

**DETEKSI KOMENTAR SPAM BAHASA INDONESIA PADA
INSTAGRAM MENGGUNAKAN METODE NAIVE BAYES
DENGAN ENHANCED SMOOTHING**

Skripsi



oleh
SAMUEL ANDRY GUNAWAN
71120130

PROGRAM STUDI INFORMATIKA FAKULTAS TEKNOLOGI INFORMASI
UNIVERSITAS KRISTEN DUTA WACANA
2019

**DETEKSI KOMENTAR SPAM BAHASA INDONESIA PADA
INSTAGRAM MENGGUNAKAN METODE NAIVE BAYES
DENGAN ENHANCED SMOOTHING**

Skripsi



Diajukan kepada Program Studi Informatika Fakultas Teknologi Informasi
Universitas Kristen Duta Wacana
Sebagai Salah Satu Syarat dalam Memperoleh Gelar
Sarjana Komputer

Disusun oleh

SAMUEL ANDRY GUNAWAN
71120130

PROGRAM STUDI INFORMATIKA FAKULTAS TEKNOLOGI INFORMASI
UNIVERSITAS KRISTEN DUTA WACANA
2019

PERNYATAAN KEASLIAN SKRIPSI

Saya menyatakan dengan sesungguhnya bahwa skripsi dengan judul:

DETEKSI KOMENTAR SPAM BAHASA INDONESIA PADA INSTAGRAM MENGGUNAKAN METODE NAIVE BAYES DENGAN ENHANCED SMOOTHING

yang saya kerjakan untuk melengkapi sebagian persyaratan menjadi Sarjana Komputer pada pendidikan Sarjana Program Studi Informatika Fakultas Teknologi Informasi Universitas Kristen Duta Wacana, bukan merupakan tiruan atau duplikasi dari skripsi kesarjanaan di lingkungan Universitas Kristen Duta Wacana maupun di Perguruan Tinggi atau instansi manapun, kecuali bagian yang sumber informasinya dicantumkan sebagaimana mestinya.

Jika dikemudian hari didapati bahwa hasil skripsi ini adalah hasil plagiasi atau tiruan dari skripsi lain, saya bersedia dikenai sanksi yakni pencabutan gelar kesarjanaan saya.

Yogyakarta, 28 Juni 2019



SAMUEL ANDRY GUNAWAN
71120130

HALAMAN PERSETUJUAN

Judul Skripsi : DETEKSI KOMENTAR SPAM BAHASA
INDONESIA PADA INSTAGRAM
MENGUNAKAN METODE NAIVE BAYES
DENGAN ENHANCED SMOOTHING

Nama Mahasiswa : SAMUEL ANDRY GUNAWAN

N I M : 71120130

Matakuliah : Skripsi (Tugas Akhir)

Kode : TIW276

Semester : Genap

Tahun Akademik : 2018/2019

Telah diperiksa dan disetujui di
Yogyakarta,
Pada tanggal 28 Juni 2019

Dosen Pembimbing I



Antonius Rachmat C., S.Kom., M.Cs.

Dosen Pembimbing II



Yuan Lukito, S.Kom., M.Cs.

HALAMAN PENGESAHAN

DETEKSI KOMENTAR SPAM BAHASA INDONESIA PADA INSTAGRAM MENGGUNAKAN METODE NAIVE BAYES DENGAN ENHANCED SMOOTHING

Oleh: SAMUEL ANDRY GUNAWAN / 71120130

Dipertahankan di depan Dewan Penguji Skripsi
Program Studi Informatika Fakultas Teknologi Informasi
Universitas Kristen Duta Wacana - Yogyakarta
Dan dinyatakan diterima untuk memenuhi salah satu syarat memperoleh gelar
Sarjana Komputer
pada tanggal 14 Juni 2019

Yogyakarta, 28 Juni 2019
Mengesahkan,

Dewan Penguji:

1. Antonius Rachmat C., S.Kom.,M.Cs.
2. Yuan Lukito, S.Kom., M.Cs.
3. Willy Sudiarto Raharjo, S.Kom.,M.Cs.
4. Restyandito, S.Kom.,MSIS, Ph.D



Dekan

(Budi Susanto, S.Kom., M.T.)

Ketua Program Studi

(Gloria Virginia, Ph.D.)

HALAMAN UCAPAN TERIMA KASIH

Puji syukur kepada Tuhan Yang Maha Esa atas berkat, anugerah, dan kasih-Nya sehingga tugas akhir yang berjudul *Deteksi Komentar Spam Bahasa Indonesia pada Instagram Menggunakan Metode Naive Bayes dengan Enhanced Smoothing* ini dapat terselesaikan dengan baik.

Penelitian tugas akhir ini dilakukan dalam rangka untuk memenuhi salah satu syarat yang diperlukan untuk memperoleh gelar sarjana (S1) pada Program Studi Informatika Fakultas Teknologi Informasi Universitas Kristen Duta Wacana.

Dalam proses pengerjaan dan penyelesaian tugas akhir ini, penulis telah menerima bimbingan dan dukungan dari berbagai, untuk itu, penulis mengucapkan terima kasih kepada :

1. Bapak Antonius Rachmat C., S.Kom., M.Cs., selaku Dosen Pembimbing I dan Bapak Yuan Lukito, S.Kom., M.Cs., selaku Dosen Pembimbing II yang telah memberikan banyak arahan, bimbingan, dan dukungan selama penyusunan dan penyelesaian tugas akhir ini.
2. Bapak Willy Sudioarto Raharjo, S.Kom., M.Cs., selaku anggota dewan penguji dan Bapak Restyandito, S.Kom., MSIS, Ph.D, selaku anggota dewan penguji atas segala saran, masukan, dan kritikan yang diberikan dalam proses penyempurnaan tugas akhir ini.
3. Bapak Budi Susanto, S.Kom., M.T., selaku Dekan Fakultas Teknologi Informasi, Ibu Gloria Virginia, Ph.D., selaku Ketua Program Studi Informatika UKDW, beserta seluruh dosen Program Studi Informatika UKDW yang telah menyampaikan ilmu pengetahuannya selama proses perkuliahan berlangsung.
4. Semua teman, kerabat, dan saudara yang telah mendukung selama proses perkuliahan dan penyelesaian tugas akhir ini.

Yogyakarta, 28 Juni 2019

Penulis

KATA PENGANTAR

Puji syukur kepada Tuhan Yang Maha Esa, karena atas berkat dan anugerahNya sehingga penulis mampu menyelesaikan tugas akhir penelitian yang berjudul *Deteksi Komentar Spam Bahasa Indonesia pada Instagram Menggunakan Metode Naive Bayes dengan Enhanced Smoothing*. Penelitian ini dilakukan sebagai salah satu syarat untuk mendapatkan gelar sarjana informatika.

Dalam proses pengerjaan dan penyelesaian tugas akhir ini, penulis telah mendapatkan begitu banyak arahan, bimbingan, dan bantuan dari berbagai belah pihak yang mendukung agar proses penelitian ini bisa berjalan dengan baik dan terselesaikan. Untuk itu penulis mengucapkan terima kasih banyak atas dukungannya.

Adapun hasil penelitian yang dilakukan beserta dengan penulisan laporan ini masih jauh dari kata sempurna. Oleh karena itu, penulis berharap agar bisa menerima kritikan yang diberikan atau saran yang membangun ke depannya, sehingga bisa menghasilkan karya berikutnya yang lebih baik lagi.

Akhir kata, penulis memohon maaf kepada pembaca apabila terdapat penggunaan kata atau kalimat yang masih kurang tepat atau kurang berkenan. Harapan penulis adalah agar hasil penelitian ini bisa memberikan manfaat kepada banyak pihak dan bagi kita sekalian. Terima kasih.

Yogyakarta, 28 Juni 2019

Penulis

INTISARI

Instagram sebagai salah satu media sosial yang paling banyak digunakan di Indonesia merupakan wadah komunikasi digital terkini. Berbagai kalangan masyarakat seperti musisi dan artis memanfaatkan teknologi ini untuk berinteraksi dengan para pengikutnya. Namun menjamurnya komentar *spam* di setiap kiriman yang ada sangat mengganggu proses komunikasi yang sedang terjadi dan menyita waktu banyak orang. Chrismanto dan Lukito (2017) dalam penelitiannya yang berjudul *Deteksi Komentar Spam Bahasa Indonesia Pada Instagram Menggunakan Naive Bayes* telah membuktikan bahwa metode tersebut mampu menghasilkan tingkat akurasi sebesar 77,25%.

Berdasarkan hasil yang didapat, dilakukan penelitian menggunakan data set yang sama untuk melihat hasil perbandingan tingkat akurasi yang didapat menggunakan metode Naive Bayes biasa terhadap metode Naive Bayes dengan *Enhanced Smoothing*. Teknik *smoothing* yang dilakukan akan menghilangkan kemungkinan *zero probability* yang ada dan berbagai macam prosedur yang dilakukan pada tahap preproses, pemilihan fitur, dan klasifikasi akan menghasilkan nilai akurasi yang lebih baik. Metode klasifikasi yang digunakan dalam penelitian ini mencakup Metode *Naive Bayes*, *Naive Bayes Laplace Smoothing*, *Naive Bayes Jelinek-Mercer Smoothing*, *Naive Bayes Jelinek-Mercer Enhanced Smoothing*, *Naive Bayes Dirichlet Smoothing*, dan *Naive Bayes Dirichlet Enhanced Smoothing*.

Hasil dari penelitian ini pun membuktikan bahwa metode Naive Bayes dengan *Enhanced Smoothing* mampu menghasilkan kenaikan tingkat akurasi sebesar 5% - 10% dengan nilai akurasi tertinggi yang didapat adalah sebesar 90,55%.

Kata Kunci— Naive Bayes, Naive Bayes Enhanced Smoothing, spam comment decection, Instagram, Indonesian spam detection

DAFTAR ISI

PERNYATAAN KEASLIAN TUGAS AKHIR	iii
HALAMAN PERSETUJUAN	iv
HALAMAN PENGESAHAN	v
HALAMAN UCAPAN TERIMA KASIH	vi
KATA PENGANTAR	vii
INTISARI.....	viii
DAFTAR ISI	ix
DAFTAR GAMBAR	xii
DAFTAR TABEL	xv
DAFTAR LAMPIRAN	xvi
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Batasan Masalah	2
1.4 Tujuan Penelitian	3
1.5 Metode Penelitian/Pendekatan.....	3
1.6 Sistematika Penulisan	4
BAB II LANDASAN TEOR.....	5
2.1 Tinjauan Pustaka	5
2.2 Landasan Teori	6
2.2.1 Komentar Spam	6
2.2.2 Instagram	8
2.2.3 Text Mining	8
2.2.4 Text Preprocessing	9
2.2.5 Term Frequency-Inverse Document Frequency (TF-IDF)	11
2.2.6 Metode Naive Bayes dengan Enhanced Smoothing	12
2.2.7 Confusion Matrix	19
BAB III PERANCANGAN SISTEM	20
3.1 Analisis Kebutuhan	20

3.1.1 Analisis Kebutuhan Perangkat Keras	20
3.1.2 Analisis Perangkat Lunak	20
3.1.3 Analisis Fungsionalitas	20
3.1.4 Analisis Data	21
3.2 Blok Diagram	22
3.2.1 Blok Diagram Penelitian	22
3.2.2 Blok Diagram Sistem	23
3.3 Flowchart	24
3.3.1 Flowchart Partisi Data Latih dan Uji	24
3.3.2 Flowchart Preprocessing Data	25
3.3.3 Flowchart Perhitungan TF-IDF	26
3.3.4 Flowchart Feature Selection	27
3.3.5 Flowchart Perhitungan Conditional Probability	27
3.3.6 Flowchart Klasifikasi dengan Perhitungan Frekuensi Kemunculan Term	28
3.3.7 Flowchart Klasifikasi dengan Menggunakan Fitur TF-IDF.....	29
3.4 Class Diagram	30
3.5 Database	31
3.5.1 Database Sistem	31
3.5.2 Data Konversi Slang	32
3.5.3 Data Konversi Emoticon	32
3.6 Rancangan Tampilan Antar Muka Sistem	33
3.6.1 Tampilan Menu Awal	33
3.6.2 Tampilan Menu Demo	34
3.6.3 Tampilan Menu Hasil Demo	35
3.6.4 Rancangan Tampilan Menu Menu Training	36
3.6.5 Rancangan Tampilan Menu Menu Testing	36
3.6.6 Rancangan Tampilan Menu Menu Evaluasi	37
3.6.7 Rancangan Tampilan Menu CRUD	38
3.7 Rancangan Proses Pengujian	38
BAB IV IMPLEMENTASI DAN ANALISIS SISTEM	40

4.1 Implementasi Sistem	40
4.1.1 Form Menu Home	40
4.1.2 Form Menu Demo	41
4.1.3 Form Menu Hasil Demo	45
4.1.4 Form Menu Data Training	47
4.1.5 Form Menu Data Testing	48
4.1.6 Form Menu Evaluasi	49
4.1.7 Form Menu CRUD	50
4.2 Analisis Sistem	52
4.2.1 Analisis Klasifikasi Komentar Topik Berbeda	52
4.2.2 Pengaruh Smoothing Factor Terhadap Akurasi Data Tidak Seimbang dengan Perhitungan Frekuensi Kemunculan Term	54
4.2.3 Hasil Analisis Tingkat Akurasi Terhadap Data Tidak Seimbang dengan Perhitungan Frekuensi Kemunculan Term	56
4.2.4 Pengaruh Smoothing Factor Terhadap Akurasi Data Seimbang dengan Perhitungan Frekuensi Kemunculan Term	57
4.2.5 Hasil Analisis Tingkat Akurasi Terhadap Data Seimbang dengan Perhitungan Frekuensi Kemunculan Term	58
4.2.6 Persentase Akurasi Data Terhadap Penggunaan Fitur TF-IDF Pada Data Tidak Seimbang	59
4.2.7 Hasil Analisis Tingkat Akurasi Terhadap Data Tidak Seimbang Dengan Fitur TF-IDF	60
4.2.8 Pengaruh Smoothing Factor Terhadap Akurasi Data Seimbang Dengan Fitur TF-IDF	61
4.2.9 Hasil Analisis Tingkat Akurasi Terhadap Data Seimbang Dengan Fitur TF-IDF	62
BAB V KESIMPULAN DAN SARAN	65
5.1 Kesimpulan	65
5.2 Saran	65
DAFTAR PUSTAKA	66
LAMPIRAN	68

DAFTAR GAMBAR

Gambar 2.1 Contoh Komentar <i>Spam</i> Pada <i>Instagram</i> (1)	7
Gambar 2.2 Contoh Komentar <i>Spam</i> Pada <i>Instagram</i> (2)	7
Gambar 3.1 Blok diagram langkah-langkah penelitian secara umum	22
Gambar 3.2 Blok diagram sistem secara umum	23
Gambar 3.3 Flowchart partisi data latih dan uji	24
Gambar 3.4 Flowchart preprocessing data	25
Gambar 3.5 Flowchart Pembobotan TF-IDF	26
Gambar 3.6 Flowchart Pemilihan Fitur	27
Gambar 3.7 Flowchart Perhitungan Conditional Probability	27
Gambar 3.8 Flowchart Klasifikasi Data dengan Perhitungan Frekuensi Kemunculan Term	28
Gambar 3.9 Flowchart Klasifikasi dengan Menggunakan Fitur TF-IDF	29
Gambar 3.10 Class Diagram Sistem	30
Gambar 3.11 Skema Basis Data	31
Gambar 3.12 Skema Basis Data (2)	31
Gambar 3.13 Konversi Slang Menjadi Kata Baku	32
Gambar 3.14 Konversi Emoticon Menjadi String	32
Gambar 3.15 Rancangan Tampilan Menu Awal	33
Gambar 3.16 Rancangan Tampilan Menu Demo	34
Gambar 3.17 Rancangan Tampilan Modal Hasil Demo	34
Gambar 3.18 Rancangan Tampilan Menu Demo Perhitungan Akurasi	35
Gambar 3.19 Rancangan Tampilan Menu Hasil Demo	35
Gambar 3.20 Rancangan Tampilan Menu Data Training	36
Gambar 3.21 Rancangan Tampilan Menu Testing	36
Gambar 3.22 Rancangan Tampilan Menu Evaluasi	37
Gambar 3.23 Rancangan Tampilan Menu CRUD	38
Gambar 3.24 Partisi Data Latih dan Uji Data Tidak Seimbang	39
Gambar 3.25 Partisi Data Latih dan Uji Data Seimbang	39
Gambar 4.1 Form Menu Home	40

Gambar 4.2 Form Menu Home	41
Gambar 4.3 Form Menu Demo – Pilihan Metode	42
Gambar 4.4 Form Menu Demo – Hasil Deteksi	42
Gambar 4.5 Form Menu Demo – Perhitungan Akurasi	43
Gambar 4.6 Form Menu Demo – Hasil Perhitungan Akurasi	43
Gambar 4.7 Form Menu Demo – Deteksi dengan Semua Metode	44
Gambar 4.8 Form Menu Demo – Contoh Hasil Deteksi dengan Semua Metode	44
Gambar 4.9 Form Menu Hasil Demo	45
Gambar 4.10 Form Menu Hasil Demo – Komentar Akurasi	45
Gambar 4.11 Form Menu Hasil Demo – Hasil Akurasi	46
Gambar 4.12 Form Menu Hasil Demo – Detail Klasifikasi	46
Gambar 4.13 Form Menu Hasil Demo – Detail Klasifikasi(2)	46
Gambar 4.14 Form Menu Data Training – Data Partisi	47
Gambar 4.15 Form Menu Data Training – Detail Partisi	47
Gambar 4.16 Form Menu Data Training – Hasil Preproses	48
Gambar 4.17 Form Menu Data Training – Hasil Pembobotan	48
Gambar 4.18 Form Menu Data Testing – Hasil Pengujian	49
Gambar 4.19 Form Menu Data Evaluasi – Hasil Perhitungan Akurasi Data	50
Gambar 4.20 Form Menu Data Evaluasi – Tampilan Visual Hasil Evaluasi	50
Gambar 4.21 Form Menu CRUD – Contoh Tampilan Proses Insert	51
Gambar 4.22 Form Menu CRUD – Contoh Tampilan Proses Delete	51
Gambar 4.23 Contoh Hasil Klasifikasi Komentar Topik Lain	53
Gambar 4.24 Contoh Penambahan Data Latih Baru	53
Gambar 4.25 Contoh Hasil Klasifikasi Data Baru setelah Penambahan Data Latih	54
Gambar 4.26 Pengaruh Smoothing Factor Terhadap Akurasi Data Tidak Seimbang dengan Perhitungan Frekuensi Kemunculan Term	55
Gambar 4.27 Persentase Akurasi Data Tidak Seimbang dengan Perhitungan Frekuensi Kemunculan Term	56
Gambar 4.28 Pengaruh Smoothing Factor Terhadap Akurasi Data Seimbang dengan Perhitungan Frekuensi Kemunculan Term	57
Gambar 4.29 Persentase Akurasi Data Tidak Seimbang dengan Perhitungan	

Frekuensi Kemunculan Term	58
Gambar 4.30 Persentase Akurasi Terhadap Pemilihan Fitur TF-IDF Data Tidak Seimbang	59
Gambar 4.31 Persentase Akurasi Maksimal Terhadap Data Tidak Seimbang Dengan Fitur TF-IDF	60
Gambar 4.32 Persentase Akurasi Terhadap Pemilihan Fitur TF-IDF Data Seimbang	61
Gambar 4.33 Persentase Akurasi Maksimal Terhadap Pemilihan Fitur TF-IDF Data Seimbang	62

©UKDWN

DAFTAR TABEL

Tabel 2.1 Hasil Proses Tokenisasi	10
Tabel 2.2 Hasil Proses Normalisasi	10
Tabel 2.3 Hasil Proses stemming	10
Tabel 2.4 Hasil Proses stopword removal	11
Tabel 2.5 Contoh Proses Perhitungan Nilai TF-IDF	12
Tabel 2.6 Contoh Proses Klasifikasi Sebuah Komentar Baru Terhadap Data Latih	16
Tabel 2.7 Data Hasil Preprocessing	17
Tabel 2.8 Proses perhitungan jumlah kemunculan kata w_k pada kelas C	17
Tabel 2.9 Hasil perhitungan Jelinek-Mercer Smoothing	18
Tabel 2.10 Confusion matrix	19
Tabel 3.1 Spesifikasi Minimum Perangkat Keras yang Digunakan	20
Tabel 4.1 Tabel Evaluasi Akhir Skenario Pertama	63
Tabel 4.2 Tabel Evaluasi Akhir Skenario Kedua	63
Tabel 4.3 Tabel Evaluasi Akhir Skenario Ketiga	63
Tabel 4.4 Tabel Evaluasi Akhir Skenario Keempat	64

DAFTAR LAMPIRAN

LAMPIRAN A	: HASIL EVALUASI SKENARIO I
LAMPIRAN B	: HASIL EVALUASI SKENARIO II
LAMPIRAN C	: HASIL EVALUASI SKENARIO III
LAMPIRAN D	: HASIL EVALUASI SKENARIO IV
LAMPIRAN E	: SCAN KARTU KONSULTASI
LAMPIRAN F	: SCAN FORMULIR REVISI SKRIPSI
LAMPIRAN G	: LISTING PROGRAM

INTISARI

Instagram sebagai salah satu media sosial yang paling banyak digunakan di Indonesia merupakan wadah komunikasi digital terkini. Berbagai kalangan masyarakat seperti musisi dan artis memanfaatkan teknologi ini untuk berinteraksi dengan para pengikutnya. Namun menjamurnya komentar *spam* di setiap kiriman yang ada sangat mengganggu proses komunikasi yang sedang terjadi dan menyita waktu banyak orang. Chrismanto dan Lukito (2017) dalam penelitiannya yang berjudul *Deteksi Komentar Spam Bahasa Indonesia Pada Instagram Menggunakan Naive Bayes* telah membuktikan bahwa metode tersebut mampu menghasilkan tingkat akurasi sebesar 77,25%.

Berdasarkan hasil yang didapat, dilakukan penelitian menggunakan data set yang sama untuk melihat hasil perbandingan tingkat akurasi yang didapat menggunakan metode Naive Bayes biasa terhadap metode Naive Bayes dengan *Enhanced Smoothing*. Teknik *smoothing* yang dilakukan akan menghilangkan kemungkinan *zero probability* yang ada dan berbagai macam prosedur yang dilakukan pada tahap preproses, pemilihan fitur, dan klasifikasi akan menghasilkan nilai akurasi yang lebih baik. Metode klasifikasi yang digunakan dalam penelitian ini mencakup Metode *Naive Bayes*, *Naive Bayes Laplace Smoothing*, *Naive Bayes Jelinek-Mercher Smoothing*, *Naive Bayes Jelinek-Mercer Enhanced Smoothing*, *Naive Bayes Dirichlet Smoothing*, dan *Naive Bayes Dirichlet Enhanced Smoothing*.

Hasil dari penelitian ini pun membuktikan bahwa metode Naive Bayes dengan *Enhanced Smoothing* mampu menghasilkan kenaikan tingkat akurasi sebesar 5% - 10% dengan nilai akurasi tertinggi yang didapat adalah sebesar 90,55%.

Kata Kunci— Naive Bayes, Naive Bayes Enhanced Smoothing, spam comment decection, Instagram, Indonesian spam detection

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Internet sebagai salah satu teknologi yang paling penting di era sekarang ini telah memungkinkan penggunanya untuk bisa saling berkomunikasi dengan satu sama lain tanpa dibatasi oleh jarak dan waktu. Dengan menggunakan *smartphone* atau komputer, pengguna bisa mengakses internet dengan mudahnya melalui *browser* yang diinstal di perangkatnya untuk bisa saling terhubung dengan pengguna lain. Salah satu cara berinteraksi antar sesama pengguna internet yang paling populer saat ini adalah dengan memakai media sosial, dan salah satunya yang paling populer dan paling sering digunakan di Indonesia adalah *Instagram*. Melalui *Instagram*, pengguna bisa mengirimkan informasi berupa gambar atau video beserta dengan keterangan atau *caption* yang diinginkan dalam bentuk teks.

Dalam proses tersebut, tidak jarang terdapat pesan-pesan teks berupa komentar terhadap kiriman seseorang yang tidak berkaitan atau tidak berhubungan sama sekali dengan topik yang sedang dibahas. Pesan teks atau komentar seperti itu dikenal dengan sebutan *spam*. Komentar berupa *spam* cukup mengganggu karena menyita waktu pengguna untuk membaca sesuatu yang tidak relevan. Tindakan yang dilakukan untuk menghilangkan komentar *spam* pun masih cukup terbatas, seperti dengan cara menghapus komentar *spam* secara satu per satu atau menggunakan fitur untuk melaporkan komentar tersebut. Apabila proses tersebut dilakukan secara manual, tentu akan sangat menyita waktu dan tenaga.

Upaya untuk menghilangkan komentar *spam* secara otomatis pun telah dilakukan melalui berbagai macam penelitian dan metode terkait, salah satunya adalah dengan menggunakan algoritma *Naive Bayes* yang diaplikasikan pada bidang *text mining*. Chrismanto dan Lukito (2017) pada penelitiannya yang berjudul *Deteksi Komentar Spam Bahasa Indonesia Pada Instagram Menggunakan Naive Bayes* telah membuktikan bahwa metode tersebut mampu mendeteksi komentar *spam* berbahasa Indonesia dengan baik. Sedangkan Aggarwal dan Kaul (2013) dalam penelitiannya yang berjudul *Enhanced*

Smoothing Methods Using Naive Bayes Classifier for Better Spam Classification, menunjukkan bahwa algoritma *Naive Bayes* yang telah dimodifikasi dengan menggunakan metode *smoothing* pada beberapa bagian tertentu mampu meningkatkan efisiensi perhitungan dan akurasi sistem. Oleh karena itu, penulis akan melakukan penelitian untuk membandingkan kinerja kedua teknik metode tersebut terhadap *dataset* komentar pengguna *Instagram* pada beberapa akun artis terkenal di Indonesia. Sistem diharapkan mampu mendeteksi komentar *spam* berbahasa Indonesia pada *Instagram* dengan baik serta mampu menghasilkan kinerja atau akurasi yang lebih tinggi dibanding menggunakan metode *Naive Bayes* biasa.

1.2 Rumusan Masalah

Berdasarkan latar belakang di atas, dapat dijabarkan perumusan masalah sebagai berikut :

1. Apakah metode *Naive Bayes* dengan *Enhanced Smoothing* mampu diimplementasikan ke dalam sistem untuk mendeteksi komentar *spam* bahasa Indonesia pada *Instagram*?
2. Seberapa besar peningkatan nilai akurasi yang dihasilkan oleh metode *Naive Bayes* dengan *Enhanced Smoothing* dibandingkan dengan metode *Naive Bayes* biasa?

1.3 Batasan Masalah

Batasan masalah yang digunakan dalam penelitian ini adalah sebagai berikut:

1. Dataset yang digunakan merupakan dataset yang dikumpulkan pada penelitian yang dilakukan oleh Chrismanto dan Lukito (2017) dengan judul *Deteksi Komentar Spam Bahasa Indonesia pada Instagram Menggunakan Naive Bayes*.
2. Teknik perhitungan yang dipakai dalam penelitian ini menggunakan metode *Naive Bayes*, *Naive Bayes Laplace Smoothing*, *Naive Bayes Jelinek-Mercer Smoothing*, *Naive Bayes Jelinek-Mercer Enhanced*

Smoothing, Naive Bayes Dirichlet Smoothing, dan Naive Bayes Dirichlet Enhanced Smoothing.

3. Sistem yang dibangun berbasis web.
4. Jumlah konversi emoticon yang digunakan adalah sebanyak 50 buah.
5. Jumlah konversi *slang* yang dipakai adalah sebanyak 457 kata.

1.4 Tujuan Penelitian

Tujuan penelitian ini adalah untuk mengukur serta membandingkan kinerja dari metode *Naive Bayes* dengan *Enhanced Smoothing* terhadap metode *Naive Bayes* biasa dalam mendeteksi komentar *spam* berbahasa Indonesia pada *Instagram* berbasis sistem web.

1.5 Metode Penelitian/Pendekatan

1. Studi Pustaka

Melalui studi pustaka penulis mencari dan mempelajari beberapa literatur yang memuat teori dan pembahasan terkait dengan metode *Naive Bayes* dengan *Enhanced Smoothing* yang digunakan untuk mendeteksi *spam* berjenis teks.

2. Pengumpulan Data dan Perancangan Sistem

Data yang digunakan berupa data set hasil penelitian Chrismanto dan Lukito (2017). Sedangkan pada tahap perancangan sistem akan dibuat rancangan *database* dan antarmuka sistem yang akan dibangun.

3. Pembuatan Sistem

Pada tahap ini akan dibangun suatu sistem sesuai dengan hasil perancangan *database* dan antarmuka.

4. Implementasi dan Pengujian Sistem

Pada bagian ini akan dilakukan pengujian terhadap sistem yang telah dibangun menggunakan metode *Naive Bayes* dengan *Enhanced Smoothing*

untuk mendeteksi komentar *spam* berbahasa Indonesia pada *Instagram* dengan output berupa hasil proses klasifikasi.

5. Analisis Hasil dan Evaluasi Sistem

Pada tahap ini dilakukan analisa dan evaluasi sistem untuk mengetahui seberapa akurat metode yang digunakan dalam melakukan pendeteksian komentar *spam* menggunakan perhitungan *confusion matrix*.

1.6 Sistematika Penulisan

Penyusunan laporan Tugas Akhir ini akan disusun dengan menggunakan sistem penulisan sebagai berikut :

Bab 1 : Pendahuluan

Berisi latar belakang, perumusan masalah, batasan masalah, tujuan penelitian, metode penelitian/pendekatan, dan sistematika penulisan.

Bab 2 : Tinjauan Pustaka

Berisi tinjauan pustaka yang memuat berbagai macam teori yang diperoleh dari beberapa sumber pustaka serta landasan teori yang memuat tentang konsep dan prinsip utama metode *Naive Bayes* dengan *Enhanced Smoothing*.

Bab 3 : Perancangan Sistem

Membahas tentang algoritma, analisis data, persiapan data sampel, perancangan basis data, perancangan input, perancangan proses, dan perancangan output.

Bab 4 : Implementasi dan Analisis Sistem

Membahas tentang analisa penerapan metode *Naive Bayes* dengan *Enhanced Smoothing* terhadap sistem rancangan yang meliputi penjelasan cara kerja, proses input dan output, serta evaluasi sistem.

Bab 5 : Kesimpulan dan Saran

Berisikan kesimpulan dari sistem yang dibuat dan saran untuk pengembangan di masa mendatang.

BAB 5

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Beberapa kesimpulan yang diambil dari hasil penelitian ini adalah sebagai berikut :

1. Sistem deteksi komentar *spam* berbahasa Indonesia di Instagram menggunakan metode *Naive Bayes* dengan *Enhanced Smoothing* mampu menghasilkan nilai akurasi yang lebih tinggi dibandingkan dengan metode *Naive Bayes* biasa dengan kenaikan sebesar 5% sampai dengan 10%.
2. Tingkat akurasi tertinggi skenario pertama, kedua, ketiga, dan keempat yang diperoleh secara berturut turut adalah sebesar 90,29% dengan metode NBL; 90,55% dengan metode NBL; 86,66% dengan metode NBJM & NBJME, dan 87,53% dengan metode NBL.
3. Pengaruh nilai koefisien smoothing terhadap metode NBJM dan NBJME adalah semakin besar nilai koefisien yang dipakai, semakin berkurang akurasi yang diperoleh. Sedangkan pada metode NBD dan NBDE tidak menunjukkan pengaruh atau perubahan yang signifikan.
4. Penggunaan fitur pembobotan cocok digunakan pada metode NBJM, NBJME, NBD, dan NBDE dengan hasil rata-rata nilai akurasi yang cukup stabil walaupun jumlah fitur yang dipakai sedikit.

5.2 Saran

Sistem bisa dikembangkan untuk mengklasifikasikan komentar *spam* secara real time dalam bentuk web service atau extentions. Bagian preproses perlu ditambahkan algoritma untuk menormalisasi string yang menggunakan karakter-karakter khusus beserta string non baku dengan algoritma auto-correction. Selain itu juga diperlukan metode atau algoritma yang mampu mengklasifikasikan sebuah komentar dengan keseluruhan term yang masuk ke dalam stopword list atau belum pernah ada sebelumnya di dalam data latih.

DAFTAR PUSTAKA

- Aggarwal, S., & Kaur, D. (2013). Enhanced Smoothing Methods Using Naive Bayes Classifier for Better Spam Classification. *International Journal of Engineering Research & Technology* , 2 (9).
- Chandra, W. N., Indrawan, G., & Sukajaya, I. N. (2016). Spam Filtering Dengan Metode Pos Tagger Dan Klasifikasi Naïve Bayes. *Jurnal Ilmiah Teknologi dan Informasia ASIA (JITIKA)* , 10 (1).
- Chrismanto, A. R., & Lukito, Y. (2017). Deteksi Komentar Spam Bahasa Indonesia Pada Instagram Menggunakan Naive Bayes. *Ultimatics* , IX (1).
- Gupta, V., & Lehal, G. (2009). A Survey of Text Mining Techniques and Applications. *Journal of Emerging Technologies in Web Intelligence* , 1 (1), 60-76.
- Help.Instagram. (n.d.). Retrieved October 15, 2017, from Instagram: <https://help.instagram.com>
- Hines, K. (n.d.). *blog.Kissmetrics*. Retrieved September 13, 2017, from Kissmetrics: <https://blog.kissmetrics.com/control-blog-comment-spam/>
- Instagram. (n.d.). Retrieved October 15, 2017, from Instagram: <https://www.instagram.com/p/BaF1M-ojI3H/?hl=en&taken-by=agnezmo>
- Instagram. (n.d.). Retrieved October 15, 2017, from Instagram: <https://www.instagram.com/p/BaETta1HnK1/?hl=en&taken-by=raisa6690>
- Jindal, N., & Liu, B. (2008). Opinion Spam and Analysis. *Proceedings of the Conference on Web Search and Web Data Mining* .
- Lengstorf, J. (2010). *Pro PHP and JQuery*. New York: Springer Science+Business Media, LLC. .
- Liu, B. (2012). *Sentiment Analysis and Opinion Mining*. Morgan & Claypool.
- Manning, C. D., Raghavan, P., & Schutze, H. (2009). *An Introduction to Information Retrieval*. Cambridge, England: Cambridge University Press.
- Narayanan, V., Arora, I., & Bhatia, A. (2013). Fast and accurate sentiment classification using an enhanced Naive Bayes model. *Intelligent Data Engineering and Automated Learning – IDEAL* .

- Nayak, A. S., Kanive, A. P., Chandavekar, N., & R., B. (2016). Survey on Pre-Processing Techniques for Text Mining. *International Journal Of Engineering And Computer Science* , 5 (6), 16875-16879.
- Prayoga, F. A., Pinandito, A., & Perdana, R. S. (2018). Rancang Bangun Aplikasi Deteksi Spam Twitter menggunakan Metode Naive Bayes dan KNN pada Perangkat Bergerak Android . *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer* , 2 (2), 554-564.
- Sebastiani, F. (2002). Machine Learning in Automated Text Categorization. *ACM Computing Surveys* , 34 (1).
- Talib, R., Hanif, M. K., Ayesha, S., & Fatima, F. (2016). Text Mining : Techniques, Applications and Issues. *International Journal of Advanced Computer Science and Applications* , 7 (11).
- Tatroe, K., MacIntyre, P., & Lerdorf, R. (2013). *Programming PHP, Third Edition*. O'Reilly Media, Inc. .
- Weiss, S., Indurkha, N., Zhang, T., & Damerau, F. (2005). *Text Mining: Predictive Methods for Analyzing Unstructured Information*.
- Witten, I. H., & Frank, E. (2005). *Data Mining Practical Machine Learning Tools and Techniques*. Morgan Kaufmann .
- Yuan, Q., Cong, G., & Thalmann, N. M. (2012). Enhancing Naive Bayes with Various Smoothing Methods for Short Text Classification . Proceedings of the 21st International Conference on World Wide Web .