

**PENERAPAN MODEL SUPPORT VECTOR MACHINE TEXT
MINING PADA REVIEW SMARTPHONE ANDROID VS
BLACKBERRY DENGAN TEKNIK OPTIMASI GENETIC
ALGORITHM**

TESIS



**Oleh :
AGUS DARMAWAN
1211601909**

PROGRAM STUDI : MAGISTER ILMU KOMPUTER (MKOM)

PROGRAM PASCASARJANA

UNIVERSITAS BUDI LUHUR

JAKARTA

2015

**PENERAPAN MODEL SUPPORT VECTOR MACHINE
TEXTMINING PADA REVIEW SMARTPHONE ANDROID VS
BLACKBERRY DENGAN TEKNIK OPTIMASI GENETIC
ALGORITHM**

Diajukan untuk memenuhi salah satu persyaratan
memperoleh gelar Magister Ilmu Komputer
(MKOM)



Oleh :

AGUS DARMAWAN

1211601909

PROGRAM STUDI : MAGISTER ILMU KOMPUTER (MKOM)

PROGRAM PASCASARJANA

UNIVERSITAS BUDI LUHUR

JAKARTA

2015

SURAT PERNYATAAN ORISINALITAS

Yang bertanda tangan di bawah ini :

Nama : Agus Darmawan
NIM : 1211601909
Program Studi : Magister Ilmu
Komputer Jenjang : Strata Dua (S2)
Konsentrasi : *Sistem Information*

Dengan ini menyatakan bahwa tesis yang telah saya buat dengan judul: "*Penerapan Model Support Vector Machine TextMining Pada Smartphone Android VS Blackberry Dengan Teknik Optimisasi Genetic Agorithm*" adalah hasil karya sendiri, dan semua sumber baik yang kutip maupun yang dirujuk telah saya nyatakan dengan benar dan tesis belum pernah diterbitkan atau dipublikasikan dimanapun dan dalam bentuk apapun.

Demikianlah surat pernyataan ini saya buat dengan sebenar-benarnya. Apabila dikemudian hari ternyata saya memberikan keterangan palsu dan atau ada pihak lain yang mengklaim bahwa tesis yang telah saya buat adalah hasil karya milik seseorang atau badan tertentu, saya bersedia diproses baik secara pidana maupun perdata dan kelulusan saya dari Program Pascasarjana Magister Ilmu Komputer Universitas budi Luhur dicabut/ dibatalkan.

Jakarta, 12 Febuari 2015

Yang menyatakan,



Agus Darmawan



PROGRAM STUDI: MAGISTER ILMU KOMPUTER (MKOM)

PROGRAM PASCASARJANA

UNIVERSITAS BUDI LUHUR

LEMBAR PENGESAHAN

Nama : Agus Darmawan

Nomor Induk Mahasiswa : 1211601909

Konsentrasi : Sistem Informasi

Jenjang : Strata 2

Judul : **PENERAPAN MODEL SUPPORT VECTOR
MACHINES TEXTMINING PADA
SMARTPHONE ANDROID VS
BLACKBERRY DENGAN TEKNIK
OPTIMASI GENETIC ALGORITHM**

Jakarta, 25 Februari 2015

Tim Penguji

Tanda tangan :

Ketua,

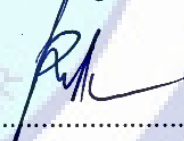
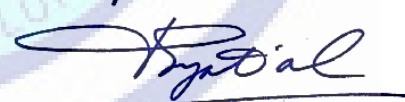
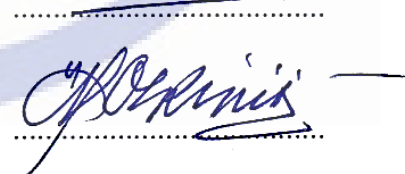
(Dr. Ir. Nazori AZ, MT)

Anggota

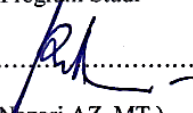
(Hadi Syahril, M.Kom)

Pembimbing

(Prof. Dr. Moedjiono, M.Sc)


.....

.....

.....

Ketua Program Studi


.....
(Dr. Ir. Nazori AZ, MT)

ABSTRAK

Pasar *smartphone* saat ini semakin banyak, penjualanya tidak hanya secara konvensional tetapi sudah merambah di *online shop*. Namun tidak semua *smartphone* memiliki kualitas yang baik untuk menunjang kebutuhan konsumen dan hal ini yang harus diperhatikan oleh para konsumen. Sebelum konsumen memutuskan untuk membeli *smartphone android* atau *belackberry* sebaiknya konsumen mengetahui dengan detail spesifikasi dan fungsi dari *smartphone* tersebut, hal ini dapat dipelajari dari testimoni dan opini atau hasil komentar *review* dari pengguna perbedaan *smartphone android* atau *belackberry*. Membaca komentar *review* tersebut secara keseluruhan dapat memakan waktu, namun jika hanya sedikit komentar *review* yang dibaca evaluasi akan menjadi bias. Dari beberapa teknik tersebut yang paling sering digunakan untuk klasifikasi data adalah *Support Vector Machines* (SVM). SVM memiliki kelebihan yaitu mampu mengidentifikasi dengan *hyperplane* terpisah yang memaksimalkan margin antara dua kelas yang berbeda. Pemilihan fitur sekaligus penyetingan parameter di SVM secara signifikan mempengaruhi hasil akurasi klasifikasi. Oleh karena itu, dalam penelitian ini digunakan penggabungan metode seleksi fitur, yaitu *Genetic Algoritm* (GA) agar bisa meningkatkan akurasi pengklasifikasi *Support Vector Machines*. Penelitian ini menghasilkan klasifikasi teks dalam bentuk positif atau negatif dari *review* produk *Smartphone*. Pengukuran berdasarkan akurasi *Support Vector Machines* sebelum dan sesudah penambahan metode seleksi fitur. Sedangkan pengukuran akurasi diukur dengan *confusion matrix* dan kurva ROC. Hasil penelitian menunjukkan peningkatan akurasi *Support Vector Machines* dari 71.00 % dengan penambahan *Genetic Algoritm* menjadi 78.02%.

Kata Kunci: Komentar, Review, SVM, *Genetic Algoritm* (GA) , *Confusion Matrix*, Kurva ROC

ABSTRAK

The smartphone market is now more and more, not only in the conventional penjualanya but have penetrated in the online shop. But not all smartphones have good quality to support the needs of consumers and it is to be noticed by the consumer. Before consumers decide to buy android smartphone or belackberry consumer should know the details of the specifications and functions of the smartphone, it can be learned from the testimony and opinion or the results of a user review comments or belackberry difference android smartphone. Reading the comments of the review as a whole can be time consuming, but if only a few comments are read reviews on evaluations will be biased. Of some of the techniques most often used for data classification is Support Vector Machines (SVM). SVM has the advantage of being able to identify the separate hyperplane that maximizes the margin between two different classes. Selection of features at once setup parameters in SVM significantly influence the results of classification accuracy. Therefore, in this study used the incorporation of feature selection methods, namely Genetic Encryption (GA) in order to improve the accuracy of the classifier Support Vector Machines. This research resulted in the classification of text in a positive or negative form of product reviews Smartphone. Measurement is based on Support Vector Machines accuracy before and after the addition of feature selection methods. While the measurement accuracy is measured by the confusion matrix and ROC curves. The results showed an increase in Support Vector Machines accuracy of 71.00% with the addition of Genetic Encryption be 78.02%

Keywords : Comments , Review, SVM , Genetic Encryption (GA) , Confusion Matrix , ROC curve

LEMBAR PERSEMBAHAN

*Setiap keberhasilan kita yang kita peroleh ,
Bukan karena dari usaha kita saja
Melainkan dari doa doa orang yang menyayangi kita
Oleh karena itu keberhasilan milik kita bersama ,
jangan pernah sombong ketika sudah berhasil
(Agus darmawan)*

Dengan mengucap puji syukur kepada Tuhan Yang Maha Esa, tesis ini kupersembahkan untuk :

1. Alm. Bapak dan Ibu tercinta yang telah membesarkan aku dan selalu membimbing, mendukung, memotivasi, memberi apa yang terbaik bagiku serta selalu mendoakan aku untuk meraih kesuksesanku serta kakak dan adik yang selalu mendoakan ku
2. Keluarga Besar Universitas Indraprasta PGRI yang turut membantu.
3. Sahabat-sahabatku yang telah menjadi curahan hatiku, yang telah memberiku semangat, aku selalu sayang kalian.
4. Syamsiah, yang selalu setia dan memberikan semangat tiada henti hentinya, aku akan selalu mencintaimu “Istriku sayang”.

KATA PENGANTAR

Puji dan syukur kehadiran Allah SWT atas berkah dan rahmat-Nya, sehingga penulisan tesis yang berjudul ini “*Penerapan Model Support Vector Machine TextMining Pada Smartphone Android VS Blackberry Dengan Teknik Optimisasi Genetic Algorithm* ” dapat terselesaikan dengan baik.

Tujuan pembuatan tesis ini adalah untuk memenuhi salah satu syarat dalam menyelesaikan jenjang Strata Dua (S-2) pada Program Studi Magister Ilmu Komputer (MKOM) Program Pascasarjana Universitas Budi Luhur

Tesis ini menggunakan *Text data mining, Support Vector Mahcine dan optimasi Genetic Algorithms*, dalam menentukan review komentar antara blackberry vs android. Penulis juga mencari dan menganalisa berbagai macam sumber referensi, baik dalam bentuk jurnal ilmiah, buku-buku literatur, internet, dll yang terkait dengan pembahasan pada tesis ini

Penulis menyadari bahwa tanpa bimbingan dan dukungan dari semua pihak dalam pembuatan tesis ini, maka penulis tidak dapat menyelesaikan tesis ini tepat pada waktunya. Untuk itu ijinkanlah penulis pada kesempatan ini untuk mengucapkan terima kasih yang sebesar-besarnya kepada :

1. Bapak Prof. Dr. Moedjiono, M.Sc., selaku Direktur Program Pascasarjana Universitas Budi Luhur sekaligus dosen pembimbing tesis yang telah membimbing dan memotivasi penulis dalam menyelesaikan penelitian tesis ini.
2. Bapak DR Ir. Nazori AZ MT, sebagai Ketua Program Studi Magister Ilmu Komputer Universitas Budi Luhur Jakarta.
3. Alm. Bapak Dr. Ir. Prabowo Pudjo Widodo, MS sebagai pembimbing tesis yang terdahulu yang sangat banyak membantu saya dalam menyelesaikan tesis ini.
4. Orang tua tercinta serta seluruh keluarga yang telah memberikan dukungan material dan moral kepada penulis.
5. “Syamsiah ”, yang selalu memberikan motivasi dan semangat serta doa setiap saat kepada penulis untuk menyelesaikan studi.

6. Seluruh staf pengajar (dosen) Program Studi Pascasarjana Magister Ilmu Komputer Universitas Budi Luhur yang telah memberikan pelajaran yang berarti bagi penulis selama menempuh studi.
7. Seluruh staf dan karyawan Program Pascasarjana Magister Ilmu Komputer Universitas Budi Luhur yang telah melayani penulis dengan baik selama kuliah.

Serta semua pihak yang terlalu banyak untuk penulis sebutkan satu persatu sehingga terwujudnya penulisan tesis ini. Penulis menyadari bahwa penulisan tesis ini masih jauh sekali dari sempurna, untuk itu penulis mohon kritik dan saran yang bersifat membangun demi kesempurnaan penulisan karya ilmiah yang penulis hasilkan untuk yang akan datang.

Akhir kata semoga tesis ini dapat bermanfaat bagi penulis dan khususnya bagi para pembaca yang berminat pada umumnya.

Jakarta, Februari 2015

Agus Darmawan
Penulis

DAFTAR ISI

	Halaman
ABSTRAK	i
ABSTRACT	ii
LEMBAR PERSEMBAHAN	iii
KATA PENGANTAR	iv
DAFTAR ISI	ii
DAFTAR GAMBAR	viii
DAFTAR TABEL	ix
DAFTAR LAMPIRAN	x
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Masalah Penelitian	3
1.2.1 Identifikasi Masalah	3
1.2.2 Batasan Masalah	4
1.2.3 Rumusan Masalah	4
1.3 Tujuan dan Manfaat Penelitian	4
1.3.1 Tujuan Penelitian	4
1.3.2 Manfaat Penelitian	5
1.4 Sistematika Penulisan	5
BAB II LANDASAN TEORI DAN KERANGKA PEMIKIRAN	6
2.1 Tinjauan Pustaka	6
2.1.1 Review Produk	6
2.1.2 Analisa Sentimen (<i>Sentiment Analysis</i>)	7
2.1.3 Pemilihan Fitur (<i>Feature Selection</i>)	8
2.1.4 Genetic Algorithm (GA)	10
2.1.5 Algoritma Support Vector Machine (SVM)	13
2.1.6 Pemilihan Parameter pada Support Vector Machine	14
2.1.7 Contoh Implementasi SVM	15
2.1.8 Validasi dan Evaluasi Model	17
2.2 Tinjauan Studi	17
2.2.1 Model Penelitian Basari	18
2.2.2 Model Penelitian Chou	20
2.2.3 Model Penelitian Ye	22
2.2.4 Penelitian Terkait	23
2.3 Tinjauan Obyek Penelitian	25
2.3.1 Review Produk <i>Smartphone</i>	25
2.3.2 Pemilihan Fitur(<i>Feature Selection</i>)	25
2.4 Kerangka Berpikir	26
2.5 Hipotesis	27
BAB III METODOLOGI PENELITIAN	28
3.1 Perancangan Penelitian	28
3.2 Pengumpulan Data	29
3.3 Pengolahan Data Awal	29

3.4	Metode Yang Diusulkan	30
3.5	Eksperimen Dan Pengujian Model	32
3.6	Evaluasi Dan Hasil	32
3.7	Jadwal Penelitian	33
BAB IV HASIL PENELITIAN DAN PEMBAHASAN		34
4.1	Hasil	34
4.1.1	Klasifikasi Teks Menggunakan Algoritma Support Vector Machine	34
4.1.2	Pengujian model dengan 10 Foldcross Validation	38
4.1.3	Optimasi Model dengan Metode Genetic Algoritma	39
4.1.4	Eksperimen Terhadap Indikator Model	39
4.2	Pembahasan	40
4.2.1	Pengukuran dengan <i>Confusion Matrix</i>	41
4.3	Desain dan Implementasi	43
4.4	Implikasi Penelitian	45
BAB V PENUTUP		47
5.1	Kesimpulan	47
5.2	Saran	47
DAFTAR PUSTAKA		48
LAMPIRAN-LAMPIRAN		51

DAFTAR GAMBAR

Gambar	Halaman
II.1. Prototipe Review Produk di amazon.com.....	6
II.2. Decision Rules padaKlasifikasi Review Produk di	6
II.3. Taksonomi Masalah Pengurangan Dimensi.....	9
II.4. Operasi Crossover dan Mutasi	11
II.5. SVM berusaha menemukan <i>Hyperplane</i> terbaik yang memisahkan kedua class negatif dan positif	14
II.6. Model yang diusulkan oleh Basari	18
II.7. Model yang diusulkan oleh Chou.....	20
II.8. Model yang diusulkan oleh Ye.....	21
II.9. Kerangka Pemikiran.....	26
III.1. Model yang diusulkan.....	30
IV.1. Desain model (SVM) menggunakan RapidMiner	38
IV.2. Desain model SVM dan GA didalam RapidMiner	40
IV.3. Kurva ROC Model (SVM) sebelum menggunakan metode GA	42
IV.4. Kurva ROC Model Support Vector Machine (SVM) setelah menggunakan metode Genetik Algorithm	43
IV.5. Diagram alir tahapan proses klasifikasi dengan Support Vector Machine (SVM) menggunakan Metode Genetic algorithm	44
IV.6. Tampilan aplikasi sentimen analisi <i>komentar</i>	45
IV.7. Tampilan aplikasi sentimen analisi <i>komentar</i>	45

DAFTAR TABEL

Tabel	Halaman
II.1. AND Problem.....	15
II.2. Penelitian yang terkait.....	22
III.1. Spesifikasi Sistem Komputer Minimum yang digunakan.....	31
III.2. Confusion Matrix	32
III.3. Jadwal Penelitian.....	32
IV.1. teks sebelum dan sesudah dilakukan proses tokenization	35
IV.2. teks sebelum dan sesudah dilakukan proses filter tokens	35
IV.3. teks sebelum dan sesudah dilakukan proses Stopwords	36
IV.4. teks sebelum dan sesudah dilakukan proses transform cases	36
IV.5. teks sebelum dan sesudah dilakukan proses Stemming	37
IV.7. Tabel dokumen yang dihitung nilai bobotnya.....	39
IV.8. Tabel indikator dan hasil pengujian	41
IV.9. Model (SVM) sebelum dan sesudah menggunakan metode GA ..	41
IV.10. Confusion Matrix (SVM) sebelum penambahan metode GA ...	42

DAFTAR LAMPIRAN

Lampiran	halaman
1 Tabel dataset kometar review positif	51
2 Tabel dataset kometar review negatif.....	58
3 Data Stoplist by Tala, F. Z	65
4 Langkah langkah SVM dalam Rafid Miner.....	69
5 Langkah langkah SVM + GA dalam Rafid Miner	72
6 Biodata Penulis.....	73



DAFTAR PUSTAKA

- [Basari et al., 2013] Basari, A. S. H dan Hussin, B., Ananta, I. G. P., dan Zeniarja, J. Opinion Mining of Movie Review using Hybrid Method of Support Vector Machine and Particle Swarm Optimization. *Procedia Engineering*, 53, 453–462. doi:10.1016/j.proeng.2013.02.059, 2013
- [Chou et al., 2014] Chou, J.-S., Cheng, M.-Y., Wu, Y.-W., dan Pham, A.-D. 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, 2014
- [Dehkharghani et al., 2014] Dehkharghani, R. dan Mercan, H. dan Javeed, A., Saygin, Y. Sentimental causal rule discovery from Twitter. *Expert Systems with Applications*, 41(10), 4950–4958. doi:10.1016/j.eswa.2014.02.024, 2014
- [Habernal et al., 2014] Habernal, I., Ptáček, T., dan Steinberger, J. (2014). Supervised sentiment analysis in Czech social media. *Information Processing & Management*, 50(5), 693–707. doi:10.1016/j.ipm.2014.05.001, 2014
- [Haddi et al., 2013] Haddi, E., Liu, X., dan Shi, Y. The Role of Text Pre-processing in Sentiment Analysis. *Procedia Computer Science*, 17, 26–32. doi:10.1016/j.procs.2013.05.005, 2013.
- [Han et al., 2007] Han, J., dan Kamber, M. *Data Mining Concepts and Techniques*. San Francisco: Diane Cerra, 2007.
- [Huang et al., 2008]. Huang, K., Yang, H., King, I., dan Lyu, M. *Machine Learning Modeling Data Locally And Globally*. Berlin Heidelberg: Zhejiang University Press, 2008.
- [Khan et al., 2014] Khan, K., Baharudin, B., dan Khan, A. Mining Opinion Components from Unstructured Reviews: A Review. *Journal of King Saud University - Computer and Information Sciences*. doi:10.1016/j.jksuci.2014.03.009, 2014.
- [L. Zhang et al., 2014] Zhang L dan Hua K dan Wang H dan Qian G., Sentiment Analysis on Reviews of Mobile Users. *Procedia Computer Science*, 34, 458–465. doi:10.1016/j.procs.2014.07.013, 2014
- [Larose, 2005] Larose, D.T ,*Data Mining Methods And Models*. Canada: John Wiley & Sons, Inc. 2005.