PENERAPAN PART-OF-SPEECH FILTERING PADA FEATURE SELECTION DALAM METODE SUPPORT VECTOR MACHINE TERHADAP ANALISA SENTIMEN TWITTER MENGENAI PEMILIHAN GUBERNUR DKI JAKARTA 2017

SKRIPSI

disusun oleh :

Fregy Damara

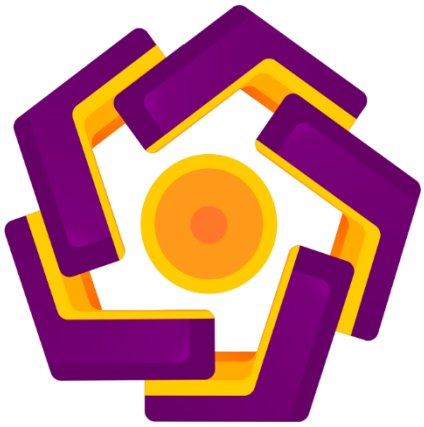
13.11.6927

PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2017

PENERAPAN PART-OF-SPEECH FILTERING PADA FEATURE SELECTION DALAM METODE SUPPORT VECTOR MACHINE TERHADAP ANALISA SENTIMEN TWITTER MENGENAI PEMILIHAN GUBERNUR DKI JAKARTA 2017

SKRIPSI

untuk memenuhi sebagian persyaratan
mencapai gelar Sarjana
pada Program Studi Informatika



disusun oleh:

Fregy Damara

13.11.6927

PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA

PERSETUJUAN

SKRIPSI

PENERAPAN PART-OF-SPEECH FILTERING PADA FEATURE SELECTION DALAM METODE SUPPORT VECTOR MACHINE TERHADAP ANALISA SENTIMEN TWITTER MENGENAI PEMILIHAN GUBERNUR DKI JAKARTA 2017

yang dipersiapkan dan disusun oleh

Fregy Damara

13.11.6927

telah disetujui oleh Dosen Pembimbing Skripsi

pada tanggal 25 Agustus 2017

Dosen Pembimbing,

Hartatik, ST, M.Cs

NIK. 190302232

PENGESAHAN

SKRIPSI

PENERAPAN PART-OF-SPEECH FILTERING PADA FEATURE SELECTION DALAM METODE SUPPORT VECTOR MACHINE TERHADAP ANALISA SENTIMEN TWITTER MENGENAI PEMILIHAN GUBERNUR DKI JAKARTA 2017

yang dipersiapkan dan disusun oleh

Fregy Damara

13.11.6927

telah dipertahankan di depan Dewan Penguji
pada tanggal 22 Agustus 2017

Susunan Dewan Penguji

Nama Penguji	Tanda Tangan
Yuli Astuti, M.Kom **NIK. 190302146**	
Bayu Setiaji, M.Kom **NIK. 190302216**	
Hartatik, ST, M.Cs **NIK. 190302232**	

Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 25 Agustus 2017

DEKAN FAKULTAS ILMU KOMPUTER

Krisnawati, S.Si, M.T.
NIK. 190302038

PERNYATAAN KEASLIAN

Saya yang bertanda tangan dibawah ini menyatakan bahwa, Skripsi ini merupakan karya saya sendiri (ASLI), dan isi dalam skripsi ini tidak terdapat karya yang pernah di ajukan oleh orang lain atau kelompok lain untuk memperoleh gelar akademis di suatu Institusi Pendidikan, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain atau kelompok lain, kecuali yang secara tertulis sebagai acuan dalam naskah ini dan disebutkan dalam daftar pustaka.

Segala sesuatu yang terkait dengan naskah dan karya yang telah dibuat adalah menjadi tanggung jawab saya sendiri

Yogyakarta, 25 Agustus 2017

Fregy Damara
13.11.6927

MOTTO

“Aut viam inveniam aut faciam”

I shall either find a way or make one

Robert Sidney, 1588

PERSEMBAHAN

Alhamdulillah segala puji syukur atas kehadirat Allah SWT yang telah memberikan rahmat dan karunia-Nya sehingga karya ini dapat terselesaikan dengan sebaik – baiknya, tidak lepas dari bantuan dan dukungan berbagai pihak.

Skripsi ini saya persembahkan dengan rasa syukur kepada saksi dan penolong seumur hidupku, Allah SWT

Untuk Bapak, Ibu, Adikku *jazakumullah khairan katsiir* atas pengertian dan harapan kalian, semoga Allah kuatkan kami untuk selalu berbakti..

Untuk Bu Hartatik *jazakillah* atas bimbingan, ilmu, alur berfikir, kerja keras dan pengorbanannya..

Semua pihak yang telah banyak membantu dalam penyusunan skripsi ini yang tidak dapat disebutkan satu persatu.

KATA PENGANTAR

Puji syukur penulis persembahkan atas kehadirat Allah SWT yang telah memberikan rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan skripsi dengan judul "*Penerapan Part-Of-Speech Filtering Pada Feature Selection Dalam Metode Support Vector Machine Terhadap Analisa Sentimen Twitter Mengenai Pemilihan Gubernur Dki Jakarta 2017"* dengan sebaik – baiknya. Tidak lupa sholawat serta salam penulis haturkan kepada junjungan umat Nabi Muhammad SAW, yang telah membawa umat Islam dari jaman jahiliyah ke jaman yang penuh ilmu pengetahuan.

Dengan selesainya skripsi ini, maka penulis mengucapkan terima kasih kepada :

1. Ibu Hartatik, ST, M.CS selaku dosen pembimbing yang telah sabar membimbing dan banyak memberikan pengarahan dan motivasi bagi penulis dalam pembuatan skripsi ini.
2. Bapak Hastari Utama, M.Cs selaku dosen yang membantu dalam penelitian skripsi ini.
3. Wahyu Setiawan, Muhammad Iqbal, Seno Aji, Anggun Putri, Yunisa A, Nabila Nadia, Muhammad Hatta Putra, Afdal Abdullah yang selalu menemani dan turut berpartisipasi dalam pengembangan penelitian ini.
4. Para Dosen dan Staff Universitas Amikom Yogyakarta yang telah banyak memberikan ilmu pengetahuan, pengalaman, dan bantuannya selama penulis kuliah hingga terselesaikannya skripsi ini.

5. Ibu, Bapak, dan Saudara – saudara penulis yang telah memberikan dukungan baik moril ataupun materil.
6. Teman-teman kuliah di kelas TI-03.
7. Teman-teman kontrakan, Arief Darmawan, Nirwan Darmawan, Kurnia Akbar, Fransiskus Paskalis, Izzul Admaza, Joko Budianto, Rizki Pramono, Jecklin Sianturi, Randy Julihartono, Ivan Julio, Ihsan Prasetyo, dan Teguh Kurniawan.
8. Semua pihak yang telah banyak membantu dalam penyusunan skripsi ini yang tidak dapat disebutkan satu persatu.

Penulis tentunya menyadari bahwa pembuatan skripsi ini masih banyak sekali kekurangan – kekurangan dan kelemahan – kelemahannya. Oleh karena itu penulis berharap kepada semua pihak agar dapat menyampaikan kritik dan saran yang membangun untuk menambah kesempurnaan skripsi ini. Semoga skripsi ini dapat bermanfaat bagi pihak terkait dan pembaca pada umumnya.

Yogyakarta, 25 Agustus 2017

Fregy Damara

13.11.6927

DAFTARI ISI

DAFTAR GAMBAR

DAFTAR TABEL

INTISARI

Di dalam penelitian ini, peneliti melakukan penelitian untuk mengetahui efek dari feature selection pada Support Vector Machine dalam melakukan klasifikasi sentimen pada tweet di Twitter. Input space yang diberikan terhadap SVM yaitu sebuah feature yang telah diproses melalui tahapan Part-of-Speech Filtering, yang berguna untuk menentukan porsi kata-kata yang sesuai untuk proses pembelajaran model dari perspektif teoritis maupun linguistik. Terdapat 4 tag yang di seleksi yaitu tag kata benda(NN), kata kerja(VB), kata sifat(JJ), dan kata keterangan(RB).

Input Space tersebut sebelumnya telah diolah melalui perhitungan bobot TF-IDF. Setelah TF-IDF masing-masing tweet diketahui, lalu akan diukur kedekatan TF-IDF tersebut dengan feature list positif, negatif, dan netral melalui perhitungan Cosine Similarity. Ketiga bobot Cosine Similarity ini yang nantinya akan di klasifikasi oleh Support Vector Machine.

Kesimpulan yang dihasilkan dari penelitian ini adalah akurasi model yang didapatkan oleh model tanpa proses POS Filtering mengungguli model dengan proses POS Filtering dengan persentase masing-masing 96.66 % dan 99.25 %. Adapun persentase akurasi prediksi yang dilakukan oleh masing-masing model yaitu sebesar 53,33% untuk pos filter dan 56,66% untuk POS Filter.

Kata Kunci : Support Vector Machine, Part Of Speech, Part Of Speech Filtering, Data Mining, Text Minning, Filtering Feature Selection, Analisis Sentimen.

ABSTRACT

In this research, the authors doing a research to determine the effects of feature selection on the Support Vector Machine in classifying sentiments on Twitter tweets. The input space given to SVM is a feature that has been processed through the Part-of-Speech Filtering stage, which is useful for determining the portion of words appropriate for the learning process model from theoretical and linguistic perspectives. There are 4 tags that are selected, that is the noun tag (NN), verb (VB), adjectives (JJ), and adverbs (RB).

Input Space has previously been processed by the calculation of TF-IDF weight. After TF-IDF of each tweet has been calculated, next step is measure the similarity of TF-IDF between each positive, negative and neutral feature list by calculating Cosine Similarity weights. These three Cosine Similarity weights will be classified by the Support Vector Machine.

In addition, the comparison of two models (POS Filter and without POS Filter) clarified that the models without POS Filtering outperformed the model with POS Filtering with the percentage of accuracy by 99,25%. The percentage of accuracy that obtained by the model with POS Filtering is 96.66%. The percentage of prediction accuracy done by each model is equal to 53,33% for filter post and 56,66% for POS Filter. This proves that the number of features in features list used in the Cosine Similarity weighting process has an effect on the classification process done by Support Vector Machine.

Keywords : *Support Vector Machine, Part Of Speech, Part Of Speech Filtering, Data Mining, Text Minning, Filtering Feature Selection, Sentiment Analysis.*