

Terbit online pada laman web jurnal: <http://jurnal.iaii.or.id>

# JURNAL RESTI

(Rekayasa Sistem dan Teknologi Informasi)

Vol. 3 No. 2 (2019) 176 - 183

ISSN Media Elektronik: 2580-0760

## Perbandingan Metode Klasifikasi Analisis Sentimen Tokoh Politik Pada Komentar Media Berita Online

Sigit Kurniawan<sup>1</sup>, Windu Gata<sup>2</sup>, Dewi Ayu Puspitawati<sup>3</sup>, Nurmalasari<sup>4</sup>, Muhamad Tabrani<sup>5</sup>, Kadinar Novel<sup>6</sup><sup>1,2,3</sup>Pasca Sarjana, Ilmu Komputer, STMIK Nusa Mandiri Jakarta<sup>4,5</sup>Sistem Informasi, STMIK Nusa Mandiri Jakarta<sup>6</sup>Teknologi Informasi, Universitas Bina Sarana Informatika<sup>1</sup>kurniawan.sgt@gmail.com, <sup>2</sup>windu@nusamandiri.ac.id, <sup>3</sup>dewiayu.puspitawati@gmail.com, <sup>4</sup>nurmalasari.nmr@nusamandiri.ac.id,<sup>5</sup>muhamad.mtb@nusamandiri.ac.id, <sup>6</sup>kadinar.ked@bsi.ac.id

### Abstract

General elections are an important part of the political process so that many political figures participate in the process. Electability is one of the concerns, various things are done to be able to increase the electability of political figures who participate in general elections. Media has become one of the important tools used to increase electability, one of which is online news media. Reader comments can be used as an assessment of political figures in the form of sentiment analysis. However, it is not easy to analyze sentiments from comments on online news media, because comments contain unstructured text, especially in Indonesian text. Text pre-processing in text mining is an important part of getting the basic information contained in the comments. This research uses Indonesian text pre-processing using the Gata Framework Textmining. Then proceed with extracting information using the Naïve Bayes classification algorithm and Support Vector Machine which are optimized using Particle Swarm Optimization. Tests carried out with both methods get the results that, Particle Swarm Optimization based on Support Vector Machine is the best method with an accuracy of 78.40% and AUC 0.850. This study found an algorithm that was effective in classifying positive and negative comments related to political figures from online news media.

Keywords: *Text Mining, Sentiment Analysis, Naïve Bayes, Support Vector Machine, Particle Swarm Optimization*

### Abstrak

Pemilihan umum adalah bagian penting dari proses politik sehingga banyak tokoh politik yang berpartisipasi dalam proses tersebut. Elektabilitas adalah salah satu yang menjadi perhatian, berbagai hal dilakukan untuk dapat meningkatkan elektabilitas tokoh politik yang ikut serta dalam pemilihan umum. Media telah menjadi salah satu alat penting yang digunakan untuk meningkatkan elektabilitas, salah satunya adalah media berita online. Komentar dari pembaca dan dapat digunakan sebagai penilaian terhadap tokoh-tokoh politik dalam bentuk analisis sentimen. Namun, tidak mudah untuk menganalisis sentimen dari berbagai komentar di media berita online, karena komentar mengandung teks yang tidak terstruktur, terutama dalam teks Indonesia. Pra-pemrosesan teks dalam penambangan teks merupakan bagian penting untuk mendapatkan informasi dasar yang terkandung dalam komentar. Penelitian ini menggunakan pra-pemrosesan teks Indonesia dengan menggunakan Gata Framework Textmining. Kemudian dilanjutkan dengan mengekstraksi informasi menggunakan metode klasifikasi *Naïve Bayes* dan *Support Vector Machine* yang dioptimalkan menggunakan *Particle Swarm Optimization*. Pengujian yang dilakukan dengan kedua metode mendapatkan hasil bahwa, *Support Vector Machine* berbasis *Particle Swarm Optimization* adalah metode terbaik dengan akurasi 78.40% dan AUC 0.850. Penelitian ini mendapatkan algoritma yang efektif dalam mengklasifikasikan komentar positif dan negatif terkait dengan tokoh-tokoh politik dari media berita online.

Kata kunci: *Text Mining, Sentimen Analisis, Naïve Bayes, Support Vector Machine, Particel Swarm Optimization*

© 2019 Jurnal RESTI

## 1. Pendahuluan

Pemilihan umum adalah bagian penting dari proses politik di Indonesia. menjadi pemilihan umum di Indonesia pada tahun 2019, tidak sedikit tokoh politik mulai bersiap-siap untuk berpartisipasi dalam proses tersebut. Berbagai upaya oleh para pemimpin politik dilakukan untuk meningkatkan elektabilitas mereka. Electability penuh popularitas, orang yang populer akan memiliki electability yang tinggi [1]. Dalam politik, peran elektabilitas selalu menantang untuk meyakinkan pemilih untuk memilih tokoh politik yang mereka sukai [2]. Kecenderungan elektabilitas adalah yang paling selama musim pemilihan kompilasi politik tidak populer dengan pemilihan umum [2] dan dapat menjadi tolok ukur untuk mengukur tingkat elektabilitas seseorang [1].

Analisis sentimen adalah proses perbandingan untuk mengungkap dan mengkategorikan pendapat yang diungkapkan dalam teks, untuk menentukan apakah tanggapan terhadap topik tertentu positif, negatif, atau netral [3][4]. Analisis sentimen juga disebut pertambangan adalah bidang studi yang menganalisis pendapat, sentimen, pendapat, penilaian, dan penilaian emosional pada entitas atribut yang dibahas dalam makalah [5]. Menganalisis sentimen dalam teks-teks politik dapat menjadi representasi dari hasil pemilihan, elektabilitas kandidat, atau perwakilan politik nyata di Negara [3].

Terkait dengan teknologi web, pengikut aktivitas politik di media online merupakan aspek yang menjadi pertimbangan untuk menganalisis sentimen elektabilitas tokoh politik melalui posting dan komentar tertulis. Penelitian yang berkaitan dengan analisis sentimen tokoh politik yang berasal dari media online banyak dilakukan oleh peneliti dengan berbagai metode yang digunakan [6][7][8][9]. Melakukan analisis sentimen yang berasal dari posting teks dan mengomentari media online tidak mudah, karena komentar yang mengandung teks memiliki ketidakteraturan, terlebih pada teks Indonesia yang memiliki fitur khas yang berbeda dari bahasa lain [10].

Penelitian ini menggunakan teknologi pembelajaran mesin untuk menganalisis sentimen elektabilitas tokoh-tokoh politik yang berasal dari data teks komentar pada media berita online di Indonesia. Penambangan teks adalah salah satu metode yang digunakan untuk mengekstrak informasi yang tidak terstruktur dari menggabungkan data teks [11], yang dikonversi menjadi nilai numerik yang akan digunakan dalam berbagai metode penambangan data [12]. Selain menggunakan metode penambangan teks, metode penambangan data dengan fungsi kelompok juga digunakan untuk data teks yang telah diekstraksi dengan tujuan mendapatkan informasi yang lengkap. Algoritma *Naïve Bayes* [13] dan *Support Vector Machine* [14] dioptimalkan dengan *Particle Swarm*

*Optimization* [15] adalah metode penambangan data untuk klasifikasi yang digunakan dalam penelitian ini.

Penelitian terkait dengan sentimen analisis dengan menggunakan algoritma-algoritma klasifikasi sudah banyak dilakukan oleh para peneliti, seperti yang dilakukan oleh Buntoro dan kawan-kawan dengan topik pada pemilihan Gubernur DKI Jakarta pada tahun 2017, penelitian tersebut menggunakan algoritma klasifikasi *Naïve Bayes* dan *Support Vector Machine*. Penelitian tersebut tetap menggunakan tahap *pre-processing* teks, seperti *tokenization*, *cleansing* dan *filtering*, namun yang menjadi kekurangan pada penelitian tersebut, adalah pada *stemmer* bahasa Indonesia dan daftar *stopword* yang masih terbatas dan harus terus dikembangkan. Hasil dari penelitian tersebut didapatkan bahwa klasifikasi sentimen analisis terhadap calon gubernur DKI Jakarta 2017, adalah pada metode klasifikasi *Naïve Bayes* [16].

Penelitian lain juga dilakukan oleh Nurirwan dan kawan-kawan dengan topik penelitian adalah terkait dengan sentimen Presiden Joko Widodo setelah pelantikan Presiden di tahun 2014. Penelitian tersebut menggunakan tahapan *pre-processing* sebagai tahap awal dengan menggunakan *tools* aplikasi Master Sastra serta menggunakan algoritma *Support Vector Machine* sebagai *classifier*. Pada penelitian tersebut menggunakan tahap *pre-processing* dengan menggunakan bantuan sebuah *tools* yang bernama Sastrawi, namun *tools* tersebut harus dilakukan instalasi terlebih dulu pada perangkat komputer yang digunakan dan memiliki beberapa konfigurasi yang sedikit rumit bagi orang-orang yang awam terhadap perangkat lunak komputer, selain menggunakan *tools* Sastrawi, penelitian tersebut juga menggunakan WEKA sebagai *tools* untuk yang digunakan untuk membantu menerapkan algoritma pada mesin pembelajaran. Hasil dari penelitian tersebut menyimpulkan bahwa proses normalisasi dan *stemming* dapat meningkatkan akurasi pada metode algoritma yang digunakan, jika dibandingkan dengan data yang hanya dilakukan normalisasi saja. Sedangkan untuk metode algoritma yang unggul pada penelitian tersebut adalah SVM, baik dengan data yang di normalisasi dan *stemming* ataupun untuk data yang hanya dilakukan normalisasi saja [6].

Penelitian oleh Della Fitriyani Budiono dan kawan-kawan dengan topik pemilihan Gubernur DKI Jakarta 2017 dengan pendekatan prediktif dan deskriptif, data yang diambil dari Twitter. Di dalam penelitian tersebut menggunakan algoritma *Naïve Bayes Multinomial* dan *Support Vector Machine* sebagai klasifikasi dataset. Penelitian tersebut masih ditemukan beberapa kesalahan terkait dengan tahap *pre-processing* teks dan pada tahap klasifikasi dengan menggunakan algoritma yang dipilih. Hasil dari penelitian tersebut kedua

algoritma tidak memiliki perbedaan nilai akurasi yang jauh [17].

Algoritma Naïve Bayes juga digunakan oleh Tirana Noor Fatyanosa dan kawan-kawan, dengan topik pembahasan pemilihan Gubernur DKI Jakarta, penelitian tersebut menggunakan klasifikasi dengan *Summation Method*, *Average on Tweet*, *Weighted Average* dan *Naïve Bayes*. Penelitian tersebut tidak berbeda dengan beberapa penelitian lain yang telah dilakukan yaitu, dimulai dengan proses *pre-processing* teks yang dilanjutkan dengan mengklasifikasi teks dengan menggunakan metode algoritma. Hasil dari penelitian tersebut dapat disimpulkan bahwa algoritma Naive Bayes cukup baik untuk mengklasifikasi teks yang bersentimen netral dan positif [18].

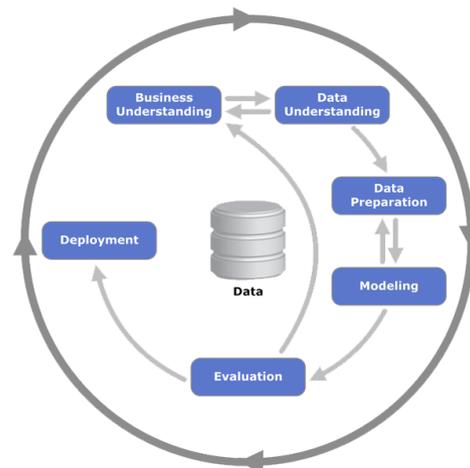
Selain pada sentimen analisis terkait dengan tokoh politik, penelitian lain terkait dengan sentimen analisis juga dilakukan pada topik penelitian pada dataset ulasan film yang diambil datanya di dunia maya, yang dilakukan oleh Shweta Rana dan Archana Singh dengan menggunakan algoritman klasifikasi *Naïve Bayes* dan *Support Vector Machine*. Pada penelitian tersebut tidak menggunakan tahap *pre-processing* pada teks, namun langsung melakukan seleksi terhadap teks ulasan dengan menggunakan Rapidminer. Hasil dari penelitian tersebut didapatkan bahwa film drama memiliki tingkat akurasi yang tinggi, dan dapat disimpulkan bahwa dari data ulasan yang digunakan pada penelitian tersebut bahwa sebagian besar pengguna lebih suka menonton film yang beraliran drama [19].

Tujuan utama dari penelitian ini adalah untuk mempelajari dan mendapatkan metode terbaik dalam mengklasifikasikan dan menganalisis sentimen yang dikombinasikan dengan tahap *pre-processing* dengan menggunakan alat bantu Gata Framework Textminig. Analisis komentar pembaca di media berita online tentang pemilihan tokoh politik yang mengikuti proses pemilihan. Penelitian ini juga diharapkan dapat memberikan solusi dalam hal analisis klasifikasi sentimen dari tokoh-tokoh politik dan mendukung pembahasan penelitian lebih lanjut tentang objek penelitian lain, seperti analisis sentimen pada tingkat kepuasan pelanggan di perusahaan ataupun analisis sentimen pada kepuasan pelayanan publik pemerintahan.

## 2. Metode Penelitian

Penelitian ini termasuk kedalam *machine learning*, lebih tepatnya pada bagian *text-mining*. Sumber data yang menjadi dataset pada penelitian ini bersumber dari komentar terhadap tokoh politik yang terdapat pada media berita online di Indonesia. Adapun yang menjadi metode penelitiannya adalah dengan menggunakan *Cross-Industry Standard Process for Data Mining*

(CRISP-DM). Metode tersebut memiliki beberapa tahapan seperti yang terlihat pada gambar 1 [20].



Gambar 1. Metode CRISP-DM

### 2.1. Business Understanding

Pada tahap *business understanding*, adalah bagian pemahaman tentang objek penelitian yang dilakukan. Pada penelitian ini pemahaman terhadap objek penelitian dilakukan dengan menggali informasi melalui beberapa situs media berita online dengan objek komentar terhadap tokoh politik. Motivasi dalam fase ini adalah berita yang disajikan biasanya berupa teks pada media *digital* yang dikelompokkan berdasarkan isi diskusi masing-masing kategori berita. Media online tidak hanya terbatas sebagai media untuk membaca artikel berita tetapi juga dapat digunakan untuk melihat masalah yang terjadi dan bahkan dapat digunakan untuk melihat elektabilitas tokoh politik. Analisis sentimen ini dilakukan untuk menemukan metode klasifikasi yang dapat membantu menentukan komentar artikel berita positif dan negatif. Pada tahap ini juga dibuat pemahaman untuk menemukan metode klasifikasi terbaik sehingga dapat membantu selama proses pengolahan data yang akan dilakukan dengan membandingkan hasil algoritma yang digunakan dan untuk meningkatkan kinerja metode klasifikasi dapat dilakukan dengan menggunakan pemilihan fitur (*feature selection*).

### 2.2. Data Understanding

Pada tahap *data understanding*, proses pengambilan data mentah dilakukan sesuai dengan atribut yang diperlukan. Data diperoleh dari situs media berita online yang terdiri dari [www.detik.com](http://www.detik.com), [www.tribunnews.com](http://www.tribunnews.com), [www.kompas.com](http://www.kompas.com), [www.merdeka.com](http://www.merdeka.com), [www.viva.co.id](http://www.viva.co.id) dan [www.kompasiana.com](http://www.kompasiana.com). Data dikumpulkan dari 16 Februari 2018, hingga 31 Mei 2018. Data primer yang diperoleh adalah 1480 data komentar dari artikel berita di media berita online, yaitu Detiknews dengan data 1393 komentar, Tribunnews dengan 8 data komentar, Kompas dengan 72 data komentar, Merdeka dengan 3

data komentar, Viva dengan 3 data komentar dan Kompasiana dengan 1 data komentar. Setelah data terkumpul maka proses pembersihan dilakukan, diperoleh data 1459 komentar. Penelitian ini menggunakan data komentar dalam Bahasa Indonesia.

### 2.3. Data Preparation

Tahap data preparation adalah tahap dengan proses persiapan data yang bertujuan untuk mendapatkan data bersih dan siap digunakan dalam penelitian. Pada tahap awal penambangan teks, tahap *pre-processing* teks akan dilakukan, pada tahap ini, peneliti akan menggunakan *Tools* Gata Framework Textmining.

Pada tahap ini peneliti akan melakukan beberapa proses preprocessing teks pada dataset komentar, antara lain: *Case folding*, *Tokenizing*, *Tagging*, *Filtering*, *Stemming*. Pembahasan tahapan-tahapan tersebut akan dijelaskan lebih mendetail pada bagian berikutnya.

### 2.4. Modeling

Merupakan tahap pemilihan teknik penambangan dengan menentukan algoritma yang akan digunakan. Penelitian ini menggunakan *tools* yang digunakan untuk melakukan pemodelan sesuai dengan teknik yang telah ditentukan, *tools* tersebut adalah RapidMiner versi 8.2. Penelitian ini menggunakan 2 algoritma klasifikasi sebagai modelnya. Algoritma klasifikasi yang digunakan yaitu, *Naïve Bayes* (NB) dan *Support Vector Machine* (SVM) yang dioptimasi dengan menggunakan *Particle Swarm Optimization* (PSO). Hasil pengetesan setiap model adalah untuk mengklasifikasikan artikel berita positif dan artikel berita negatif untuk mendapatkan nilai akurasi terbaik pada setiap algoritma.

### 2.5. Evaluation

Tahap *evaluation* bertujuan untuk menentukan kegunaan model yang telah berhasil dibuat pada langkah *modeling* sebelumnya. Penelitian ini menggunakan tahap evaluasi dengan *10 fold-cross validation*. Proses validasi terdiri dari dua subproses yaitu, data pelatihan (*training set*) dan data pengujian (*testing set*). Subproses pelatihan digunakan untuk melatih model algoritma yang sudah ditentukan sebelumnya pada tahap *modeling* dengan data pelatihan yang ada. Setelah model algoritma dilatih pada tahap subproses pelatihan, selanjutnya akan model tersebut akan dilakukan pengujian pada subproses pengujian.

Selanjutnya, setelah data pengujian dilakukan, proses silang dilakukan yaitu, data pengujian digunakan sebagai data pelatihan dan sebaliknya data pelatihan sebagai data pengujian, proses tersebut dilakukan sebanyak 10 kali untuk setiap bagian, sehingga setiap bagian dari 10 bagian telah diuji untuk model algoritma yang digunakan.

### 2.5. Deployment

Tahap *deployment* adalah tahap yang digunakan untuk membuat sebuah model implementasi yang dibuat dalam sebuah *tools* yang dapat dibangun dengan berbagai jenis pemrograman. Pembuatan model implementasi ini menggunakan hasil dari proses eksperimen dan evaluasi sebagai sumber data acuan.

## 3. Hasil dan Pembahasan

Penelitian ini menggunakan data yang telah diambil dari komentar terkait dengan tokoh politik yang bersumber dari 6 media berita online seperti yang telah disebutkan pada bagian *data understanding* diatas. Adapun data yang diambil secara keseluruhan adalah berjumlah 1480 data komentar. Kemudian data tersebut pada tahap awal akan dilakukan *data cleansing* dan dilakukan pemberian label dengan label positif atau negatif pada setiap komentar. Dari tahap awal tersebut didapatkan data sebanyak 1459 komentar yang telah berlabel maka, data tersebutlah yang akan menjadi dataset pada penelitian ini.

Dalam mengolah data untuk mendapatkan model yang sesuai dengan kasus penelitian ini yaitu, sentimen analisis tokoh politik dengan menggunakan algoritma klasifikasi *Naïve Bayes* dan *Support Vector Machine* digunakanlah sebuah *tools* Rapidminer versi 8.2. Karena penelitian ini termasuk ke dalam bagian dari *text-mining* maka, akan ada tahap yang harus dilakukan lebih dulu sebelum bisa didapatkan sebuah model yang baik pada studi kasus sentimen analisis komentar tokoh politik.

### 3.1. Pre-processing teks komentar

Pembahasan pada tahapan ini adalah proses awal pengolahan dataset sebelum dapat diproses untuk klasifikasi dengan algoritma *Naïve Bayes* (NB) dan *Support Vector Machine* (SVM) yang dioptimasi dengan *Particle Swarm Optimization* (PSO). Penelitian ini menggunakan beberapa tahapan *pre-processing* untuk dataset teks komentar, berikut ini adalah tahapan tersebut:

#### a. Case Folding

Tahap ini adalah proses untuk mengkonversi teks yang tidak beraturan dalam penggunaan huruf di penulisan teks komentar, sehingga teks komentar yang ditulis tidak konsisten. Proses *case folding* ini berfungsi untuk mengubah huruf pada teks komentar kedalam bentuk standar yaitu, menjadi huruf kecil seluruhnya. Hasil dari proses *case folding* tersebut dapat dilihat pada tabel 1.

#### b. Tokenizing

Tahap *Tokenizing* adalah tahap pemotongan *string input* berdasarkan tiap kata yang menyusunnya. Hasil dari proses tersebut dapat dilihat pada tabel 2.

Tabel 1. Perbandingan Teks Komentar Sebelum dan Sesudah di Proses *Case Folding*

Sebelum	Sesudah
DARIMANA LOGIKANYA SI AHY PALING CUCOK DAN ELEKTABILITAS TINGGI UTK CAWAPRES JOKOWI.. ANAK MASIH HIJAU DAN TIDAK PUNYA PENGALAMAN POLITIK DAN EKSEKUTIF NOL BESAR !!! NGACO..!!! #2019TETAPJOKOWI	darimana logikanya si ahy paling cucok dan elektabilitas tinggi utk cawapres jokowi.. anak masih hijau dan tidak punya pengalaman politik dan eksekutif nol besar !!! ngaco..!!! #2019tetapjokowi

Tabel 2. Perbandingan Teks Komentar Sebelum dan Sesudah di Proses *Tokenizing*

Sebelum	Sesudah
darimana logikanya si ahy paling cucok dan elektabilitas tinggi utk cawapres jokowi.. anak masih hijau dan tidak punya pengalaman politik dan eksekutif nol besar !!! ngaco..!!! #2019tetapjokowi	darimana logikanya si ahy paling cucok dan elektabilitas tinggi utk cawapres jokowi.. anak masih hijau dan tidak punya pengalaman politik dan eksekutif nol besar !!! ngaco..!!! #2019tetapjokowi

c. *Tagging*

Menghilangkan penyebutan nama pengguna yang disebutkan (@username), menghapus hastag (#hastag), menghapus tanda baca, menghapus nomor karena hanya data teks yang digunakan, menghapus tautan (http://), menghapus karakter khusus seperti simbol-simbol atau *emoticon*, menghilangkan kata-kata asing karena hanya mengambil kata-kata yang berbahasa Indonesia saja. Untuk kata yang di depannya terdapat kata “tidak” akan di normalisasi dengan menggunakan tanda garis bawah “\_” agar memiliki arti yang jelas, misalnya “tidak baik” menjadi “tidak\_baik” karena kata “tidak” bermakna negatif sedangkan kata “baik” bermakna positif. Hasil dari proses *tagging* tersebut dapat dilihat pada tabel 3.

Tabel 3. Perbandingan Teks Komentar Sebelum dan Sesudah di Proses *Tagging*

Sebelum	Sesudah
darimana logikanya si ahy paling cucok dan elektabilitas tinggi utk cawapres jokowi.. anak masih hijau dan tidak punya pengalaman politik dan eksekutif nol besar !!! ngaco..!!! #2019tetapjokowi	darimana logikanya si ahy paling cucok dan elektabilitas tinggi utk cawapres jokowi anak masih hijau dan tidak_punya pengalaman politik dan eksekutif nol besar ngaco

d. *Filtering*

Tahap *Filtering* adalah tahap mengambil kata-kata penting dari hasil *token*. Bisa menggunakan algoritma *stoplist* (membuang kata kurang penting) atau *wordlist* (menyimpan kata penting). Pada tahap ini untuk kata-kata slang (dalam Bahasa Indonesia) harus dinormalisasi ke dalam bentuk standarnya, misalnya “utk” menjadi “untuk”. Hasil dari proses *filtering* tersebut dapat dilihat pada tabel 4.

Tabel 4. Perbandingan Teks Komentar Sebelum dan Sesudah di Proses *Filtering*

Sebelum	Sesudah
darimana logikanya si ahy paling cucok dan elektabilitas tinggi utk cawapres Jokowi anak masih hijau dan tidak_punya pengalaman politik dan eksekutif nol besar ngaco	darimana logikanya si ahy cocok elektabilitas cawapres Jokowi hijau tidak_punya pengalaman politik eksekutif nol ngaco

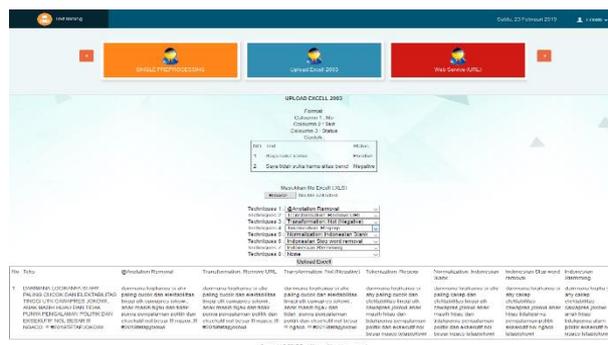
e. *Stemming*

Proses *stemming* adalah proses untuk mencari akar dari kata hasil dari proses *filtering*. Pencarian akar sebuah kata atau biasa disebut dengan kata dasar dapat memperkecil hasil indeks tanpa harus menghilangkan makna. Hasil dari proses *stemming* ini dapat dilihat pada tabel 5.

Tabel 5. Perbandingan Teks Komentar Sebelum dan Sesudah Proses *Stemming*

Sebelum	Sesudah
darimana logikanya si ahy cocok elektabilitas cawapres Jokowi hijau tidak_punya pengalaman politik eksekutif nol ngaco	darimana logika si ahy cocok elektabilitas cawapres Jokowi hijau tidak_punya pengalaman politik eksekutif nol ngaco

Tahapan-tahapan *pre-processing* teks komentar diatas diproses dengan menggunakan *tools* Gata Framework Textmining yang dapat diakses melalui <http://www.gataframework.com/textmining/>, untuk menggunakan *tools* tersebut dapat dalam *single* teks yang dapat langsung di masukan pada kolom yang tersedia di halaman *tools* tersebut, namun juga dapat dalam bentuk *file excel* yang kemudian diunggah. Berikut ini adalah tampilan halaman *pre-processing* teks dengan menggunakan *tools* Gata Framework Textmining, pada gambar 2.



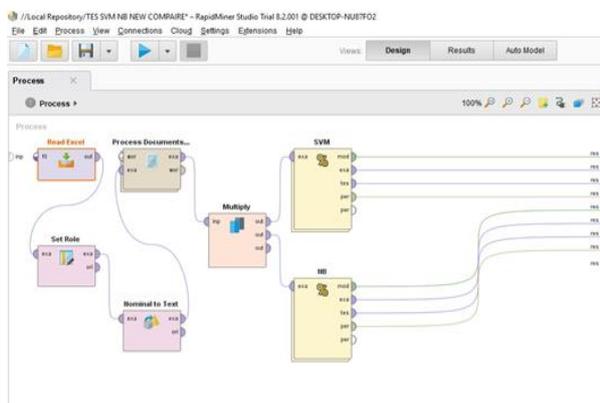
Gambar 2. Tampilan tools Gata Framework Textmining

3.2. Pembuatan Model Klasifikasi

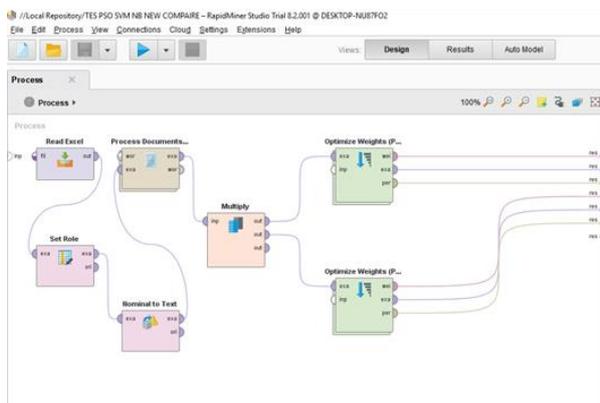
Tahap ini pada penelitian adalah untuk membuat model dengan menggunakan algoritma klasifikasi untuk dataset teks komentar yang sudah melalui tahap *pre-processing*. Tahap ini menggunakan dua algoritma klasifikasi yaitu, *Naive Bayes* (NB) dan *Support Vector Machine* (SVM) yang dioptimalkan dengan *Particle Swarm Optimization* (PSO). Penelitian ini

menggunakan *tools* Rapidminer versi 8.2 untuk mengolah dataset teks komentar yang sudah melalui tahap *data preparation* dengan *pre-processing* teks.

Tahap pertama dari proses ini adalah data teks komentar akan diunggah ke dalam *tools* dengan menggunakan file excel yang kemudian akan di proses dengan algoritma *Naive Bayes* (NB) dan *Support Vector Machine* (SVM) untuk mendapatkan hasil awal dari masing-masing algoritma tersebut, seperti yang dapat dilihat pada gambar 3. Setelah tahapan pertama dilakukan maka, penelitian ini melanjutkan dengan melakukan perbandingan kedua algoritma dengan menambahkan algoritma optimalisasi dengan menggunakan *Particle Swarm Optimization* (PSO). Langkah penggunaan PSO dalam proses modelling ini bertujuan untuk meningkatkan nilai akurasi hasil dari klasifikasi algoritma NB dan SVM, proses tersebut dapat dilihat pada gambar 4.



Gambar 3. Desain Model Menggunakan Algoritma NB dan SVM

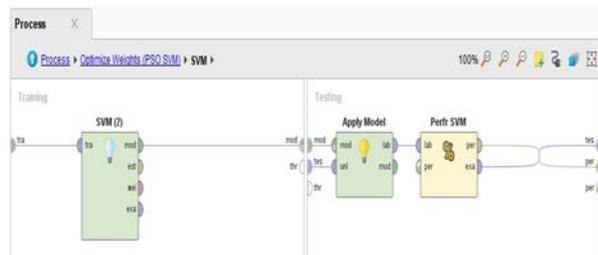


Gambar 4. Desain Model Menggunakan NB dan SVM dengan PSO

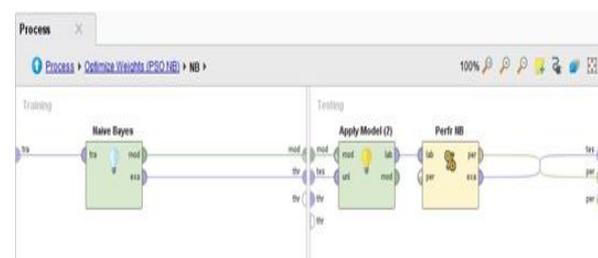
### 3.3. Evaluasi Model Klasifikasi

Tahapan evaluasi bertujuan untuk menentukan nilai kegunaan dari model yang telah berhasil dibuat pada langkah sebelumnya. Untuk evaluasi digunakan *10 fold-cross validation*. Pada pengujian ini, data digunakan adalah data bersih yang telah melalui *pre-processing*. Data tersebut diambil dari operator *Read Excel*, hal ini dilakukan karena dataset disimpan dalam bentuk Excel. *Process documents from files* untuk

mengkonversi file menjadi dokumen. Proses validasi terdiri dari data pelatihan dan data pengujian. Pada tahap ini juga menggunakan *Set Role* yang berfungsi untuk menentukan *field* pada kelas kemudian menggunakan *Particle Swarm Optimization* (PSO) agar akurasi yang dihasilkan lebih tinggi dari hasil pemodelan yang telah dilakukan sebelumnya. Desain model klasifikasi yang di evaluasi menggunakan PSO dapat dilihat pada gambar 5 dan 6.



Gambar 5. Desain Evaluasi *10 Fold Cross-Validation* untuk algoritma NB dengan PSO



Gambar 6. Desain Evaluasi *10-Fold Cross-Validation* untuk Algoritma SVM dengan PSO

Pada tahap ini juga menggunakan *Role Set* yang berfungsi untuk menentukan bidang di kelas kemudian menggunakan *Particle Swarm Optimization* (PSO) sehingga akurasi yang dihasilkan lebih tinggi dari hasil pemodelan sebelumnya. Kategori Klasifikasi memiliki nilai: 1). 0,90 - 1,00 = klasifikasi sangat baik. 2). 0,80 - 0,90 = klasifikasi yang baik. 3). 0,70 - 0,80 = klasifikasi yang adil. 4). 0,60 - 0,70 = klasifikasi buruk. 5). 0,50 - 0,60 = kegagalan. Sedangkan untuk perbandingan akurasi dan *Area Under Curve* (AUC) hasil dari kedua algoritma yang telah digunakan dapat dilihat pada tabel 6.

Tabel 6. Perbandingan Nilai Akurasi dan AUC

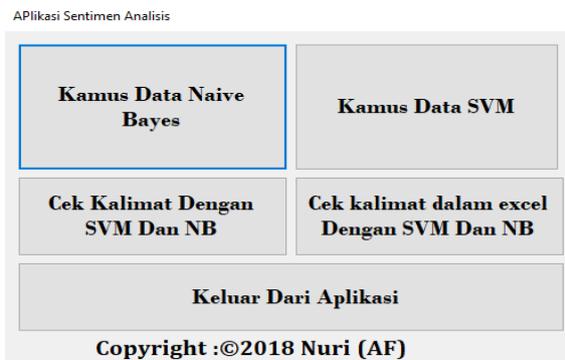
Algoritma	Akurasi	AUC
<i>Support Vector Machine</i>	76.09%	0.848
<i>Support Vector Machine</i> + PSO	78.40%	0.850
<i>Naive Bayes</i>	68.21%	0.672
<i>Naive Bayes</i> + PSO	74.98%	0.708

### 3.4. Model Implementasi

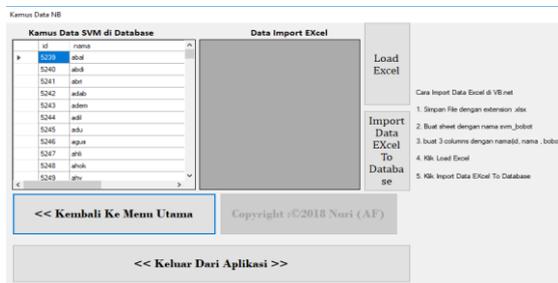
Tahap ini adalah merupakan tahapan terakhir dari penelitian ini yaitu, pembuatan model implementasi dari model algoritma klasifikasi yang sudah dilakukan sebelumnya. Model implementasi ini menggunakan hasil dari pengolahan dataset komentar tokoh politik

yang telah di proses dengan algoritma klasifikasi NB dan SVM yang dioptimalisasi menggunakan PSO. Kemudian untuk membuat model implementasinya penelitian ini menggunakan pemrograman VB.net dan database MySQL.

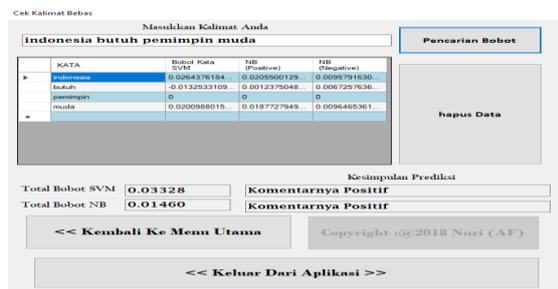
Pada model implementasi tersebut penelitian ini membuat empat menu utama yang dapat dilihat pada gambar 7 yang merupakan menu utama, gambar 8 menampilkan menu kamus data *Naive Bayes* dan *Support Vector Machine* yang didapat dari hasil perhitungan dengan menggunakan Rapidminer yang di masukan ke dalam database MySQL, gambar 9 merupakan gambar model implementasi untuk pengecekan kalimat secara masukan manual dan gambar 10 merupakan menu pengecekan kalimat yang di upload melalui file excel.



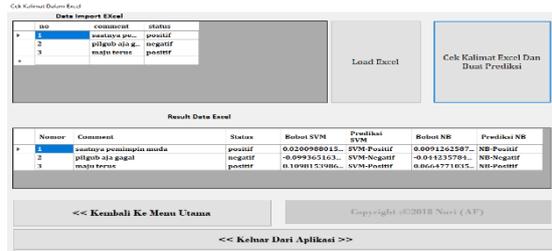
Gambar 7. Desain Menu Utama Model Implementasi



Gambar 8. Tampilan Menu Kamus Data



Gambar 9. Tampilan Menu Pengecekan Kalimat dengan Masukan Manual



Gambar 10. Tampilan Menu Pengecekan Kalimat dengan File Excel.

#### 4. Kesimpulan

Dari hasil penelitian yang telah dilakukan terkait dengan klasifikasi sentimen analisis untuk elektabilitas dari tokoh-tokoh politik pada komentar media berita online, dapat disimpulkan bahwa penggunaan Gata Framework Textmining sangat membantu dalam tahap *pre-processing* teks komentar bahasa Indonesia. Penggunaan Algoritma *Support Vector Machine* yang dioptimalisasikan dengan *Particle Swarm Optimization* (PSO) memiliki tingkat akurasi yang lebih baik daripada *Naive Bayes* yang dioptimalisasikan dengan *Particle Swarm Optimization* (PSO). Sementara *Particle Swarm Optimization* (PSO) memainkan peran penting dalam meningkatkan akurasi algoritma *Support Vector Machine* berbasis *Particle Swarm Optimization* (PSO) dapat menjadi solusi untuk klasifikasi dalam sentimen analisis elektabilitas tokoh politik. Berdasarkan hasil di atas, penelitian ini juga melakukan fase *deployment* berdasarkan akurasi algoritma *Support Vector Machine* berdasarkan *Particle Swarm Optimization* (PSO) menggunakan VB.net dan MySQL.

Selain kesimpulan di atas, penelitian ini juga berupaya memberikan saran untuk penelitian lebih lanjut, seperti melakukan pengujian dengan menggunakan algoritma yang berbeda dari penelitian ini dan tidak hanya melakukan analisis sentimen pada komentar yang berkaitan dengan tokoh politik atau media berita online namun juga bisa dengan topik yang berbeda misal seperti topik pelayanan pendidikan, pelayanan penjualan, pelayanan publik, dan lainnya.

#### Daftar Rujukan

- [1] Aryo P.W., 2018. Selektifitas Media Sosial Dalam Meningkatkan Elektabilitas Figur. *ETTISAL Journal of Communication*, 3(2), pp. 153-161.
- [2] Taibbi, M., 2019. *In 2019, Let's Finally Retire 'Electability' A favored totem of the campaign press has long been useless, and now is actively misleading* [Online] (Updated 2 Jan 2019) Tersedia di: <https://www.rollingstone.com/politics/politics-features/what-is-electability-774196/>
- [3] Jannti, R., Mahendra, R., Wardhana, W, C. Adriani, M, 2018. Stance Classification Towards Political Figures on Blog Writing. In: *IEEE, 2018 International Conference on Asian Language Processing (IALP)*. Bandung, Indonesia 15-17 November 2018. Indonesia.
- [4] Oxford Disctionaries, 2018. Sentiment Analysis. [Online] Tersedia di: [https://en.oxforddictionaries.com/definition/sentiment\\_analysis/](https://en.oxforddictionaries.com/definition/sentiment_analysis/)

- [5] Liu, B., 2015. *Sentiment Analysis Opinions, Sentiment, and Emotion in Text*. Chicago, USA: Cambridge University Press.
- [6] Saputra, N., Adji, T. B., Permanasari, A. E., 2015. Analisis Sentimen Data Presiden Jokowi dengan Preprocessing Normalisasi dan Stemming menggunakan Metode Naive Bayes dan SVM. *Jurnal Dinamika Informatika*, 5(1).
- [7] Li, Quanzhi., Nourbakhsh, A. Fang, Rui. Liu, Xiaomo. Shah, Sameena, 2016. Sentiment Analysis of Political Figures across News and Social Media. In: Behavioral-Cultural Modeling, & Prediction and Behavior Representation in Modeling and Simulation, 2016 *International Conference on Social Computing*. UDC Center, Washington DC, USA.
- [8] Budiharto W., Meiliana M., 2018. Prediction and analysis of Indonesia Presidential Election from Twitter using Sentiment Analysis. *Journal of Big Data* 5(1), pp.1-10.
- [9] Aliandu, Paulina, 2013. Twitter Used by Indonesian President: An Sentiment Analysis of Timeline. 2013 *Information Systems International Conference (ISICO)*. Surabaya, Indonesia 2-4 December 2013. Indonesia.
- [10] Apriyanti, T., Wulandari, H., Safitri, M., & Dewi, N., 2016. Translating Theory of English into Indonesian and Vice-Versa. *Indonesian Journal of English Language Studies* 2(1), pp. 38-59.
- [11] R. Feldman and J. Sanger, 2007. *The Text Mining Handbook: Advances Approaches in Analyzing Unstructured Data*. New York, NY: Cambridge University Press.
- [12] Goyal, Shubham, 2016. Sentimental Analysis of Twitter Data using Text Mining and Hybrid Classification Approach. *International Journal of Advance Research, Ideas and Innovations in Technology* 2(5), pp.1-9.
- [13] Vadivukarassi, M., Puviarasan, N., Aruna, P., 2017. Sentimental Analysis of Tweets Using Naive Bayes Algorithm. *World Applied Sciences Journal* 35(1), pp.54-59.
- [14] Al-Amrani, Y., Lazaar, M., El Kadiri, E.K., 2018. Sentiment Analysis Using Hybrid Method of Support Vector Machine And Decision Tree. *Journal of Theoretical and Applied Information Technology* 96(7), pp.1886-1895.
- [15] Wahyudi, M., Kristiyanti, D.A., 2016. Sentiment Analysis of Smartphone Product Review Using Support Vector Machine Algorithm-Based Particle Swarm Optimization. *Journal of Theoretical and Applied Information Technology* 91(1), pp.189-201.
- [16] Buntoro, G. A., 2016. Analisis Sentimen Calon Gubernur DKI Jakarta 2017 Di Twitter. *Integer Journal* Maret 1(1), pp. 32-41.
- [17] Budiono, D. F., Nugroho, A. S., Doewes, A., 2017. Twitter Sentiment Analysis of DKI Jakarta's Gubernatorial Election 2017 with Predictive and Descriptive Approaches. In: IEEE, 2017 *International Conference on Computer, Control, Informatics and Its Applications*. Jakarta, Indonesia 23-26 October 2017. Indonesia.
- [18] Fatyanosa, T. N., & Bachtiar, F. A., 2017. Classification method comparison on Indonesian social media sentiment analysis. In: IEEE, 2017 *International Conference on Sustainable Information Engineering and Technology (SIET)*. Malang, Indonesia 24-25 November 2017. Indonesia.
- [19] Rana, S., Singh, A., 2016. Comparative Analysis of Sentiment Orientation Using SVM and Naïve Bayes Techniques. In: IEEE, *International Conference on Next Generation Computing Technologies*. Dehradun, India 14-16 October 2016. India.
- [20] Kenneth Jensen, 2016. *IBM SPSS Modeler CRISP-DM Guide*. [Online] Tersedia di: <ftp://public.dhe.ibm.com/software/analytics/spss/documentation/modeler/18.0/en/ModelerCRISPDM.pdf>