

**ANALISIS SENTIMEN TERHADAP BERITA ARTIS BERDASARKAN
OPINI PUBLIK MENGGUNAKAN ALGORITMA SUPPORT VECTOR
MACHINE DAN PARTICLE SWARM OPTIMIZATION**



RINGKASAN TESIS

**NORMA YUNITA
14001105**

**PROGRAM PASCASARJANA MAGISTER ILMU KOMPUTER
SEKOLAH TINGGI MANAJEMEN INFORMATIKA DAN KOMPUTER
NUSA MANDIRI
JAKARTA
2015**

HALAMAN PENGESAHAN

Tesis ini diajukan oleh:

Nama : Norma Yunita
NIM : 14001105
Program Studi : Magister Ilmu Komputer
Jenjang : Strata Dua (S2)
Konsentrasi : Manajemen Information System (MIS)
Judul Tesis : *Analisis Sentimen Terhadap Berita Artis Berdasarkan Opini Publik Menggunakan Algoritma Support Vector Machine Dan Particle Swarm Optimization*

Telah berhasil dipertahankan dihadapan Dewan Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Magister Ilmu Komputer (M. Kom) pada program Pascasarjana Magister Ilmu Komputer Sekolah Tinggi Manajemen Informatika dan Komputer Nusa Mandiri (STMIK Nusa Mandiri).

Jakarta, 10 Oktober 2015
Pascasarjana Magister Ilmu Komputer
STMIK Nusa Mandiri
Direktur

Prof. Dr. Ir. R. Eko Indrajit, MSC, MBA

DEWAN PENGUJI

Penguji I : Dr. Sularso Budilaksono, M.Kom

Penguji II : Drs. Bambang Setiarso, MA

Penguji III/ : Dr. Mochamad Wahyudi, MM, M.Kom, M.Pd

Pembimbing



DAFTAR ISI

HALAMAN SAMPUL	i
HALAMAN JUDUL	ii
HALAMAN PERNYATAAN ORISINALITAS	iii
HALAMAN PENGESAHAN	iv
KATA PENGANTAR	v
HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI	
KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS.....	vii
ABSTRAK	viii
ABSTRAC	ix
DAFTAR ISI	x
DAFTAR TABEL	xi
DAFTAR GAMBAR	xii
DAFTAR LAMPIRAN	xiii
BAB 1. PENDAHULUAN	1
1.1.Latar Belakang	1
1.2.Permasalahan Penelitian	3
1.3.Manfaat dan Tujuan Penelitian	4
1.4.Sistematika Penulisan	5
BAB 2. LANDASAN TEORI	6
2.1. Tinjauan Pustaka	6
2.2. Tinjauan Studi Penelitian Terdahulu yang Terkait	13
2.3. Tinjauan Objek Penelitian	19
BAB 3. METODOLOGI PENELITIAN	23
3.1. Perancangan Penelitian	23
3.2. Pengumpulan Data	24
3.3. Pengolahan Data Awal	25
3.4. Metode yang diusulkan	26
3.5. Eksperimen dan Hasil Pengujian	26
3.6. Evaluasi dan Validasi Hasil	27
BAB 4. HASIL PENELITIAN DAN PEMBAHASAN	28
4.1. Hasil Penelitian	28
4.2. Analisis Evaluasi Hasil dan Validasi Model	32
4.3. Pembahasan	37
4.4. Desain dan Implementasi	38
4.5. Implikasi Penelitian	41
BAB 5. PENUTUP	42
1.1. Kesimpulan	42
1.2. Saran	42
DAFTAR REFERENSI	43
DAFTAR RIWAYAT HIDUP	46

Di masa saat ini informasi sangat mudah diperoleh terutama informasi atau kabar tentang artis tertentu, banyak pembaca yang rela berjam-jam membaca untuk mendapatkan informasi terkini dari artis tertentu. Media untuk mendapatkan informasi tentang artis sangat banyak seperti melalui media tv, radio, koran, majalah, *website* dan lain-lain. Tetapi dari sebagian besar media tersebut *website* merupakan media yang sangat *fleksible* karena dapat diakses di berbagai macam tempat yang terkoneksi jaringan *Internet*, informasi yang disediakan sangatlah *up to date* dan juga setiap orang bisa mengomentari artikel yang terkait. Informasi yang menyebar sangat cepat dan diiringi dengan kebebasan mengeluarkan pendapat dapat menimbulkan berbagai jenis opini, baik opini negatif atau positif. Opini negatif disini mengandung arti bahwa kata atau opini yang dapat menimbulkan permusuhan, penghinaan, perdebatan dan perselisihan di dunia maya. Sedangkan opini positif yaitu kata atau opini yang sifatnya positif dan tidak menimbulkan permusuhan, penghinaan, perdebatan, dan perselisihan di dunia maya.

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).

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).

Support Vector Machine adalah metode yang bekerja dengan baik pada set data dengan dimensi yang tinggi dan hanya sejumlah data terpilih sajalah yang berkontribusi untuk membentuk model yang digunakan dalam klasifikasi yang akan dipelajari. SVM hanya menyimpan sebagian kecil saja dari data latih untuk digunakan pada saat prediksi (Prasetyo, 2014). Selain memiliki banyak kemampuan yang telah disebutkan diatas, metode SVM juga memiliki kelemahan yaitu sulitnya menentukan nilai parameter yang optimal (Ilhan & Tezel 2013; Raghavendra. N & Deka, 2014; Zhao, Fu, Ji, Tang, & Zhou, 2011).

Particle Swarm Optimization (PSO) merupakan teknik komputasi evolusioner yang mampu menghasilkan solusi secara global optimal dalam ruang pencarian melalui interaksi individu dalam segerombolan partikel. Setiap partikel menyampaikan informasi berupa posisi terbaiknya kepada partikel yang lain dan menyesuaikan posisi dan kecepatan masing-masing berdasarkan informasi yang diterima mengenai posisi yang terbaik tersebut (Shuzhou & Bo, 2011). PSO memiliki kemampuan pencarian global yang kuat, juga dapat membantu mencari parameter yang optimum secara cepat (Wang, et al., 2014). PSO banyak digunakan untuk memecahkan masalah optimasi serta pada seleksi fitur (Liu, et al., 2011).

Berdasarkan pada latar belakang, maka penelitian ini metode *Support Vector Machine* 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 www.newsmedia.co.id, www.kapanlagi.com dan www.tribunnews.com yang terdiri dari beberapa review atau pendapat masyarakat mengenai artis. Peneliti mengambil sample data secara simple random sebanyak 150 review positif dan 150 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 Swarm Optimization* (PSO). Sedangkan klasifikasi yang digunakan adalah *Support Vector Machine*, dimana akurasinya dihitung berdasarkan epsilon, c, kernel type dan population size. Pengujian 10 *fold cross validation* akan

dilakukan, akurasi algoritma diukur dengan Confution Metrix dan hasil olahan akan ditampilkan dalam bentuk kurva ROC dan accuracy. Software yang digunakan untuk mengolah data klasifikasi adalah RapidMiner sebagai alat bantu dalam mengukur akurasi data eksperimen.

Penelitian ini menghasilkan akurasi dalam bentuk *Confusion Matrix* dan Kurva ROC. Adapun akurasi yang dihasilkan pada algoritma *Support Vector Machine* sebesar 73.33% dan AUC sebesar 0.774, sedangkan *Support Vector Machine* dan *Particle Swarm Optimmmization* dengan akurasi 76.00% dan AUC sebesar 0.794. Dengan demikian dapat disimpulkan bahwa penerapan optimasi dapat meningkatkan akurasi dari 73.33% ke 76.00%. Model di *Support Vector Machine* dan *Particle Swarm Optimization* dapat memberikan solusi terhadap permasalahan klasifikasi review opini publik berita artis agar lebih akurat dan optimal.

DAFTAR REFERENSI

- Andini (2013). Klasifikasi Dokumen Text Menggunakan Algoritma Naive Bayes Dengan Bahasa Pemrograman Java. *Jurnal Teknologi Informasi & Pendidikan*. 2086-4981
- 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.
- 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.
- Chen et al. (2009), *Feature selection for text classification with Naïve Bayes*. *Expert Systems with Applications*, 36, 5432-5435
- 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.
- 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). *The 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.

- Kurniawan et al (2012), *Klasifikasi Konten Berita Dengan Metode Text Mining*. Teknologi Informasi. 14-19
- Liu, Bing. 2012. *Sentiment Analysis And Opinion Mining*. Chicago: Morgan & Claypool Publisher.
- 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-analysisurvey.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 Classifier 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 Making*. 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.