

ANALISA SENTIMEN TERHADAP TOKOH PUBLIK MENGGUNAKAN METODE NAÏVE BAYES CLASSIFIER DAN SUPPORT VECTOR MACHINE

Page | 230

Deni Rusdian¹, Didi Rosiyadi²
Pasca Sarjana STMIK Nusa Mandiri
Jalan Kramat Raya No.18, Senen, Jakarta Pusat
¹14002115@nusamandiri.ac.id, ²didi.rosiyadi@gmail.com

Abstrak— Twitter merupakan salah satu sosial media yang banyak digunakan oleh masyarakat dalam kehidupan sehari-hari. Salah satunya adalah twitter digunakan sebagai tempat untuk menyampaikan setiap pendapat atau opini terhadap suatu hal seperti pendapat atau opini terhadap tokoh publik. Dalam penelitian ini akan dilakukan analisa sentimen terhadap tokoh publik yang diungkapkan masyarakat melalui jejaring sosial twitter. Ada beberapa tahap dalam melakukan sentiment analysis diantaranya adalah pengumpulan data menggunakan API Twitter, memberikan label kepada setiap tweet secara manual, Pre Processing Data dan POS Tagging. Untuk mendapatkan hasil ekstraksinya, penelitian ini menggunakan metode Naïve Bayes Classifier dan Support Vector Machine agar hasil tersebut dapat dibandingkan keakurasiannya. Hasil dari penelitian ini didapat bahwa algoritma Support Vector Machine memiliki tingkat keakurasi sebesar 73.96% sedangkan untuk algoritma Naïve Bayes Classifier memiliki tingkat keakurasi sebesar 71.94% dengan menggunakan dataset yang didata.

Keywords—Analisa Sentimen, Twitter, Naïve Bayes Classifier, Support Vector Machine

Abstract— Twitter is one of the social media that is widely used by people in everyday life. One of them is twitter is used as a place to convey any opinions or opinions on a matter such as opinions or opinions on public figures. In this study, a sentiment analysis of public figures will be carried out by the public through twitter social networking. There are several stages in conducting analyst sentiment including data collection using the Twitter API, labeling each tweet manually, Pre Processing Data and POS Tagging. To get the results of the extraction, this study uses the Naïve Bayes Classifier method and Support Vector Machine so that the results can be compared for accuracy. The results of this study found that the Support Vector Machine algorithm has an accuracy rate of 73.96% while for the Naïve Bayes Classifier algorithm has an accuracy rate of 71.94% using a dataset that is recorded.

Keywords—Analisa Sentimen, Twitter, Naïve Bayes Classifier, Support Vector Machine

I. PENDAHULUAN

Pada saat ini informasi sangat mudah diperoleh terutama informasi atau opini terhadap seorang tokoh public, banyak masyarakat yang memberikan opini mereka melalui situs di internet. Situs tempat bertukar opini dapat berupa forum, portal berita maupun jejaring sosial. Twitter menjadi salah satu jejaring sosial yang digunakan masyarakat untuk menyampaikan opini mereka terhadap seorang tokoh public. Pada tahun 2015 lalu Indonesia telah menjadi salah satu negara yang memiliki pengguna aktif twitter terbesar didunia dengan jumlah tweet nya yang mencapai 500juta tiap harinya[6].

Informasi mengenai opini masyarakat dapat tersebar dengan sangat mudah. Opini yang disampaikan beragam mulai dari hasil kinerjanya, tindakan yang dia lakukan maupun keputusan yang diambilnya. Jenis opininya pun berupa opini negates yang dapat menimbulkan permusuhan, penghinaan maupun perdebatan di dunia maya. Ada pula yang berbentuk opini positif yang memberikan pujian, sanjungan maupun penghormatan terhadap tokoh public tersebut.

Meluasnya penggunaan internet telah meningkatkan jumlah informasi yang disimpan dan diakses melalui web dalam kecepatan yang sangat cepat [1]. Jumlah informasi yang tersimpan semakin hari akan semakin bertambah banyak sehingga informasi tersebut berubah menjadi data yang besar. Data yang besar tentu memiliki sebuah

informasi yang berguna dan dapat dimanfaatkan untuk masa depan. Salah satu cara mencari informasi yang tersembunyi dari data yang besar adalah dengan menggunakan opinion mining. Fokus dari penelitian opinion mining adalah melakukan analisa opini dari suatu dokumen teks [15].

II. METODOLOGI

A. *Analisa Sentimen*

Analisa Sentimen 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 [9]. 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 [2]. Analisa sentimen dapat juga berupa bentuk perasaan marah, gembira maupun sedih.

Langkah-langkah yang umumnya ditemukan pada klasifikasi teks analisa sentimen adalah

- **Definisikan domain dataset**
Pengumpulan dataset yang melingkupi suatu domain, misalnya dataset review film, dataset review produk, dan lain sebagainya.
- ***Pre-processing***
Tahap pemrosesan awal yang umumnya dilakukan dengan proses Tokenization, Stopwords removal, dan Stemming.
- ***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.
- ***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.
- ***Classification***
Proses klasifikasi umumnya menggunakan pengklasifikasi seperti Naïve Bayes, Support

Vector Machine, dan lain sebagainya. f. Interpretation/Evaluation Tahap evaluasi biasanya menghitung akurasi, recall, precision, dan F-1.

B. *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 linguistic [8].

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 [6].

Text mining adalah penambahan 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 [4].

Text mining merupakan teknik yang digunakan untuk menangani masalah klasifikasi, clustering, information extraction dan information retrieval [3].

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 [16].

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 [17].

C. *Review Tokoh Publik*

Opini terhadap tokoh publik sangat banyak dan mudah ditemukan di dunia maya tetapi untuk menjadikan opini tersebut sebuah informasi yang sangat bermanfaat sangatlah sulit. Untuk mempermudah pengamatan terhadap tokoh publik maka digunakan sosial media twitter untuk pengumpulan opini yang dilontarkan oleh masyarakat sehingga dapat terkumpul bermacam-macam opini .

D. Naïve Bayes Classifier

Naïve Bayes Classifier adalah salah satu metode klasifikasi dengan probabilitas sederhana yang mengaplikasikan Teorema Bayes dengan asumsi ketidaktergantungan (independen) yang tinggi. Penggunaan metode Naive Bayes Classifier pada penelitian ini didasarkan pada banyaknya dataset yang dipakai sehingga membutuhkan suatu metode yang mempunyai performansi yang cepat dalam pengklasifikasian serta keakuratan yang cukup tinggi [9]. Keuntungan penggunaan Naive Bayes Classifier adalah metode ini hanya membutuhkan jumlah data pelatihan (training data) yang kecil untuk menentukan estimasi parameter yang diperlukan dalam proses pengklasifikasian [9].

Metode Naive Bayes Classifier menempuh dua tahap dalam proses klasifikasi teks, yaitu tahap pelatihan dan tahap klasifikasi. Pada tahap pelatihan dilakukan proses terhadap sampel data yang sedapat mungkin dapat menjadi representasi data tersebut. Selanjutnya adalah penentuan probabilitas prior bagi tiap kategori berdasarkan sampel data. Pada tahap klasifikasi ditentukan nilai kategori dari suatu data berdasarkan term yang muncul dalam data yang diklasifikasi. Teorema Naïve Bayes dapat dinyatakan dalam persamaan 2.1

$$P(X_k|Y) = \frac{P(Y|X_k)}{\sum_i P(Y|X_i)} \dots \dots \dots (2.1)$$

Dimana, keadaan Probabilitas Xk di dalam Y dapat dihitung dari keadaan Probabilitas Y di dalam Xk dibagi dengan jumlah dari semua probabilitas Y di dalam semua Xi.

Untuk dapat mengklasifikasikan suatu tweet, dalam penelitian ini penulis menggunakan metode Naive Bayes Classifier untuk klasifikasi teks, seperti yang dilakukan [13] pada persamaan 2.2.

$$P(V1|C = c) = \frac{CountTerms(v1, docsv(c))}{AllTerms(docs(c))} (2.2)$$

Dimana v1 dalam penelitian ini adalah satu kata tertentu dalam tweet, sedangkan CountTerms?v1 , docsv?c?? menunjuk pada jumlah kemunculan suatu kata berlabel c (“positif” atau “negatif” atau “netral”) . AllTerms1docs?c?2 menunjuk pada jumlah semua kata berlabel c yang ada pada dataset. Untuk menghindari adanya nilai nol pada probabilitas, maka diberlakukan Laplace (add-one) smoothing. Tujuan daripada smoothing adalah untuk mengurangi probabilitas dari hasil/keluaran yang terobeservasi, dan juga sekaligus meningkatkan/menambah probabilitas hasil/keluaran yang belum terobservasi [1], sehingga persamaan menjadi sebagai berikut:

$$P(V1|C = c) = \frac{CountTerms(v1, docsv(c)) + 1}{AllTerms(docs(c)) + |V|} (2.3)$$

Dimana |V| menunjuk pada jumlah semua kata dalam tweet yang ada di dataset.

E. Support Vector Machine

SVM merupakan algoritma klasifikasi yang memiliki tujuan untuk menemukan fungsi pemisah (hyperplane) dengan margin paling besar, sehingga dapat memisahkan dua kumpulan data secara optimal [8].

SVM pada awalnya digunakan untuk klasifikasi data numerik, tetapi ternyata SVM juga sangat efektif dan cepat untuk menyelesaikan masalah-masalah data teks. Data teks cocok untuk dilakukan klasifikasi dengan algoritma SVM karena sifat dasar teks yang cenderung mempunyai dimensi yang tinggi, dimana terdapat beberapa fitur yang tidak relevan, tetapi akan cenderung berkolerasi satu sama lain dan umumnya akan disusun dalam kategori yang terpisah secara linear [18].

SVM memiliki kelebihan yaitu mampu mengidentifikasi hyperplane terpisah yang memaksimalkan margin antara dua kelas yang berbeda [4]. Namun SVM memiliki kekurangan terhadap masalah pemilihan parameter atau fitur yang sesuai [2]. Pemilihan fitur sekaligus penyetingan parameter di SVM secara signifikan mempengaruhi hasil akurasi lasifikasi [19].

III. PEMBAHASAN

A. Pengumpulan Data

Pengumpulan data opini dari twitter dilakukan dengan bantuan aplikasi twiny yang menggunakan bahasa pemrograman python. Dengan aplikasi twiny kita dapat menarik data twitter dengan range waktu tertentu, keyword tertentu maupun diexport kedalam format tertentu. Pada penelitian ini menggunakan opini sebanyak 100 yang bernilai negatif dan 100 opini yang bernilai positif yang dilabeli secara manual. Untuk contoh twit yang telah dilabeli dapat dilihat pada tabel dibawah ini.

TABEL 1.
CONTOH OPINI TWITTER

Text	Sentimen
@Fahrihamzah @htbns2003 Saya pilih Jokowi yang kata Anda "Mukidi" saya sudah lihat dunia mas... Dan Indonesia mengarah jadi negara maju spt negara lain... Apa saya bloom? Loe apa udah pintar? Coba tanya kebanyakan org" Indonesia yg di Luar negeri .. pil	Positif
Saya rasa klo pa @mohmahfudmd jadi pendamping pa @jokowi maka politisi2 lawan pa Jokowi akan mikir 2x klo ngomong sembarangan nyerang pemerintah tanpa bukti & tanpa aturan.	Positif

@detikfinance @detikcom Caranya patut ditiru bpk jokowi....Tolong freeport diledakan ! Biar ga balik lagi!	Positif
@kompascom Wakilnya sapa aja deh, yg penting Jokowi 2 periode	Positif
@jokowi Maju terus garuda indonesia	Positif
Fitnah dengan Isu PKI, anti Islam, dan pro Cina sudah jelas merupakan FITNAH YG NYATA dari pembenci Jokowi yang modal politiknya memang cuma fitnah. Pendukung @jokowi jangan berhenti menangkal fitnah2 berikutnya yg pasti terus dilakukan sampai Pilpres 2019 https://twitter.com/alvano91/status/1018395183027662855	Negatif
@LawanPoLitikJKW #2019GantiPresiden demi kepentingan bangsa dan negara, jangan lg biarkan jokowi selalu kerasukan elektabilitas shg terus meraya layak mjd pemimpin negara tanpa mnyadari negara kian sekarat.Bangun kembali kejayaan Indonesia bersama @prabow	Negatif
@geloraco Jika dibalik, terBUKTI saat ini, @jokowi lemah maka ekonomi kita saat ini lemah!#2019GantiPresiden	Negatif
Lucu semua ya para Pejabat di era Pak Jokowi. Harga telur naik Piala dunia yg disalahkan. Yang satu lagi tanda tangan tp tidak tahu yg ditanda tangani #Meulaboh4Jokowi2019#2019GantiPresiden pic.twitter.com/0dY0LdAbMX	Negatif
Bahkan jika HRS sekalipun bergabung dengan PDIP dan disandingkan sebagai Cawapres Jokowi. Gue tak akan memilihnya. Tekat gue sudah bulat #2019GantiPresiden https://twitter.com/maspiyuu/status/1019762017358376960	Negatif

Untuk label positif berarti masyarakat yang melakukan tweet tersebut setuju atau mendukung tokoh publik yang sedang dibicarakan, begitupun sebaliknya apabila label bernilai negative maka masyarakat tersebut tidak setuju atau tidak mendukung tokoh publik yang sedang bicarakananya.

B. Pre-processing

Sebelum memasuki proses *Pre-processing* penelitian ini melakukan *sub-process* dimana pada aplikasi *Rapidminer* operator yang dapat digunakan adalah *subprocess*. Didalam operator *subprocess*, tahapan yang dilalui adalah melakukan penghapusan *Uniform Resource Locator* atau yang lebih dikenal dengan nama URL dengan menggunakan operator *Remove URL* dan

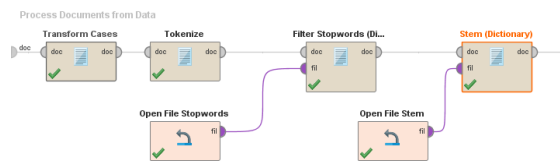
konfigurasi dengan menggunakan *Regular Expression (Regex)* `http\S+/\S+co\S+` maka akan dihasilkan seluruh URL pada setiap tweet akan menghilang.

Sub-process selanjutnya pada penelitian ini adalah dengan menghapus *emoticon* dan merubahnya menjadi sebuah kata yang ada pada setiap tweet sehingga opini yang disampaikan oleh masyarakat melalui media sosial twitter menjadi lebih mudah untuk di analisa.

TABEL 2.
CONTOH EMOTICON

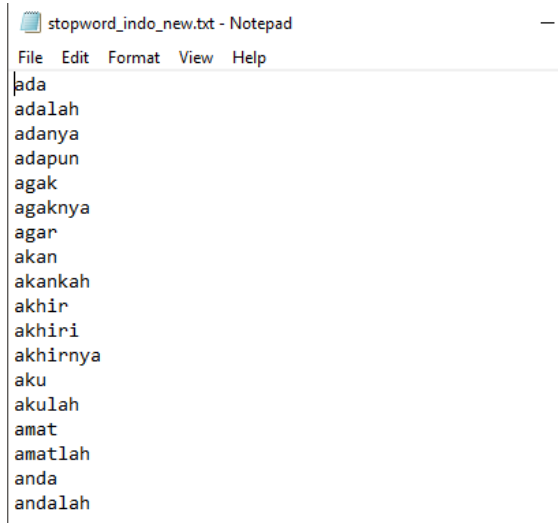
Emoticon	Kata
☹	Sedih
)”:	Sedih
(^^)	Senang
:*	Senang
<3	Senang

Setelah *sub-process* selesai, tahapan selanjutnya, data yang sudah disiapkan kemudian dilakukan *pre-processing* dimana opini yang ada akan mendapatkan proses *tokenizing*, *stopwords* dan *steming*. Pada aplikasi *rapid miner*, proses *pre-processing* terdapat pada operator *preprocess dokumen*. Berikut ini adalah *pre-processing* opini pada penelitian ini



Gbr. 1. Pre processing data

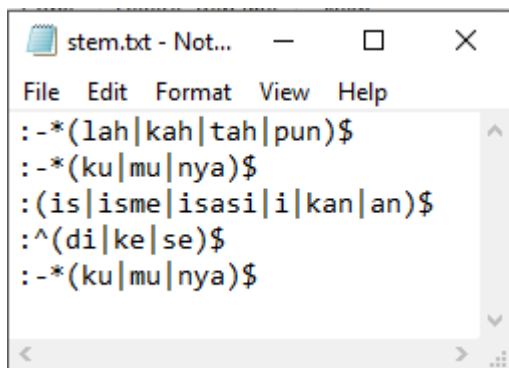
Operator *Transform Case* berguna untuk merubah seluruh text yang ada menjadi huruf kecil atau non capital kemudian operator *tokenize* yang berguna untuk memecah setiap kalimat menjadi perkata untuk selanjutnya diproses oleh operator *filter stopwords* dimana operator *filter stopwords* berguna untuk menghilangkan kata-kata yang tidak dibutuhkan dalam proses selanjutnya. Dibawah ini adalah contoh kata yang termasuk dalam file *stopwords*



Gbr. 2 Contoh Kata File Stopwords

List kata – kata yang tidak dibutuhkan tersebut berada pada satu file yang harus dibuka terlebih dahulu dengan menggunakan operator *open file stopwords*.

Proses selanjutnya dalam *pre-processing* adalah melakukan stemming atau dengan kata lain menghilangkan imbuhan di setiap kata – kata yang telah diproses sebelumnya sehingga keluaran dari proses *pre-processing* ini sudah dapat digunakan untuk dilakukan perhitungan dengan menggunakan algoritma *naive bayes classifier* maupun *support vector machine*. Untuk proses *stemming* pada penelitian ini menggunakan web aplikasi secara online di situs gataframework.com kemudian ditambahkan dengan *stemming* menggunakan file regexp seperti pada gambar dibawah ini

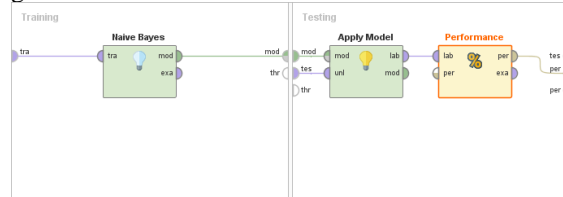


Gbr 3. File Regexp untuk Stemming

C. Naive Bayes Classifier

Pada penelitian ini menggunakan dua algoritma yang berbeda yaitu *Naive Bayes Classifier* dan *Support Vector Machine*. Hal ini digunakan untuk melihat perbandingan tingkat akurasi kedua algoritma tersebut untuk melakukan proses analisa sentiment. Dengan bantuan

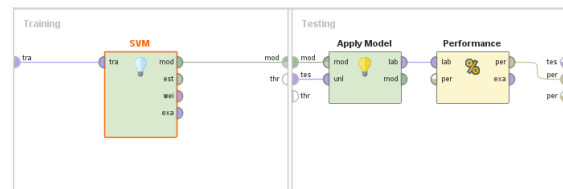
aplikasi rapidminer maka digunakan operator *Naive Bayes Classifier* untuk percobaan pertama seperti pada gambar dibawah ini.



Gbr 4. Naive Bayes Classifier Operator

D. Support Vector Machine

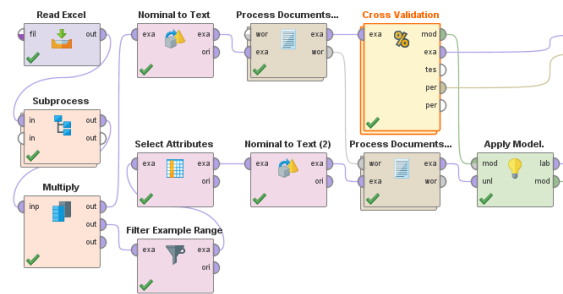
Untuk perbandingan algoritma sebelumnya maka data yang ada diproses kembali dengan menggunakan algoritma yang berbeda yaitu *Support Vector Machine* seperti pada gambar dibawah ini



Gbr 5. Support Vector Machine Operator

IV. HASIL DAN PEMBAHASAN

Setelah menjalani serangkaian proses *pre-processing* kemudian dilakukan analisa menggunakan algoritma *Naive Bayes Classifier* dan *Support Vector Machine* sehingga bentuk kumpulan operator pada aplikasi *Rapidminer* menjadi seperti gambar di bawah ini.



Gbr 6. Rangkaian proses pada Rapidminer

Pada algoritma *Naive Bayes Classifier*, nilai akurasinya berada pada angka 71.94% seperti pada gambar dibawah ini.

accuracy: 71.94% +/- 8.66% (micro average: 72.00%)

	true Positif	true Negatif	class precision
pred. Positif	75	31	70.75%
pred. Negatif	25	69	73.40%
class recall	75.00%	69.00%	

Gbr 7. Akurasi Naive Bayes Classifier

Sedangkan pada percobaan dengan menggunakan algoritma *Support Vector Machine*, tingkat akurasi yang didapatkan sebesar 73.96% seperti pada gambar dibawah ini

accuracy: 73.96% +/- 7.25% (micro average: 74.00%)

	true Positif	true Negatif	class precision
pred. Positif	63	15	80.77%
pred. Negatif	37	85	69.67%
class recall	63.00%	85.00%	

Gbr 6. Akurasi Support Vector Machine

V. KESIMPULAN

Dengan perbedaan hasil perbandingan kedua algoritma yaitu *Naïve Bayes Classifier* dan *Support Vector Machine* maka disimpulkan bahwa algoritma *Support Vector Machine* lebih baik dalam menganalisa sentimen daripada algoritma *Naïve Bayes Classifier* dengan tingkat akurasi sebesar 73.96% sedangkan untuk *Naïve Bayes Classifier* adalah 71.94%.

REFERENSI

- [1] Arguello, J., 2013. *Naïve Bayes Text Classification* (https://ils.unc.edu/courses/2013_fall/inls613_001/lectures/04NaiveBayesClassification.pdf). Diakses tanggal 12 April 2014). The University of North Carolina.
- [2] Basari et al. (2013). Opinion Mining of Movie Review using Hybrid Method of Support Vector Machine and Particle Swarm Optimization. *Procedia Engineering*, 53, 453-462.
- [3] Berry, M.W. & Kogan, J. 2010. *Text Mining Application and theory*. WILEY : United Kingdom.
- [4] Chou, J.-S., Cheng, M.-Y., Wu, Y.-W., & Pham, A.-D. (2014). Optimizing parameters of Support Vector Machine using fast messy genetic algorithm for dispute classification. *Expert Systems with Applications*, 41(8), 3955–3964. doi:10.1016/j.eswa.2013.12.035.
- [5] Feldman, Ronen and Sanger, James. 2007. *The Text Mining Handbook Advanced Approaches in Analyzing Unstructured Data*. Cambridge University Press, New York.Francisco: Diane Cerra.
- [6] Hadi, A. F., W, D. B. C., Hasan, M., & Penelitian, A. D. (2017). *Text Mining Pada Media Sosial Twitter Studi Kasus : Masa Tenang Pilkada Dki 2017 Putaran 2*.
- [7] Hashimi, Hussein, Alaaeldin Hafez, & Hassan Mathkour. (2014). Selection criteria for text mining approaches. *Computers in Human Behavior*. 729-733.
- [8] Jiawei, H., Kamber, M., & Pei, J. (2012). *Data Mining: Concepts and Techniques Third Edition*. Waltham, MA: Morgan Kaufmann.
- [9] Kao, A & Potet, S. (2007) *Natural Language Processing and Text Mining*. Springer Verlag, London, England.
- [10] Larose, D. T. 2006. *Naïve Bayes Estimation and Bayesian Networks*, in *Data Mining Methods and Models*, John Wiley & Sons, Inc., Hoboken, NJ, USA. doi: 10.1002/0471756482.ch5.
- [11] Liu, Bing. 2012. *Sentiment Analysis And Opinion Mining*. Chicago: Morgan & ClaypoolPublisher.
- [12] Moraes, R., Valiati, J. F., & Gavião Neto, W. P. (2013). Document-level sentiment.
- [13] Ricci, F.; Rokach, L.; et al. 2011. *Recommender Systems Handbook*. Berlin : Springer.
- [14] Rocha, Leonardo et al (2013). Temporal contexts: effective text classification in evolving document collection. *Information Systems*. 388-409.

- [15] Statsoft. (2015). *Naive Bayes Classifier Introductory Overview*. Retrieved April 22, 2015, from Statsoft Web Site: <http://www.statsoft.com/textbook/naivebayes-classifier>.
- [16] Statsoft. (2015). *Naive Bayes Classifier Introductory Overview*. Retrieved April 22, 2015, from Statsoft Web Site: <http://www.statsoft.com/textbook/naivebayes-classifier>.
- [17] Yunina, Norma. (2016). Analisis Sentimen Berita Artis Dengan Menggunakan Algoritma Support Vector Machine dan Particle Swarm Optimization. *Jurnal Sistem Informasi STMIK Antar Bangsa*. Vol.V No.2. 104-112.
- [18] Zhai, C., & Aggarwal, C. C. (2012). *Mining Text Data*. New York: Springer.
- [19] 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.