PENINGKATAN OPTIMASI SENTIMEN DALAM PELAKSANAAN PROSES PEMILIHAN PRESIDEN BERDASARKAN OPINI PUBLIK DENGAN MENGGUNAKAN ALGORITMA *NAÏVE BAYES* DAN *PARICLE SWARM OPTIMIZATION*

Betesda
betesdasinaga@gmail.com
Universitas Dirgantara Marsekal Suryadarma

Abstract- *The development of increasingly advanced IT in the process of presidential elections. When the Presidential election of 2014 yesterday has a lot of people use the phrase does not educate inappropriate to be delivered among the public. Pros and cons indeed occur among people are so warm that they pour on the internet. This happens because when getting warm diperbincangan 2014 presidential election yesterday happened pengkubu-kubuan two candidates. Society can not adjust the development of IT process well. Naive Bayes is widely used for classification problems in data mining and machine learning for its simplicity and accuracy of classification impressive. Naive Bayes classifier has been shown to be very effective to solve the problem of large scale for text categorization with high accuracy. In addition to having many capabilities mentioned above, however this method has a drawback in the assumptions that are difficult to fulfill, namely the independence of the feature. Particle Swarm Optimization (PSO) is an evolutionary computation technique which is able to produce globally optimal solution in the search space through the interaction of individuals in a swarm of particles. PSO is widely used to solve optimization problems as well as the feature selection. Accuracy is generated on Naive Bayes algorithm amounted to 63.85% and AUC by 0523, while Naive Bayes and Particle Swarm Optimmization with an accuracy of 71.15% and the AUC of 0.600. It can be concluded that the application of optimization can improve the accuracy of 63.85% to 71.15%. Naive Bayes Model and Particle Swarm Optimization can provide solutions to the problems of classification review of public opinion news of the election in order to more accurately and optimally.*

Keywords:Public Opinion, Classification, Naive Bayes, Particle Swarm Optimization, Text Mining.

I. PENDAHULUAN

Perkembangan IT semakin maju dalam proses Pilpres. Saat pemilihan Presiden tahun 2014 kemarin telah banyak masyarakat menggunakan kalimat tidak men-didik yang tidak pantas untuk disampaikan dipublik. Pro dan kontra memang terjadi begitu hangat dikalangan masyarakat yang mereka tuangkan di internet. Hal ini terjadi semakin hangat diperbinca-ngan karena saat pilpres 2014 kemarin terjadi pengkubukubuan dua calon tersebut. Masyarakat tidak dapat menyesuaikan perkembangan proses IT dengan baik. Penyalahgunaan perkem-bangan infrastruktur teknologi saat ini sedang mengalami keterpurukan di-karenakan banyak masyarakat yang memposting kalimat tidak mendidik dalam suatu situs. Mereka menggunakan media elektronik untuk mencari informasi mengenai proses pilpres dan mulai memberikan aspirasi calon pilpres mereka masing-masing dari kedua partai tersebut. Seperti diketahui, sejumlah organisasi masyarakat sipil, seperti *Southeast Asia*

Freedom of Expression Network (Safenet), Elsam, dan Kontras, mendesak pemerintah segera merevisi UU ITE, khususnya pasal 27 ayat 3 tentang penghinaan dan pencemaran nama baik lewat media massa. Pasal itu, seringkali digunakan banyak pihak untuk menuntut secara pidana para pengkritiknya melalui dunia maya. Direktur Jenderal Aplikasi Informatika (Aptik) Kementerian Komunikasi dan Informatika (Kominfo) Ashwin Sasongko mengatakan, pemerintah akan merevisi Undang-Undang (UU) Informasi dan Transaksi Elektronik (ITE) tahun 2014 mendatang. Menurutnya, revisi UU No 11/2008 tentang Informasi dan Transaksi Elektronik (UU ITE) mengenai ancaman pidana. UU ITE sendiri terbit pada 25 Maret 2008 dengan cakupan meliputi globalisasi, perkembangan teknologi informasi, dan keinginan untuk mencerdaskan kehidupan bangsa. Meski mengandung banyak sisi positif, UU ITE dianggap banyak pihak memiliki sejumlah pasal karet dan kejanggalan. Positifnya UU ITE memberikan peluang bagi bisnis baru di Indonesia karena penyelenggaraan sistem elektronik diwajibkan berbadan hukum dan berdomisili di Indonesia. UU ITE, juga dapat mengantisipasi kemungkinan penyalahgunaan internet yang merugikan, memberikan perlindungan hukum terhadap transaksi dan sistem elektronik, dan memberikan perlindungan hukum terhadap kegiatan ekonomi misalnya *e-tourism*, *e-learning*, imple-mentasi EDI, dan transaksi dagang. UU itu juga memungkinkan kejahatan yang dilakukan oleh seseorang di luar Indonesia dapat diadili. Selain itu, UU ITE juga membuka peluang kepada pemerintah untuk mengadakan program pemberdayaan internet. Beberapa tahun terakhir, pengguna internet telah berkembang sangat pesat. Banyak forum, blog, jejaring sosial, situs web *e-commerce*, dan laporan berita berfungsi sebagai bentuk untuk mengekspresikan pendapat, yang dapat dimanfaatkan untuk memahami pendapat masyarakat umum dan konsumen pada peristiwa sosial, politik, strategi perusahaan, preferensi produk, dan reputasi pemantauan (Saleh, 2011). Penggunaan situs dengan *world wide web* menghadirkan fitur *hyperlink* yang memberikan kemudahan kepada pembaca untuk menelusuri informasi-informasi lanjutan mengenai topik dalam sebuah berita di lokasi yang berbeda. Penggunaan situs memungkinkan penyebaran lebih cepat, aktual, lebih murah, dan ramah lingkungan (Andri, 2014).

Analisa sentimen atau *opinion mining* adalah studi komputasi mengenai pendapat, perilaku dan emosi seseorang terhadap entitas. Entitas tersebut dapat menggambarkan individu, kejadian atau topik. Topik tersebut kemungkinan besar dapat berupa review (Medhat, Hassan dan Korashy, 2014). Analisis sentimen akan mengelompokkan polaritas dari teks yang ada dalam kalimat atau dokumen untuk mengetahui pendapat yang dikemukakan dalam kalimat atau dokumen tersebut apakah bersifat positif, negatif atau netral (Pang dan Lee, 2008).

Terdapat beberapa penelitian yang sudah dilakukan dalam melakukan klasifikasi sentimen terhadap *review* yang tersedia secara *online* diantaranya, Analisa sentimen pada opini *review* film menggunakan pengklasifikasi *Support Vector Machine* dan *Particle Swarm Optimization* (Basari, et al, 2013). Sentimen analisis terhadap *social media* (Habernal, Ptáček, Steinberger, 2015). Sentimen analisis terhadap konten berita (Kurniawan, et al, 2012). Pengklasifikasian sentimen pada *review* restoran di internet yang ditulis dalam bahasa Canton menggunakan pengklasifikasi *Naïve Bayes* dan *Support Vector Machine* (Zhang, et al, 2011), sedangkan penulis

akan melakukan penelitian terhadap sentimen analisis opini publik berita pilpres dengan menggunakan algoritma *Naive Bayes* dan *Particle Swarm Optimization*.

Naive Bayes banyak digunakan untuk klasifikasi masalah dalam *data mining* dan *machine learning* karena kesederhanaan dan akurasi klasifikasi yang mengesankan (D. Farid et al, 2014). *Naive Bayes Classifier* telah terbukti sangat efektif untuk memecahkan masalah skala besar untuk kategorisasi teks dengan akurasi yang tinggi (Kumar, Zayaraz, 2015). Selain memiliki banyak kemampuan yang telah disebutkan diatas, Namun metode ini memiliki kelemahan dalam asumsi yang sulit dipenuhi, yaitu *independensi feature* kata (Hamzah, 2012).

Particle Swarm Optimization (PSO) merupakan teknik komputasi evolusioner yang mampu menghasilkan solusi secara global optimal dalam ruang pencarian melalui interaksi individu dalam segerombolan partikel (Shuzhou & Bo, 2011). PSO banyak digunakan untuk memecahkan masalah optimasi serta pada seleksi fitur (Liu, et al., 2011).

Untuk itu, penelitian ini menggunakan *Particle Swarm Optimization* sebagai seleksi fitur untuk review opini publik berita pilpres dengan *Naïve Bayes*.

II. LANDASAN TEORI

2.1. Analisa Sentimen (*Sentiment Analysis*)

Sentiment Analysis atau *opinion mining* mengacu pada bidang yang luas dari pengolahan bahasa alami, komputasi linguistik dan *text mining* yang bertujuan menganalisa pendapat, sentimen, evaluasi, sikap, penilaian dan emosi seseorang apakah pembicara atau penulis berkenaan dengan suatu topik, produk layanan, organisasi, individu, ataupun kegiatan tertentu (Liu, 2011).

Tujuan dari analisa sentimen adalah untuk menentukan perilaku atau opini dari seorang penulis dengan memperhatikan suatu topik tertentu. Perilaku bisa mengindikasikan alasan, opini atau penilaian, kondisi kecenderungan (Basari, A., 2013). *Sentiment analysis* juga dapat menyatakan perasaan emosional sedih, gembira, atau marah.

Menurut Moraes (Moraes et al., 2013) langkah-langkah yang umumnya ditemukan pada klasifikasi teks analisa sentimen adalah:

1. Definisikan *domain dataset*
   Pengumpulan dataset yang melingkupi suatu *domain*, misalnya *dataset review* film, *dataset review* produk, dan lain sebagainya.
2. *Pre-processing*
   Tahap pemrosesan awal yang umumnya dilakukan dengan proses *Tokenization, Stopwords removal,* dan *Stemming.*
3. *Transformation*
   Proses representasi angka yang dihitung dari data tekstual. *Binary representation* yang umumnya digunakan dan hanya menghitung kehadiran atau ketidakhadiran sebuah kata di dalam dokumen. Berapa kali sebuah kata muncul di dalam suatu dokumen juga digunakan sebagai skema pembobotan dari data tekstual. Proses yang umumnya digunakan yaitu TF-IDF, *Binary transformation*, dan *Frequency transformation.*
4. *Feature Selection*
   Pemilihan fitur (*feature selection*) bisa membuat pengklasifikasi lebih efisien/efektif dengan mengurangi jumlah data untuk dianalisa dengan mengidentifikasi fitur yang relevan yang selanjutnya akan diproses.

Metode pemilihan fitur yang biasanya digunakan adalah Expert. Knowledge, Minimum Frequency, Information gain, Chi-Square, dan lain sebagainya.

5. *Classification*
   Proses klasifikasi umumnya menggunakan pengklasifikasi seperti Naïve Bayes, Support Vector Machine, dan lain sebagainya.
6. *Interpretation/Evaluation*
   Tahap evaluasi biasanya menghitung akurasi, recall, precision, dan F-1

2.2. Text Mining

Text mining atau *text analytics* adalah istilah yang mendeskripsikan sebuah teknologi yang mampu menganalisis data teks semi-terstruktur maupun tidak terstruktur, hal inilah yang membedakannya dengan *data mining* dimana *data mining* mengolah data yang sifatnya terstruktur. Pada dasarnya, *text mining* merupakan bidang interdisiplin yang mengacu pada perolehan informasi (*information retrieval*), *data mining*, pembelajaran mesin (*machine learning*), statistik, dan komputasi linguistik (Jiawei, Kamber, dan Pei, 2012).

Text mining umumnya mencakup kategorisasi informasi atau teks, mengelompokkan teks, ekstraksi entitas atau konsep, pengembangan dan perumusan taksonomi umum. *Text mining* berkenaan dengan informasi terstruktur atau tekstual ekstraksi informasi yang bermakna dan pengetahuan dari jumlah besar teks (Hashimi, Alaaeldin, dan Hassan, 2014).

Text mining adalah penambangan yang dilakukan oleh komputer untuk mendapatkan sesuatu yang baru, sesuatu yang tidak diketahui sebelumnya atau menemukan kembali informasi yang tersirat secara implisit, yang berasal dari informasi yang diekstrak secara otomatis dari sumber-sumber data teks yang berbeda-beda (Feldman dan Sanger, 2007). *Text mining* merupakan teknik yang digunakan untuk menangani masalah klasifikasi, *clustering, information extraction* dan *information retrival* (Berry dan Kogan, 2010).

Text mining dapat menganalisis dokumen, mengelompokkan dokumen berdasarkan kata-kata yang terkandung di dalamnya, serta menentukan kesamaan di antara dokumen untuk mengetahui bagaimana mereka berhubungan dengan variabel lainnya (Statsoft, 2015).

Dari ke lima pendapat ahli diatas, maka dapat disimpulkan bahwa *text mining* adalah informasi terstruktur yang digunakan untuk menganalisis atau mengelompokkan dokumen atau teks dari sejumlah besar dokumen atau teks.

Beberapa tahun terakhir, penggunaan dan penelitian mengenai *text mining* telah banyak mendapat perhatian dan aktif dilakukan seiring dengan semakin banyaknya data teks yang diperoleh dari berbagai jaringan sosial, web, dan aplikasi lainnya. Sebagian besar informasi teks yang disimpan tersebut seperti artikel berita, makalah, buku, perpustakaan digital, pesan email, blog, dan halaman web.

2.3. Review Opini Publik Berita Pilpres

Ulasan atau review yang terdapat di internet sangat banyak namun tidak diolah menjadi sebuah informasi yang bermanfaat. Kini konsumen semakin meningkat sehingga mereka dapat memberikan opini dan pengalaman yang tersedia secara online (Horrigan, 2008).

Review pilpres yang digunakan dalam penelitian ini, menggunakan data yang bersumber dari pemilu.com.

Mayoritas rakyat Indonesia tahu bahwa pemilihan umum (pemilu) akan diselenggarakan pada April 2014, tetapi beberapa orang masuih memiliki beberapa pertanyaan mengenai pemilu. Mungkin mereka ingin tahu logo partai politik, profil parlemen atau calon presiden, atau mungkin ingin tahutentang mekanisme bagaimana voting mereka akan dihitung menjadi kursi parlemen yang terpilih. Untungnya ada sebuah aplikasi untuk Android yang menyajikan banyak informasi tentang pemilu. Nama app tersebut adalah PEMILU (http://www.kejut.com/pemilu).

2.4. Algoritma *Particle Swarm Optimization* (PSO)

Particle Swarm Optimization sebagai teknik evolusi pertama dikemukakan oleh J Kennedy dan R. Eberhart pada tahun 1995. Seorang Psikolog dan Insinyur Listrik di Amerika Serikat (Kenndey dalam Lie et al, 2014).

Particle Swarm Optimization (PSO) untuk memecahkan masalah optimasi global dalam bentuk algoritma metaheuristik paralel. Dalam beberapa tahun terakhir, banyak algoritma metaheuristik telah maju. PSO adalah salah satu dari mereka, sangat efektif untuk memecahkan masalah ini. Tapi PSO memiliki beberapa kekurangan seperti konvergensi prematur dan terjebak dalam minimum lokal. Untuk mengatasi kekurangan ini, banyak varian PSO telah diusulkan. (Chen et al, 2009).

Algoritma *Particle Swarm Optimization* (PSO) adalah teknik optimasi berdasarkan populasi yang terinspirasi oleh perilaku sosial dari pergerakan burung atau ikan (*bird flocking* atau *fish schooling*). PSO sebagai alat optimasi menyediakan prosedur pencarian berbasis populasi dimana masing-masing individu yang disebut partikel mengubah posisi mereka terhadap waktu (Rosita, Yudhi, dan Rully, 2012). Pada sistem PSO, masing-masing partikel terbang mengitari ruang pencarian multi dimensional (*multidimensional search space*) dan menyesuaikan posisinya berdasarkan pengalaman pribadinya dan pengalaman partikel di sebelahnya. Tiap partikel memiliki posisi $x_i = (x_{i1}, x_{i2}, ..., x_{iN})$ dan kecepatan $v_i = (v_{i1}, v_{i2}, ..., v_{iN})$ pada ruang pencarian berdimensi *N*, dimana *i* menyatakan partikel ke-i dan *N* menyatakan dimensi ruang pencarian datau jumlah variabel yang belum diketahui pada sistem persamaan nonlinear. Inisialisasi algoritma PSO dimulai dengan menetapkan posisi awal partikel secara acak (solusi) dan kemudian mencari nilai optimal dengan memperbarui posisinya. Seperti yang telah dijelaskan di atas, setiap iterasi masing-masing partikel memperbarui posisinya mengikuti dua nilai terbaik, yaitu solusi terbaik yang telah didapat oleh masing-masing partikel (pbest) dan solusi terbaik pada populasi (gbest). Setelah mendapatkan dua nilai terbaik, posisi dan kecepatan partikel diperbarui dengan menggunakan persamaan berikut:

$$v_k^i = wv_i^k + c_1 r_1 (pbest_i^k - x_i^k) + c_2 r_2 (gbest^k - x_i^k)$$

$$x_i^{k+1} = x_i^k + v_i^{k+1}$$

dimana v_i^k adalah kecepatan partikel ke *i* pada iterasi ke *k*, dan x_i^k adalah solusi (posisi) partikel ke *i* pada iterasi ke *k*, c_1, c_2 adalah konstanta positif, dan r_1, r_2 adalah dua variabel acak terdistribusi *uniform* antara 0 sampai 1. Pada persamaan di atas, *w* adalah bobot inersi yang menunjukkan pengaruh perubahan kecepatan dari vektor lama ke vektor yang baru.

2.5. Algoritma Naive Bayes

Naive Bayes adalah metode sederhana dan banyak digunakan untuk

supervised learning. Salah satu algoritma dengan pembelajaran tercepat dan dapat menangani sejumlah *feature* atau *class.* Meskipun sederhana dalam model tapi Naive Bayes bekerja dengan baik untuk ssetiap masalah. *Naive Bayes* berasal dari teorema Bayes untuk menghitung banyaknya label *class* dari *instance* baru, karena semua feature dianggap independen yang dapat memberikan nilai pada class. (Lee, Hwan, 2015).

Naive Bayes Classifier telah terbukti sangat efektif untuk memecahkan masalah skala besar untuk kategorisasi teks dengan akurasi yang tinggi (Kumar, Zayaraz, 2015). *Naive Bayes* memungkinkan klasifikasi berdasarkan asumsi kondisi tersendiri antara prediksi attributes diberikan *class*. Untuk itu *Naive Bayes* adalah klasifikasi yang benar-benar kompeten, bekerja cukup baik dalam tugas-tugas klasifikasi sehingga banyak peneliti yang mencoba untuk meningkatkan performa *Naive Bayes* (Bermejo et al, 2014).

Metode *Naïve Bayes Classification* (NBC) Metode NBC menempuh dua tahap dalam proses klasifikasi teks, yaitu tahap pelatihan dan tahap klasifikasi (Awaludin, 2015). 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. Selanjut-nya adalah penentuan probabilitas prior bagi tiap kategori berdasarkan sampel dokumen. Pada tahap klasifikasi ditentu-kan nilai kategori dari suatu dokumen berdasarkan term yang muncul dalam dokumen yang diklasifikasi. Lebih kon-kritnya jika diasumsikan dimiliki koleksi dokumen D={di|i=1,2,…|D|}={d1,d2,…, d|D|} dan koleksi kategori V={vj|j=1,2,… |V|}={v1,v2,…,v|V|}. Klasifikasi NBC dilakukan dengan cara mencari probabilitas P(V=vj | D=di), yaitu probabilitas category vj jika diketahui dokumen di . Dokumen di dipandang sebagai tuple dari kata-kata dalam dokumen, yaitu , yang frekuensi kemunculannya diasumsikan sebagai variable random dengan distribusi probabilitas Bernoulli.

Selanjutnya klasifikasi dokumen adalah mencari nilai maksimum dari:
VMAP = argmax (| , ,...,) 1 2 vj V j a a an P v ∈ .....(2.3)
Teorema Bayes menyatakan tentang probabilitas bersyarat menyatakan:
P(B|A)= P(A) P(A | B)P(B) .... (2.Dengan menerapkan teorema Bayes persamaan (2.3) dapat ditulis :
VMAP = (, ,..,) (, ,.., |) ( ) argmax 1 2 1 2 vj V n n j j P a a a P a a a v P v ∈(2.5)
Karena nilai (, ,...,) 1 2 n P a a a untuk semua vj besarnya sama maka nilainya dapat diabaikan, sehingga persamaan (2.5) menjadi :
VMAP = argmax (, ,..., |) ( ) 1 2 vj V n j j P a a a v P v ∈ .....(2.6)
Dengan mengasumsikan bahwa setiap kata dalam adalah independent, maka (, ,..., |) 1 2 n j P a a a v dalam persamaan (2.6) dapat ditulis sebagai berikut:
(, ,..., |) 1 2 n j P a a a v = ∏i i j P(a | v) .....(2.7)

$$P(Y = y_k|X_1, \ldots, X_n) = \frac{P(Y = y_k)\prod_i P(X_i|Y = y_k)}{\sum_j P(Y = y_j)\prod_i P(X_i|Y = y_k)}$$

Sehingga persamaan (2.6) dapat ditulis:
VMAP = ∈ ∏i j i j argmax P(v) P(a | v) vj V .....(2.8)
Nilai P(vj) ditentukan pada saat pelatihan, yang nilainya didekati dengan :
P(vj) = Contoh docJ ....(2.9)
Dimana docJ adalah banyaknya dokumen yang memiliki kategori j dalam pelatihan, sedangkan Contoh banyaknya dokumen

dalam contoh yang digunakan untuk pelatihan. Untuk nilai (|) k j P w v , yaitu probabilitas kata wk dalam kategori j ditentukan dengan

(|) k j P w v = n vocabulary nk + +1 .....(2.10)

dimana nk adalah frekuensi munculnya kata wk dalam dokumen yang ber kategori vj , sedangkan nilai n adalah banyaknya seluruh kata dalam dokumen berkategori vj, dan vocabulary adalah banyaknya kata dalam contoh pelatihan (Hamzah, 2012).

Sedangkan Naive Bayes menurut (Muralidharan dan Sugumaran, 2012) yaitu klasifikasi algoritma berdasarkan Bayes rules, dengan asumsi atribute X_1, ..., X_n disebut dengan conditional independent untuk semuanya, dengan memberikan nilai Y. Nilai tersebut diasumsikan untuk menyederhanakan representasi dari P(X/Y) dan problem estimasi dari data training.

Berikut contoh, dimana $X=(X_1,X_2)$, berikut persamaan matematika:

$P(X/Y=P(X_1X_2/Y)=P(X_1/X_2,Y)=P(X_2/Y)=P(X_1/Y)P(X_2/Y)$.....(2.11)

Lebih umumnya ketika X mengandung n attribute yang termasuk conditional independent dari satu yang diberikan nilai Y.

$$P(X_1,\ldots,X_n|Y)=\prod_{i=1}^{n}P(X_i|Y)$$

.....(2.12)

Memberitahukan bahwa ketika nilai Y dan X adalah variabel boolean, yang hanya 2^n parameter dibutuhkan untuk menetapkan $P(X_i=X_{ik}|Y=y_j)$. Ekspresi

$$P(Y=y_k|X_1,\ldots,X_n)=\frac{P(Y=y_k)P(X_1,\ldots,X_n|Y=y_k)}{\sum_j P(Y=y_j)P(X_1,\ldots,X_n|Y=y_j)}$$

untuk kemungkinan nilai Y termasuk ke dalam persamaan Bayes Rule berikut ini:.....(2.13)

Dimana jumlah diberikan untuk untuk nilai y_j pada Y yang memungkinkan. Sekarang asumsi X_i adalah *conditional independent* diberikan Y. Dapat ditulis-kan pada persamaan berikut:

.....(2.14)

Persamaan di atas merupakan dasar dari Naive Bayes Classifier. Pemberian *instance* yang baru $X_{new}=(1,...,X_n)$, persamaan tersebut mengkalkulasi proba-bilitas nilai Y yang akan diberikan nilai, diberikan observasi attribute nilai X_{new} dan diberikan distribusi P(Y) dan $P(X_i/Y)$ dan estimasi dari data training. Hal tersebut memungkinkan nilai Y di-temukan, maka dengan ini, persamaan Naive Bayes Classifier menjadi:

$$y \leftarrow \arg\max_{y_k} \frac{P(Y=y_k)\prod_i P(X_i|Y=y_k)}{\sum_j P(Y=y_j)\prod_i P(X_i|Y=y_k)}$$

.....(2.15)

Berdasarkan penjelasan di atas Naive Bayes menurut penulis adalah salah satu metode *supervised leraning* yang ber-fungsi untuk menangani fitur dan kelas pada pemecahan masalah dengan klasifikasi teks yang besar sehingga menghasilkan nilai akurasi yang tinggi.

2.6. Tinjauan Studi Penelitian Terkait

Berikut merupakan ringkasan dari penelitian terkait yang dijadikan peneliti sebagai panduan untuk penelitian ini:

Tabel II.1 Tinjauan Studi Penelitian Terkait

Judul	Peneliti	Classifier and Feature Selection	Akurasi	Hasil
Senti-lexicon and improved Naïve Bayes algorithms for sentiment analysis of restaurant reviews	Kang	NB, SVM + unigrams +bigrams	81.3%	Peneliti menghasilkan bahwa SVM memiliki kerja yang optimal
Performance of KNN and SVM classifiers on full word Arabic articles	Hmeidi et al	SVM, KNN, TF +IDF	Recall KNN=0.96, SVM=1	Peneliti menghasilkan pengujian data dengan SVM +TFIDF memiliki akurasi tertinggi

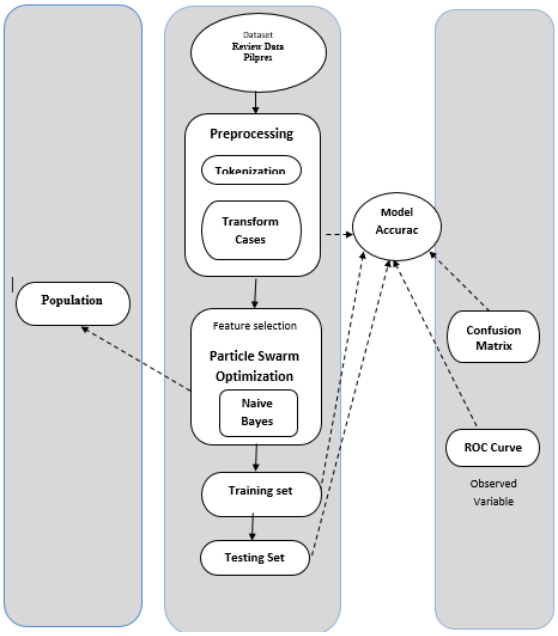
An improved K-nearest-neighbor algorithm for text categorization	Jiang	KNN	83%	Peneliti mengusulkan algoritma INNTC yang memiliki akurasi tertinggi
Chinese text classification by the Naïve Bayes Classifier and the associative classifier with multiple confidence threshold values	Hwa Lu et al	NBC, CAR	91.58%	Peneliti mengusulkan metode yang baru sehingga terbukti dapat meningkatkan akurasi dibandingkan metode classifier yang telah ada.
Opinion Mining of Movie Review using Hybrid Method of Support Vector Machine and *Particle Swarm Optimization*	Basari et al	Hybrid SVM PSO	77%	Peneliti menggunakan PSO sebagai optimasi dengan 10 Fold-Cross Validation
A novel hybrid system for feature selection based on an improvedgravitational search algorithm and k-NN method	Xiang dan Han	GSA, k-NN	83.9%	Peneliti banyak melakukan eksperimen terhadap beberapa banyak dataset, yang diambil contoh dari pengujian k-NN dan GA, ternyata menghasilkan akurasi yang cukup optimal.
Penyalahgunaan Perkembangan Infrastruktur Teknologi Dalam Pelaksanaan Proses Pemilihan Presiden Berdasarkan Opini Publik Dengan Menggunakan Algoritma Naive Bayes dan *Particle Swarm Optimization*	Peneliti	NB PSO	Sedang Diteliti	Peneliti melakukan pengujian terhadap review opini publik tentang berita *Pilpres*, menggunakan Algoritma NB berbasis PSO

Sumber: (Hasil Penelitian, 2015)

2.7. Kerangka Pemikiran

Berdasarkan pada latar belakang, maka penelitian ini akan melakukan penerapan metode Naive Bayes dalam pengklasifikasian. Dari klasifikasi tersebut akan dioptimasi kembali oleh fitur seleksi PSO agar nilai akurasi yang didapat menjadi lebih optimal dan baik. Peneliti mengambil data dari sumber http://www.pemilu.com/jokowi-vs-prabowo-pilpres-2014/ yang terdiri dari beberapa *review* atau pendapat masyarakat mengenai *Pilpres*. Peneliti mengambil sample data secara simple random sebanyak 130 review positif dan 130 *review* negatif. Sebelum data diklasifikasi, terlebih dahulu dilakukan preprocessing antara lain: *Tokenization* dan *Transform Cases.* Dalam pembobotan yang peneliti lakukan adalah Term Frequency Invers Document Fruquency (TF-IDF) dan pemilihan seleksi fitur menggunakan *Particle Swarn Optimization* (PSO). Sedangkan klasifikasi yang digunakan adalah Naive Bayes (NB). Software yang digunakan untuk mengolah data klasifikasi adalah RapidMiner sebagai alat bantu dalam mengukur akurasi data eksperimen. RapidMiner sangat terkemuka di dunia dan tidak perlu dipertanyakan lagi sebagai sistem sumber terbuka untuk *data mining*. RapidMiner umumnya dikenal dengan YALE (*Yey Another Learning Envirotment)* adalah perangkat lunak *open source* untuk *knowledge discovery* dan *data mining* merupakan mesin pembelajaran algoritma yang dikembangkan oleh *University of Dortmund, Germany* pada tahun 2001. RapidMiner memiliki lebih dari 400 prosedur (operator) *data mining*, termasuk operator untuk masukan, keluaran, data *pre-processing* dan visualisasi. Ribuan aplikasi telah banyak dikembangkan di lebih 40 negara, baik dalam dunia bisnis maupun penelitian (Rapid-I Gmbh.2010).

Berikut merupakan kerangka pemikiran pada penelitian tesis ini:



Gambar II.1 Kerangka Pemikiran

III. METODOLOGI PENELITIAN

3.1. Perancangan Penelitian

Pada dasarnya, penelitian merupakan suatu investigasi yang terorganisasi, yang dilakukan untuk menyajikan suatu

informasi dan memecahkan masalah. Metode penelitian yang digunakan penulis menggunakan metode penelitian eksperimen. Adapun metode penelitian yang penulis gunakan melalui beberapa tahapan sebagai berikut:

1. Pengumpulan Data
   Data yang digunakan untuk melakukan eksperimen dikumpulkan melalui website *newsmedia.co.id, kapanlagi.com* dan *tribunnews.com*, kemudian data opini publik berita artis tersebut diseleksi dan dikumpulkan ke dalam notepad untuk diolah dalam pengujian data.
2. Pengolahan Data awal
   Memilih metode yang akan digunakan pada saat pengujian data. Metode yang dipilih, berdasarkan penelitian yang terdahulu. Penulis menggunakan Metode Algoritma *Support Vector Machine*.
3. Metode yang Diusulkan
   Metode yang diusulkan penulis ditambahkan optimasi agar dapat meningkatkan nilai akurasi. Optimasi yang digunakan yaitu *Particle Swarm Optimization* (PSO).
4. Eksperimen dan Pengujian Metode
   Eksperimen yang dilakukan peneliti, menggunakan framework RapidMiner 6.4 untuk mengolah data sehingga menghasilkan nilai akurasi yang akurat dan untuk pengujian metode penulis membuat aplikasi menggunakan bahasa pemrograman PHP dan HTML.
5. Evaluasi dan Validasi Hasil Evaluasi
   Evaluasi berfungsi untuk mengetahui akurasi dari model algoritma yang diusulkan. Validasi digunakan untuk melihat perbandingan hasil akurasi dari model yang digunakan dengan hasil yang telah ada sebelumnya. Teknik validasi yang digunakan adalah *Cross Validation*. Akurasi algoritma akan diukur menggunakan *Confusion Matrix* dan hasil perhitungan akan ditampilkan dalam bentuk *Curve ROC (Receiver Operating Characteristic).*

3.2. Pengolahan Data Awal

Teks Mining adalah suatu proses yang bertujuan untuk menemukan informasi atau tren terbaru yang sebelumnya tidak terungkap, dengan memproses dan menganalisa data dalam jumlah besar. Dalam menganalisa sebagian atau keseluruhan *unstructured text, text mining* mencoba untuk mengasosiasikan satu bagian teks dengan yang lainnya berdasarkan aturan-aturan tertentu (Kunaifi, 2009). Teks yang belum diolah biasanya memiliki karakteristik dimensi yang tinggi, terdapat *noise* pada data dan terdapat struktur teks yang tidak baik. Untuk itu, dalam pengolahan data awal, teks mining harus melalui beberapa tahapan yang disebut dengan *preprocessing*. Tahapan-tahapan tersebut yaitu:

1. *Tokenization*
   Proses memootong setiap kata dalam teks dan mengubah huruf dalam dokumen menjadi huruf kecil. Hanya huruf yang diterima, sedangkan karakter khusus atau tanda baca akan dihilangkan. Jadi hasil dari proses tokenization adalah kata-kata yang merupakan penyusun kalimat atau string yang dimasukan tanpa ada tanda baca.
2. *Transform Cases*
   Merubah seluruh huruf menjadi huruf kecil atau kapital semua.

IV. PEMBAHASAN

4.1. Hasil Penelitian

Data training yang digunakan pada saat pengujian data diambil dari *pemilu.com*. Pengujian data, dilakukan dengan menggunakan review berita pilpres (260 *data training*, yang terdiri dari 130 review negatif dan 130 review positif) kemudian dilakukan *testing* dan *training dataset* sehingga didapatkan

accuracy dan AUC. Berikut akan dijelaskan lebih rinci mengenai hasil penelitian yang diperoleh.

Berikut merupakan tahapan-tahapan dalam melakukan pengolahan data yaitu:

1. Pengumpulan Data
   Review berita artis, masing-masing dikelompokkan dengan cara disimpan ke dalam satu folder yaitu folder positif dan folder negatif, kemudian tiap dokumennya diberikan ekstensi .txt sehingga dapat dibuka dengan aplikasi Notepad.

2. Pengolahan Data Awal *(Preprocessing)*
   Berikut merupakan tahapan yang dilakukan dalam *preprocessing:*
   a. *Tokenization*
      Dalam proses *tokenization* ini, semua kata yang ada di dalam tiap dokumen dikumpulkan dan dihilangkan tanda baca, serta dihilangkan jika terdapat simbol, karakter khusus atau apapun yang bukan huruf.

Tabel IV.1 Perbandingan teks sebelum dan sesudah dilakukan proses *Tokenization*

Teks sebelum dilakukan proses *tokenization*	hanya org2 yang ingin negaranya hancur milih probowo track racordnya gak jelas, mantan jendral yg kena pecat , antek2 suharto, para2 partai yang bermasalah bersembunyi di balik ketiaknya inilah negara kita yng salah diagungkan yg baik di singkirakan,.,,...:)
Teks setelah dilakukan proses *tokenization*	hanya org yang ingin negaranya hancur milih probowo track racordnya gak jelas mantan jendral yg kena pecat antek suharto para partai yang bermasalah bersembunyi di balik ketiaknya inilah negara kita yng salah diagungkan yg baik di singkirakan

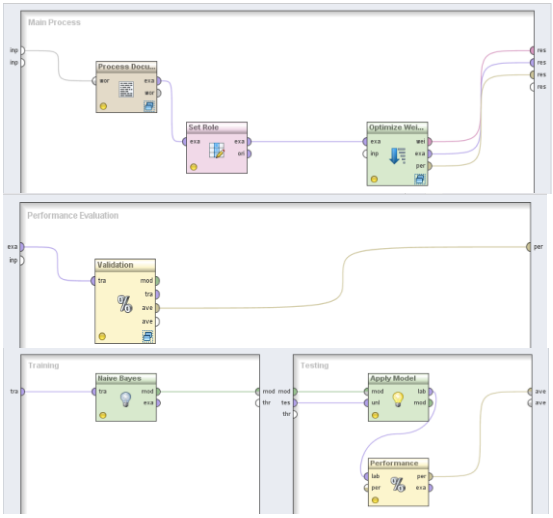
   b. *Transform Cases*
      Dalam proses *transform cases* ini, semua huruf dirubah menjadi huruf kecil semua atau huruf kapital semua.

Tabel IV.2. Perbandingan teks sebelum dan sesudah dilakukan proses *Transform Cases*

Teks sebelum dilakukan proses *transform cases*	Prabowo adalah Leader,Jokowi adalah manager,mana yang dibutuhkan bangsa saat ini? saya kira "Leader"lah yang cocok saat ini karena sekarang ini bangsa kita kurang dihargai oleh bangsa lain,dibutuhkan orang pemimpin yang tegas bukan saja merakyat..ingat nomor satu saudara/i sekalian kalau kita ingin jadi bangsa yang bangkit yaang sudah tidur beberapa tahun...SALAM MACAN ASIA
Teks setelah dilakukan proses *transform cases*	prabowo adalah leader jokowi adalah manager mana yang dibutuhkan bangsa saat ini saya kira leader lah yang cocok saat ini karena sekarang ini bangsa kita kurang dihargai oleh bangsa lain dibutuhkan orang pemimpin yang tegas bukan saja merakyat ingat nomor satu saudara i sekalian kalau kita ingin jadi bangsa yang bangkit yaang sudah tidur beberapa tahun salam macan asia

4.2. Hasil Pengujian Model *Naïve Bayes* dan *Particle Swarm Optimization.*

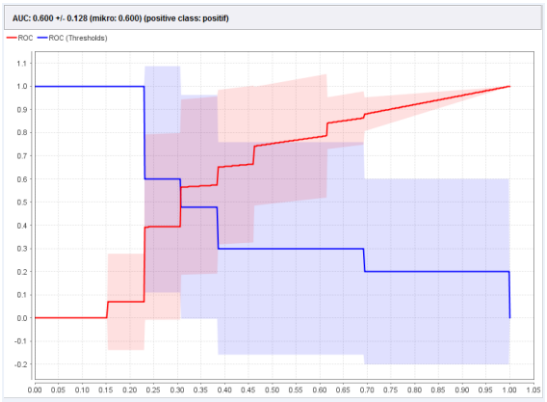
Validation adalah proses untuk mengevaluasi keakuratan prediksi dari model. Validasi digunakan untuk memperoleh prediksi menggunakan model yang ada dan kemudian membandingkan hasil tersebut dengan hasil yang sudah diketahui, ini mewakili langkah paling penting dalam proses membangun sebuah model (Mabrur dan Lubis, 2012). Pada penelitian penentuan hasil review opini publik berita pilpres menggunakan algoritma *Naive Bayes* berbasis *Particle Swarm Optimization* pada *framework* Rapid Miner sebagai berikut:



Gambar 4.1 Model Pengujian *Naive Bayes* berbasis *Particle Swarm Optimization*

Hasil pengujian data training metode *Naive Bayes* berbasis *Particle Swarm Optimization* menggunakan *Set Role* yang berfungsi untuk menentukan field pada kelas kemudian diberikan optimasi menggunakan *Particle Swarm Optimization* agar akurasi yang dihasilkan lebih tinggi. Pengukuran akurasi tersebut, akan dijabarkan melalui Kurva ROC dan *Confusion Matrix* di bawah ini:

1. Kurva ROC



Gambar 4.2 Kurva ROC *Naive Bayes* Berbasis *Particle Swarm Optimization*

Kurva ROC yang dihasilkan berdasarkan pengujian data pada gambar di atas, menunjukan bahwa ada peningkatan pada **akurasi** menggunakan *Naive Bayes* berbasis *Particle Swarm Optimization* sebesar **71.15%** dan AUC sebesar **0.600.**

2. Confusion Matrix

Tabel 4.3 *Confusion Matrix Naive Bayes* berbasis *Particle Swarm Optimization*

Accuracy: 71.15% +/- 9.77% (mikro: 71.15%)			
	True Positif	*True* Negatif	*Class Precission*
Prediksi Negatif	89	34	72.36%
Prediksi Positif	41	96	70.07%
Class Recall	68.46%	73.85%	

Sumber: Hasil Penelitian (2016)

Acc (Accuracy)= $\frac{TP+TN}{TP+TN+FP+FN} = \frac{89+96}{34+89+41+96} = \frac{185}{260} = 0.71$

Data training yang digunakan terdiri dari **130** data review positif mengenai review opini publik berita pilpres dan **130** data review negatif pada review opini publik berita pilpres. Data review negatif, setelah melalui beberapa tahap pengolahan pada RapidMainer dengan model *Naive Bayes* berbasis *Particle Swarm Optimization*, diklasi-fikasikan untuk review positif yang sesuai prediksi sebanyak **89** data, kemudian **34** data yang diprediksi positif namun masuk kedalam kategori review negatif. Sedangkan untuk data review positif, yang diprediksi bahwa data tersebut negatif adalah **41**, dan untuk prediksi review negatif yang masuk dalam prediksi review negatif adalah **96** data, hasil akurasi yang muncul adalah **71.15%.**

4.3. Pembahasan

Berdasarkan pengujian yang telah dilakukan terhadap review opini publik berita pilpres dengan menggunakan metode *Naive Bayes*, *Naive Bayes* berbasis *Particle Swarm Optimization*. Penerapan *Particle Swarm Optimization* (PSO) terbukti dapat meningkatkan akurasi pada klasifikasi review opini publik berita pilpres untuk mengidentifikasi antara review positif dan review negatif. Apabila sudah memiliki model klasifikasi teks pada review maka akan lebih memudahkan pembaca untuk mengetahui review positif dan review negatif. Berdasarkan data review yang sudah diolah melalui RapidMiner, kemudian hasilnya terpisah menjadi kata-kata. Kata-kata tersebut, masing-masing memiliki bobot sehingga dapat dilihat kata mana saja yang

berhubungan dengan sentimen yang sering muncul dan memiliki bobot tertinggi. Dengan demikian dapat diketahui review tersebut termasuk ke dalam review berita pilpres positif dan review berita pilpres negatif. Dalam penelitian ini, hasil perhitungan metode *Naive Bayes* (NB) memiliki Accuracy sebesar 63.85% dan AUC sebesar 0.523 sedangkan Metode *Naive Bayes* berbasis *Particle Swarm Optimization* (PSO) menghasilkan Accuracy sebesar 71.15% dan AUC sebesar 0.600. Hal ini menunjukan bahwa penggunaan optimasi *Particle Swarm Optimization* dapat meningkatkan nilai akurasi.

V. PENUTUP

5.1. Kesimpulan

Berdasarkan pemaparan yang telah dijelaskan pada bab sebelumnya, penelitian ini menghasilkan akurasi dalam bentuk *Confusion Matrix* dan Kurva ROC. Adapun akurasi yang dihasilkan pada algoritma *Naive Bayes* sebesar 63.85% dan AUC sebesar 0.523, sedangkan *Naive Bayes* dan *Particle Swarm Optimmization* dengan akurasi 71.15% dan AUC sebesar 0.600. Dengan demikian dapat disimpulkan bahwa penerapan optimasi dapat meningkatkan akurasi dari 63.85% ke 71.15%. Model *Naive Bayes* dan *Particle Swarm Optimization* dapat memberikan solusi terhadap permasalahan klasifikasi review opini publik berita pilpres agar lebih akurat dan optimal.

5.2. Saran

Agar penelitian ini bisa ditingkatkan, berikut adalah saran-saran yang diusulkan:

1. Untuk penerapan lebih lanjut, algoritma *Naive Bayes* dan *Particle Swarm Optimization* dapat digunakan sebagai moderator pendeteksi kelayakan *posting* pada sebuah forum atau berita *online*.
2. Menggunakan data review dari domain yang berbeda, misalnya review restoran, review film, review travel dan lain sebagainya.
3. Menggunakan pengklasifikasi lain yang mungkin di luar *Supervised learning*. Sehingga bisa dilakukan penelitian yang berbeda dari umumnya yang sudah ada.

DAFTAR REFERENSI

Andini (2013). Klasifikasi Dokumen Text menggunakan Algoritma Naive Bayes Dengan Bahasa Pemrograman Java. Jurnal Teknologi Informasi & Pendidikan. 2086-4981

Awaludin, M. (2015). Penerapan Metode Distance Transform Pada Linear Discriminant Analysis Untuk Kemunculan Kulit Pada Deteksi Kulit. *Journal of Intelligent Systems*, *1*(1), 49–55.

Basari, A. S. H., Hussin, B., Ananta, I. G. P., & Zeniarja, J.(2013). *Opinion Mining of Movie Review using HybridMethod of Support Vector Machine and Particle SwarmOptimization*. Procedia Engineering, 53, 453-462.doi:10.1016/j.proeng.2013.02.059.

Berry, M.W. & Kogan, J. (2010). *Text Mining Aplication and theory*. WILEY : United Kingdom.

Feldman, Ronen and Sanger, James. (2007). *The Text Mining Handbook Advanced Approaches in Analyzing Unstructured Data*. Cambridge University Press, New York.Francisco: Diane Cerra.

Habernal, Ptáček, Steinberger. (2015). Reprint of "*Supervised sentiment analysis in Czech social media*". *Information Processing & Management*, 50, 693-707

Haddi, E., Liu, X., & Shi, Y. (2013). T*he Role of Text Pre-processing in Sentiment Analysis*. Procedia Computer Science, 17, 26–32. doi:10.1016/j.procs.2013.05.005

Han, J., & Kamber, M. (2007). *Data Mining Concepts and Techniques*. San Ilhan & Tezel 2013; Raghavendra. N & Deka, 2014; Zhao, Fu, Ji, Tang, & Zhou, 2011

Hashimi, Hussein, Alaaeldin Hafez, & Hassan Mathkour. (2014). *Selection criteria for text mining approaches*. Computers in Human Behavior. 729-733

Jiawei, H., Kamber, M., & Pei, J. (2012). *Data Mining: Concepts and Techniques Third Edition*. Waltham, MA: Morgan Kaufmann.

Kaplan, A., & Haenlein, M. (2010). *Users of the world, unite! The challenges and opportunities of social media*. Business Horizons, 53, 59–68.

Kunaifi, Aang.(2009). Klasifikasi Email Berbahasa Indonesia menggunakan Text Mining dan Algoritma KMeans. Surabaya: Politeknik Elektronika Negeri Surabaya.

Liu, Bing. (2012). *Sentiment Analysis And Opinion Mining*. Chicago: Morgan & ClaypoolPublisher.

Liu, H., Tian, H., Chen, C., & Li, Y. (2013). *Electrical Power and Energy Systems An experimental investigation of two Wavelet-MLP hybrid frameworks for wind speed prediction using GA and PSO optimization*. *International Journal of Electrical Power & Energy Systems*, *52*, 161–173.

M.R. Saleh, M.T. Martín-Valdivia, A. Montejo-Ráez, L.A. Ureña-López, *Experiments with SVM to classify opinions in different domains*, Expert Syst. Appl. 38 (2011) 14799–14804.

Moraes, R., Valiati, J. F., & Gavião Neto, W. P. (2013). Document-level sentiment.

Mostafa, Mohamed M. (2013). *More than words: social network's text mining for consumer brand sentiments.* Expert Systems With Applications. 4241-4251

Nugroho, (2007). Pengantar Support Vector Machine.

Pang, B. & Lee, L. 2008. *Subjectivity Detection and Opinion Identification. Opinion Mining and Sentiment Analysis*. Now Publishers Inc. [Online]. Tersedia di: http://www.cs.cornell.edu/home/llee/opinion-mining-sentiment-analysissurvey.html.

Prasetyo, Heri. (2014). Data Mining Mengolah Data menjadi Informasi. Yogyakarta: Andi Offset.

Ramesh (2015). *An Advanced Multi Class Instance Selection Based Support Vector Machine for Text Classification.* Procedia Computer Science. 1124-1130.

Rocha, Leonardo et al (2013). *Temporal contexts: effective text classification in evolving document collection.* Information Systems. 388-409

Rozi, Hadi, Achmad. (2012), Implementasi Opinion Mining (Analisis Sentimen) untuk Ekstraksi Data Opini Publik pada Perguruan Tinggi. *Jurnal EECCIS Vol. 6, No. 1, Juni 2012. Systems with Applications*, 40(2), 621–633. doi:10.1016/j.eswa.2012.07.059

Statsoft. (2015). *Naive Bayes Clasifier Introductory Overview*. Retrieved April 22, 2015, from Statsoft Web Site: http://www.statsoft.com/textbook/naivebayes-classifier

Vercellis, C. (2009). *Business Intelligence Data Mining And Optimization For Decision Makin*g .United Kingdom: A John Wiley And Sons, Ltd.,Publication.

Wang, X., Wen, J., Zhang, Y., & Wang, Y. (2014). *Optik Real estate price forecasting based on SVM optimized by PSO*. Optik - International Journal for Light and Electron Optics, *125*(3), 1439–1443.

Witten, I. H., Frank, E., & Hall, M. A. (2011). *Data Mining Practical Machine Learning Tools and Techniques* (Third., p. 665).

Yao, Zhi-Min. (2012), *An Optimized NBC Approach in Text Classification*. Physics Procedia, 24, 1910-1914

Zhai, C., & Aggarwal, C. C. (2012). *Mining Text Data.* New York: Springer.

Zhao, M., Fu, C., Ji, L., Tang, K., & Zhou, M. (2011). *Feature selection and parameter optimization for support vector machines: A new approach based on genetic algorithm with feature chromosomes*. Expert Systems with Applications, 38(5), 5197–5204. doi:10.1016/j.eswa.2010.10.041.