

**Klasifikasi Komentar untuk Pendukung Capres dari Portal Berita  
Menggunakan Metode Bernoulli Naive Bayes**

Skripsi



Diajukan oleh:

YOSEF RESI AUGMANTO

71150011

PROGRAM STUDI INFORMATIKA  
FAKULTAS TEKNOLOGI INFORMASI  
UNIVERSITAS KRISTEN DUTA WACANA  
YOGYAKARTA

2019

**Klasifikasi Komentar untuk Pendukung Capres dari Portal Berita  
Menggunakan Metode Bernoulli Naive Bayes**

Skripsi



Diajukan kepada Fakultas Teknologi Informasi Program Studi Informatika  
Universitas Kristen Duta Wacana  
Sebagai salah satu syarat dalam memperoleh gelar Sarjana Komputer

Diajukan oleh:

YOSEF RESI AUGMANTO

71150011

PROGRAM STUDI INFORMATIKA  
FAKULTAS TEKNOLOGI INFORMASI  
UNIVERSITAS KRISTEN DUTA WACANA YOGYAKARTA

2019

## PERNYATAAN KEASLIAN SKRIPSI

Saya menyatakan dengan sesungguhnya bahwa skripsi dengan judul:

### **KLASIFIKASI KOMENTAR UNTUK PENDUKUNG CAPRES DARI PORTAL BERITA MENGGUNAKAN METODE BERNOULLI NAIVE BAYES**

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 kesarjanaannya 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 kesarjanaannya saya.

Yogyakarta, 24 Juni 2019



YOSEF RESI AUGMANTO

71150011

## HALAMAN PERSETUJUAN

Judul Skripsi : KLASIFIKASI KOMENTAR UNTUK  
PENDUKUNG CAPRES DARI PORTAL BERITA  
MENGUNAKAN METODE BERNOULLI NAIVE  
BAYES

Nama Mahasiswa : YOSEF RESI AUGMANTO

NIM : 71150011

Matakuliah : Skripsi (Tugas Akhir)

Kode : TIW276

Semester : Genap

Tahun Akademik : 2018/2019

Telah diperiksa dan disetujui di  
Yogyakarta,  
Pada tanggal 24 Juni 2019

Dosen Pembimbing I



Danny Sebastian, S.Kom., M.M., M.T.

Dosen Pembimbing II



Kristian Adi Nugraha, S.Kom., M.T.

## HALAMAN PENGESAHAN

### KLASIFIKASI KOMENTAR UNTUK PENDUKUNG CAPRES DARI PORTAL BERITA MENGGUNAKAN METODE BERNOULLI NAIVE BAYES

Oleh: YOSEF RESI AUGMANTO / 71150011

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, 24 Juni 2019  
Mengesahkan,

Dewan Penguji:

1. Danny Sebastian, S.Kom., M.M., M.T.
2. Kristian Adi Nugraha, S.Kom., M.T.
3. Budi Susanto, S.Kom., M.T.
4. Antonius Rachmat C., S.Kom., M.Cs.



Dekan



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

Ketua Program Studi



(Gloria Virginia, Ph.D.)

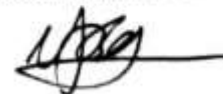
## UCAPAN TERIMA KASIH

Penyusunan skripsi ini tidak terlepas dari dukungan berbagai pihak. Peneliti secara khusus mengucapkan terima kasih yang sebesar-besarnya kepada semua pihak yang telah membantu. Penulis banyak menerima bimbingan, petunjuk, dan bantuan serta dorongan dari berbagai pihak yang bersifat moral maupun material. Pada kesempatan ini penulis ingin menyampaikan rasa terima kasih kepada:

1. Tuhan Y.M.E. atas berkat, rahmat, dan karunia-Nya sehingga penulis dapat menyelesaikan penelitian ini.
2. Kedua orang tua yang selalu memberikan kasih sayang, semangat, dan mendoakan penulis.
3. Bapak Danny Sebastian, S.Kom, M.M., M.T. dan Bapak Kristian Adi Nugraha, S.Kom, M.T. selaku dosen pembimbing atas waktu, saran, dan bimbingan yang diberikan kepada penulis sehingga laporan dan program tugas akhir dapat selesai dengan baik.
4. Bapak Budi Susanto, S.Kom., M.T selaku Dekan Fakultas Teknologi Informasi Universitas Kristen Duta Wacana.
5. Segenap staf dan karyawan yang telah bekerja keras memberikan pelayanan kepada mahasiswa / mahasiswi Universitas Kristen Duta Wacana
6. Samuel Eddijanto dan Michell Bernardi yang sudah membantu proses labelling data penulis.
7. Teman – teman lain yang sudah memberikan berbagai bantuan dalam pengerjaan tugas akhir ini.

Serta masih banyak lagi pihak-pihak yang tidak bisa penulis sebutkan satu persatu yang memberikan berbagai bantuan kepada penulis. Semoga Tuhan senantiasa membalas semua kebaikan yang telah diberikan. Semoga penelitian ini dapat bermanfaat bagi penulis, maupun bagi orang lain.

Yogyakarta, 24 Mei 2019



Yosef Resi Augmanto

## INTISARI

### KLASIFIKASI KOMENTAR UNTUK PENDUKUNG CAPRES DARI PORTAL BERITA MENGGUNAKAN METODE BERNOULLI NAIVE BAYES

Seiring kemajuan teknologi informasi dan komunikasi, penyebaran informasi atau berita di Indonesia sekarang banyak bermula dari media *online*. Masyarakat pun dapat dengan bebas mengemukakan pendapatnya pada platform sosial media maupun portal berita *online*. Klasifikasi komentar secara otomatis diperlukan untuk membantu melihat sikap masyarakat terhadap suatu berita atau fenomena. Tujuan dari perancangan sistem ini adalah membantu mengklasifikasikan komentar pada berita dengan kategori PEMILU 2019. Adapun komentar akan dibagi menjadi 2 kelompok, yaitu komentar pendukung paslon 01 Joko Widodo – Ma'aruf Amin dan komentar pendukung paslon 02 Prabowo Subianto – Sandiaga Uno.

Pada penelitian ini, penulis menggunakan data dari situs portal berita *online* detik.com. Data diambil menggunakan *grabber*, kemudian dilakukan pemberian label menggunakan sistem *crowdsourcing*. Selanjutnya, dilakukan proses *preprocessing* seperti *case folding*, dan *stemming*, serta aturan-aturan lain seperti penghapusan kata singkatan, penghapusan *emoticon* dan angka, dan pengolahan kata negasi. Proses klasifikasi dilakukan menggunakan metode *Bernoulli Naive Bayes*. Luaran berupa klasifikasi untuk sebuah komentar yang dimasukkan ke dalam sistem.

Pengujian menunjukkan rata-rata akurasi yang dihasilkan sebesar 67,3%. Seleksi fitur sebesar 10% dari fitur teratas memberikan nilai *accuracy* dan *F-Measure* tertinggi.

**Kata Kunci:** [klasifikasi teks, *bernoulli naive bayes*, komentar berita]

## DAFTAR ISI

PERNYATAAN KEASLIAN SKRIPSI.....	iii
HALAMAN PERSETUJUAN.....	iv
HALAMAN PENGESAHAN.....	v
UCAPAN TERIMA KASIH.....	vi
INTISARI.....	vii
DAFTAR ISI.....	viii
DAFTAR TABEL.....	x
DAFTAR GAMBAR.....	xi
BAB 1 PENDAHULUAN.....	1
1.1. Latar Belakang Masalah.....	1
1.2. Perumusan Masalah.....	2
1.3. Batasan Masalah.....	2
1.4. Tujuan Penelitian.....	2
1.5. Manfaat Penelitian.....	3
1.6. Metode / Pendekatan.....	3
1.7. Sistematika Penulisan.....	4
BAB 2 TINJAUAN PUSTAKA.....	6
2.1. Tinjauan Pustaka.....	6
2.2. Landasan Teori.....	8
BAB 3 PERANCANGAN SISTEM.....	15
3.1. Spesifikasi Sistem.....	15
3.2. Perancangan Sistem.....	16
3.3. ERD.....	27
3.4. Kamus Data.....	28



3.5. Langkah Pengujian.....	34
<b>BAB 4 IMPLEMENTASI SISTEM .....</b>	<b>35</b>
4.1. Implementasi Sistem .....	35
4.2. Hasil Pengujian Sistem .....	45
4.3. Analisa Hasil Pengujian Sistem .....	48
<b>BAB 5 KESIMPULAN DAN SARAN .....</b>	<b>52</b>
5.1 Kesimpulan .....	52
5.2 Saran.....	52
<b>DAFTAR PUSTAKA .....</b>	<b>53</b>
<b>LAMPIRAN A .....</b>	<b>55</b>

©UKYDWN

## DAFTAR TABEL

Tabel 2.1 Hasil Penelitian (Rachmat C. & Lukito, 2016) dengan <i>10-fold cross validation</i> .....	6
Tabel 2.2 Hasil penelitian Zhao, et al. (2014) untuk <i>SVM</i> dan <i>Naive Bayes</i> .....	7
Tabel 2.3 Contoh dokumen setelah <i>preprocessing</i> .....	10
Tabel 2.4 Contoh dokumen jika dihitung menggunakan <i>Bernoulli Model</i> .....	13
Tabel 2.5 <i>Confusion matrix</i> .....	14
Tabel 3.1 Struktur kolom pada tabel berita.....	29
Tabel 3.2 Struktur kolom pada tabel komentar .....	29
Tabel 3.3 Struktur kolom pada tabel label_komentar .....	30
Tabel 3.4 Struktur kolom pada tabel user .....	30
Tabel 3.5 Struktur kolom pada tabel data_training .....	30
Tabel 3.6 Struktur kolom pada tabel dokumen_latih .....	31
Tabel 3.7 Struktur dokumen_latih_input .....	31
Tabel 3.8 Struktur tabel dokumen_proses.....	32
Tabel 3.9 Struktur tabel dokumen_seleksi_fitur .....	32
Tabel 3.10 Struktur kolom tabel feature_pos .....	32
Tabel 3.11 Struktur kolom tabel feature_neg.....	33
Tabel 3.12 Struktur kolom tabel terms.....	33
Tabel 3.13 Struktur kolom tabel vote.....	33
Tabel 4.1 Nilai <i>accuracy</i> dari semua percobaan .....	45
Tabel 4.2 Nilai <i>precision</i> dari semua percobaan pada kelas 01 .....	46
Tabel 4.3 Nilai <i>precision</i> dari semua percobaan pada kelas 02 .....	46
Tabel 4.4 Nilai <i>recall</i> dari semua percobaan pada kelas 01.....	46
Tabel 4.5 Nilai <i>recall</i> dari semua percobaan pada kelas 02.....	47
Tabel 4.6 Nilai <i>f-measure</i> dari semua percobaan pada kelas 01 .....	47
Tabel 4.7 Nilai <i>f-measure</i> dari semua percobaan pada kelas 02 .....	47
Tabel 4.8 Rata-rata nilai evaluasi pada kelas 01 .....	48
Tabel 4.9 Rata-rata nilai evaluasi pada kelas 02 .....	48
Tabel 4.10 50 <i>feature set</i> pertama dari kelas 01 dan kelas 02.....	49

## DAFTAR GAMBAR

<i>Gambar 2.1</i> Proses tokenisasi .....	8
<i>Gambar 2.2</i> Proses <i>stemming</i> .....	9
<i>Gambar 3.1</i> Blok diagram sistem .....	16
<i>Gambar 3.2</i> Arsitektur Sistem .....	18
<i>Gambar 3.3</i> <i>Flowchart</i> alur kerja <i>grabber</i> data .....	19
<i>Gambar 3.4</i> Rancangan antarmuka aplikasi web untuk <i>labelling</i> .....	20
<i>Gambar 3.5</i> Rancangan antarmuka sistem klasifikasi bagian pelatihan model ....	21
<i>Gambar 3.6</i> Rancangan antarmuka sistem klasifikasi bagian prediksi data .....	21
<i>Gambar 3.7</i> Rancangan antarmuka sistem klasifikasi bagian evaluasi model .....	22
<i>Gambar 3.8</i> <i>Flowchart</i> proses <i>preprocessing</i> .....	24
<i>Gambar 3.9</i> <i>Flowchart</i> proses <i>feature selection</i> .....	25
<i>Gambar 3.10</i> <i>Flowchart</i> proses klasifikasi metode <i>Bernoulli Naive Bayes</i> .....	26
<i>Gambar 3.11</i> <i>Flowchart procedure</i> buat_vektor .....	26
<i>Gambar 3.12</i> <i>Flowchart procedure</i> hitung_probabilitas .....	27
<i>Gambar 3.13</i> Skema ERD dari <i>database</i> sistem.....	27
<i>Gambar 4.1</i> Halaman depan sistem <i>labelling</i> .....	35
<i>Gambar 4.2</i> Halaman sistem <i>labelling</i> setelah <i>login</i> .....	36
<i>Gambar 4.3</i> Tampilan antarmuka tab Dokumen Training sistem klasifikasi .....	39
<i>Gambar 4.4</i> Tampilan antarmuka tab Prediksi sistem klasifikasi.....	40
<i>Gambar 4.5</i> Tampilan antarmuka tab Hasil Prediksi sistem klasifikasi .....	41
<i>Gambar 4.6</i> <i>Flowchart</i> urutan proses klasifikasi .....	41
<i>Gambar 4.7</i> Grafik nilai <i>precision</i> dan <i>recall</i> kelas 01 .....	49
<i>Gambar 4.8</i> Grafik nilai <i>precision</i> dan <i>recall</i> kelas 02.....	49
<i>Gambar 4.9</i> Grafik nilai <i>accuracy</i> dan <i>F-measure</i> .....	50

## INTISARI

### KLASIFIKASI KOMENTAR UNTUK PENDUKUNG CAPRES DARI PORTAL BERITA MENGGUNAKAN METODE BERNOULLI NAIVE BAYES

Seiring kemajuan teknologi informasi dan komunikasi, penyebaran informasi atau berita di Indonesia sekarang banyak bermula dari media *online*. Masyarakat pun dapat dengan bebas mengemukakan pendapatnya pada platform sosial media maupun portal berita *online*. Klasifikasi komentar secara otomatis diperlukan untuk membantu melihat sikap masyarakat terhadap suatu berita atau fenomena. Tujuan dari perancangan sistem ini adalah membantu mengklasifikasikan komentar pada berita dengan kategori PEMILU 2019. Adapun komentar akan dibagi menjadi 2 kelompok, yaitu komentar pendukung paslon 01 Joko Widodo – Ma'aruf Amin dan komentar pendukung paslon 02 Prabowo Subianto – Sandiaga Uno.

Pada penelitian ini, penulis menggunakan data dari situs portal berita *online* detik.com. Data diambil menggunakan *grabber*, kemudian dilakukan pemberian label menggunakan sistem *crowdsourcing*. Selanjutnya, dilakukan proses *preprocessing* seperti *case folding*, dan *stemming*, serta aturan-aturan lain seperti penghapusan kata singkatan, penghapusan *emoticon* dan angka, dan pengolahan kata negasi. Proses klasifikasi dilakukan menggunakan metode *Bernoulli Naive Bayes*. Luaran berupa klasifikasi untuk sebuah komentar yang dimasukkan ke dalam sistem.

Pengujian menunjukkan rata-rata akurasi yang dihasilkan sebesar 67,3%. Seleksi fitur sebesar 10% dari fitur teratas memberikan nilai *accuracy* dan *F-Measure* tertinggi.

**Kata Kunci:** [klasifikasi teks, *bernoulli naive bayes*, komentar berita]

# BAB 1

## PENDAHULUAN

### 1.1. Latar Belakang Masalah

Seiring kemajuan teknologi informasi dan komunikasi yang semakin cepat, jumlah pengguna internet pun terus bertambah. Indonesia termasuk salah satu negara dengan pengguna internet terbesar di dunia. Jumlah pengguna internet di Indonesia mencapai 143 juta jiwa pada tahun 2017, atau setara dengan 54,68% (APJII, 2017). Tidak heran penyebaran informasi / berita di Indonesia banyak bermula dari media *online*. Situs berita merupakan salah satu agen penyebar informasi yang dominan di internet. Dari keseluruhan pembaca berita aktif, hanya sekitar 17% yang berlangganan surat kabar tercetak (JAKPAT, 2016). Situs berita yang paling banyak dikunjungi adalah detik.com disusul dengan kompas.com (detikcom, KOMPAS.com & LIPUTAN6 are Indonesia's leading news websites, 2018).

Dalam sebuah artikel di portal berita *online*, pembaca dapat menambahkan sebuah komentar. Untuk sebuah artikel yang populer, jumlah komentar dapat mencapai lebih dari 1000 komentar. Dalam berita tentang pemilu 2019, komentar pembaca ini dapat dimanfaatkan sebagai data untuk melihat seberapa banyak dan bagaimana komentar pendukung dari masing-masing pasangan calon. Masalah yang timbul adalah kita tidak bisa mengklasifikasi komentar secara manual dengan tenaga manusia karena jumlahnya yang sangat banyak. Karena banyaknya data ditambah dengan banyaknya komentar sampah (e.g. bot, spam), diperlukan klasifikasi secara otomatis.

Melalui penelitian ini, akan diajukan sebuah solusi berupa sistem klasifikasi otomatis untuk mengklasifikasikan komentar pada artikel berita pemilu 2019 di portal berita *online*. Klasifikasi akan dibagi menjadi 2 kelas, yaitu pendukung pasangan calon 01 Joko Widodo – Ma'aruf Amin (Kelas 01) dan 02 Prabowo Subianto – Sandiaga Uno (Kelas 02). Untuk klasifikasi teks dengan *machine learning*, ada beberapa algoritma klasifikasi yang dapat digunakan, seperti *Naive*

*Bayes*, *k-Nearest Neighbor (KNN)*, *SVM*, dan *Maximum Entropy*. Penelitian ini akan menggunakan pendekatan berbasis *machine learning* dengan algoritma varian dari *Naive Bayes*, yaitu *Bernoulli Naive Bayes*. *Naive Bayes* digunakan karena algoritmanya yang sederhana dan tidak membutuhkan banyak data untuk pelatihan (Shayaa, et al., 2018), serta akurasinya yang cukup tinggi (Vaghela & Jadav, 2016). Hasil penelitian kemudian akan dievaluasi menggunakan parameter *accuracy*, *precision*, *recall*, dan *F-Measure*.

## 1.2. Perumusan Masalah

Berdasarkan penjabaran pada latar belakang, maka rumusan masalah adalah sebagai berikut :

- a. Bagaimana penerapan metode *Bernoulli Naive Bayes* untuk sistem klasifikasi komentar berita pemilu 2019 di portal berita *online*
- b. Bagaimana *accuracy*, *precision*, *recall*, dan *F-Measure* dari metode *Bernoulli Naive Bayes* untuk kasus komentar berita pemilu 2019 pada situs portal berita *online*

## 1.3. Batasan Masalah

Beberapa batasan masalah pada penelitian ini adalah sebagai berikut:

- a. Tidak mendeteksi sarkasme / satire
- b. Komentar sampah (i.e. bot, spam) dan netral akan dibuang
- c. Data akan diambil secara berkala selama penelitian dari situs [www.detik.com](http://www.detik.com) dengan kategori PEMILU 2019 sebanyak 200 artikel berita yang masing-masing diambil komentarnya sebanyak 30 komentar.
- d. Sentimen akan dihasilkan untuk masing-masing komentar
- e. Pengambilan data menggunakan *grabber* yang ditulis menggunakan *framework Scrapy*
- f. Proses *stemming* dan penghapusan *stopword* menggunakan *library Sastrawi*

## 1.4. Tujuan Penelitian

Adapun tujuan yang ingin dicapai dalam penelitian adalah sebagai berikut:

- a. Membangun sistem klasifikasi untuk komentar berita pemilu 2019 pada portal berita *online* dengan metode *Bernoulli Naive Bayes*
- b. Menguji metode *Bernoulli Naive Bayes* dalam mengklasifikasi komentar berita pemilu 2019 menggunakan parameter *accuracy*, *precision*, *recall*, dan *F-Measure*

### 1.5. Manfaat Penelitian

Adapun manfaat yang didapat pada penelitian ini adalah:

- a. Membantu penulis memahami penerapan *machine learning* secara praktis
- b. Membantu penulis memahami penerapan metode *Bernoulli Naive Bayes* untuk mengklasifikasikan sentimen komentar pada portal berita *online*.
- c. Membantu memberikan solusi untuk permasalahan klasifikasi komentar pada portal berita *online*

### 1.6. Metode / Pendekatan

#### 1.6.1. Studi literatur

Studi literatur dilakukan dengan membaca dan mempelajari jurnal, artikel, prosiding, dan bahan lain dengan kata kunci *text mining*, klasifikasi teks, *naive bayes*, dan *natural language processing*.

#### 1.6.2. Perancangan sistem

Sistem yang dirancang dapat dibagi menjadi 3 sistem terpisah, yaitu aplikasi berbasis web untuk melakukan pemberian label data, aplikasi berbasis desktop untuk melakukan pelatihan, pengujian, dan evaluasi model, dan aplikasi *grabber* data. Untuk pelatihan model, pengguna dapat memilih akan menggunakan data yang sudah diberi label sebelumnya atau memasukkan data sendiri. Aplikasi

memberikan pilihan untuk melatih model dengan jumlah fitur berbeda. Model ini kemudian dapat diujikan pada data uji yang dimasukkan oleh pengguna.

### 1.6.3. Pengambilan data

Data akan diambil menggunakan *grabber* yang dibuat dengan bahasa *Python*. *Grabber* akan dijalankan pada komputer penulis dan data akan diambil secara berkala dari situs detik.com melalui *URL* <https://www.detik.com/pemilu>. Setelah data diambil, akan dilakukan pemberian label secara manual melalui aplikasi web yang sudah dibuat. Pemberian label dilakukan dengan menampilkan artikel berita dan komentar, kemudian memilih opsi “Paslon 01”, “Paslon 02”, atau “Tidak Tahu” untuk setiap komentar pada setiap artikel berita. Responden untuk sistem pelabelan data berjumlah 3 orang. Data kemudian akan divalidasi oleh penulis.

### 1.6.4. Pengujian

Pengujian dilakukan dengan metode *k-fold cross validation*, yaitu membagi data yang sudah diberi label sejumlah  $k$  bagian, kemudian menjadikan  $k-1$  bagian sebagai data latih dan satu bagian sisanya sebagai data uji. Pengujian diulangi sebanyak  $k$  kali hingga semua data mendapat bagian menjadi data uji kemudian diambil rata-ratanya. Pengujian dilakukan sebanyak 5 kali, yaitu untuk *feature selection* 10%, 25%, 50%, 75%, dan 100%. Setelah melakukan pengujian, akan dihitung nilai *accuracy*, *precision*, *recall*, dan *F-Measure* setiap pengujian.

### 1.6.5. Pelaporan

Hasil pengujian dan evaluasi kemudian akan dianalisis dan dilaporkan dalam bentuk Tugas Akhir / skripsi.

## 1.7. Sistematika Penulisan

Penelitian ini akan ditulis dalam 5 bab. Bab 1 berisi tentang latar belakang masalah, batasan, serta pentingnya penelitian tentang klasifikasi sentimen komentar. Bab 2 memuat tinjauan pustaka dari penelitian sebelumnya dan landasan



teori yang digunakan dalam penelitian. Bab 3 memuat rancangan sistem yang akan digunakan untuk pengambilan data, pelabelan data, dan pengujian.

Adapun bab 4 berisi pelatihan dan pengujian data menggunakan metode *Bernoulli Naive Bayes* serta hasilnya. Hasil pengujian akan ditampilkan dalam tabel dan akan dihitung angka *accuracy*, *precision*, *recall*, dan *F-Measure*. Bab 5 akan memuat kesimpulan yang didapat dari hasil pengujian serta saran untuk penelitian selanjutnya.

©UKDW

## BAB 5

### KESIMPULAN DAN SARAN

#### 5.1 Kesimpulan

Dari penelitian ini dapat disimpulkan bahwa metode *Bernoulli Naive Bayes* belum dapat digunakan untuk melakukan klasifikasi komentar untuk pendukung capres pada portal berita *online* dengan baik. Nilai *precision* yang tinggi pada kelas 01 disebabkan karena kelas 01 memiliki jumlah *feature set* yang lebih banyak dan kata-kata yang umum keluar di kedua kelas. Hal ini membuat sistem hanya memilih sedikit data yang memiliki probabilitas tinggi untuk masuk dalam kelas 01. Nilai *recall* yang tinggi pada kelas 02 disebabkan karena jumlah *feature set*-nya lebih sedikit dan kata-kata yang lebih spesifik sehingga sistem lebih sensitif untuk mendeteksi kelas 02.

Penggunaan *feature selection* 10% memiliki hasil *accuracy* dan *F-measure* yang paling baik jika dibandingkan dengan 25%, 50%, dan 100%. Semakin banyak fitur yang digunakan, nilai *precision* pada kelas 01 semakin meningkat, sedangkan nilai *recall*-nya semakin menurun. Sebaliknya, pada kelas 02 nilai *precision* semakin menurun dan nilai *recall* semakin meningkat. Hal ini disebabkan karena semakin banyak fitur yang digunakan, maka semakin banyak juga fitur-fitur yang kurang penting.

#### 5.2 Saran

Ada beberapa hal yang masih dapat dikembangkan dalam penelitian ini, terutama terkait dengan data yang digunakan. Jumlah data bersih sebaiknya ditambah dan rentang waktu pengambilan data diperluas agar sistem dapat menggeneralisasi dengan lebih baik. Kemudian data netral juga dapat digunakan untuk membuat klasifikasi menjadi *multiclass*. Pada metode penelitian, dapat ditambahkan survei untuk melihat fitur mana saja yang penting pada setiap kelas. Kepada fitur-fitur ini kemudian dilakukan penambahan bobot TF-IDF. Untuk penelitian selanjutnya, dapat dipertimbangkan untuk menggunakan *rough set model*, serta integrasi langsung sistem ke dalam aplikasi web.

## DAFTAR PUSTAKA

- Amalia, R., Bijaksana, M. A., & Darmantoro, D. (2018). Negation handling in sentiment classification using rule-based adapted from Indonesian language syntactic for Indonesian text in Twitter. *Journal of Physics: Conference Series*, 971.
- APJII. (2017). *Penetrasi dan Perilaku Pengguna Internet Indonesia 2017*. Jakarta: APJII.
- Ariadi, D., & Fithriasari, K. (2015). Klasifikasi Berita Indonesia Menggunakan Metode Naive Bayesian Classification dan Support Vector Machine dengan Confix Stripping Stemmer. *Jurnal Sains dan Seni ITS*. Vol. 4, No. 2, 2337-3520. doi:10.12962/j23373520.v4i2.10966
- Asmi, A., & Ishaya, T. (2012). Negation Identification and Calculation in Sentiment Analysis. *IMMM 2012 : The Second International Conference on Advances in Information Mining and Managementnt*.
- detikcom, KOMPAS.com & LIPUTAN6 are Indonesia's leading news websites. (2018). Retrieved from Roy Morgan: <http://www.roymorgan.com/findings/7622-top-10-indonesian-news-websites-march-2018-201806172227>
- Gurusamy, V., & Kannan, S. (2014). Preprocessing Techniques for Text Mining. Retrieved from [https://www.researchgate.net/publication/273127322\\_Preprocessing\\_Techniques\\_for\\_Text\\_Mining](https://www.researchgate.net/publication/273127322_Preprocessing_Techniques_for_Text_Mining)
- JAKPAT. (2016). *What's the News Today? Indonesian's News Consumption Survey 2016*. Retrieved September 4, 2018, from blog.jakpat.net: <https://blog.jakpat.net/whats-the-news-today-indonesians-news-consumption-survey-2016/>
- Rachmat C., A., & Lukito, Y. (2016). Klasifikasi Sentimen Komentar Politik dari Facebook Page Menggunakan Naive Bayes. *Jurnal Informatika Dan Sistem*

*Informasi*. Vol. 2, No. 2, 26-34. Retrieved from <https://journal.uc.ac.id/index.php/JUISI/article/view/239>

Rachmat, A., & Lukito, Y. (2016). SENTIPOL: DATASET SENTIMEN KOMENTAR PADA KAMPANYE. *Konferensi Nasional Teknologi Informasi dan Komunikasi (KNASTIK 2016)*.

Sadida, R., Azkia, M., Candra, B., Rezeki, N., & Rendy, M. (2017). Perancangan Sistem Analisis Sentimen Masyarakat Pada Sosial Media dan Portal Berita. *Seminar Nasional Teknologi Informasi dan Multimedia*.

Shayaa, S., Jaafar, N., Bahri, S., Sulaiman, A., Wai, P., Chung, Y., . . . Al-Garadi, M. (2018). Sentiment Analysis of Big Data: Methods, Applications, and Open Challenges. *IEEE Access*. Vol. 6, 37807-37827. doi:10.1109/access.2018.2851311

Utami, L. A. (2017). Analisis Sentimen Opini Publik Berita Kebakaran Hutan Melalui Komparasi Algoritma Support Vector Machine dan K-Nearest Neighbour Berbasis Particle Swarm Optimization. *Jurnal Pilar Nusa Mandiri Volume 13 No.1*, 103-112.

Vaghela, V., & Jadav, B. (2016). Analysis of Various Sentiment Classification Techniques. *International Journal of Computer Applications*. Vol. 140, 22-27. Retrieved from <https://pdfs.semanticscholar.org/34e7/03d61e29213f82e3a3e6776c17933219d63f.pdf>

Zhao, Y., Dong, S., & Li, L. (2014). Sentiment Analysis on News Comments Based on Supervised Learning Method. *International Journal of Multimedia and Ubiquitous Engineering*. Vol. 9, No. 7, 333-346. doi:10.14257/ijmue.2014.9.7.28