Kode 5.14 Menghapus Kata dengan Stopwords Manual

Setelah dilakukan penghapusan dan dilakukan proses klasifikasi, akurasi yang dihasilkan secara keseluruhan meningkat tetapi masih dengan nilai akurasi yang kurang dari 70%.

5.3.2.4. Percobaan 4

Hasil akurasi klasifikasi pada percobaan 3 meningkat setelah dilakukan eliminasi kata-kata yang tidak representatif pada setiap kelas. Namun, setelah dilakukan analisis kembali terhadap hasil klasifikasi, terdapat kata yang sama di beberapa kelas yang berbeda. Sehingga, dibutuhkan analisis lebih lanjut terhadap kata yang merepresentasikan masing-masing kelas. Berikut adalah percobaan 4 yang akan dilakukan.



Gambar 5.7 Algoritma Percobaan 4

Untuk melihat kata-kata apa saja yang merepresentasikan setiap kelas, kita menggunakan Kode 5.15 di bawah ini. Kata-kata yang ditampilkan berjumlah 20 teratas berdasarkan jumlah kata terbanyak dari setiap kelas.

```

dtm <- TermDocumentMatrix(docs)
m <- as.matrix(dtm)
v <- sort(rowSums(m),decreasing=TRUE)
d <- data.frame(word = names(v),freq=v)
head(d, 20)
  
```

Kode 5.15 Menampilkan 20 Kata Teratas

Berikut kutipan analisis hasil yang didapat dari kelas penipuan, pemerasan, dan narkoba. Perbandingan analisis hasil

20 kata teratas dapat dilihat pada **LAMPIRAN B.2** untuk lebih lengkapnya.

Tabel 5.7 Kutipan Perbandingan 20 Kata Teratas

PENIPUAN		PEMERASAN		NARKOBA	
word	Freq	word	freq	word	freq
penipuan	671	pemerasan	110	narkoba	131
uang	248	preman	57	oki	33
sms	190	uang	50	sabu	30
hadiah	181	motor	33	shabu	24
tolong	138	ibu	32	ganja	21
rekening	125	kasih	30	bandar	19
prog	123	polisi	30	cengkareng	18
bank	119	rumah	28	warga	16
nomor	113	tolong	25	desa	13
konten	111	mobil	22	pedagang	13
korban	110	masalah	19	jatim	12
pihak	95	meminta	19	jenis	12
berikan	94	pihak	19	kecamatan	11
penipu	93	terima	19	anak	10
transfer	92	datang	18	kab	10
cirebon	90	masyarakat	18	kec	10
kirim	88	minta	18	tolong	10
smp	86	jalan	17	ditangkap	9
pelaku	82	surat	17	gembong	9
barang	81	keluarga	16	lampung	9

Kutipan hasil perbandingan 20 kata teratas di atas menunjukkan bahwa kata – kata dengan *highlight* berwarna kuning merupakan kata-kata yang berada pada lebih dari satu kelas. Untuk menghindari hal tersebut, pada tahap representasi dilakukan percobaan dengan menggunakan *dictionary* dari

penggabungan *topwords* dari semua kelas di atas dengan kode program yang digunakan yaitu Kode 5.16. Proses penentuan kata – kata yang digunakan beserta akurasi yang dihasilkan dapat dilihat pada **LAMPIRAN B.3**.

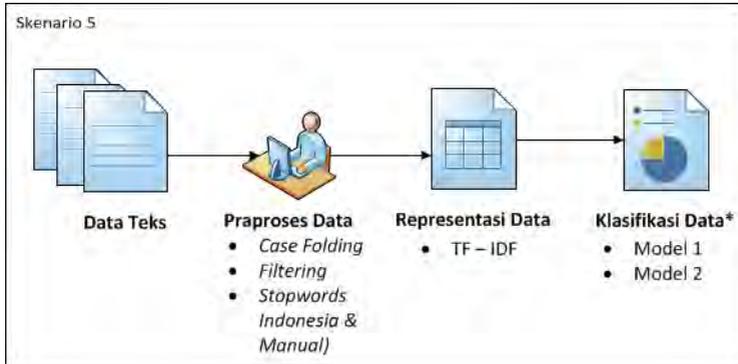
```
crime_train <- DocumentTermMatrix(train_corpus,
control=list(dictionary = c("akun", "anak",
"anggota", "angkot", "aparatus", "apbd", "arah",
"atm", "bandar", "bank", "bbm", "begal", "blok",
"bupati", "bus", "copet", "daerah", "dana",
"desa", "email", "facebook", "foto", "ganja",
"gembong", "hadiah", "http", "hukum", "ilegal",
"jalan", "jenis", "judi", "juta", "kasus",
"keamanan", "kejadian", "keluarga", "kepala",
" kirim", "komplek", "konten", "koperasi",
"korban", "kumpul", "malam", "marak", "masalah",
"masyarakat", "media", "meresahkan", "mobil",
"motor", "negara", "nomor", "pasir",
"pelacuran", "pelaku", "pembegalan",
"pemerintah", "pemilik", "pencemaran",
"pencurian", "pedagang", "penipuan",
"projudian", "pihak", "pin", "preman",
"prostitusi", "proyek", "rawan", "rekening",
"rumah", "sabu", "sms", "sosial", "surat",
"taman", "tambang", "tanah", "tim", "timah",
"tindak", "tni", "togel", "transfer", "uang",
"warga")))
```

Kode 5.16 Menggunakan *Dictionary* Manual Pada Pembobotan

Dari beberapa percobaan yang dilakukan, ditemukan akurasi optimal yaitu pada kelas penipuan dan non penipuan sebesar 90.90%. Akurasi dengan nilai inilah yang diharapkan dari model klasifikasi.

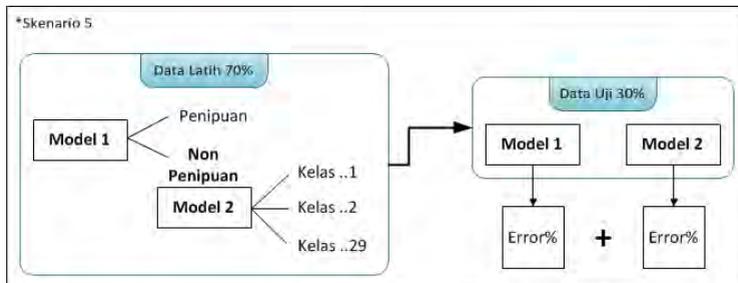
5.3.2.5. Percobaan 5

Dengan didapatkannya akurasi optimal sebesar 90.90% pada percobaan 4, maka perlu dilakukan konfigurasi data dan algoritma kembali untuk mendapatkan model klasifikasi yang optimal. Berikut percobaan pengembangan yang akan dilakukan.



Gambar 5.8 Algoritma Percobaan 5

Pada percobaan 5 ini dibentuk dua model klasifikasi, model satu yaitu klasifikasi kelas penipuan dan non penipuan, sedangkan model dua yaitu klasifikasi 29 kelas selain kelas penipuan seperti yang tergambar pada Gambar 5.9



Gambar 5.9 Lanjutan Algoritma Percobaan 5

Data yang digunakan pada setiap model berbeda. Data model 1 menggunakan semua data dengan label penipuan dan non penipuan. Sedangkan untuk data model 2 merupakan data non penipuan pada model 1 tetapi dengan label 29 kelas selain penipuan. Setelah dataset siap untuk diolah, pertama dilakukan proses klasifikasi pada model 1 dengan tahap seperti pada bab 5.3.1. Kemudian dilakukan kembali klasifikasi pada model 2 dengan pemilihan data model melalui Kode 5.17 berikut.

```
model2 <-
susun[which(susun$model1=="nonpenipuan"),]
```

Kode 5.17 Menseleksi Data Mode

Proses yang sama dilakukan kembali untuk model 2, dengan tahap akhir yaitu melakukan evaluasi akurasi total dari kedua model. Untuk mendapatkan hasil akurasi, digunakan Kode 5.18 di bawah ini.

```
nbc1 <-
100*sum(predicted==actual)/length(predicted)
nbc2 <-
100*sum(predicted2==actual2)/length(predicted2)
#Total Accuracy
((length(test_docs)*nbc1)+(length(test_docs2)*nbc2))/(length(test_docs)+length(test_docs2))
```

Kode 5.18 Menghitung Nilai Akurasi Total

Hasil yang didapatkan menunjukkan model 1 memiliki nilai akurasi sebesar 91.788% dan model 2 sebesar 31.858% dengan akurasi total 67.901%.

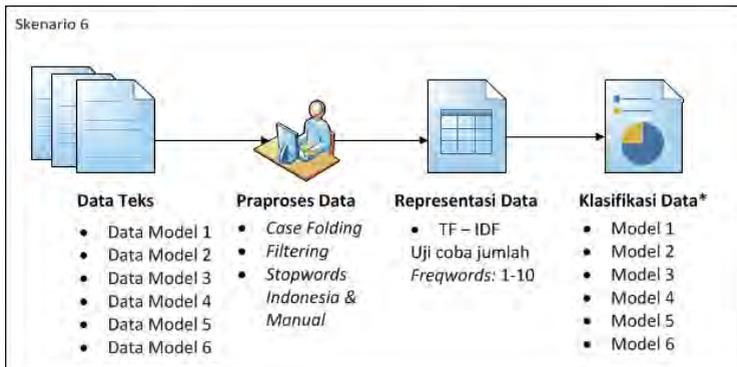
5.3.2.6. Percobaan 6

Klasifikasi model bersusun yang dijalankan pada percobaan 5 menghasilkan akurasi total 67.901% dengan rincian 91.788% untuk model 1 dan 31.858% untuk model 2. Akurasi model 1 sudah mencapai akurasi optimal, sedangkan untuk model 2 masih jauh dari optimal. Jika dibandingkan, kedua model ini memiliki jumlah kelas yang sangat jauh berbeda. Model 1 hanya mengklasifikasikan 2 kelas, sedangkan model 2 mengklasifikasikan 29 kelas. Melihat hal ini, perlu untuk memecah kembali klasifikasi ke dalam model – model yang lebih banyak dengan mempertimbangkan keseimbangan data yang digunakan.

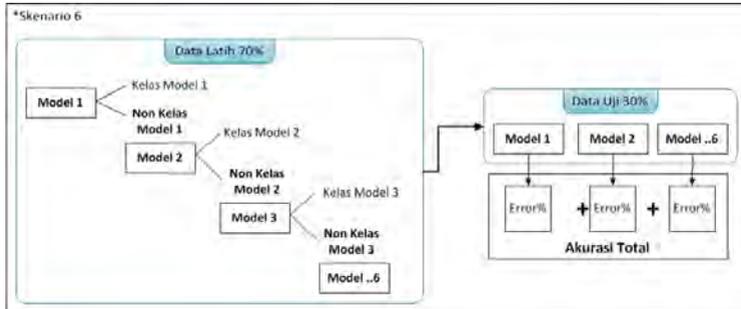
Pada percobaan 6 ini terbentuk 6 model dengan masing-masing model terdiri dari kelas sebagai berikut.

Tabel 5.8 Daftar Kelas Setiap Model Pada Percobaan 6

Model	Kelas Model	Jumlah Kelas	Jumlah Laporan
Model 1	penipuan, non penipuan	2	1136
Model 2	pemerasan, narkoba, perjudian, curat, non model 1	5	753
Model 3	implikasi kontijensi, himbauan, korupsi, lain-lain, non model 2	5	419
Model 4	curas, penggelapan, kesusilaan, penghinaan, pencurian ringan, non model 3	6	276
Model 5	pembunuhan, jabatan, penyusunan, pemalsuan surat, curanmor, penganiayaan, non model 4	8	145
Model 6	hutang piutang, pengerusakan, lantasi, akibat orang laka, pembakaran, perkosaan, penadahan, penculikan, perdagangan manusia, kebakaran	10	50



Gambar 5.10 Algoritma Percobaan 6



Gambar 5.11 Lanjutan Algoritma Percobaan 6

Algoritma yang dijalankan pada percobaan 6 ini terdiri dari enam kali proses klasifikasi, seperti yang digambarkan pada gambar Gambar 5.10 dan Gambar 5.11.

Proses klasifikasi pada percobaan 6 tidak jauh berbeda dengan percobaan – percobaan sebelumnya. Namun, pada tahap representasi data dilakukan percobaan terhadap *frequent words* yang digunakan yaitu 1 hingga 10. Hal ini melihat jumlah laporan untuk setiap model tidak sama sehingga pemakaian nilai *frequent words* pasti sangat berpengaruh terhadap hasil klasifikasi.

Kesimpulan yang didapat dari hasil klasifikasi yaitu setiap model memiliki akurasi terbaik pada *frequent words* masing-masing. Terlihat polanya bahwa semakin sedikit jumlah laporan dan semakin banyak kelas yang digunakan, maka akurasi akan cenderung mengalami penurunan. Sementara *frequent words* yang digunakan akan semakin kecil seiring dengan berkurangnya jumlah dokumen. Jika model – model dengan akurasi terbaik digabungkan, maka akan menghasilkan performa dengan total akurasi 75.5396%.

5.3.2.7. Percobaan 7

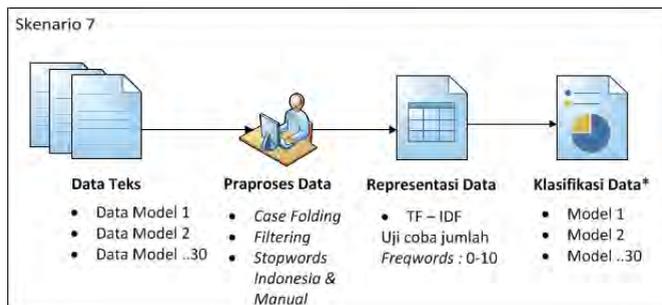
Selain jumlah laporan yang berpengaruh terhadap hasil klasifikasi, pada percobaan 6 menunjukkan bahwa dengan jumlah kelas yang sedikit dapat menghasilkan akurasi yang lebih optimal. Oleh karena itu, pada percobaan 7 ini akan

dibuat model – model dengan jumlah setiap model terdiri dari 2 kelas. Sehingga terbentuk 29 model dengan masing-masing model terdiri dari kelas sebagai berikut.

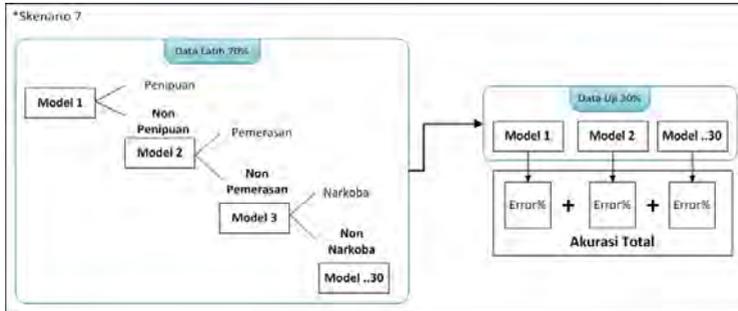
Tabel 5.9 Daftar Kelas Setiap Model Pada Percobaan 7

Model	Kelas	Jumlah Laporan
1	Penipuan, non penipuan	1136
2	Pemerasan, non pemerasan	753
3	Narkoba, non narkoba	653
4	Perjudian, non perjudian	584
5	Lain-lain, non lain-lain	524
6	Curat, non curat	469
7	Implikasi kontijensi, non implikasi kontijensi	419
8	Himbauan, non himbauan	370
9	Korupsi, non korupsi	323
10	Curas, non curas	276
11	Penggelapan, non penggelapan	246
12	Kesusilaan, non kesusilaan	216
13	Penghinaan, non penghinaan	190
14	pencurianringan	166
15	Pembunuhan, non pembunuhan	145
16	Jabatan, non jabatan	127
17	Penyuapan, non penyuapan	110
18	Pemalsuan surat, non pemalsuan surat	94
19	Curanmor, non	79

Model	Kelas	Jumlah Laporan
	curanmor	
20	Penganiayaan, non penganiayaan	65
21	Hutang piutang, non hutang piutang	50
22	Pengerusakan, non pengerusakan	40
23	Lantas, non lantas	30
24	Akibat orang luka, non akibat orang luka	23
25	Pembakaran, non pembakaran	17
26	Perkosaan, non perkosaan	12
27	Penadahan, non penadahan	8
28	Penculikan, non penculikan	5
29	Perdagangan manusia, kebakaran	3



Gambar 5.12 Algoritma Percobaan 7



Gambar 5.13 Lanjutan Algoritma Percobaan 7

Algoritma yang dijalankan pada percobaan 7 ini terdiri dari 29 kali proses klasifikasi, seperti yang digambarkan pada Gambar 5.12 dan Gambar 5.13.

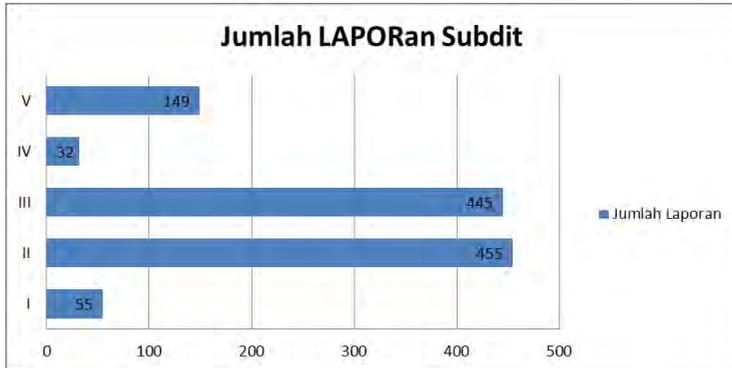
Seperti halnya pada percobaan 6, proses klasifikasi dilakukan percobaan menggunakan *frequent words* dengan jumlah berbeda – beda yaitu 0 hingga 10.

Hasil menunjukkan bahwa dengan menggunakan sedikit kelas pada model klasifikasi, menghasilkan nilai akurasi yang lebih optimal. Namun, ada beberapa kelas yang tidak dapat melakukan klasifikasi karena keterbatasan anggota kelas. Keterbatasan anggota kelas ini membuat model tidak mampu bekerja secara optimal. Sehingga, akurasi yang dihasilkan pun tidak maksimal.

5.3.2.8 Percobaan Subdit

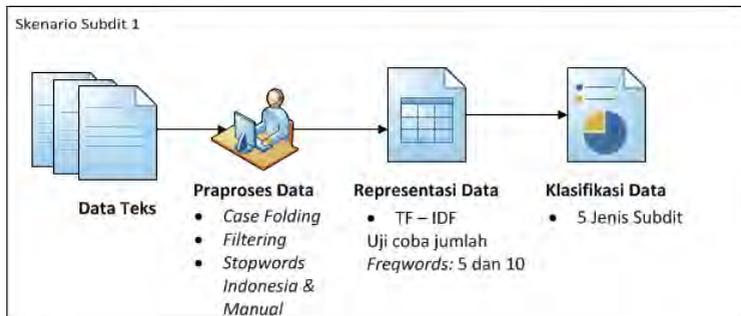
Percobaan subdit menggunakan label subdit sebagai kelas klasifikasi. Label subdit terdiri dari 5 kelas dengan jenis anggota kelas seperti yang ditunjukkan pada

Tabel 4.3. Gambar 5.14 berikut merupakan distribusi dari pelabelan kelas subdit.



Gambar 5.14 Distribusi Laporan Subdit

Proses klasifikasi percobaan subdit yang ditunjukkan pada Gambar 5.15 tidak berbeda dengan yang dijelaskan pada bab 5.3.1. Uji coba juga dilakukan pada *frequent words* dengan nilai 5 dan 10.



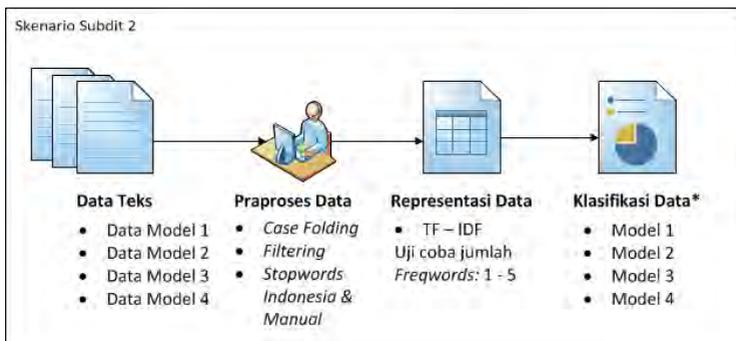
Gambar 5.15 Skenario Algoritma Subdit 1

Hasil menunjukkan nilai akurasi lebih tinggi didapat ketika menggunakan *frequent words* 5 sebesar 70.96% dibanding 10 dengan nilai 69.2%. Kemudian dilakukan optimalisasi model kembali dengan melakukan konfigurasi algoritma. Gambar 5.16 dan Gambar 5.17 merupakan percobaan perbaikan yaitu dengan membentuk model klasifikasi bersusun berjumlah 4 model. Setiap model melakukan klasifikasi sebanyak 2 kelas.

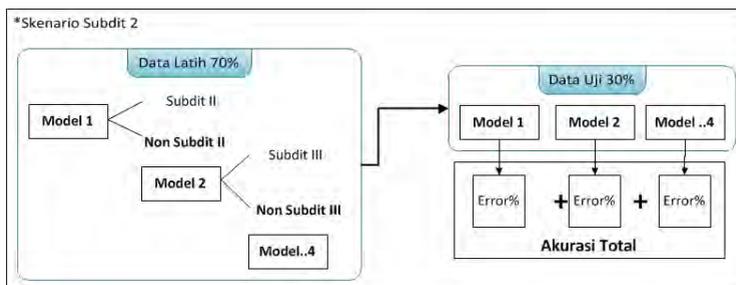
Tabel di bawah ini menjelaskan anggota kelas dari setiap model.

Tabel 5.10 Daftar Kelas Setiap Model Pada Percobaan Subdit

Model	Kelas Model	Jumlah Laporan
Model 1	subdit II, non subdit II	1136
Model 2	subdit III, non subdit III	681
Model 3	subdit V, non subdit V	236
Model 4	subdit I, subdit IV	87



Gambar 5.16 Algoritma Percobaan Subdit 2



Gambar 5.17 Lanjutan Algoritma Percobaan Subdit 2

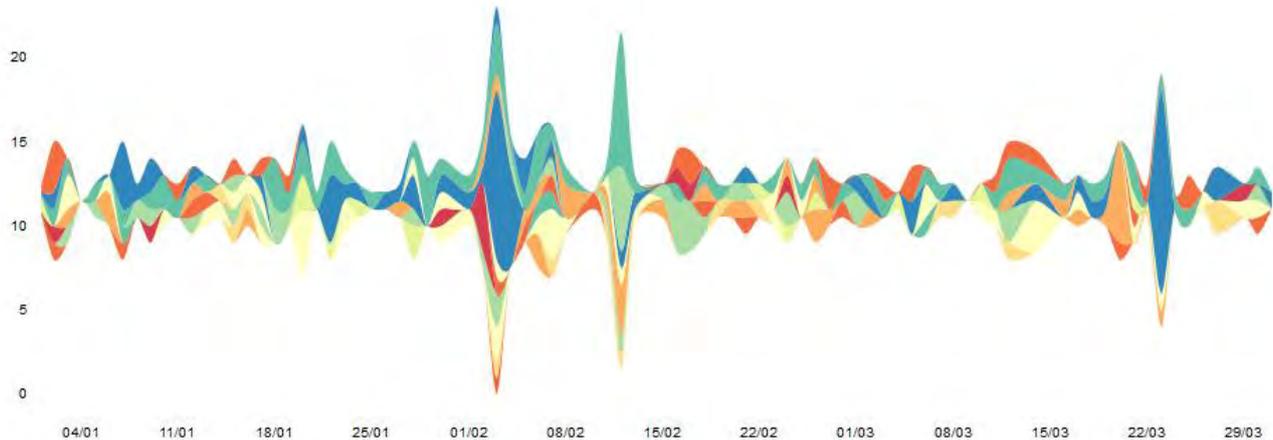
Hasilnya memperlihatkan bahwa model mampu belajar lebih baik dibanding dengan percobaan awal. Jika dilihat lebih jauh *precision*, *recall*, dan *f – measure*, nilai yang dihasilkan sudah cukup baik. Sehingga, model bisa dikatakan optimal dan akan semakin optimal seiring dengan penambahan jumlah data.

5.4. Visualisasi

Visualisasi merupakan bentuk konversi data ke dalam format visual sehingga karakteristik dan relasi antar data dapat dianalisis. Visualisasi data adalah salah satu teknik terbaik dan menarik untuk eksplorasi data. Manusia akan memiliki kemampuan yang baik untuk menganalisis sejumlah besar informasi yang dipresentasi secara visual. Pada penelitian ini dihasilkan beberapa bentuk visualisasi, yaitu *streamgraph* dan *wordcloud*. Berikut ini adalah penjelasan bagaimana data dikonversi ke dalam bentuk visual.

5.4.1. *Streamgraph*

Streamgraph atau grafik alir menggambarkan serangkaian jumlah data yang berubah dari waktu ke waktu. Perubahan ini ditunjukkan dengan ketebalan setiap lapisan yang dihasilkan. Pada penelitian ini, *streamgraph* yang dihasilkan memiliki 30 lapis dengan warna dan ketebalan berbeda pada setiap lapisannya. *Streamgraph* ini menggambarkan jumlah laporan kriminalitas untuk setiap jenisnya pada setiap triwulan.



Gambar 5.18 Hasil Visualisasi *Streamgraph* Pada Triwulan 4

Visualisasi ini dilakukan dengan aplikasi R Studio menggunakan *package* *streamgraph* [30] dan *dplyr* [31]. Data yang digunakan adalah data laporan kriminalitas LAPOR! dengan jumlah 1136 pada Januari 2014 – September 2015. Untuk membuat visualisasi ini digunakan Kode 5.19 berikut.

```

library(streamgraph)
library(dplyr)
library(pbbapply)

timeline <- read.csv("E:/MATA KULIAH/Skripsi/bahan
TA/LAPOR!/streamgraph/triwulan5.csv")

timeline %>%

filter(kelas %in% c("akibatorangluka", "curanmor",
"curas", "curat", "himbauan", "hutangpiutang",
"implikasikontijensi", "jabatan", "kebakaran",
"kesusilaan", "korupsi", "lainlain", "lantas",
"narkoba", "pemalsuansurat", "pembakaran", "pembunuhan",
"pemerasan", "penadahan", "penculikan", "pencurianringan",
"penganiayaan", "pengerusakan",
"penggelapan", "penghinaan", "penipuan", "penyuapan",
"perdaganganmanusia", "perjudian", "perkosaan"))%>%

group_by(tanggal, kelas) -> some
streamgraph(some, "kelas", "laporan", "tanggal") %>%
sg_axis_x(tick_interval=1,
tick_units="tanggal", "%d/%m")%>%
sg_legend(show=TRUE, label="Jenis kejahatan: ")

```

Kode 5.19 Visualisasi *Streamgraph*

5.4.2. *Wordcloud*

Wordcloud merupakan representasi visual dari data teks. Semakin sering suatu kata digunakan, maka semakin besar pula ukuran kata tersebut ditampilkan dalam *wordcloud*. Pada penelitian ini, digambarkan *wordcloud* dari setiap kelas pada jenis kejahatan yang berjumlah 30. Untuk membuat visualisasi ini digunakan kode program seperti pada Bab 5.3.2.3 Percobaan 3. Berikut salah satu contoh *wordcloud* yang dihasilkan.

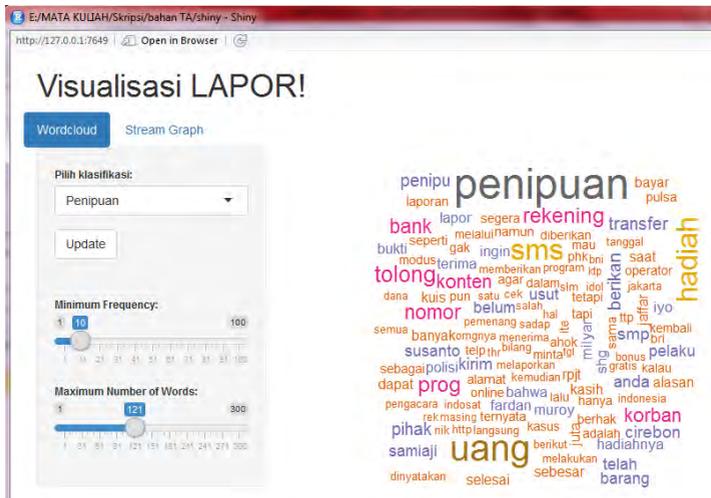


Gambar 5.19 Hasil Visualisasi *Wordcloud* Pada Kelas Korupsi

5.4.3. *Shiny*

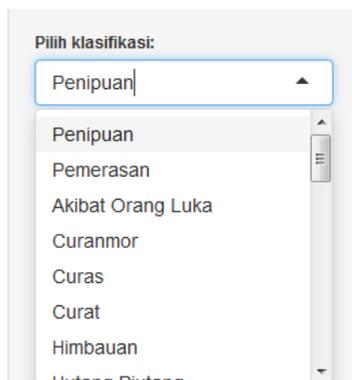
Shiny merupakan aplikasi yang digunakan untuk visualisasi data dalam bentuk web yang interaktif. Dalam penelitian ini, dengan *shiny* dilakukan visualisasi dengan menggabungkan kedua visualisasi di atas menjadi satu tampilan berbasis web. Untuk membuat visualisasi ini digunakan aplikasi R Studio menggunakan *package shiny* [32]. Kode program untuk membuat visualisasi ini terdapat pada **LAMPIRAN A.2**.

Dalam visualisasi *shiny* ini dibuat beberapa tab yang masing-masing menampilkan hasil visualisasi. Tab pertama menampilkan hasil visualisasi dengan *wordcloud*. Sedangkan tab kedua menampilkan hasil visualisasi dengan *streamgraph*.

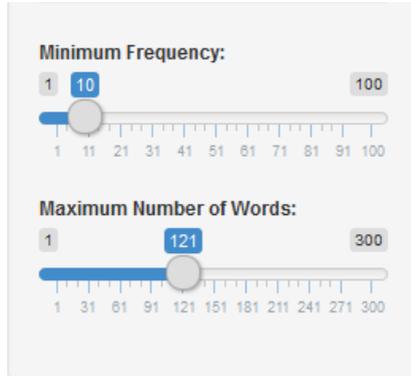


Gambar 5.20 Tab *Shiny* yang Menampilkan Visualisasi *Wordcloud*

Pada tab pertama ini, berisi hasil visualisasi menggunakan *wordcloud*. Terdapat 2 bagian, *sidebar panel* dan *main panel*. Pada *sidebar panel* terdapat beberapa pilihan jenis kejahatan serta pilihan konfigurasi *minimum frequency* dan *maximum number of words* yang akan ditampilkan pada *main panel*.

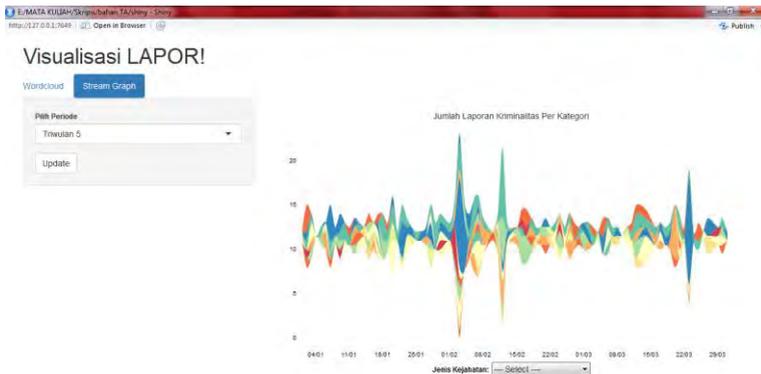


Gambar 5.21 Menu Pilihan Jenis Kejahatan Pada *Shiny*

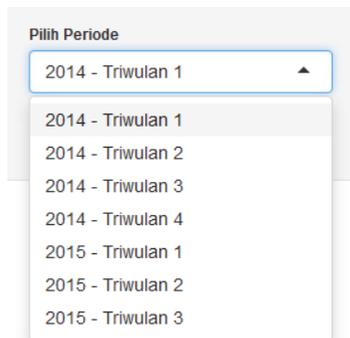


Gambar 5.22 Menu Konfigurasi *Minimum Frequency* dan *Maximum Number of Words* Pada *Shiny*

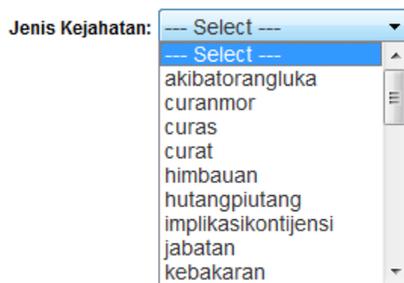
Tab kedua ini berisikan visualisasi dengan menggunakan *streamgraph*. Sama dengan tab pertama, tab kedua ini terdiri dari 2 bagian. Pada *sidebar panel* terdapat beberapa pilihan periode waktu yang akan ditampilkan, yaitu dalam triwulan. Pada *main panel* akan ditampilkan grafik *streamgraph* ditambah dengan menu *dropdown* yang berfungsi untuk memilih jenis kejahatan manakah yang akan ditampilkan.



Gambar 5.23 Tab *Shiny* yang Menampilkan Visualisasi *Streamgraph*



Gambar 5.24 Menu Pilihan Periode



Gambar 5.25 Menu Pilihan Jenis Kejahatan

Halaman ini sengaja dikosongkan

BAB VI HASIL DAN PEMBAHASAN

Pada bab ini akan dibahas mengenai hasil dari model klasifikasi yang telah dibangun pada bab sebelumnya. Hasil klasifikasi ini meliputi perhitungan akurasi, *precision*, *recall*, dan *f-measures*, dengan pembahasan yang dilakukan mencakup analisis model, performa klasifikasi, visualisasi, serta pembahasan kategori klasifikasi.

6.1 Analisis Model Klasifikasi

Tujuan dari klasifikasi adalah membangun model prediktif yang mampu mengklasifikasikan dokumen secara otomatis sehingga diketahui kategori kelas dari sejumlah data. Model klasifikasi pada penelitian ini dibangun dengan prinsip algoritma *Naïve Bayes Classifier* (NBC). Untuk mendapatkan hasil klasifikasi terbaik, dibutuhkan konfigurasi yang tepat terhadap algoritma klasifikasi. Pada penelitian ini dilakukan 7 kali percobaan untuk menemukan model klasifikasi yang paling optimal. Proses penemuan model tersebut ditunjukkan pada Gambar 5.1 dengan jenis perlakuan sebagai berikut.

- Percobaan 1 : Dokumen tanpa penghapusan *stopwords*
- Percobaan 2 : Dokumen dengan penghapusan *stopwords* Indonesia + uji coba jumlah *frequent words* (5, 10, 50)
- Percobaan 3 : Dokumen dengan penghapusan *stopwords* Indonesia + eliminasi kata I + uji coba jumlah *frequent words* (5, 10, 50)
- Percobaan 4 : Dokumen dengan penghapusan *stopwords* Indonesia + eliminasi kata II + *dictionary* dari gabungan *topwords*
- Percobaan 5 : Dokumen dengan penghapusan *stopwords* Indonesia + eliminasi kata II + 2 kali proses klasifikasi

- Percobaan 6 : Dokumen dengan penghapusan *stopwords* Indonesia + eliminasi kata II + uji coba jumlah *frequent words* (1-10) + 6 kali proses klasifikasi
- Percobaan 7 : Dokumen dengan penghapusan *stopwords* Indonesia + eliminasi kata II + uji coba jumlah *frequent words* (0-10) + 29 kali proses klasifikasi

6.1.1 Hasil Percobaan 1

Percobaan 1 menjalankan algoritma dasar dari klasifikasi tanpa dilakukan konfigurasi tambahan. Hasil yang didapat menunjukkan nilai akurasi sebesar 7.625%. Dengan nilai akurasi tersebut tentu bukan merupakan model yang optimal. Setelah dilakukan analisis terhadap hasil klasifikasi, ternyata anggota dari setiap kelas memiliki jumlah yang tidak berimbang, seperti yang ditunjukkan pada Gambar 5.4.

6.1.2 Hasil Percobaan 2

Percobaan 2 menjalankan algoritma pengembangan dari percobaan 1. Untuk menyelesaikan permasalahan tersebut, dilakukan pengelompokan kelas – kelas berdasarkan jumlah anggota, yang dapat dilihat pada bab 5.3.2.2. Dengan jumlah anggota yang tidak sama, *frequent words* yang digunakan tentu akan berpengaruh terhadap hasil. Oleh karena itu, dilakukan uji coba terhadap jumlah *frequent words*, yaitu 5, 10, dan 50. Hasil yang didapat menunjukkan bahwa akurasi yang dihasilkan meningkat. Namun, dengan hasil akurasi yang tidak lebih dari 65% tentu masih perlu dilakukan perbaikan algoritma.

Tabel 6.1 Hasil Akurasi Percobaan 2

Jenis Data	<i>Frequent words</i>		
	5	10	50
Data dari kelas yang beranggotakan lebih dari	12.26994	13.80368	6.134969

Jenis Data	Frequent words		
	5	10	50
10			
Data dari kelas yang beranggotakan lebih dari 20	55.81395	53.48837	43.52159
Data dari kelas yang beranggotakan lebih dari 30	52.67176	51.52672	44.65649
Data dari kelas yang beranggotakan lebih dari 30 dan tanpa memakai kelas penipuan.	49.65986	46.2585	24.4898
Data dari kelas yang beranggotakan lebih dari 30 dan tanpa memakai kelas lain-lain dan kelas himbauan.	63.15789	60.08772	48.68421

6.1.3 Hasil Percobaan 3

Percobaan 3 melakukan analisis terhadap kata – kata apa yang mendominasi setiap kelas. Pada *wordcloud* yang dihasilkan ditemukan adanya kata konjungsi dan frasa yang tidak representatif terhadap kategori kejahatan. Kata – kata tersebut kemudian dieleminasi menggunakan fungsi *stopwords* pada aplikasi R Studio. Dari beberapa jenis data, hasil menunjukkan bahwa akurasi semakin meningkat dengan beberapa lainnya juga mengalami penurunan.

Tabel 6.2 Hasil Akurasi Percobaan 3

Jenis Data	Frequent words		
	5	10	50
Data dari kelas yang beranggotakan lebih dari 10	14.11043	14.41718	7.055215
Data dari kelas yang beranggotakan lebih	52.34899	49.32886	46.6443

Jenis Data	Frequent words		
	5	10	50
dari 20			
Data dari kelas yang beranggotakan lebih dari 30	18.21705	17.82946	10.85271
Data dari kelas yang beranggotakan lebih dari 30 dan tanpa memakai kelas penipuan.	42.65734	46.85315	18.18182
Data dari kelas yang beranggotakan lebih dari 30 dan tanpa memakai kelas lain-lain dan kelas himbauan.	36.28319	35.39823	15.04425

6.1.4 Hasil Percobaan 4

Percobaan 4 masih melakukan penghapusan terhadap kata – kata yang mengganggu hasil klasifikasi. Hal ini dilakukan karena ditemukan kata yang sama di beberapa kelas berbeda. Untuk menyelesaikannya terlebih dahulu dilakukan analisis dengan membandingkan 20 kata teratas yang muncul pada setiap kelas. Setelah itu dilakukan penggabungan kata yang kemudian digunakan sebagai *dictionary* pada tahap pembobotan. Hasil yang didapat pada Tabel 6.3 menunjukkan bahwa dari beberapa percobaan yang dilakukan, ditemukan akurasi optimal sebesar 90.90% yaitu pada data kelas penipuan dan non penipuan. Akurasi dengan nilai seperti inilah yang diharapkan akan dihasilkan dari model yang nantinya berhasil ditemukan.

Tabel 6.3 Hasil Akurasi Percobaan 4

Jenis Data	Akurasi
Penipuan dan non Penipuan	90.90909
Data dari 30 kelas	6.744868

Jenis Data	Akurasi
Data dari kelas yang beranggotakan lebih dari 10	14.11043
Data dari kelas yang beranggotakan lebih dari 20	62.08054
Data dari kelas yang beranggotakan lebih dari 30	15.11628
Data dari kelas yang beranggotakan lebih dari 30 dan tanpa memakai kelas lain-lain dan himbauan	70.17544

6.1.5 Hasil Percobaan 5

Percobaan 5 melakukan konfigurasi algoritma pengembangan dengan membentuk 2 model klasifikasi. Model klasifikasi pertama menggunakan model yang ditemukan pada percobaan 4 yaitu kelas penipuan dan non penipuan. Sedangkan algoritma kedua menggunakan kelas non penipuan pada model 1 dengan melakukan pelabelan terhadap 29 kelas selain kelas penipuan. Hasil yang didapat menyatakan model 1 memiliki nilai akurasi sebesar 91.788%. Sedangkan model 2 menghasilkan nilai sebesar 31.858% dengan akurasi total 67.901%.

6.1.6 Hasil Percobaan 6

Percobaan 6 masih tetap melakukan pengembangan terhadap model klasifikasi. Setelah dianalisis hasil percobaan 5, model 1 yang hanya mengklasifikasi dua kelas memiliki hasil yang lebih optimal dibanding dengan model 2 yang mengklasifikasi 29 kelas. Oleh karena itu, pada percobaan ini dilakukan pembagian kelas – kelas berdasarkan jumlah anggota kelas sehingga terbentuk 6 model. Dilakukan uji coba kembali terhadap *frequent words* dengan jumlah 1 hingga 10. Hasilnya percobaan ini dapat dilihat pada tabel di bawah ini. Dapat

disimpulkan bahwa jika jumlah dokumen yang digunakan semakin sedikit dengan kelas yang semakin banyak, akurasi yang dihasilkan akan cenderung mengalami penurunan. Sementara *frequent words* yang digunakan akan semakin kecil seiring dengan berkurangnya jumlah dokumen.

Tabel 6.4 Hasil Akurasi Percobaan 6

FW	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
1	89.736	62.832	69.048	56.627	44.186	13.333
2	90.323	61.062	63.492	51.807	39.535	13.333
3	90.616	69.469	67.460	51.807	25.581	13.333
4	92.962	66.814	67.460	51.807	25.581	13.333
5	92.082	69.912	67.460	51.807	25.581	13.333
6	91.202	59.735	67.460	51.807	25.581	13.333
7	90.616	57.965	67.460	51.807	25.581	13.333
8	90.323	57.965	67.460	51.807	25.581	13.333
9	84.457	57.965	67.460	51.807	25.581	13.333
10	83.578	57.965	67.460	51.807	25.581	13.333

6.1.7 Hasil Percobaan 7

Percobaan 7 membentuk model klasifikasi bersusun dengan jumlah 29 model dengan masing – masing model hanya mengklasifikasi 2 kelas. Susunan model yang digunakan diawali dari kelas dengan jumlah anggota terbanyak. Kelas terbanyak adalah kelas penipuan, sehingga pada model pertama melakukan klasifikasi kelas penipuan dan non penipuan. Kemudian model selanjutnya menggunakan data non penipuan yang akan dibagi menjadi dua kelas dengan label jenis kejahatan dengan jumlah terbanyak kedua setelah dan begitu seterusnya. Hasil yang didapat pada Tabel 6.5 menunjukkan 27 dari 29 kelas menghasilkan nilai lebih dari 80%. Bisa dikatakan bahwa model mampu mempelajari data

dengan baik. Sehingga, dapat disimpulkan bahwa pada percobaan 7 ini merupakan model klasifikasi terbaik.

Tabel 6.5 Hasil Akurasi Percobaan 7

Model	<i>Frequent Words</i>	Akurasi
1	6	92.9619
2	5	88.9381
3	3	96.4286
4	3	98.2857
5	5	89.172
6	3	94.3262
7	3	92.0635
8	3	89.1892
9	2	95.8763
10	1	93.9759
11	0	91.8919
12	2	89.2308
13	0	89.4737
14	0	90
15	0	95.3488
16	0	94.7368
17	1	100
18	0	85.7143
19	0	83.3333
20	0	84.2105
21	0	86.6667
22	0	91.6667
23	0	88.8889
24	1	85.7143
25	1	80

Model	<i>Frequent Words</i>	Akurasi
26	0	75
27	1	100
28	1	100
29	-	-

6.1.8 Hasil Percobaan Subdit

Percobaan subdit menggunakan label subdit sebagai kelas klasifikasi. Percobaan yang dilakukan terhadap model klasifikasi yaitu pertama menggunakan model 1 kali proses klasifikasi, kedua menggunakan model bersusun seperti yang dilakukan pada percobaan 7.

Hasil dari model pertama yaitu akurasi dengan nilai 70.96% untuk *frequent words* 5 dan 69.2% untuk *frequent words* 10. Kemudian dilakukan optimalisasi model kembali dengan hasil yang didapat pada Tabel 6.6.

Tabel 6.6 Hasil Akurasi Percobaan Subdit

FQ	Model 1	Model 2	Model 3	Model 4
1	85.04399	72.05882	69.01408	53.84615
2	84.75073	74.01961	71.83099	-
3	85.6305	79.90196	60.56338	-
4	88.56305	77.94118	54.92958	-
5	86.80352	72.05882	54.92958	-

Seperti yang dijelaskan sebelumnya bahwa model dengan akurasi yang baik belum tentu mencerminkan performa model yang baik pula. Dibawah ini dijelaskan nilai *precision*, *recall*, dan *f-measure* pada model di atas.

Tabel 6.7 Nilai *Precision*, *Recall*, dan *F - Measure* dari Percobaan Subdit

No	Kelas	<i>Precision</i>	<i>Recall</i>	<i>F - Measure</i>
1	II	0.9274	0.7931	0.855008

No	Kelas	<i>Precision</i>	<i>Recall</i>	<i>F - Measure</i>
2	III	0.8392	0.8696	0.85413
3	V	0.6863	0.8974	0.777781
4	I	0.5909	0.8125	0.666655
5	IV	0.2500	0.1000	0.142857

Berdasarkan hasil yang didapat seperti ditunjukkan pada Tabel 6.7, semua model mampu melakukan klasifikasi. Hanya saja tidak semua mampu melakukan klasifikasi dengan optimal. Model klasifikasi subdit II dan II mampu melakukan klasifikasi dengan nilai *f - measure* masing – masing 85.5% dan 85.4%. Disusul dengan model subdit V sebesar 77.7% dan subdit I sebesar 66.6%. Sedangkan model IV hanya mampu mengklasifikasi model dengan akurasi 14.2%. Hal ini dikarenakan jumlah data yang digunakan terbatas, yaitu hanya sejumlah 87 data.

6.2 Analisis Performa Model Klasifikasi

Dari analisis model klasifikasi yang dilakukan pada bab 6.1, didapatkan bahwa model klasifikasi terbaik ada pada percobaan 7. Jika dibandingkan dengan percobaan 1 hingga 6, percobaan 7 memiliki akurasi paling optimal dengan nilai lebih dari 80%. Namun begitu, akurasi optimal tidak menunjukkan bahwa model memiliki performa yang baik. Tabel 6.8 berikut menunjukkan nilai akurasi, *precision*, *recall*, dan *F - measure* dari masing-masing kelas.

Tabel 6.8 *Precision*, *Recall*, dan *F - Measure* Percobaan 7

No	Kelas	<i>Precision</i>	<i>Recall</i>	<i>F - Measure</i>
1	Penyuapan	1	1	1.00000
2	Perjudian	0.9500	0.9048	0.92685
3	Penipuan	0.8678	0.9292	0.89745
4	Pengerusakan	1	0.8	0.88889
5	Korupsi	0.8462	0.8462	0.84620

No	Kelas	Precision	Recall	F - Measure
6	Narkoba	0.6667	1	0.80002
7	Pencurian ringan	0.6250	0.7143	0.66667
8	Lantas	0.5000	1	0.66667
9	Perkosaan	0.5000	1	0.66667
10	Penganiayaan	0.7500	0.6000	0.66667
11	Penghinaan	0.5556	0.7143	0.62503
12	Curas	0.5714	0.6667	0.61538
13	Curat	0.3333	1	0.49996
14	Jabatan	0.3333	1	0.49996
15	Penggelapan	0.3750	0.6000	0.46154
16	Pemerasan	0.2647	1	0.41860
17	Curanmor	0.2000	1	0.33333
18	Implikasi kontijensi	0.1818	0.6667	0.28569
19	Lainlain	0	0	0.00000
20	Himbauan	NA	NA	NA
21	Kesusilaan	NA	NA	NA
22	Pembunuhan	NA	NA	NA
23	Pemalsuan surat	NA	NA	NA
24	Hutang piutang	NA	NA	NA
25	Akibat orang luka	NA	NA	NA
26	Pembakaran	NA	NA	NA
27	Penadahan	NA	NA	NA
28	Penculikan	-	-	-
29	Perdagangan manusia	-	-	-
30	Kebakaran	-	-	-

Berdasarkan hasil yang ditunjukkan pada Tabel 6.8 bahwa tidak semua kelas dapat dijadikan sebagai pengklasifikasi. Kelas 1 hingga kelas 18 mampu melakukan klasifikasi hingga dapat menghasilkan performa yang cukup baik. Model klasifikasi bekerja sangat akurat dalam mengklasifikasikan anggota dari kelas penyipuan dengan benar. Disusul dengan kelas perjudian, penipuan, pengerusakan, korupsi, dan narkoba dengan nilai f -measure lebih dari 80%. Kelas – kelas tersebut mewakili 51.49% dari data yang digunakan. Sedangkan untuk kelas pencurian ringan, lantasi, perkosaan, penganiayaan, penghinaan, curas, curat, jabatan, penggelapan, pemerasan, curanmor, dan implikasi kontijensi memiliki nilai f -measure kurang dari 80%.

Hal ini menunjukkan model kurang tepat dalam melakukan klasifikasi. Botha dan Barnard [33] menyatakan bahwa ketepatan klasifikasi dipengaruhi oleh beberapa faktor, diantaranya:

1. Ukuran fragmen teks yang diidentifikasi
2. Jumlah data latihan yang digunakan
3. Fitur klasifikasi
4. Algoritma yang digunakan
5. Kemiripan kata

Sementara untuk kelas 19 hingga 29 tidak mampu untuk melakukan proses klasifikasi. Jika dilihat lebih dalam pada Tabel 6.9, kelas – kelas tersebut memiliki keterbatasan jumlah laporan, khususnya untuk jumlah data latihan dan data uji.

Tabel 6.9 Jumlah Dataset Pada Kelas 19 hingga 29

Kelas	Jumlah	Data Latih	Data Uji
Lainlain	55	39	17
Himbauan	47	33	14
Kesusilaan	26	18	8
Pembunuhan	18	13	5
Pemalsuan surat	15	11	5

Kelas	Jumlah	Data Latih	Data Uji
Hutang piutang	10	7	3
Akibat orang luka	6	4	2
Pembakaran	5	4	2
Penadahan	3	2	1
Penculikan	2	-	-
Perdagangan manusia	2	-	-
Kebakaran	1	-	-

Pada Tabel 6.9 terlihat bahwa jumlah data yang digunakan sangat terbatas yaitu tidak lebih dari 50 data setiap kelas. Sehingga tidak dipungkiri performa yang dihasilkan jauh dari optimal. Menurut Botha & Barnard, dengan menambah jumlah data latihan yang digunakan, umumnya tingkat akurasi yang dihasilkan akan semakin meningkat pula [33].

6.3 Pembahasan Kategori Klasifikasi

Kategori kriminalitas yang diperoleh dari Reserse Kriminal dan Umum POLDA Jatim terdiri dari tiga kelompok, yaitu konvensional, implikasi kontijensi, dan transnasional. Setiap kelompok tersebut terdiri dari 33, 18, dan 4 kategori kejahatan untuk masing – masing kelompok dengan rincian dapat dilihat pada Bab 4.4. Pada penelitian ini laporan yang digunakan hanya merepresentasikan 30 kategori kejahatan dengan rincian 25 jenis kejahatan kriminal umum, 3 jenis kejahatan kriminal khusus, dan untuk laporan yang tidak termasuk ke dalam kategori tersebut dikelompokkan ke dalam kategori lain – lain dan himbauan.

Proses pelabelan data dilakukan berdasar pada Kitab Undang – Undang Hukum Pidana (KUHP). Ahli hukum juga turut memberikan justifikasi pada pelabelan data tersebut. Hasil dari pelabelan data menunjukkan jumlah anggota setiap kelas tidak seimbang. Dengan jumlah total laporan 1136, 1 kelas memiliki jumlah anggota lebih dari 100, 5 kelas memiliki jumlah anggota antara 50 hingga 100, dan 25 kelas lainnya hanya

memiliki jumlah anggota di bawah 50. Untuk lebih lengkapnya dapat dilihat pada Gambar 5.4.

Selain menggunakan label kategori kejahatan, label subdit dapat menjadi alternatif dalam mengklasifikasi laporan kriminalitas. Laporan – laporan yang telah terklasifikasi dapat langsung diberikan dan ditindaklanjuti oleh bagian yang menangani jenis laporan tersebut. Distribusi pelabelan subdit dapat dilihat pada Gambar 5.14.

Dari kedua jenis pelabelan di atas, jika dibandingkan lebih baik menggunakan label subdit dalam kasus data penelitian ini. Semua model pada percobaan subdit dapat menjadi pengklasifikasi walaupun dengan performa yang tidak semuanya baik. Sedangkan untuk model dengan 30 kategori, tidak semua dapat menjadi pengklasifikasi karena keterbatasan data. Sehingga, dapat disimpulkan bahwa model klasifikasi akan memiliki performa baik tidak hanya bergantung pada algoritma yang digunakan saja. Jumlah data yang digunakan pun turut mempengaruhi tingkat performa model. Dengan menambah jumlah data, model akan belajar lebih baik dan akurasi yang dihasilkan umumnya juga akan meningkat.

6.4 Analisis Hasil Visualisasi

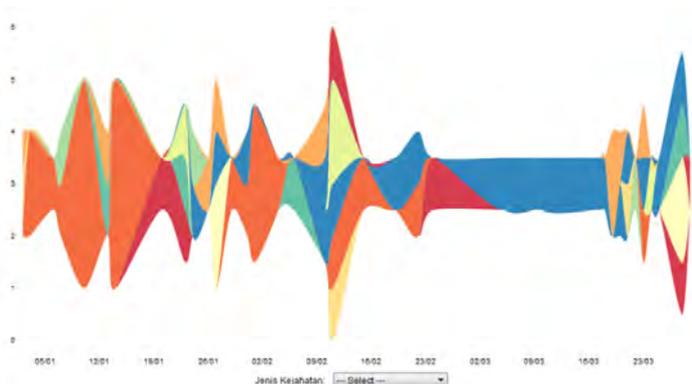
Visualisasi dihasilkan pada penelitian ini berupa grafik *streamgraph* dan *wordcloud* yang ditampilkan pada aplikasi berbasis web yang interaktif yaitu shiny. Visualisasi *streamgraph* menunjukkan jumlah laporan kriminalitas untuk setiap periode dalam bentuk grafik alir. Periode yang digunakan dalam memvisualisasikan laporan kriminalitas menggunakan triwulan dalam rentan waktu Januari 2014 – hingga September 2015, seperti yang ditunjukkan pada *timeline streamgraph* di bawah ini. Hasil untuk masing – masing grafik setiap triwulan dapat dilihat pada **LAMPIRAN E**.

2014			2015			
Jan		Des	Jan		Sept	
73	132	78	106	355	206	164

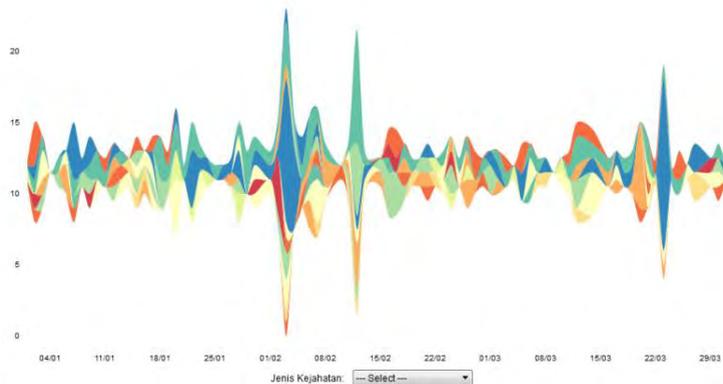
Gambar 6.1 Timeline Stram Graph 21 Bulan

Dari setiap *streamgraph* yang dihasilkan per triwulan, memiliki karakteristik yang berbeda – beda. Gambar 6.1 memperlihatkan jumlah laporan yang tersampaikan cenderung meningkat dari tahun sebelumnya. Jumlah laporan per triwulan yang tersampaikan jika dibandingkan dengan sebelumnya pun juga mengalami peningkatan. Pada triwulan 1, 2, dan, 3 tahun 2015, jumlah laporan mengalami peningkatan berturut – turut sebesar 282, 68, 86 dari triwulan 1, 2, dan, 3 tahun 2014.

Dari sisi jenis kejahatan terbanyak yang tersampaikan, secara keseluruhan memiliki perbedaan pada setiap periode. Grafik alir memperlihatkan dengan komposisi warna dan ketebalan yang berbeda pada setiap periodenya. Untuk jenis kejahatan yang tidak sama sekali tersampaikan pada periode tertentu, digambarkan dengan garis.

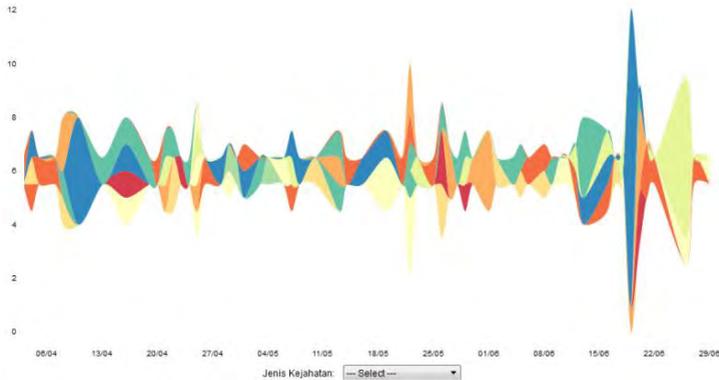


Gambar 6.2 Pola Laporan pada Periode 2014 – Triwulan 1

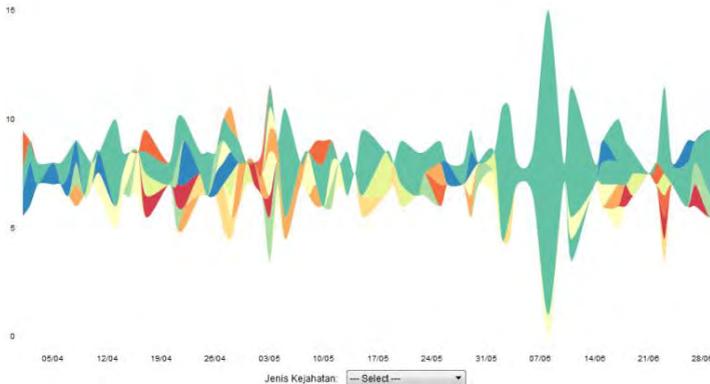


Gambar 6.3 Pola Laporan pada Periode 2015 – Triwulan 1

Jika dibandingkan, pola laporan untuk triwulan 1 dari kedua tahun memiliki perberbedaan. Dari sisi jumlah data yang tersampaikan, mengalami peningkatan dari tahun 2014 ke 2015. Hal ini dapat dilihat pada sumbu y dari setiap grafik dimana untuk tahun 2014 maksimal data yang tersampaikan berjumlah 6, sedangkan untuk tahun 2015 mencapai lebih dari 20 laporan yang tersampaikan pada periode tersebut. Dari sisi jenis kejahatan yang dilaporkan, terlihat bahwa lebih sedikit jenis yang ada pada tahun 2014 dibanding tahun 2015. Laporan yang mendominasi pada tahun 2014 yaitu korupsi, penipuan, dan akibat orang luka yang ditandai dengan warna biru, orange, dan merah, yang pada tahun 2015 tidak ada satupun jenis akibat orang luka yang tersampaikan juga laporan korupsi yang menurun jumlahnya. Namun, tetap untuk laporan penipuan yang masih tergolong ke dalam jenis laporan terbanyak yang disampaikan, disusul dengan laporan pemerasan.



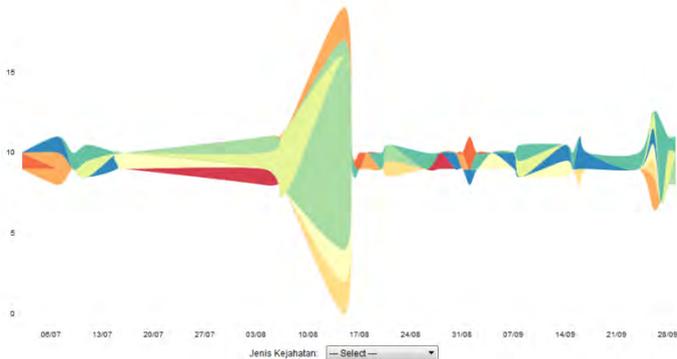
Gambar 6.4 Pola Laporan pada Periode 2014 – Triwulan 2



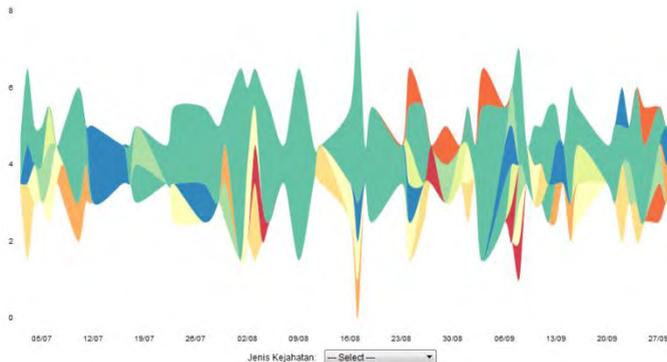
Gambar 6.5 Pola Laporan pada Periode 2015 – Triwulan 2

Perbandingan pola laporan untuk triwulan 2, dari sisi jumlah data yang tersampaikan, mengalami peningkatan dimana untuk tahun 2014 maksimal data yang tersampaikan berjumlah 12, sedangkan untuk tahun 2015 mencapai lebih dari 15 laporan. Dari sisi jenis kejahatan yang dilaporkan, terlihat bahwa dari kedua tahun tidak ada jenis kejahatan yang terlalu mendominasi. Hanya saja yang paling mencolok yaitu warna hijau pada tahun 2015 yaitu laporan penipuan, dimana laporan tersebut ada di sepanjang periode triwulan 2 tahun 2015 yang pada tahun 2014 komposisi warna hijau lebih sedikit. Disusul

dengan warna biru yaitu laporan pemerasan, yang mana lebih banyak komposisinya pada tahun 2014 dibanding 2015, seperti halnya warna orange yaitu korupsi yang komposisinya semakin mneurun.



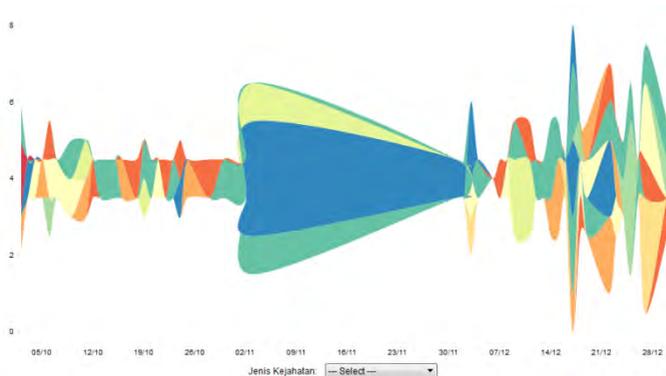
Gambar 6.6 Pola Laporan pada Periode 2014 – Triwulan 3



Gambar 6.7 Pola Laporan pada Periode 2015 – Triwulan 3

Pola laporan pada triwulan 3 dari kedua tahun berbeda dari periode 2 triwulan sebelumnya. Pada tahun 2014, jumlah laporan yang tersampaikan rata – rata 3 laporan untuk setiap tanggal. Namun, pada pertengahan periode yaitu bulan Agustus, terdapat laporan yang tersampaikan berjumlah 12 yaitu laporan implikasi kontijensi, yang pada tahun 2015 laporan ini tidak satupun tersampaikan. Pada tahun 2015 tetap

laporan penipuan yang mendominasi di sepanjang periode, bahkan semakin mendominasi dari periode sebelumnya.



Gambar 6.8 Pola Laporan pada Periode 2014 – Triwulan 4

Sementara periode triwulan 4 tidak dapat dibandingkan, karena tidak adanya data untuk tahun 2015. Pada tahun 2014, dimana merupakan periode akhir tahun, laporan pemerasan terlihat mendominasi pada pertengahan periode ndisusul dengan perjudian yang ditunjukkan dengan warna biru dan orange, dengan laporan penipuan yang masih mendominasi di sepanjang tahun walau dengan jumlah yang tidak banyak.

Secara ringkas dapat disimpulkan, pada tahun 2014 jenis laporan yang disampaikan cenderung tidak merata. Selalu ada beberapa jenis laporan yang mendominasi di setiap periodenya. Seperti halnya korupsi yang mendominasi pada periode triwulan I, implikasi kontijensi pada terlihat dominan sekali pada pertengahan triwulan III, dan pemerasan pada periode triwulan IV. Berbeda dengan tahun 2015, seiring semakin banyaknya data yang tersampaikan, jenis laporan yang disampaikan cenderung merata dengan perbedaan jumlah yang tidak begitu signifikan untuk setiap jenisnya. Jenis kejahatan penipuan yang mendominasi di sepanjang periode pada tahun ini.

Sementara visualiasi dengan *wordcloud* memperlihatkan representasi kata – kata untuk setiap jenis kejahatan. Gambar *wordcloud* setiap jenis kejahatan dapat dilihat pada **LAMPIRAN C**. Tabel 6.10 merupakan 10 kata teratas yang dihasilkan dari setiap jenis kejahatan. Pada pengerjaan percobaan klasifikasi pada Bab 5.3.2.4 , kata – kata ini diujicobakan sebagai *dictionary* pada proses pembobotan.

Tabel 6.10 Hasil Visualisasi *Wordcloud*

Jenis Kejahatan	10 Kata Teratas
Penipuan	uang, sms, hadiah, tolong, rekening, prog, bank, nomor, konten, transfer
Lain – lain	Barat, tni, uang, hutan, timur, dinas, ketua, tua, anak, desa
Pemerasan	preman, uang, motor, ibu, kasih, polisi, rumah, tolong, mobil, meminta
Narkoba	sabu, shabu, ganja, bandar, cengkareng, warga, desa, pengedar, jenis, gembong
Perjudian	judi, desa, togel, perjudiyan, masyarakat, warga, kota, tempat, malam, marak
Curat	rumah, bank, pencurian, polisi, mandiri, pelaku, bandung, lapor, malam, mobil
Implikasi Kontijensi	anak, komplek, warga, ganja, masyarakat, meresahkan, malam, polisi, daerah, kepolisian
Himbauan	lapor, anak, polsek, hukum, polisi, warga, aktivitas, aparat, jalan, rumah
Korupsi	dana, desa, kabupaten, bupati, negara, kasus, kepala, kecamatan, masyarakat, apbd
Penggelapan	curas, motor, terjadi, pelaku, jalan, daerah, begal, malam, pembegalan
Curas	timah, illegal, uang, BBM, koperasi, anggota, mobil, tambang, tanah, hukum
Kesusilaan	barat, pin, tempat, anak, taman, polisi, tolong, prostitusi, warga, pelacuran
Penghinaan	akun, baik, facebook, pencemaran, foto, nomor, bukti, lanjuti, media, pelaku

Jenis Kejahatan	10 Kata Teratas
Pencurian Ringan	jalan, minta, arah, bandung, copet, motor, baru, bus, ditindak, ibu
Pembunuhan	barat, polisi, tni, kab, tengah, bunuh, dinas, dunia, hutan, malam
Jabatan	tanah, dilakukan, uang, kasus, oknum, ajang, hukum, kepala, pegawai, pejabat
Penyuapan	suap, data, kasih, uang, lampung, pesawat, bangunan, desa, jelas, jutaan
Pemalsuan Surat	tanah, surat, pemalsuan, tgl, anak, lurah, adik, alamat, bukti, cerai
Curanmor	polisi, pencurian, terjadi, daerah, warga, bawa, jalan, laporan, perumahan, ban
Penganiayaan	adik, kasus, anak, sekolah, hukum, jatuh, pelaku, warga, ayah, kekerasan
Hutang Piutang	Rumah, hutang, kredit, marah, sertifikat, bpr, tanda, bentak, dilelang, kendaraan
Pengerusakan	rumah, desa, jalan, kecamatan, mobil, barang, pagi, pecah, pelaku, perusakan
Lantas	Polisi, jalan, arah, melewati, sim, stnk, kepolisian, lapor, tilang, bebas
Akibat Orang Luka	paku, anak, belakang, humasy, pelaku, polisi, psr, alamatnya, anggotanya, asli
Penadahan	bersubsidi, solar, penimbunan, bbm, beking, gudang, industri, tank, truk, unit
Penculikan	keluarga, dki, minta, tahu, terkahir, tolong, alamat, alias, bernama, dilacak
Perdagangan Manusia	penipuan, biro, dugaan, manusia, perdagangan, sekretaris, administrasi, dilapor, diperlama, ditulis

DAFTAR LAMPIRAN

Berikut ini adalah lampiran dari penjelasan dan pengerjaan pada penelitian tugas akhir.

KODE LAMPIRAN	LAMPIRAN
A	Kode Program
B	Daftar <i>Stopwords</i> Bahasan Indonesia
C	<i>Wordcloud</i> Jenis Kejahatan
D	Deskripsi dan Sanksi Jenis Kejahatan
E	<i>Streamgraph</i> Laporan Kriminalitas

LAMPIRAN A

KODE PROGRAM

A.1 Kode Program Percobaan

A.1.1 Kode Program Percobaan 1

```
tag <- read.csv("E:/MATA KULIAH/Skripsi/bahan
TA/LAPOR!/20151004/mydata.csv")

# Load the the tm library
library(tm)

# Load the data as a corpus
crime_corpus <-
Corpus(VectorSource(tag$isiLaporan))

# Convert the text to lower case
clean_corpus <- tm_map(crime_corpus,
content_transformer(tolower))
# Remove numbers
clean_corpus <- tm_map(clean_corpus, removeNumbers)
# Remove punctuations
clean_corpus <- tm_map(clean_corpus,
removePunctuation)
# Remove stopwords
clean_corpus <- tm_map(clean_corpus, removewords,
stopwords("english"))
#Eliminate extra white spaces
clean_corpus <- tm_map(clean_corpus,
stripwhitespace)

#Tokenizing the corpus
dtm <-
DocumentTermMatrix(clean_corpus,control=list(weight
ing=weightTfIdf))
dtm <-
DocumentTermMatrix(clean_corpus,control=list(weight
ing=weightBin))

#divide corpus into training and test data
train_size <- round(0.7*nrow(tag))
set.seed(141321)
```

```

train_indices <- sample(seq_len(nrow(tag)),
size=train_size)
train_corpus <- clean_corpus[ train_indices]
test_corpus <- clean_corpus[-train_indices]

dtm_train <- dtm[train_indices, ]

#Identify frequently used words
frequent_words <- findFreqTerms(dtm_train, 5)

#Create document-term matrices using frequent words
crime_train <- DocumentTermMatrix(train_corpus,
control=list(dictionary =
frequent_words))
crime_test <- DocumentTermMatrix(test_corpus,
control=list(dictionary =
frequent_words))

#Convert count information to "Yes", "No"
convert_count <- function(x) {
  y <- ifelse(x > 0, 1,0)
  y <- factor(y, levels=c(0,1), labels=c("No",
"Yes"))
}

#Convert document-term matrices
crime_train <- apply(crime_train, 2, convert_count)
crime_test <- apply(crime_test, 2, convert_count)

#the naive bayes
library(e1071)

#Create a Naive Bayes classifier object
crime_classifier <- naiveBayes(crime_train,
factor(tag$mode6[train_indices]))
class(crime_classifier) #cuma nampilin naive bayes

#Evaluate the performance on the test data
predictnbc <- predict(crime_classifier, crime_test)
actualnbc <- tag$tag[-train_indices]

# Calculating the accuracy in percents
levels(predictnbc) <- levels(actualnbc)
accuracy <-
100*sum(predictnbc==actualnbc)/length(test_corpus)

```

A.1.2 Kode Program Percobaan 2

```

tag <- read.csv("E:/MATA KULIAH/Skripsi/bahan
TA/LAPOR!/20151004/mydata.csv")

```

```

# Load the the library
library(tm)

# Load the data as a corpus
crime_corpus <-
Corpus(VectorSource(tag$isiLaporan))

# Convert the text to lower case
clean_corpus <- tm_map(crime_corpus,
content_transformer(tolower))
# Remove numbers
clean_corpus <- tm_map(clean_corpus, removeNumbers)
# Remove punctuations
clean_corpus <- tm_map(clean_corpus,
removePunctuation)
# Eliminate extra white spaces
clean_corpus <- tm_map(clean_corpus,
stripwhitespace)
# Remove stopword
clean_corpus <- tm_map(clean_corpus, removewords,
stopwords("english"))
stopwordID <-
"https://raw.githubusercontent.com/nurandi/nurandi.
net/master/data/stopwords-id.txt"
cStopwordID <- readLines(stopwordID);
clean_corpus <- tm_map(clean_corpus, removewords,
c(cStopwordID))
#Eliminate extra white spaces
clean_corpus <- tm_map(clean_corpus,
stripwhitespace)

# tokenizing the corpus
dtm <-
DocumentTermMatrix(clean_corpus,control=list(weight
ing=weightTfIdf))

#divide corpus into training and test data
train_size <- round(0.7*nrow(tag))
set.seed(141321)
train_indices <- sample(seq_len(nrow(tag)),
size=train_size)
train_corpus <- clean_corpus[ train_indices]
test_corpus <- clean_corpus[-train_indices]

dtm_train <- dtm[train_indices, ]

#Identify frequently used words
frequent_words <- findFreqTerms(dtm_train, 5)
frequent_words <- findFreqTerms(dtm_train, 10)
frequent_words <- findFreqTerms(dtm_train, 50)

```

```

#Create document-term matrices using frequent words
crime_train <- DocumentTermMatrix(train_corpus,
  control=list(dictionary =
  frequent_words))
crime_test <- DocumentTermMatrix(test_corpus,
  control=list(dictionary =
  frequent_words))

#Convert count information to "Yes", "No"
convert_count <- function(x) {
  y <- ifelse(x > 0, 1,0)
  y <- factor(y, levels=c(0,1), labels=c("No",
  "Yes"))
  y
}

#Convert document-term matrices
crime_train <- apply(crime_train, 2, convert_count)
crime_test <- apply(crime_test, 2, convert_count)

#the naive bayes
library(e1071)

#Create a Naive Bayes classifier object
crime_classifier <- naiveBayes(crime_train,
  factor(tag$mode6[train_indices]))
class(crime_classifier) #cuma nampilin naive bayes

#Evaluate the performance on the test data
predictnbc <- predict(crime_classifier, crime_test)
actualnbc <- tag$tag[-train_indices]

# Calculating the accuracy in percents
levels(predictnbc) <- levels(actualnbc)
accuracy <-
100*sum(predictnbc==actualnbc)/length(test_corpus)

```

A.1.3 Kode Program Percobaan 3

```

tag <- read.csv("E:/MATA KULIAH/Skripsi/bahan
TA/LAPOR!/20151004/mydata.csv")

# Load the the library
library(tm)

# Load the data as a corpus
crime_corpus <-
Corpus(VectorSource(tag$isiLaporan))

# Convert the text to lower case

```

```

clean_corpus <- tm_map(crime_corpus,
content_transformer(tolower))
# Remove numbers
clean_corpus <- tm_map(clean_corpus, removeNumbers)
# Remove punctuations
clean_corpus <- tm_map(clean_corpus,
removePunctuation)
# Eliminate extra white spaces
clean_corpus <- tm_map(clean_corpus,
stripwhitespace)
# Remove stopwords
clean_corpus <- tm_map(clean_corpus, removewords,
stopwords("english"))
stopwordID <-
"https://raw.githubusercontent.com/nurandi/nurandi.
net/master/data/stopwords-id.txt"
cstopwordID <- readLines(stopwordID);
clean_corpus <- tm_map(clean_corpus, removewords,
c(cstopwordID))
clean_corpus <- tm_map(clean_corpus, removewords,
c("nama", "hari", "yth", "sdh", "jadi", "org",
"dpt", "krn", "bpk", "lagi", "akan", "sampai",
"kan", "setelah", "pak", "atas", "orang", "saya",
"dgn", "atau", "juga", "bisa", "kepada", "tsb",
"sangat", "mereka", "karena", "dan", "yang", "ini",
"dengan", "ada", "tidak", "untuk", "dari", "nya",
"tdk", "oleh", "kami", "dia", "mohon", "itu",
"tersebut", "pada", "sudah"))
#Eliminate extra white spaces
clean_corpus <- tm_map(clean_corpus,
stripwhitespace)

# tokenizing the corpus
dtm <-
DocumentTermMatrix(clean_corpus,control=list(weight
ing=weightTfIdf))

#divide corpus into training and test data
train_size <- round(0.7*nrow(tag))
set.seed(141321)
train_indices <- sample(seq_len(nrow(tag)),
size=train_size)
train_corpus <- clean_corpus[ train_indices]
test_corpus <- clean_corpus[-train_indices]

dtm_train <- dtm[train_indices, ]

#Identify frequently used words
frequent_words <- findFreqTerms(dtm_train, 5)
frequent_words <- findFreqTerms(dtm_train, 10)
frequent_words <- findFreqTerms(dtm_train, 50)

```

```

#Create document-term matrices using frequent words
crime_train <- DocumentTermMatrix(train_corpus,
  control=list(dictionary =
  frequent_words))
crime_test <- DocumentTermMatrix(test_corpus,
  control=list(dictionary =
  frequent_words))

#Convert count information to "Yes", "No"
convert_count <- function(x) {
  y <- ifelse(x > 0, 1,0)
  y <- factor(y, levels=c(0,1), labels=c("No",
  "Yes"))
  y
}

#Convert document-term matrices
crime_train <- apply(crime_train, 2, convert_count)
crime_test <- apply(crime_test, 2, convert_count)

#the naive bayes
library(e1071)

#Create a Naive Bayes classifier object
crime_classifier <- naiveBayes(crime_train,
  factor(tag$mode6[train_indices]))
class(crime_classifier) #cuma nampilin naive bayes

#Evaluate the performance on the test data
predictnbc <- predict(crime_classifier, crime_test)
actualnbc <- tag$tag[-train_indices]

# Calculating the accuracy in percents
levels(predictnbc) <- levels(actualnbc)
accuracy <-
100*sum(predictnbc==actualnbc)/length(test_corpus)

```

A.1.4 Kode Program Percobaan 4

```

tag <- read.csv("E:/MATA KULIAH/Skripsi/bahan
TA/LAPOR!/20151004/mydata.csv")

# Load the the library
library(tm)

# Load the data as a corpus
crime_corpus <-
Corpus(VectorSource(tag$isiLaporan))

# Convert the text to lower case

```

```

clean_corpus <- tm_map(crime_corpus,
content_transformer(tolower))
# Remove numbers
clean_corpus <- tm_map(clean_corpus, removeNumbers)
# Remove punctuations
clean_corpus <- tm_map(clean_corpus,
removePunctuation)
# Eliminate extra white spaces
clean_corpus <- tm_map(clean_corpus,
stripwhitespace)
# Remove stopwords
clean_corpus <- tm_map(clean_corpus, removewords,
stopwords("english"))
stopwordID <-
"https://raw.githubusercontent.com/nurandi/nurandi.
net/master/data/stopwords-id.txt"
cstopwordID <- readLines(stopwordID);
clean_corpus <- tm_map(clean_corpus, removewords,
c(cstopwordID))
clean_corpus <- tm_map(clean_corpus, removewords,
c("nama", "hari", "yth", "sdh", "jadi", "org",
"dpt", "krn", "bpk", "lagi", "akan", "sampai",
"kan", "setelah", "pak", "atas", "orang", "saya",
"dgn", "atau", "juga", "bisa", "kepada", "tsb",
"sangat", "mereka", "karena", "dan", "yang", "ini",
"dengan", "ada", "tidak", "untuk", "dari", "nya",
"tdk", "oleh", "kami", "dia", "mohon", "itu",
"tersebut", "pada", "sudah"))
#Eliminate extra white spaces
clean_corpus <- tm_map(clean_corpus,
stripwhitespace)

# tokenizing the corpus
dtm <-
DocumentTermMatrix(clean_corpus,control=list(weight
ing=weightTfIdf))

#divide corpus into training and test data
train_size <- round(0.7*nrow(tag))
set.seed(141321)
train_indices <- sample(seq_len(nrow(tag)),
size=train_size)
train_corpus <- clean_corpus[ train_indices]
test_corpus <- clean_corpus[-train_indices]

dtm_train <- dtm[train_indices, ]

#Identify frequently used words
frequent_words <- findFreqTerms(dtm_train, 5)
frequent_words <- findFreqTerms(dtm_train, 10)
frequent_words <- findFreqTerms(dtm_train, 50)

```

```

#Create document-term matrices using frequent words
crime_train <- DocumentTermMatrix(train_corpus,
control=list(dictionary = c("akun", "anak",
"anggota", "angkot", "aparatus", "apbd", "arah",
"atm", "bandar", "bank", "bbm", "begal", "blok",
"bupati", "bus", "copet", "daerah", "dana", "desa",
"email", "facebook", "foto", "ganja", "gembong",
"hadiah", "http", "hukum", "ilegal", "jalan",
"jenis", "judi", "juta", "kasus", "keamanan",
"kejadian", "keluarga", "kepala", "kirim",
"komplek", "konten", "koperasi", "korban",
"kumpul", "malam", "marak", "masalah",
"masyarakat", "media", "meresahkan", "mobil",
"motor", "negara", "nomor", "pasir", "pelacuran",
"pelaku", "pembegalan", "pemerintah", "pemilik",
"pencemaran", "pencurian", "pengedar", "penipuan",
"perjudian", "pihak", "pin", "preman",
"prostitusi", "proyek", "rawan", "rekening",
"rumah", "sabu", "sms", "sosial", "surat", "taman",
"tambang", "tanah", "tim", "timah", "tindak",
"tni", "togel", "transfer", "uang", "warga")))

crime_test <- DocumentTermMatrix(test_corpus,
control=list(dictionary = c("akun", "anak",
"anggota", "angkot", "aparatus", "apbd", "arah",
"atm", "bandar", "bank", "bbm", "begal", "blok",
"bupati", "bus", "copet", "daerah", "dana", "desa",
"email", "facebook", "foto", "ganja", "gembong",
"hadiah", "http", "hukum", "ilegal", "jalan",
"jenis", "judi", "juta", "kasus", "keamanan",
"kejadian", "keluarga", "kepala", "kirim",
"komplek", "konten", "koperasi", "korban",
"kumpul", "malam", "marak", "masalah",
"masyarakat", "media", "meresahkan", "mobil",
"motor", "negara", "nomor", "pasir", "pelacuran",
"pelaku", "pembegalan", "pemerintah", "pemilik",
"pencemaran", "pencurian", "pengedar", "penipuan",
"perjudian", "pihak", "pin", "preman",
"prostitusi", "proyek", "rawan", "rekening",
"rumah", "sabu", "sms", "sosial", "surat", "taman",
"tambang", "tanah", "tim", "timah", "tindak",
"tni", "togel", "transfer", "uang", "warga")))

#Convert count information to "Yes", "No"
convert_count <- function(x) {
  y <- ifelse(x > 0, 1,0)
  y <- factor(y, levels=c(0,1), labels=c("No",
"Yes"))
  y
}

#Convert document-term matrices

```

```

crime_train <- apply(crime_train, 2, convert_count)
crime_test <- apply(crime_test, 2, convert_count)

#the naive bayes
library(e1071)

#Create a Naive Bayes classifier object
crime_classifier <- naiveBayes(crime_train,
factor(tag$mode6[train_indices]))
class(crime_classifier) #cuma nampilin naive bayes

#Evaluate the performance on the test data
predictnbc <- predict(crime_classifier, crime_test)
actualnbc <- tag$tag[-train_indices]

# Calculating the accuracy in percents
levels(predictnbc) <- levels(actualnbc)
accuracy <-
100*sum(predictnbc==actualnbc)/length(test_corpus)

```

A.1.5 Kode Program Percobaan 5

```

susun <- read.csv("E:/MATA KULIAH/Skripsi/bahan
TA/LAPOR!/20151004/susun.csv")

# Load the the tm library
library(tm)

#divide corpus into training and test data
train_size <- round(0.7*nrow(susun))
set.seed(60282)
train_indices <- sample(seq_len(nrow(susun)),
size=train_size)
train_corpus <- susun[ train_indices, ]
test_corpus <- susun[-train_indices, ]

# Load the data train as a corpus
train_docs <-
Corpus(VectorSource(train_corpus$isiLaporan))

# Clean corpus
train_docs <- tm_map(train_docs,
content_transformer(tolower))
train_docs <- tm_map(train_docs, removeNumbers)
train_docs <- tm_map(train_docs, removePunctuation)
train_docs <- tm_map(train_docs, stripwhitespace)
train_docs <- tm_map(train_docs, removewords,
c("nama", "hari", "yth", "sdh", "jadi", "org",
"dpt", "krn", "bpk", "lagi", "akan", "sampai",
"kan", "setelah", "pak", "atas", "orang", "saya",

```

```

"dgn", "atau", "juga", "bisa", "kepada", "tsb",
"sangat", "mereka", "karena", "dan", "yang", "ini",
"dengan", "ada", "tidak", "untuk", "dari", "nya",
"tdk", "oleh", "kami", "dia", "mohon", "itu",
"tersebut", "pada", "sudah", "iii"))

# tokenizing the corpus
dtmtrain <- DocumentTermMatrix(train_docs,
control=list(weighting=weightTfIdf)) #TfIdf
frequent_words1 <- findFreqTerms(dtmtrain, 5)
dtm_train <- DocumentTermMatrix(train_docs,
control=list(dictionary = frequent_words1))

#Convert count information to "Yes", "No"
convert_count <- function(x) {
  y <- ifelse(x > 0, 1,0)
  y <- factor(y, levels=c(0,1), labels=c("No",
"Yes"))
  y
}

#Convert document-term matrices
dtm_train <- apply(dtm_train, 2, convert_count)
df_train <- as.data.frame(data.matrix(dtm_train),
stringsAsfactors = FALSE)

#the naive bayes
library(e1071)

#_____ MODEL 1 _____

#Create a Naive Bayes classifier object
model1 <- naiveBayes(df_train,
factor(susun$model1[train_indices]))

#_____ MODEL 2 _____

#choose data for model 2
susun2 <-
susun[which(susun$model1=="nonpenipuan"),]

#divide corpus into training and test data
train_size2 <- round(0.7*nrow(susun2))
set.seed(60282)
train_indices2 <- sample(seq_len(nrow(susun2)),
size=train_size2)
train_corpus2 <- susun2[ train_indices2, ]
test_corpus2 <- susun2[-train_indices2, ]

# Load the data train as a corpus
train_docs2 <-

```

```

Corpus(VectorSource(train_corpus2$isiLaporan))

# Clean corpus
train_docs2 <- tm_map(train_docs2,
content_transformer(tolower))
train_docs2 <- tm_map(train_docs2, removeNumbers)
train_docs2 <- tm_map(train_docs2,
removePunctuation)
train_docs2 <- tm_map(train_docs2, stripwhitespace)
train_docs2 <- tm_map(train_docs2, removewords,
c("nama", "hari", "yth", "sdh", "jadi", "org",
"dpt", "krn", "bpk", "lagi", "akan", "sampai",
"kan", "setelah", "pak", "atas", "orang", "saya",
"dgn", "atau", "juga", "bisa", "kepada", "tsb",
"sangat", "mereka", "karena", "dan", "yang", "ini",
"dengan", "ada", "tidak", "untuk", "dari", "nya",
"tdk", "oleh", "kami", "dia", "mohon", "itu",
"tersebut", "pada", "sudah"))

# tokenizing the corpus
dtmtrain2 <- DocumentTermMatrix(train_docs2,
control=list(weighting=weightTfIdf)) #TfIdf
frequent_words2 <- findFreqTerms(dtmtrain2, 1)
dtm_train2 <- DocumentTermMatrix(train_docs2,
control=list(dictionary = frequent_words2))

#Convert document-term matrices
dtm_train2 <- apply(dtm_train2, 2, convert_count)
df_train2 <- as.data.frame(matrix(dtm_train2),
stringsAsFactors = FALSE)

#Create a Naive Bayes classifier object
model2 <- naiveBayes(df_train2,
factor(susun2$tag[train_indices2]))

#_____ TESTING MODEL 1 _____

#Load the data test as a corpus
test_docs <-
Corpus(VectorSource(test_corpus$isiLaporan))

# Clean corpus
test_docs <- tm_map(test_docs,
content_transformer(tolower))
test_docs <- tm_map(test_docs, removeNumbers)
test_docs <- tm_map(test_docs, removePunctuation)
test_docs <- tm_map(test_docs, stripwhitespace)
test_docs <- tm_map(test_docs, removewords,
c("nama", "hari", "yth", "sdh", "jadi", "org",
"dpt", "krn", "bpk", "lagi", "akan", "sampai",
"kan", "setelah", "pak", "atas", "orang", "saya",
"dgn", "atau", "juga", "bisa", "kepada", "tsb",

```

```

"sangat", "mereka", "karena", "dan", "yang", "ini",
"dengan", "ada", "tidak", "untuk", "dari", "nya",
"tdk", "oleh", "kami", "dia", "mohon", "itu",
"tersebut", "pada", "sudah", "iii"))

dtm_test <- DocumentTermMatrix(test_docs, control =
list(dictionary = names(df_train)))
dtm_test <- apply(dtm_test, 2, convert_count)
df_test <- as.data.frame(data.matrix(dtm_test),
stringsAsfactors = FALSE)

#Evaluate the performance on the test data
predicted <- predict(model1, df_test)
actual <- susun$model1[-train_indices]

levels(predicted) <- levels(actual)
nbc1 <-
100*sum(predicted==actual)/length(predicted)
100*sum(predicted!=actual)/length(predicted)

# _____ TEST MODEL 2 _____

#Load the data test as a corpus
test_docs2 <-
Corpus(VectorSource(test_corpus2$isiLaporan))

# Clean corpus
test_docs2 <- tm_map(test_docs2,
content_transformer(tolower))
test_docs2 <- tm_map(test_docs2, removeNumbers)
test_docs2 <- tm_map(test_docs2, removePunctuation)
test_docs2 <- tm_map(test_docs2, stripwhitespace)
test_docs2 <- tm_map(test_docs2, removewords,
c("nama", "hari", "yth", "sdh", "jadi", "org",
"dpt", "krn", "bpk", "lagi", "akan", "sampai",
"kan", "setelah", "pak", "atas", "orang", "saya",
"dgn", "atau", "juga", "bisa", "kepada", "tsb",
"sangat", "mereka", "karena", "dan", "yang", "ini",
"dengan", "ada", "tidak", "untuk", "dari", "nya",
"tdk", "oleh", "kami", "dia", "mohon", "itu",
"tersebut", "pada", "sudah", "iii"))

dtm_test2 <- DocumentTermMatrix(test_docs2, control =
list(dictionary = names(df_train2)))
dtm_test2 <- apply(dtm_test2, 2, convert_count)
df_test2 <- as.data.frame(data.matrix(dtm_test2),
stringsAsfactors = FALSE)

#Evaluate the performance on the test data
predicted2 <- predict(model2, df_test2) #length
predicted2 = 341
actual2 <- susun2$tag[-train_indices2] #length =

```

```

341 (masih ada penipuan) > akurasi = 9.97

levels(predicted2) <- levels(actual2)
nbc2 <-
100*sum(predicted2==actual2)/length(predicted2)
100*sum(predicted2!=actual2)/length(predicted2)

#_____ AKURASI TOTAL _____

((length(test_docs)*nbc1)+(length(test_docs2)*nbc2)
)/(length(test_docs)+length(test_docs2))

```

A.2 Shiny

A.2.1 Server.R

```

function(input, output, session) {
  terms <- reactive({
    # Change when the "update" button is pressed...
    input$update
    # ...but not for anything else
    isolate({
      withProgress({
        setProgress(message = "Processing
corpus...")
        getTermMatrix(input$klasifikasi)
      })
    })
  })

  # Make the wordcloud drawing predictable during a
  session
  wordcloud_rep <- repeatable(wordcloud)

  output$plot <- renderPlot({
    v <- terms()
    wordcloud_rep(names(v), v, scale=c(4,0.5),
min.freq = input$freq,
max.words=input$max,
colors=brewer.pal(8, "Dark2"))})

  colors <- brewer.pal(2, "RdBu")

  Periode <- reactive({
    switch(input$Periode,
      "Triwulan 1" = timeline <-
read.csv("C:/Users/user/R/TA/triwulan1.csv"),
      "Triwulan 2" = timeline <-
read.csv("C:/Users/user/R/TA/triwulan2.csv"),

```

```

"Triwulan 3" = timeline <-
read.csv("C:/Users/user/R/TA/triwulan3.csv"),
"Triwulan 4" = timeline <-
read.csv("C:/Users/user/R/TA/triwulan4.csv"),
"Triwulan 5" = timeline <-
read.csv("C:/Users/user/R/TA/triwulan5.csv"),
"Triwulan 6" = timeline <-
read.csv("C:/Users/user/R/TA/triwulan6.csv"),
"Triwulan 7" = timeline <-
read.csv("C:/Users/user/R/TA/triwulan7.csv")
})

output$plot1 <- renderStreamgraph({
  timeline <- Periode()
  timeline %>%

  filter(kelas %in% c("akibatorangluka", "curanmor",
"curas", "curat", "himbauan", "hutangpiutang",
"implikasikontijensi", "jabatan", "kebakaran",
"kesusilaan", "korupsi", "lainlain", "lantas",
"narkoba", "pemalsuansurat", "pembakaran",
"pembunuhan", "pemerasan", "penadahan",
"penculikan", "pencurianringan", "pengaiayaan",
"pengerusakan", "penggelapan", "penghinaan",
"penipuan", "penyuapan", "perdaganganmanusia",
"perjudian", "perkosaan"))%>%

  group_by(tanggal, kelas) -> some
  streamgraph(some, "kelas", "laporan", "tanggal")
  %>%
  sg_axis_x(tick_interval=1, tick_units="tanggal",
"%d/%m")%>%
  sg_legend(show=TRUE, label="Jenis Kejahatan: ")
  })
}

```

A.2.2 Ui.R

```

library(streamgraph)
library(dplyr)
library(RColorBrewer)

fluidPage(
  # Application title
  headerPanel("Visualisasi LAPOR!"),

  # Show word cloud
  tabsetPanel(type = c("pills"),
    tabPanel("wordcloud",
      sidebarPanel(

```

```

klasifikasi:", selectInput("klasifikasi", "Pilih
choices = klasifikasi),
actionButton("update", "Update"),
hr(),
sliderInput("freq", "Minimum
Frequency:", min = 1, max = 100, value = 20),
sliderInput("max", "Maximum Number
of Words:",
min = 1, max = 300, value = 100)
),
mainPanel(
plotOutput("plot")
)
), tabPanel("Stream Graph",
sidebarPanel(
selectInput("Periode", "Pilih
Periode", c("Triwulan 1", "Triwulan 2", "Triwulan
3", "Triwulan 4", "Triwulan 5", "Triwulan 6",
"Triwulan 7")),
actionButton("update", "Update")),
mainPanel(
sg_title("Jumlah Laporan
Kriminalitas Per Kategori"),
streamgraphOutput("plot1"))
)))

```

A.2.3 Global.R

```

library(tm)
library(wordcloud)
library(memoise)

# The list of valid crime
klasifikasi <- list("Penipuan" = "penipuan",
"Pemerasan" = "pemerasan", "Akibat Orang Luka" =
"akibatorangluka", "Curanmor" = "curanmor", "Curas"
= "curas", "Curat" = "curat", "Himbauan" =
"himbauan", "Hutang Piutang" = "hutangpiutang",
"Implikasi Kontijensi" = "implikasikontijensi",
"Jabatan" = "jabatan", "Kebakaran" = "kebakaran",
"Kesusilaan" = "kesusilaan", "Korupsi" = "korupsi",
"Lain-Lain" = "lainlain", "Lantas" = "lantas",
"Narkoba" = "narkoba", "Pemalsuan Surat" =
"pemalsuansurat", "Pembakaran" = "pembakaran",
"Pembunuhan" = "pembunuhan", "Penadahan" =
"penadahan", "Penculikan" = "penculikan",
"Pencurian Ringan" = "pencurianringan",
"Penganiayaan" = "penganiayaan", "Pengerusakan" =
"pengerusakan", "Penggelapan" = "penggelapan",
"Penghinaan" = "penghinaan", "Penyuapan" =

```

```

"penyuapan", "Perdagangan Manusia" =
"perdaganganmanusia", "Perjudian" = "perjudian",
"Perkosaan" = "perkosaan")

# Using "memoise" to automatically cache the
results

getTermMatrix <- memoise(function(klasifikasi) {
# Careful not to let just any name slip in here; a
# malicious user could manipulate this value.

if (!(klasifikasi %in% klasifikasi))
  stop("unknown klasifikasi")

text <-
readLines(sprintf("C:/Users/user/R/TA/./%s.txt.gz",
klasifikasi), encoding="UTF-8")

docs <- Corpus(VectorSource(text))
inspect(docs)
toSpace <- content_transformer(function (x ,
pattern ) gsub(pattern, " ", x))

docs <- tm_map(docs, toSpace, "/")
docs <- tm_map(docs, toSpace, "@")
docs <- tm_map(docs, toSpace, "\\|")
docs <- tm_map(docs, content_transformer(tolower))
docs <- tm_map(docs, removeNumbers)

stopwordID <-
"https://raw.githubusercontent.com/nurandi/nurandi.
net/master/data/stopwords-id.txt"
  cStopwordID <- readLines(stopwordID);
  docs <- tm_map(docs, removeWords,
c(stopwords("english"), cStopwordID))
  docs <- tm_map(docs, removeWords,
c("bapak", "ibu", "jemmy", "ucok", "adik", "evrina", "oki",
"riyanti", "untk", "mks", "dlm", "kab", "dilakukan", "k
asmin", "chatib", "arthasari", "zumairi", "karna",
"terjadi", "udh", "unt", "nama", "hari", "yth",
"sdh", "jadi", "org", "dpt", "krn", "bpk", "lagi",
"akan", "sampai", "kan", "setelah", "pak", "atas",
"orang", "saya", "dgn", "atau", "juga", "bisa",
"kepada", "tsb", "sangat", "mereka", "karena",
"dan", "yang", "ini", "dengan", "ada", "tidak",
"untuk", "dari", "nya", "tdk", "oleh", "kami",
"dia", "mohon", "itu", "tersebut", "pada",
"sudah"))
  docs <- tm_map(docs, removePunctuation)
  docs <- tm_map(docs, stripwhitespace)

```

```
dtm <- TermDocumentMatrix(docs, control =  
list(minwordLength = 1))  
m <- as.matrix(dtm)  
  
sort(rowSums(m), decreasing = TRUE)  
})
```

LAMPIRAN B

DAFTAR *STOPWORDS* BAHASA INDONESIA

B.1 Daftar *Stopwords* Indonesia

Tabel B.1 Daftar *Stopwords* Indonesia

1	ada	begitulah	entah	kenapa	meskipun	sedikit	setelah
2	adanya	begitupun	entahlah	kepada	semula	sedikitnya	telah
3	adalah	sebegitu	terhadap	kepadanya	mungkin	segala	tentang
4	adapun	belum	terhadapnya	ketika	mungkinkah	segalanya	tentu
5	agak	belumah	hal	seketika	nah	segera	tentulah
6	agaknya	sebelum	hampir	khususnya	namun	sesegera	tentunya
7	agar	sebelumnya	hanya	kini	nanti	sejak	tertentu
8	akan	sebenarnya	hanyalah	kinilah	nantinya	sejenak	seterusnya
9	akankah	berapa	harus	kiranya	nyaris	sekali	tapi

10	akhirnya	berapakah	haruslah	sekiranya	oleh	sekalian	tetapi
11	aku	berapalah	harusnya	kita	olehnya	sekalipun	setiap
12	akulah	berapapun	seharusnya	kitalah	seorang	sesekali	tiap
13	amat	betulkah	hendak	kok	seseorang	sekaligus	setidaknya
14	amatlah	sebetulnya	hendaklah	lagi	pada	sekarang	tidak
15	anda	biasa	hendaknya	lagian	padanya	sekarang	tidakkah
16	andalah	biasanya	hingga	selagi	padahal	sekitar	tidaklah
17	antar	bila	sehingga	lah	paling	sekitarnya	toh
18	diantaranya	bilakah	ia	lain	sepanjang	sela	waduh
19	antara	bisa	ialah	lainnya	pantas	selain	wah
20	antaranya	bisakah	ibarat	melainkan	sepantasnya	selalu	wahai
21	diantara	sebisanya	ingin	selaku	sepantasnyalah	seluruh	sewaktu
22	apa	boleh	inginkah	lalu	para	seluruhnya	walaupun
23	apaan	bolehkah	inginkan	melalui	pasti	semakin	walaupun
24	mengapa	bolehlah	ini	terlalu	pastilah	sementara	wong

25	apabila	buat	inikah	lama	per	sempat	yaitu
26	apakah	bukan	inilah	lamanya	pernah	semua	yakni
27	apalagi	bukankah	itu	selama	pula	semuanya	yang
28	apatah	bukanlah	itukah	selama	pun	sendiri	
29	atau	bukannya	itulah	selamanya	merupakan	sendirinya	
30	ataukah	cuma	jangan	lebih	rupanya	seolah	
31	ataupun	percuma	jangan	terlebih	serupa	seperti	
32	bagai	dahulu	janganlah	bermacam	saat	sepertinya	
33	bagaikan	dalam	jika	macam	saatnya	sering	
34	sebagai	dan	jikalau	semacam	sesaat	seringnya	
35	sebagainya	dapat	juga	maka	saja	serta	
36	bagaimana	dari	justru	makanya	sajalah	siapa	
37	bagaimanapun	daripada	kala	makin	saling	siapakah	
38	sebagaimana	dekat	kalau	malah	bersama	siapapun	
39	bagaimanakah	demi	kalaulah	malahan	sama	disini	

40	bagi	demikian	kalaupun	mampu	sesama	disinilah	
41	bahkan	demikianlah	kalian	mampukah	sambil	sini	
42	bahwa	sedemikian	kami	mana	sampai	sinilah	
43	bahwasanya	dengan	kamilah	manakala	sana	sesuatu	
44	sebaliknya	depan	kamu	manalagi	sangat	sesuatunya	
45	banyak	di	kamulah	masih	sangatlah	suatu	
46	sebanyak	dia	kan	masihkah	saya	sesudah	
47	beberapa	dialah	kapan	semasih	sayalah	sesudahnya	
48	seberapa	dini	kapankah	masing	se	sudah	
49	begini	diri	kapanpun	mau	sebab	sudahkah	
50	beginian	dirinya	dikarenakan	maupun	sebabnya	sudahlah	
51	beginikah	terdiri	karena	semaunya	sebuah	supaya	
52	beginilah	dong	karenanya	memang	tersebut	tadi	
53	sebegini	dulu	ke	mereka	tersebutlah	tadinya	
54	begitu	enggak	kecil	merekalah	sedang	tak	

55	begitukah	enggaknya	kemudian	meski	sedangkan	tanpa	
----	-----------	-----------	----------	-------	-----------	-------	--

B.2 Analisis Hasil 20 Kata Teratas

Tabel B.2 Analisis 20 Kata Teratas Kelas Penipuan, Pemerasan, Narkoba, Perjudian, Curat

PENIPUAN		PEMERASAN		NARKOBA		PERJUDIAN		CURAT	
word	Freq	word	freq	word	freq	word	freq	word	freq
penipuan	671	pemerasan	110	narkoba	131	perjudian	102	curat	50
uang	248	preman	57	oki	33	judi	59	rumah	29
sms	190	uang	50	sabu	30	desa	48	bank	16
hadiah	181	motor	33	shabu	24	togel	25	pencurian	16
tolong	138	ibu	32	ganja	21	perjudiyan	24	polisi	15
rekening	125	kasih	30	bandar	19	masyarakat	22	mandiri	14
prog	123	polisi	30	cengkareng	18	warga	21	pelaku	14
bank	119	rumah	28	warga	16	hajatan	19	bandung	13
nomor	113	tolong	25	desa	13	kota	19	script	13
konten	111	mobil	22	pengedar	13	tanggamus	19	lapor	12

PENIPUAN		PEMERASAN		NARKOBA		PERJUDIAN		CURAT	
word	Freq	word	freq	word	freq	word	freq	word	freq
korban	110	masalah	19	jatim	12	tempat	19	terjadi	11
pihak	95	meminta	19	jenis	12	agung	17	laporan	10
berikan	94	pihak	19	kecamatan	11	rumah	16	malam	10
penipu	93	terima	19	anak	10	malam	15	timur	10
transfer	92	datang	18	kab	10	marak	15	daerah	9
cirebon	90	masyarakat	18	kec	10	kabupaten	14	kalbar	9
kirim	88	minta	18	tolong	10	kecamatan	14	mobil	9
smp	86	jalan	17	ditangkap	9	polri	14	motor	9
pelaku	82	surat	17	gembong	9	kec	13	besar	8
barang	81	keluarga	16	lampung	9	lampung	13	kejadian	8

Tabel B.3 Analisis 20 Kata Teratas Kelas Implikasi Kontijensi, Himbauan, Korupsi, Curas, Penggelapan

IMPLIKASI KONTIJENSI		HIMBAUAN		KORUPSI		CURAS		PENGSELAPAN	
word	freq	word	freq	word	freq	word	freq	word	freq
implikasikontijensi	49	himbauan	47	korupsi	100	curas	30	penggelapan	36

IMPLIKASI KONTIJENSI		HIMBAUAN		KORUPSI		CURAS		PENGGELAPAN	
word	freq	word	freq	word	freq	word	freq	word	freq
anak	15	lapor	11	dana	72	motor	21	timah	25
komplek	12	anak	10	desa	43	terjadi	16	ilegal	17
warga	12	polsek	10	kabupaten	34	pelaku	14	uang	11
ganja	11	hukum	9	tanggamus	33	jalan	13	bbm	10
masyarakat	11	polisi	8	kalbar	31	daerah	12	koperasi	10
meresahkan	11	warga	8	bupati	24	begal	11	anggota	9
malam	10	aktivitas	7	negara	24	malam	10	mobil	8
polisi	10	aparatus	7	kasus	23	kasih	9	tambang	8
daerah	9	kemayoran	7	kepala	23	pembegalan	8	tanah	8
kepolisian	9	membuat	7	lampung	20	polisi	8	hukum	7
melaporkan	9	semarang	7	kecamatan	17	korban	7	laporan	7
rumah	9	jalan	6	masyarakat	17	masyarakat	7	menjadi	7
aparatus	8	rumah	6	apbd	16	menggunakan	7	panin	7
bapak	8	tim	6	kota	16	meresahkan	7	pasir	7
kumpul	8	tindak	6	pemerintah	16	terima	7	polri	7

IMPLIKASI KONTIJENSI		HIMBAUAN		KORUPSI		CURAS		PENGSELAPAN	
word	freq	word	freq	word	freq	word	freq	word	freq
tindakan	8	dilakukan	5	proyek	16	tindakan	7	trms	7
blok	7	http	5	uang	16	keamanan	6	ukpppp	7
desa	7	jakarta	5	ntt	15	kepolisian	6	bank	6
negara	7	kasih	5	polda	13	rawan	6	juta	6

Tabel B.4 Analisis 20 Kata Teratas Kelas Kesusilaan, Penghinaan, Pencurian Ringan, Pembunuhan, Jabatan

KESUSILAAN		PENGHINAAN		PENCURIANRINGAN		PEMBUNUHAN		JABATAN	
word	freq	word	freq	word	freq	word	freq	word	freq
kesusilaan	26	penghinaan	30	pencurianringan	21	barat	45	jabatan	18
barat	23	akun	13	jalan	6	pasaman	24	tanah	12
pin	13	baik	13	minta	6	sumatra	24	dilakukan	10
tempat	13	facebook	10	arah	5	pembunuhan	22	kab	9
anak	11	pencemaran	8	bandung	5	air	20	mks	7
sumatra	11	foto	7	copet	5	bangis	20	uang	7
taman	11	nomor	7	motor	5	polisi	15	chatib	6

KESUSILAAN		PENGHINAAN		PENCURIANRINGAN		PEMBUNUHAN		JABATAN	
word	freq	word	freq	word	freq	word	freq	word	freq
pasaman	10	bukti	6	baru	4	banggai	12	kasus	6
polisi	10	lanjuti	6	bus	4	tni	12	oknum	6
tolong	10	media	6	ditindak	4	kab	10	sorong	6
air	9	pelaku	6	ibu	4	pok	10	ajang	5
bangis	9	pemilik	6	pagi	4	tengah	10	hukum	5
indonesia	8	sosial	6	polisi	4	bunuh	9	kasmin	5
prostitusi	8	alamat	5	sms	4	dinas	8	kepala	5
warga	8	berita	5	tas	4	dunia	8	pegawai	5
negara	7	bitung	5	tolong	4	hutan	8	pejabat	5
pelacuran	7	dilakukan	5	aja	3	kec	8	pns	5
riyanti	7	email	5	angkot	3	kehutanan	8	sdr	5
sya	7	https	5	atm	3	marpaung	8	terjadi	5
tni	7	kasih	5	baju	3	malam	7	tersangka	5

Tabel B.5 Analisis 20 Kata Teratas Kelas Penyuapan, Pemalsuan Surat, Curanmor, Penganiayaan

PENYUAPAN		PEMALSUAN SURAT		CURANMOR		PENGANIAYAAN	
word	freq	word	freq	word	freq	word	freq
penyuapan	16	pemalsuansurat	15	motor	19	penganiayaan	21
suap	16	tanah	13	curanmor	17	adik	12
jemmy	9	surat	11	polisi	7	kasus	7
data	7	jombang	7	pencurian	6	anak	6
kasih	6	kulit	7	terjadi	6	sekolah	5
menase	6	medan	6	daerah	5	ucok	5
uang	6	pemalsuan	6	warga	5	bapak	4
lampung	5	tgl	6	bawa	4	hukum	4
pesawat	5	anak	5	jalan	4	jatuh	4
bangunan	3	lurah	5	kasih	4	pelaku	4
desa	3	utk	5	laporan	4	warga	4
jelas	3	adik	4	perumahan	4	ayah	3
jutaan	3	alamat	4	ban	3	kekerasan	3
karna	3	bukti	4	bapak	3	oknum	3

PENYUAPAN		PEMALSUAN SURAT		CURANMOR		PENGANIAYAAN	
word	freq	word	freq	word	freq	word	freq
karol	3	byk	4	cengkareng	3	pemerintah	3
kpl	3	cerai	4	kawan	3	polres	3
minta	3	hewan	4	kehilangan	3	polsek	3
oknum	3	ipar	4	konkrit	3	pot	3
pnnya	3	jatim	4	pelaku	3	seakan	3
polisi	3	kab	4	polsek	3	terima	3

B.3 Proses Penentuan Daftar Kata untuk *Dictionary*

Berikut terdapat 4 percobaan yang diujicobakan pada model klasifikasi penipuan dan non penipuan.

Tabel B.6 Uji Coba *Dictionary* 1

Percobaan	Daftar Kata	Akurasi
1	"akun", "anak", "anggota", "angkot", "aparot", "apbd", "arah", "atm", "bandar", "bank", "bbm", "begal", "blok", "bupati", "bus", "copet", "daerah", "dana", "desa", "email", "facebook", "foto", "ganja", "gembong", "hadiah", "http", "hukum", "ilegal", "jalan",	90.90909

Percobaan	Daftar Kata	Akurasi
	"jenis", "judi", "juta", "kasus", "keamanan", "kejadian", "keluarga", "kepala", "kirim", "komplek", "konten", "koperasi", "korban", "kumpul", "malam", "marak", "masalah", "masyarakat", "media", "meresahkan", "mobil", "motor", "negara", "nomor", "pasir", "pelacuran", "pelaku", "pembegalan", "pemerintah", "pemilik", "pencemaran", "pencurian", "pengedar", "penipuan", "perjudian", "pihak", "pin", "preman", "prostitusi", "proyek", "rawan", "rekening", "rumah", "sabu", "sms", "sosial", "surat", "taman", "tambang", "tanah", "tim", "timah", "tindak", "tni", "togel", "transfer", "uang", "warga"	
2	"akun", "anak", "anggota", "angkot", "aparatus", "apbd", "arah", "atm", "bandar", "bank", "barang", "bbm", "begal", "berita", "besar", "blok", "bukti", "bupati", "bus", "copet", "daerah", "dana", "desa", "email", "facebook", "foto", "ganja", "gembong", "hadiah", "http", "hukum", "ilegal", "jalan", "jenis", "judi", "juta", "kasus", "keamanan", "kejadian", "keluarga", "kepala", "kirim", "komplek", "konten", "koperasi", "korban", "kumpul", "malam", "marak", "masalah", "masyarakat", "media", "meresahkan", "minta", "mobil", "motor", "negara", "nomor", "pasir", "pelacuran", "pelaku", "pembegalan", "pemerintah", "pemilik", "pencemaran", "pencurian", "pengedar", "penipu", "penipuan", "perjudian", "pihak", "pin", "preman", "prostitusi", "proyek", "rawan", "rekening", "rumah", "sabu", "sms", "sosial", "surat", "taman", "tambang", "tanah", "tim", "timah", "tindak", "tni", "togel", "transfer", "uang", "warga"	90.61548
3	"akun", "anak", "anggota", "angkot", "aparatus", "apbd", "arah", "atm", "bandar", "baik", "bank", "barang", "bbm", "begal", "berita", "blok", "bukti", "bupati", "bus", "copet", "dana", "email", "facebook", "foto", "ganja", "gembong", "hadiah", "http", "ilegal",	87.09677

Percobaan	Daftar Kata	Akurasi
	"jalan", "jenis", "judi", "juta", "keamanan", "keluarga", "kepala", "kepolisian", "kirim", "komplek", "konten", "koperasi", "korban", "kumpul", "mandiri", "marak", "masyarakat", "media", "meresahkan", "minta", "mobil", "motor", "negara", "nomor", "pasir", "pelacuran", "pelaku", "pembegalan", "pemerintah", "pencemaran", "pencurian", "pengedar", "penipu", "perjudian", "pihak", "pin", "preman", "prostitusi", "proyek", "rawan", "rekening", "rumah", "sabu", "script", "smp", "sms", "sosial", "surat", "taman", "tambang", "tanah", "timah", "tni", "togel", "transfer", "uang", "warga"	
4	"akun", "alamat", "anak", "anggota", "angkot", "aparatus", "apbd", "arah", "atm", "baik", "bandar", "bank", "barang", "baru", "bbm", "begal", "berita", "besar", "blok", "bukti", "bupati", "bus", "copet", "daerah", "dana", "desa", "email", "facebook", "foto", "ganja", "gembong", "hadiah", "http", "hukum", "ilegal", "jalan", "jenis", "judi", "juta", "kasus", "keamanan", "kejadian", "keluarga", "kepala", "kepolisian", "kirim", "komplek", "konten", "koperasi", "korban", "kumpul", "lapor", "laporan", "malam", "mandiri", "marak", "masalah", "masyarakat", "media", "meresahkan", "minta", "mobil", "motor", "negara", "nomor", "pagi", "pasir", "pelacuran", "pelaku", "pembegalan", "pemerintah", "pemilik", "pencemaran", "pencurian", "pengedar", "penipu", "perjudian", "pihak", "pin", "polda", "polisi", "polri", "polsek", "preman", "prostitusi", "proyek", "rawan", "rekening", "rumah", "sabu", "script", "smp", "sms", "sosial", "surat", "taman", "tambang", "tanah", "tas", "terima", "tim", "timah", "timur", "tindak", "tni", "togel", "tolong", "transfer", "uang", "warga"	86.21701

Hasil percobaan di atas menunjukkan bahwa percobaan 1 menghasilkan akurasi tertinggi. Oleh karena itu, selanjutnya dilakukan proses klasifikasi dengan data lainnya dan menghasilkan akurasi sebagai berikut.

Tabel B.7 Hasil Uji Coba *Dictionary* 1

Penipuan & Non Penipuan	30 Kelas	Lebih dari 10	Lebih dari 20	Lebih dari 30	Lebih dari 30 non Himbauan & Lain – lain
90.90909	6.744868	14.11043	62.08054	15.11628	70.17544

Untuk perbandingan daftar kata di atas, dilakukan percobaan kembali menggunakan kata tanpa unsur penipuan. Berikut daftar kata setelah mengalami eliminasi kata yang berunsur penipuan.

Tabel B.8 Uji Coba *Dictionary* 2

"akun", "anak", "anggota", "angkot", "aparatus", "apbd", "arah", "atm", "bandar", "bbm", "begal", "blok", "bupati", "bus", "copet", "daerah", "dana", "desa", "email", "facebook", "foto", "ganja", "gembong", "http", "hukum", "ilegal", "jalan", "jenis", "judi", "juta", "kasus", "keamanan", "kejadian", "keluarga", "kepala", "komplek", "koperasi", "korban", "kumpul", "malam", "marak", "masalah", "masyarakat", "media", "meresahkan", "mobil", "motor", "negara", "nomor", "pasir", "pelacuran", "pelaku", "pembegalan", "pemerintah", "pemilik", "pencemaran", "pencurian", "pengedar", "perjudian", "pihak", "pin", "preman", "prostitusi", "proyek", "rawan", "rumah", "sabu", "sosial", "surat", "taman", "tambang", "tanah", "tim", "timah", "tindakan", "tni", "togel", "warga"

Tabel B.9 Hasil Uji Coba *Dictionary 2*

30 Kelas	Lebih dari 10	Lebih dari 20	Lebih dari 30 tanpa Himbauan & Lain – lain
29.20354	36.01896	53.00546	66.22807

Setelah dilakukan percobaan menggunakan daftar kata yang tidak mengandung unsur penipuan, hasil yang didapat menunjukkan bahwa akurasi tidak lebih baik dari percobaan dengan menggunakan kata penipuan. Sehingga, daftar kata yang digunakan untuk percobaan 4 yaitu daftar kata pada percobaan 1.

C.5 Himbauan



C.6 Hutang Piutang



C.9 Kebakaran



C.10 Kesusilaan



C.11 Korupsi



C.12 Lain-Lain



C.15 Pemalsuan Surat



C.16 Pembakaran



C.21 Pencurian Ringan



C.22 Penganiayaan



C.25 Penghinaan



C.26 Penipuan



LAMPIRAN D

DESKRIPSI DAN SANKSI JENIS KEJAHATAN

D.1 Kebakaran

(KUHP Pasal 188) : –Barang siapa karena kesalahan (kealpaan) menyebabkan kebakaran, ledakan atau banjir, diancam dengan pidana penjara paling lama lima tahun atau pidana kurungan paling lama satu tahun atau pidana denda paling banyak empat ribu lima ratus rupiah, jika karena perbuatan itu timbul bahaya umum bagi barang, jika karena perbuatan itu timbul bahaya bagi nyawa orang lain, atau jika karena perbuatan itu mengakibatkan orang mati.”

D.2 Pembakaran

(KUHP Pasal 187) : –Barang siapa dengan sengaja menimbulkan kebakaran, ledakan atau banjir, diancam:

1. dengan pidana penjara paling lama dua belas tahun, jika karena perbuatan tersebut di atas timbul bahaya umum bagi barang;
2. dengan pidana penjara paling lama lima belas tahun, jika karena perbuatan tersebut di atas timbul bahaya bagi nyawa orang lain; dengan pidana penjara seumur hidup atau selama waktu tertentu paling lama dua puluh tahun, jika karena perbuatan tersebut di atas timbul bahaya bagi nyawa orang lain dan mengakibatkan orang mati.”

D.3 Jabatan

(KUHP Pasal 415) : –Seorang pejabat atau orang lain yang ditugaskan menjalankan suatu jabatan umum terus-menerus atau untuk sementara waktu, yang dengan sengaja menggelapkan uang atau surat berharga yang disimpan karena jabaimnya, atau membiarkan uang atau surat berharga itu diambil atau digelapkan oleh orang lain, atau menolong sebagai pembantu dalam melakukan perbuatan tersebut, diancam dengan pidana penjsra paling lama tujuh tahun.”

D.4 Pemalsuan Surat

(KUHP Pasal 263) : —~~Barang~~ siapa membuat surat palsu atau memalsukan surat yang dapat menimbulkan sesuatu hak, perikatan atau pembebasan hutang, atau yang diperuntukkan sebagai bukti daripada sesuatu hal dengan maksud untuk memakai atau menyuruh orang lain memakai surat tersebut seolah-olah isinya benar dan tidak dipalsu, diancam jika pemakaian tersebut dapat menimbulkan kerugian, karena pemalsuan surat, dengan pidana penjara paling lama enam tahun. Diancam dengan pidana yang sama, barang siapa dengan sengaja memakai surat palsu atau yang dipalsukan seolah-olah sejati, jika pemakaian surat itu dapat menimbulkan kerugian.”

D.5 Penipuan

(KUHP Pasal 378) : —~~Barang~~ siapa dengan maksud untuk menguntungkan diri sendiri atau orang lain secara melawan hukum, dengan memakai nama palsu atau martabat palsu, dengan tipu muslihat, ataupun rangkaian kebohongan, menggerakkan orang lain untuk menyerahkan barang sesuatu kepadanya, atau supaya memberi hutang rnaupun menghapuskan piutang diancam karena penipuan dengan pidana penjara paling lama empat tahun.”

D.6 Penggelapan

(KUHP Pasal 372) : —~~Barang~~ siapa dengan sengaja dan melawan hukum memiliki barang sesuatu yang seluruhnya atau sebagian adalah kepunyaan orang lain, tetapi yang ada dalam kekuasaannya bukan karena kejahatan diancam karena penggelapan, dengan pidana penjara paling lama empat tahun atau pidana denda paling banyak sembilan ratus rupiah.”

D.7 Sumpah Palsu

(KUHP Pasal 242) :

1. Barang siapa dalam keadaan di mana undang-undang menentukan supaya memberi keterangan di atas sumpah atau mengadakan akibat hukum kepada keterangan yang demikian, dengan sengaja memberi keterangan palsu di

atas sumpah, baik dengan lisan atau tulisan, secara pribadi maupun oleh kuasanya yang khusus ditunjuk untuk itu, diancam dengan pidana penjara paling lama tujuh tahun.

2. Jika keterangan palsu di atas sumpah diberikan dalam perkara pidana dan merugikan terdakwa atau tersangka, yang bersalah diancam dengan pidana penjara paling lama sembilan tahun.
3. Disamakan dengan sumpah adalah janji atau penguatan diharuskan menurut aturan-aturan umum atau yang menjadi pengganti sumpah. Pidana pencabutan hak berdasarkan pasal 35 No. 1 - 4 dapat dijatuhkan.

D.8 Utang Piutang

(KUHP Pasal 396) : –Seorang pengusaha yang dinyatakan dalam keadaan pailit atau yang diizinkan melepaskan budel oleh pengadilan, diancam karena merugikan pemiutang dengan pidana penjara paling lama satu tahun empat bulan:

1. jika pengeluarannya melewati batas;
2. jika yang bersangkutan dengan maksud untuk menanggukuhkan kepailitannya telah meminjam uang dengan syarat-syarat yang memberatkan sedang diketahuinya bahwa pinjaman itu tiada mencegah kepailitan; jika dia tak dapat memperlihatkan dalam keadaan tak diubah bukubuku dan surat- surat untuk catatan menurut pasal 6 Kitab Undangundang Hukum Dagang dan tulisan- tulisan yang harus disimpannya menurut pasal itu.”

D.9 Jabatan

(KUHP Pasal 415) : —Serang pejabat atau orang lain yang ditugaskan menjalankan suatu jabatan umum terus-menerus atau untuk sementara waktu, yang dengan sengaja menggelapkan uang atau surat berharga yang disimpan karena jabaimnya, atau membiarkan uang atau surat berharga itu diambil atau digelapkan oleh orang lain, atau menolong

D - 4

sebagai pembantu dalam melakukan perbuatan tersebut, diancam dengan pidana penjara paling lama tujuh tahun.”

D.10 Curat

(KUHP Pasal 363) :

(1) Diancam dengan pidana penjara paling lama tujuh tahun:

1. pencurian ternak;
2. pencurian pada waktu ada kebakaran, letusan, banjir gempa bumi, atau gempa laut, gunung meletus, kapal karam, kapal terdampar, kecelakaan kereta api, huru-hara, pemberontakan atau bahaya perang;
3. pencurian di waktu malam dalam sebuah rumah atau pekarangan tertutup yang ada rumahnya, yang dilakukan oleh orang yang ada di situ tidak diketahui atau tidak dikehendaki oleh yang berhak;
4. pencurian yang dilakukan oleh dua orang atau lebih;
5. pencurian yang untuk masuk ke tempat melakukan kejahatan, atau untuk sampai pada barang yang diambil, dilakukan dengan merusak, memotong atau memanjat, atau dengan memakai anak kunci palsu, perintah palsu atau pakaian jabatan palsu.

(2) Jika pencurian yang diterangkan dalam butir 3 disertai dengan salah satu hal dalam butir 4 dan 5, maka diancam dengan pidana penjara paling lama sembilan tahun.

D.11 Curas

(KUHP Pasal 365) :

(1) Diancam dengan pidana penjara paling lama sembilan tahun pencurian yang didahului, disertai atau diikuti dengan kekerasan atau ancaman kekerasan, terhadap orang dengan maksud untuk mempersiapkan atau mempermudah pencurian, atau dalam hal tertangkap tangan, untuk memungkinkan melarikan diri sendiri atau peserta lainnya, atau untuk tetap menguasai barang yang dicuri.

(2) Diancam dengan pidana penjara paling lama dua belas tahun:

6. jika perbuatan dilakukan pada waktu malam dalam sebuah rumah atau
7. pekarangan tertutup yang ada rumahnya, di berjalan;
8. jika perbuatan dilakukan oleh dua orang atau lebih dengan bersekutu;
9. jika masuk ke tempat melakukan kejahatan dengan merusak atau memanjat atau dengan memakai anak kunci palsu, periniah palsu atau pakaian jabatan palsu.

10. jika perbuatan mengakibatkan luka-luka berat.

(3) Jika perbuatan mengakibatkan kematian maka diancam dengan pidana penjara paling lama lima belas tahun.

Diancam dengan pidana mati atau pidana penjara seumur hidup atau selama waktu tertentu paling lama dua puluh tahun, jika perbuatan mengakhintkan luka berat atau kematian dan dilakukan oleh dua orang atau lebih dengan bersekutu, disertai pula oleh salah satu hal yang diterangkan dalam no. 1 dan 3.

D.12 Curanmor

(KUHP Pasal 362) : —Bang siapa mengambil suatu barang (kendaraan bermotor), yang seluruhnya atau sebagian kepunyaan orang lain, dengan maksud untuk memiliki (kendaraan bermotor) secara melawan hukum, diancam karena pencurian, dengan pidana penjara paling lama lima tahun atau pidana denda paling banyak sembilan ratus rupiah.”

D.13 Pembunuhan

(KUHP Pasal 339) : —Pembunuhan yang diikuti, disertai atau didahului oleh suatu perbuatan pidana, yang dilakukan dengan maksud untuk mempersiapkan atau mempermudah pelaksanaannya, atau untuk melepaskan diri sendiri maupun peserta lainnya dari pidana dalam hal tertangkap tangan, ataupun untuk memastikan penguasaan barang yang diperolehnya secara melawan hukum, diancam dengan pidana penjara seumur hidup atau selama waktu tertentu, paling lama dua puluh tahun.”

(KUHP Pasal 340) : —Barang siapa dengan sengaja dan dengan rencana terlebih dahulu merampas nyawa orang lain, diancam karena pembunuhan dengan rencana, dengan pidana rnati atau pidana penjara seumur hidup atau selama waktu tertentu, paling lama dua puluh tahun.”

D.14 Pemerasan/ Ancaman

(KUHP Pasal 368) :

(1) Barang siapa dengan maksud untuk menguntungkan diri sendiri atau orang lain secara melawan hukum, memaksa seorang dengan kekerasan atau ancaman kekerasan untuk memberikan barang sesuatu, yang seluruhnya atau sebagian adalah kepunyaan orang itu atau orang lain, atau supaya membuat hutang maupun menghapuskan piutang, diancam karena pemerasan dengan pidana penjara paling lama sembilan bulan.

(2) Ketentuan pasal 365 ayat kedua, ketiga, dan keempat berlaku bagi kejahatan ini.

D.15 Perjudian

(KUHP Pasal 303) :

(1) Diancam dengan pidana penjara paling lama sepuluh tahun atau pidana denda paling banyak dua puluh lima juta rupiah, barang siapa tanpa mendapat izin:

1. dengan sengaja menawarkan atau memberikan kesempatan untuk permainan judi dan menjadikannya sebagai pen- carian, atau dengan sengaja turut serta dalam suatu perusahaan untuk itu;
2. dengan sengaja menawarkan atau memberi kesempatan kepada khalayak umum untuk bermain judi atau dengan sengaja turut serta dalam perusahaan untuk itu, dengan tidak peduli apakah untuk menggunakan kesempatan adanya sesuatu syarat atau dipenuhinya sesuatu tata-cara;
3. turut serta menjadikan permainan judi sebagai mata pencaharian

(2) Kalau yang bersalah melakukan kejahatan tersebut dalam menjalankan pencariannya, maka dapat dicabut hak nya untuk menjalankan pencarian itu.

(3) Yang disebut permainan judi adalah tiap-tiap permainan, di mana pada umumnya kemungkinan mendapat untung bergantung pada peruntungan belaka, juga karena pemainnya lebih terlatih atau lebih mahir. Di situ termasuk segala pertarungan tentang keputusan perlombaan atau permainanlain-lainnya yang tidak diadakan antara mereka yang turut berlomba atau bermain, demikian juga segala pertarungan lainnya.

D.16 Merusak Barang

(KUHP Pasal 406) :

- (1) Barangsiapa dengan sengaja dan secara melawan hukum menghancurkan, merusak, membuat tak dapat dipakai atau menghilangkan barang sesuatu yang seluruhnya atau sebagian milik orang lain, diancam dengan pidana penjara paling lama dua tahun delapan bulan atau pidana denda paling banyak empat ribu lima ratus rupiah.
- (2) Diancam dengan pidana yang sama orang yang dengan sengaja dan secara melawan hukum membunuh, merusakkan, membuat tak dapat digunakan atau menghilangkan hewan, yang seluruhnya atau sebagian milik orang lain.

D.17 Penculikan

(KUHP Pasal 328) : ~~Barang~~ siapa membawa pergi seorang dari tempat kediamannya atau tempat tinggalnya sementara dengan maksud untuk menempatkan orang itu secara melawan hukum di bawah kekuasaannya atau kekuasaan orang lain, atau untuk menempatkan dia dalam keadaan sengsara, diancam karena penculikan dengan pidana penjara paling lama dua belas tahun.”

D.18 Penghinaan

(KUHP Pasal 310) :

(1) Barang siapa sengaja menyerang kehormatan atau nama baik seseorang dengan menuduhkan sesuatu hal, yang maksudnya terang supaya hal itu diketahui umum, diancam karena pencemaran dengan pidana penjara paling lama sembilan bulan atau pidana denda paling banyak empat ribu lima ratus rupiah.

(2) Jika hal itu dilakukan dengan tulisan atau gambaran yang disiarkan, dipertunjukkan atau ditempelkan di muka umum, maka diancam karena pencemaran tertulis dengan pidana penjara paling lama satu tahun empat bulan atau pidana denda paling banyak empat ribu lima ratus rupiah.

(3) Tidak merupakan pencemaran atau pencemaran tertulis, jika perbuatan jelas dilakukan demi kepentingan umum atau karena terpaksa untuk membela diri.

D.19 Aniaya Ringan

(KUHP Pasal 352) :

(1) Kecuali yang tersebut dalam pasal 353 dan 356, maka penganiayaan yang tidak menimbulkan penyakit atau halangan untuk menjalankan pekerjaan jabatan atau pencarian, diancam, sebagai penganiayaan ringan, dengan pidana penjara paling lama tiga bulan atau pidana denda paling banyak empat ribu lima ratus rupiah. Pidana dapat ditambah sepertiga bagi orang yang melakukan kejahatan itu terhadap orang yang bekerja padanya, atau menjadi bawahannya.

(2) Percobaan untuk melakukan kejahatan ini tidak dipidana.

D.20 Akibat Orang Luka

(KUHP Pasal 359) : —Barang siapa karena kesalahannya (kealpaannya) menyebabkan orang lain mati, diancam dengan pidana penjara paling lama lima tahun atau pidana kurungan paling lama satu tahun.”

(KUHP Pasal 360) :

(1) Barang siapa karena kesalahannya (kealpaannya) menyebabkan orang lain mendapat luka-luka berat, diancam dengan pidana penjara paling lama lima tahun atau pidana kurungan paling lama satu tahun.

(2) Barang siapa karena kesalahannya (kealpaannya) menyebabkan orang lain luka-luka sedemikian rupa sehingga timbul penyakit atau halangan menjalankan pekerjaan jabatan atau pencarian selama waktu tertentu, diancam dengan pidana penjara paling lama sembilan bulan atau pidana kurungan paling lama enam bulan atau pidana denda paling tinggi empat ribu lima ratus rupiah.

D.21 Pencurian Ringan

(KUHP Pasal 362) : —Barang siapa mengambil barang sesuatu, yang seluruhnya atau sebagian kepunyaan orang lain, dengan maksud untuk dimiliki secara melawan hukum, diancam karena pencurian, dengan pidana penjara paling lama lima tahun atau pidana denda paling banyak sembilan ratus rupiah.”

D.22 Penadahan

(KUHP Pasal 480) : —Diancam dengan pidana penjara paling lama empat tahun atau pidana denda paling banyak sembilan ratus rupiah:

1. barang siapa membeli, menyewa, menukar, menerima gadai, menerima hadiah, atau untuk menarik keuntungan, menjual, menyewakan, menukarkan, menggadaikan, mengangkut, menyimpan atau menyembunyikan sesuatu benda, yang diketahui atau sepatutnya harus diduga bahwa diperoleh dari kejahatan penadahan;
2. barang siapa menarik keuntungan dari hasil sesuatu benda, yang diketahuinya atau sepatutnya harus diduga bahwa diperoleh dari kejahatan.”

D.23 Menerima Suap

(KUHP Pasal 418) : —Seorang pejabat yang menerima hadiah atau janji padahal diketahui atau sepatutnya harus diduganya, bahwa hadiah atau janji itu diberikan karena kekuasaan atau kewenangan yang berhubungan dengan jabatannya, atau yang menurut pikiran orang yang memberi hadiah atau janji itu ada hubungan dengan jabatannya diancam dengan pidana penjara paling lama enam tahun atau pidana denda paling banyak empat ribu lima ratus rupiah.”

D.24 Penganiayaan

(KUHP Pasal 351) :

- (1) Penganiayaan diancam dengan pidana penjara paling lama dua tahun delapan bulan atau pidana denda paling banyak empat ribu lima ratus rupiah,
- (2) Jika perbuatan mengakibatkan luka-luka berat, yang bersalah diancam dengan pidana penjara paling lama lima tahun.
- (3) Jika mengakibatkan mati, diancam dengan pidana penjara paling lama tujuh tahun.
- (4) Dengan penganiayaan disamakan sengaja merusak kesehatan.
- (5) Percobaan untuk melakukan kejahatan ini tidak dipidana.

D.25 Perkosaan

(KUHP Pasal 285) : —Barang siapa dengan kekerasan atau ancaman kekerasan memaksa seorang wanita bersetubuh dengan dia di luar perkawinan, diancam karena melakukan perkosaan dengan pidana penjara paling lama dua belas tahun.”

D.26 Kesusilaan

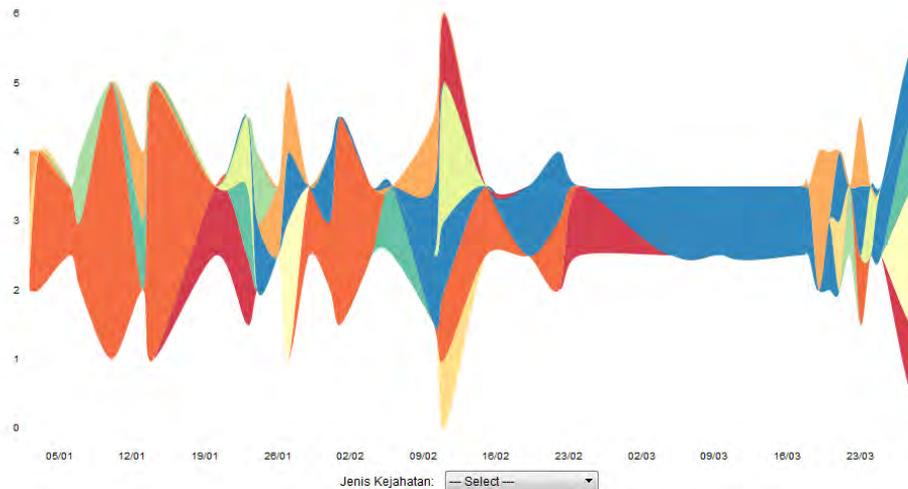
(KUHP Pasal 281) : —Diancam dengan pidana penjara paling lama dua tahun delapan bulan atau pidana denda paling banyak empat ribu lima ratus rupiah:

1. barang siapa dengan sengaja dan terbuka melanggar kesusilaan;
2. barang siapa dengan sengaja dan di depan orang lain yang ada di situ bertentangan dengan kehendaknya, melanggar kesusilaan.”

LAMPIRAN E

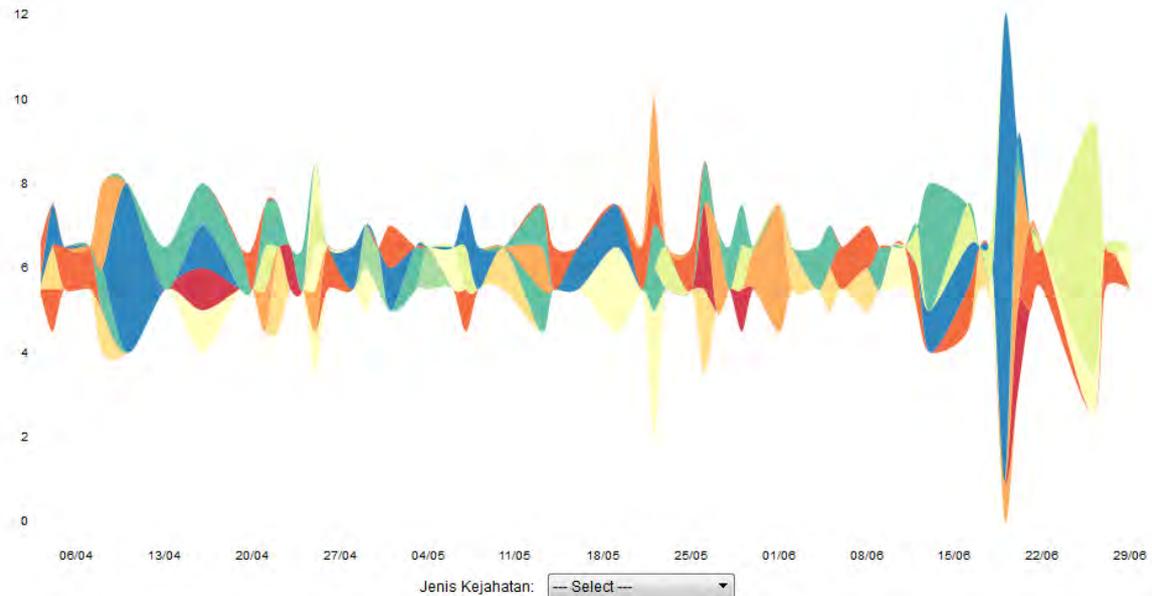
STREAMGRAPH LAPORAN KRIMINALITAS

E.1 2014 - Triwulan 1 (Januari – Maret 2014)



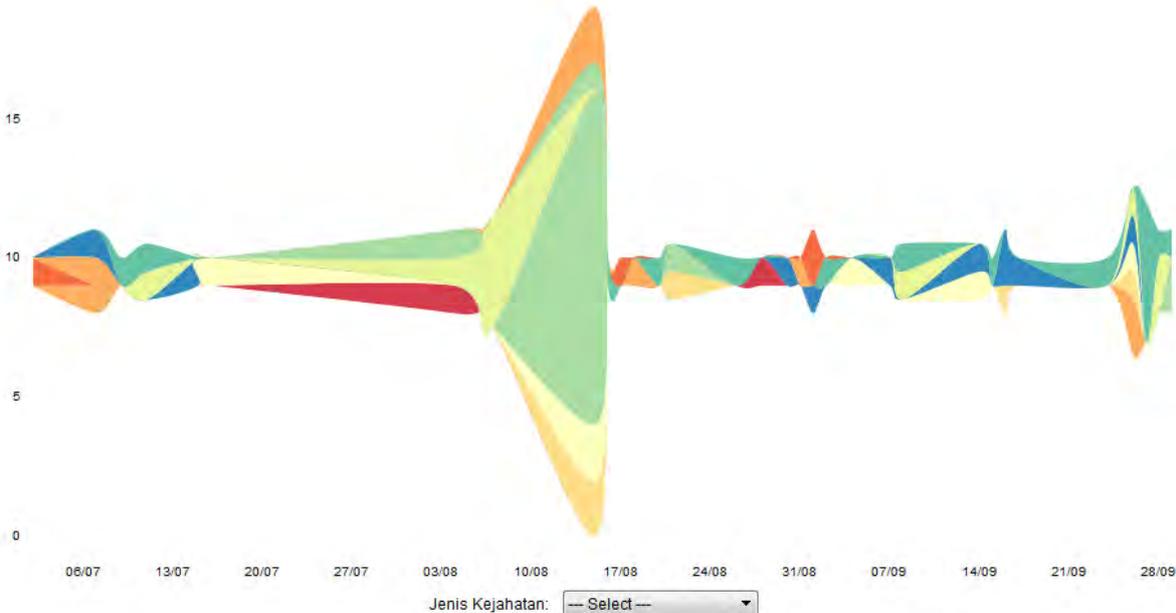
Gambar E.1 *Streamgraph* 2014 – Triwulan 1

E.2 2014 - Triwulan 2 (April – Juni 2014)



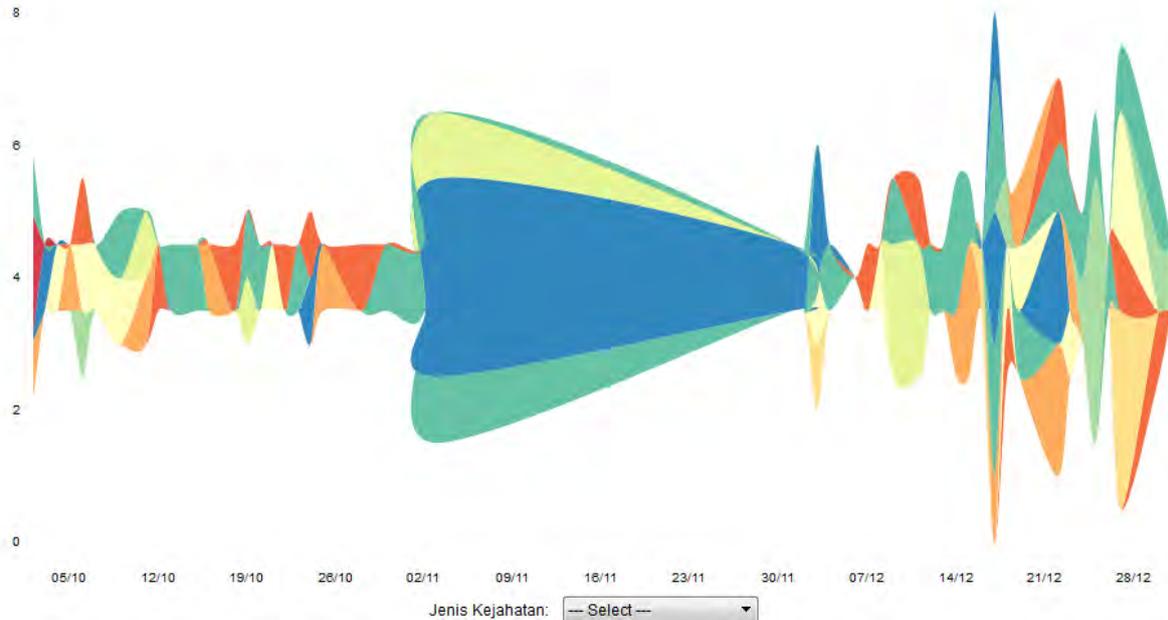
Gambar E.2 Streamgraph 2014 – Triwulan 2

E.3 2014 - Triwulan 3 (Juli – September 2014)



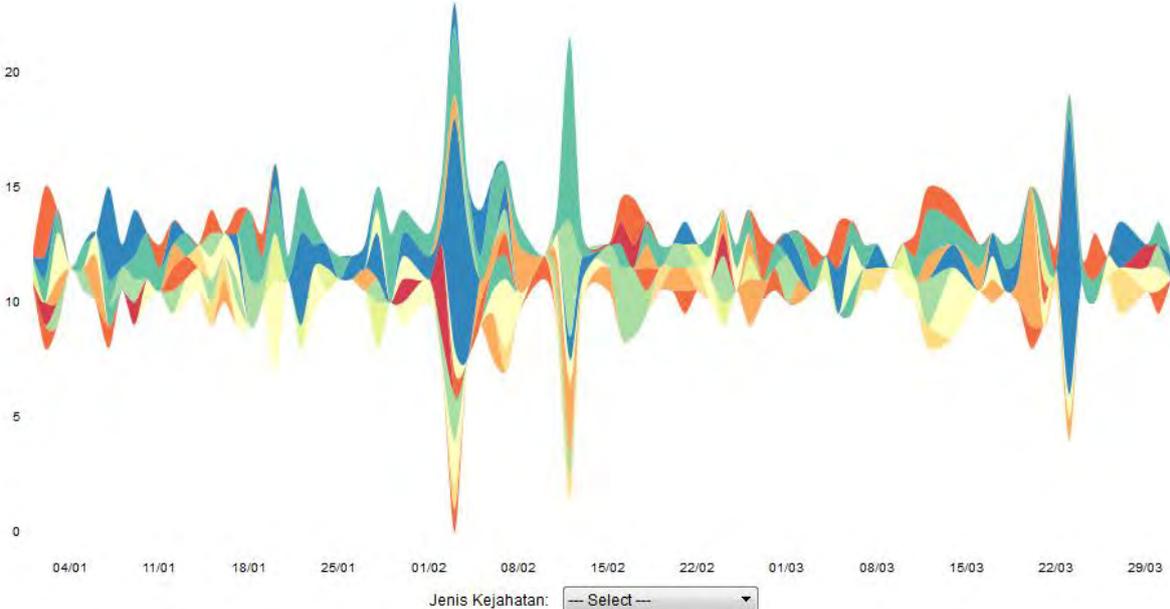
Gambar E.3 Streamgraph 2014 – Triwulan 3

E.4 2014 - Triwulan 4 (Oktober – Desember 2014)



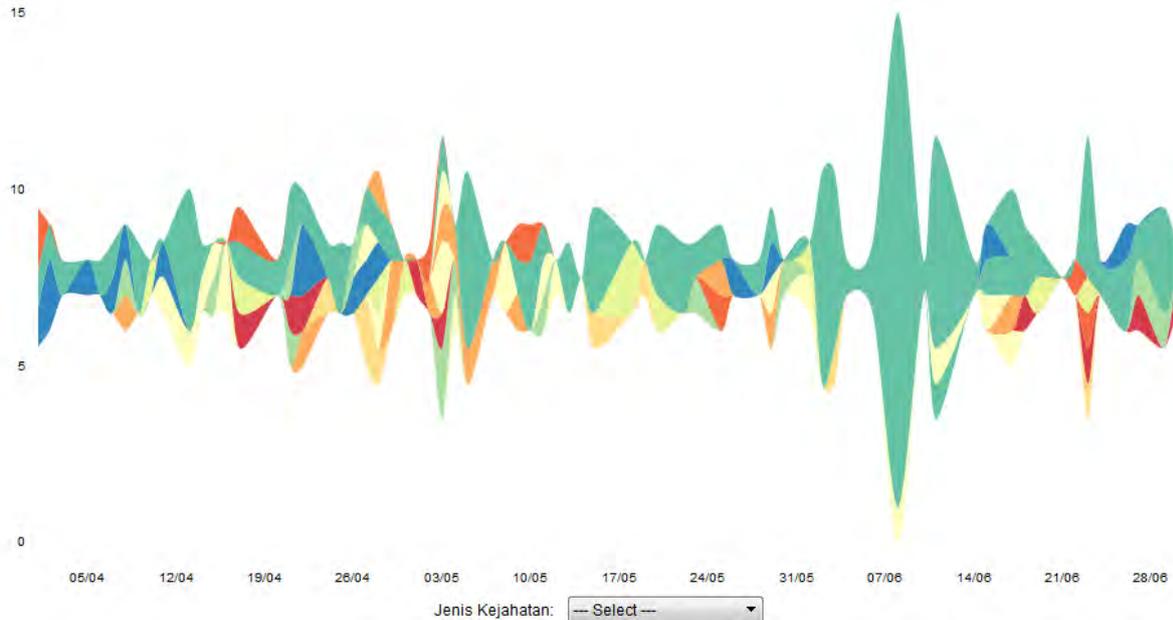
Gambar E.4 Streamgraph 2014 – Triwulan 4

E.5 2015 - Triwulan 1 (Januari – Maret 2015)



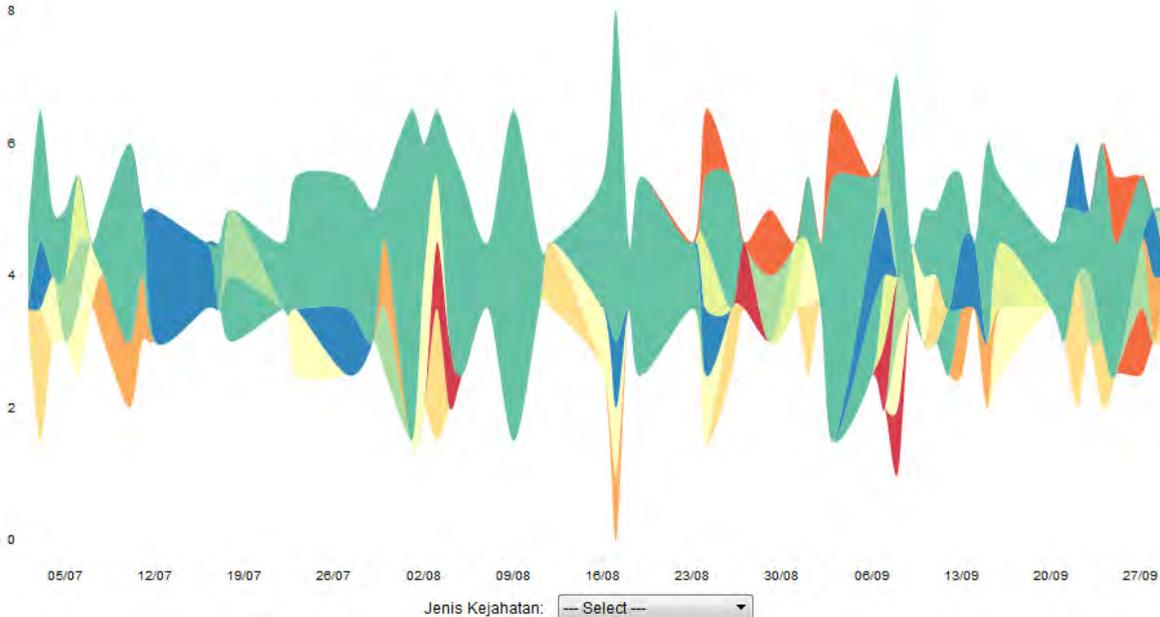
Gambar E.5 Streamgraph 2015 – Triwulan 1

E.6 2015 - Triwulan 2 (April – Juni 2015)



Gambar E.6 Streamgraph 2015 – Triwulan 2

E.7 2015 - Triwulan 3 (Juli – September 2015)



Gambar E.7 Streamgraph 2015 – Triwulan 3

BAB VII

KESIMPULAN DAN SARAN

Bab ini berisi kesimpulan dan saran terkait pengerjaan tugas akhir. Pada bagian kesimpulan akan disimpulkan bagaimana hasil dari penelitian tugas akhir yang telah dilakukan. Sedangkan pada bagian saran berisi rekomendasi perbaikan yang berguna untuk pengembangan penelitian kedepannya.

7.1. Kesimpulan

Berdasarkan proses – proses pengerjaan tugas akhir yang telah dilakukan, dihasilkan beberapa kesimpulan yang dapat diambil, diantaranya:

1. Kategori 5 subdit lebih baik digunakan pada data penelitian ini daripada kategori 30 jenis kejahatan. Semua model pada percobaan subdit dapat menjadi pengklasifikasi walaupun dengan performa yang tidak semuanya baik. Sedangkan untuk model dengan 30 kategori, tidak semua dapat menjadi pengklasifikasi karena keterbatasan data.
2. Model klasifikasi terbaik didapat setelah melakukan 7 percobaan pengembangan algoritma klasifikasi. Pada percobaan 7, algoritma klasifikasi dibentuk model bersusun dengan melakukan proses klasifikasi sebanyak 29 kali dan 4 kali pada percobaan subdit. Model klasifikasi bekerja akurat pada kelas penyupaan dengan nilai $f - measure$ 100%. Disusul dengan kelas perjudian, penipuan, pengerusakan, korupsi, dan narkoba dengan nilai $f - measure$ lebih dari 80%. Sedangkan untuk kelas pencurian ringan,

lantas, perkosaan, penganiayaan, penghinaan, curas, curat, jabatan, penggelapan, pemerasan, curanmor, dan implikasi kontijensi memiliki nilai $f - measure$ kurang dari 80%. Sementara untuk 12 kelas lainnya tidak mampu untuk melakukan proses klasifikasi. Hal ini dikarenakan jumlah data yang digunakan sangat terbatas yaitu tidak lebih dari 50 data setiap kelas. Sedangkan pada percobaan subdit, performa yang dihasilkan yaitu subdit II dan II mampu melakukan kasifikasi dengan nilai $f - measure$ masing – masing 85.5% dan 85.4%. Disusul dengan model subdit V sebesar 77.7% dan subdit I sebesar 66.6%. Sedangkan model IV hanya mampu mengklasifikasi model dengan $f - measure$ 14.2%.

3. Visualisasi yang dibentuk pada penelitian ini berupa grafik *streamgraph* dan *wordcloud* yang ditampilkan pada aplikasi berbasis web yang interakhif yaitu *shiny*. Visualisasi *streamgraph* menunjukkan jumlah laporan kriminalitas untuk setiap jenis dari periode ke periode dalam bentuk grafik alir. Pada tahun 2014 jenis laporan yang disampaikan cenderung tidak merata dan selalu ada jenis laporan yang mendominasi. Berbeda dengan tahun 2015, seiring semakin banyaknya data yang tersampaikan, jenis laporan yang disampaikan cenderung merata dengan perbedaan jumlah yang tidak signifikan. Sementara visualisasi *wordcloud* menampilkan kata – kata yang terepresentasi pada setiap kategori kejahatan. Kata – kata ini diujicobakan sebagai *dictionary* pada pembobotan di tahap pencarian model klasifikasi terbaik.

7.2. Saran

Dari pengerjaan tugas akhir ini masih terdapat beberapa hal yang perlu diperbaiki. Oleh karena itu, berikut beberapa saran yang dapat dipertimbangan untuk pengembangan kedepan yang lebih baik.

1. Mempertimbangkan untuk menggunakan metode klasifikasi lainnya agar dapat membandingkan metode mana yang mampu bekerja optimal dan memberikan akurasi terbaik.
2. Jumlah data perlu ditambah agar model dapat belajar lebih baik dan akurasi yang dihasilkan juga meningkat.
3. Penambahan jenis fitur yang digunakan, misalnya dengan menggunakan fitur *stemming* agar kualitas data yang digunakan semakin baik.

DAFTAR PUSTAKA

- [1] LAPOR! (2015, September) Tentang LAPOR! [Online].
https://www.lapor.go.id/lapor/tentang_lapor
- [2] Chyntia Megawati, "ANALISIS ASPIRASI DAN PENGADUAN DI SITUS LAPOR! ," Universitas Indonesia, Depok, Skripsi 2015.
- [3] Edd Dumbill. (2012, Januari) Defining Big Data. [Online].
<http://www.forbes.com/sites/edddumbill/2014/05/07/%20defining-big-data/>
- [4] Kartini Kartono, *Patologi Sosial*. Jakarta: Raja Grafindo Persada, 1999.
- [5] Sulthon Maarif, "Clustering & Visualisasi Geografi tentang Tingkat Kriminalitas Konvensional Dengan Metode Algoritma K-Means Menggunakan R Di Provinsi Jawa Timur (Studi Kasus : Polda Jatim)," ITS, Surabaya, Thesis 2015.
- [6] Light , Donald , Suzanne Keller, and Craig Calhoun, *Sociology*, 5th ed. New York: Alfred A. Knopf, 1989.
- [7] Republik Indonesia, *Peraturan Kepala Kepolisian Negara Republik Indonesia Nomor 22 Tahun 2010 Tentang Susunan Organisasi dan Tata Kerja Pada Tingkat Kepolisian Daerah*. Jakarta: Kepolisian Negara Republik Indonesia, 2010.
- [8] Mohammad Kusnarto. (2013) KEJAHATAN | SOSIOLOGI SEBAGAI PELETAK DASAR TOLERANSI ANTAR UMAT BERAGAMA. [Online].
<https://mohkusnarto.wordpress.com/kejahatan/>
- [9] R Soesilo, *KITAB UNDANG-UNDANG HUKUM PIDANA (KUHP)*. Bogor: Politea, 1976.
- [10] R. Soesilo, *KITAB UNDANG-UNDANG HUKUM*

PIDANA (KUHP) Pasal 1 Ayat 4. Bogor: Politea, 1976.

- [11] Yuriy A. Voronin. (2000) National Criminal Justice Reference Service. [Online]. <https://www.ncjrs.gov/App/Publications/abstract.aspx?ID=184773>
- [12] Bahriatul Jannah, Mahmudah Marzuki, Ninik Annisa, Nor Hiqmah, and Sri Suparti, "Studi Efektifitas dan Responsivitas Pelayanan LAPOR! ", PIRAC, Depok, Laporan Penelitian 2014.
- [13] R Feldman and J Sanger, *The Text Mining Handbook : Advanced Approaches in Analyzing Unstructured Data*. New York: Cambridge University Press, 2007.
- [14] H Jiawei, M Kamber, and J Pei, *Data Mining: Concepts and Techniques*. Waltham: MK: Morgan Kaufmann, 2012.
- [15] G Miner et al., *Practical Text Mining and Statistical Analysis for Non-Structured Text Data*. Oxford: Elsevier, 2012.
- [16] Statsoft. (2015) Text Mining Introductory Overview. [Online]. <http://www.statsoft.com/textbook/text-mining>
- [17] T McEnery and A Wilson, *Corpus Linguistics*, 2nd ed.: Edinburgh University Press, 2001.
- [18] Ghiyats Syafiq, "Data Mining dan Knowledge Discovery in Database," Institut Sains Terapan dan Teknologi, Surabaya, 2010.
- [19] C Zhai and C Aggarwal, *Mining Text Data*. New York: Springer, 2012.
- [20] G. S Linoff and M.J. Berry, *Data Mining Techniques: For Marketing Sales, and Customer Relationship Management Third Edition*. Indianapolis: IN: Wiley Publishing, Inc, 2011.
- [21] Sun Microsystems USA. (2015, Desember) JavaMail TM API Design Specification.
- [22] Muhamad Rachli, "Email Filtering menggunakan Naive

- Bayesian," Jurusan Teknik Elektro Institut Teknologi Bandung, Bandung, Tugas Akhir 2007.
- [23] Aida Indriani, "Klasifikasi Data Forum dengan menggunakan Metode Naïve Bayes Classifier," Seminar Nasional Aplikasi Teknologi Informasi (SNATI), Yogyakarta, Paper ISSN: 1907 - 5022, 2014.
- [24] Prasetyo Anugroho, Idris Winarno, and Nur Rosyid, "KLASIFIKASI EMAIL SPAM DENGAN METODE NAIVE BAYES CLASSIFIER MENGGUNAKAN JAVA PROGRAMMING," Politeknik Elektronika Negeri Surabaya, Surabaya, Tugas Akhir.
- [25] Peter Flach, *Machine Learning: The Art and Science of Algorithms that Make Sense of Data.:* Cambridge University Press, 2012.
- [26] Ingo Feinerer and Kurt Hornik. (2015, July) CRAN - package `tm`. [Online]. <https://cran.r-project.org/web/packages/tm/tm.pdf>
- [27] David Meyer and Evgenia Dimitriadou. (2015, August) CRAN - package `e1071`. [Online]. <https://cran.r-project.org/web/packages/e1071/e1071.pdf>
- [28] Max Kuhn and Steve Weston. (2016, April) CRAN - package `caret`. [Online]. <https://cran.r-project.org/web/packages/caret/caret.pdf>
- [29] Ian Fellows. (2015, February) CRAN - `wordcloud`. [Online]. <https://cran.r-project.org/web/packages/wordcloud/wordcloud.pdf>
- [30] Contributor Code of GitHub. (2015, December) GitHub - `htmlwidget` for creating streamgraph visualizations in R. [Online]. <https://github.com/hrbrmstr/streamgraph>
- [31] Hadley Wickham and Romain Francois. (2015, September) CRAN - package `dplyr`. [Online]. <https://cran.r-project.org/web/packages/dplyr/dplyr.pdf>
- [32] Winston Chang and Joe Cheng. (2016, March) CRAN - package `shiny`. [Online]. <https://cran.r->

project.org/web/packages/shiny/shiny.pdf

- [33] G. R., & Barnard, E Botha, "Factors that affect the accuracy of text-based," *Computer Speech and Language*, vol. 26, pp. 307 - 320, 2012.
- [34] Z Xiang, Z Schwartz, J Gerdes Jr, and M Uysal, "What can big data and text analytics tell us about hotel guestexperience and satisfaction?," *International Journal of Hospitality Management*, vol. 44, pp. 120-130, 2015.

BIODATA PENULIS



Rizka Amalia Putri dilahirkan di Kediri pada tanggal 2 Januari 1994 yang merupakan anak pertama dari dua bersaudara. Penulis telah menempuh pendidikan formal dari tingkat dasar sampai tingkat atas, yaitu SD Jember Lor 1 Jember, SMP Negeri 1 Jember, dan SMA Negeri 1 Jember.

Pasca kelulusan penulis pada jenjang SMA pada tahun 2012, penulis melanjutkan pendidikan ke jenjang perguruan tinggi di Jurusan Sistem Informasi Fakultas Teknologi Informasi Institut Teknologi Sepuluh Nopember Surabaya dan terdaftar sebagai mahasiswa dengan NRP 5212100072.

Selama menjadi mahasiswa, penulis aktif dalam kegiatan kemahasiswaan dan mengikuti beberapa kepanitiaan di tingkat jurusan hingga institusi. Hal ini dibuktikan penulis dengan menjadi menjadi Staf, Asisten Dirjen, dan Sekretaris Kementerian di Kementerian Perekonomian Badan Eksekutif Mahasiswa ITS selama tiga tahun kepengurusan. Selain itu, penulis juga menyempatkan diri menjadi asisten dosen di beberapa matakuliah. Penulis juga pernah melaksanakan kerja praktik di perusahaan minyak dan gas pada PT Pertamina Marketing Operation Region IV Semarang selama 2 bulan pada tahun 2015.

Di akhir tahun perkuliahannya, penulis mengambil konsentrasi pada bidang minat Akuisisi Data dan Diseminasi Informasi dengan topik tugas akhir di bidang *text mining* dan visualisasi. Untuk keperluan penelitian, penulis dapat dihubungi melalui e-mail : riz.amaliaputri@gmail.com