PENGARUH STEMMING DAN STOPWORDS TERHADAP AKURASI ANALISIS SENTIMEN MENGGUNAKAN METODE SUPPORT VECTOR MACHINE

SKRIPSI

disusun oleh

Aditya Wiha Pradana

15.11.9295

PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2019

PENGARUH STEMMING DAN STOPWORDS TERHADAP AKURASI ANALISIS SENTIMEN MENGGUNAKAN METODE SUPPORT VECTOR MACHINE

SKRIPSI

untuk memenuhi sebagian persyaratan
mencapai gelar Sarjana
pada Program Studi Informatika

disusun oleh

Aditya Wiha Pradana

15.11.9295

PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2019

PERSETUJUAN

SKRIPSI

PENGARUH STEMMING DAN STOPWORDS TERHADAP AKURASI ANALISIS SENTIMEN MENGGUNAKAN METODE SUPPORT VECTOR MACHINE

yang dipersiapkan dan disusun oleh

Aditya Wiha Pradana

15.11.9295

telah disetujui oleh Dosen Pembimbing Skripsi

pada tanggal 23 Juli 2019

Dosen Pembimbing,

Mardhiya Hayaty, S.T, M.Kom

NIK. 190302108

PENGESAHAN

SKRIPSI

PENGARUH STEMMING DAN STOPWORDS TERHADAP AKURASI ANALISIS SENTIMEN MENGGUNAKAN METODE SUPPORT VECTOR MACHINE

yang dipersiapkan dan disusun oleh
Aditya Wiha Pradana

15.11.9295

telah dipertahankan di depan Dewan Penguji
pada tanggal 23 Agustus 2019

Susunan Dewan Penguji

Nama Penguji	Tanda Tangan
Sri Ngudi Wahyuni, S.T, M.Kom **NIK. 190302060**	
Rumini, M.Kom **NIK. 190302246**	
Mardhiya Hayaty, S.T, M.Kom **NIK. 190302108**	

Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 5 September 2019

DEKAN FAKULTAS ILMU KOMPUTER

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

PERNYATAAN

Saya yang bertandatangan dibawah ini menyatakan bahwa, skripsi ini merupakan karya saya sendiri (ASLI), dan isi dalam skripsi ini tidak terdapat karya yang pernah diajukan oleh orang lain untuk memperoleh gelar akademis di suatu institusi pendidikan tinggi manapun, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis dan/atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

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

Yogyakarta, 23 Agustus 2019

Aditya Wiha Pradana

NIM. 15.11.9295

MOTTO

"Door Duisternis Tot Licht"

- R.A. Kartini

"それが悪魔であろうと聖人であろうと、私の名前は世界中で聞かれます！"

- Zorojuro

"The fool doth think he is wise, but the wise man knows himself to be a fool."

- William Shakespeare

"I don't care that they stole my idea … I care that they don't have any of their own."

- Nikola Tesla

PERSEMBAHAN

Alhamudlillah segala puji syukur atas kehadirat Allah SWT yang telah memberikan limpahan rahmat dan karunia-Nya sehingga penelitian ini dapat dilakukan dan diselesaikan dengan sebaik-baiknya. Saya juga ucapkan terimakasih untuk dukungan dan bantuan semua pihak yang membantu selesainya penelitan ini. Ucapan terimakasih saya persembahkan kepada :

1. Kedua orang tua dan semua keluarga yang senantiasa memberikan dukungan dan do'a kepada saya.
2. Ibu Mardhiya Hayaty, S.T, M.Kom selaku dosen pembimbing yang telah banyak memberikan bimbingan, pelajaran, serta ilmu yang sangat bermanfaat.
3. Teman-teman kelas 15-S1IF-12 yang menjadi teman seperjuangan dari awal perkuliahan hingga saat ini.
4. Sahabat-sahabat Nongki, yang banyak memberikan dukungan, kritik dan saran yang membantu proses penulisan skripsi ini.
5. Sobat Rodi, yang telah berjuang bersama dalam proses penulisan skripsi dari awal sampai selesai.
6. Teman-teman kos ijo, yang sudah menemani dan bersenang-senang bersama penulis selama berkuliah.

KATA PENGANTAR

Assalamualaikum Wr.Wb.

Puji syukur kepada allah SWT yang telah memberikan rahmat hidayah, karunia dan kesehatan, sehingga penulis dapat menyelesaikan laporan skripsi yang berjudul "Pengaruh Stemming dan Stopwords Terhadap Akurasi Analisis Sentimen Menggunakan Metode Support Vector Machine". Keberhasilan dalam menyelesaikan pembuatan laporan skripsi ini adalah berkat bantuan dan dukungan dari berbagai pihak. Maka dari itu pada kesempatan kali ini penulis ingin mengucapkan terimakasih yang sebesar-besarnya kepada:

1. Bapak Prof. Dr. M. Suyanto, M.M. selaku ketua yayasan Universitas Amikom Yogyakarta
2. Ibu Krisnawati, S.Si., M.T. selaku Dekan Fakultas Ilmu Komputer dan Bapak Sudarmawan, S.T., M.T. selaku Kepala Program Studi S1 Informatika.
3. Ibu Mardhiya Hayaty, S.T, M.Kom Selaku dosen pembimbing yang telah memberikan pengarahan, bimbingan dan motivasi selama proses penyusunan skripsi hingga selesai.
4. Segenap Dosen dan Karyawan Universitas Amikom Yogyakarta yang telah memberikan ilmu pengetahuan dan pengalamanya.
5. Kepada kedua orang tua penulis yang telah memberikan dukungan terbaiknya selama kuliah.
6. Kepada Teman-teman angkatan 2015 khususnya kelas 11-S1-TI12 yang telah berjuang bersama.

Penulis menyadari sepenuhnya bahwa laporan skripsi ini masih sangat jauh dari kesempurnaan, itu semua tidak lepas dari keterbatasan pengetahuan dan kemampuan dari penulis sendiri. untuk itu, penulis mengharapkan kritik dan saran

yang bersifat membangun guna mencapai kesempurnaan yang selalu penulis harapkan sehingga dapat bermanfaat bagi penulis, serta pihak-pihak yang membutuhkan.

Yogyakarta, 22 Oktober 2018

Aditya Wiha Pradana

15.11.9295

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

INTISARI

Preprocessing adalah tugas penting untuk analisis sentimen karena informasi tekstual membawa banyak data yang tidak terstruktur dan mengandung *noise*. *Stemming* maupun penghapusan *stopword* adalah teknik *preprocessing* yang cukup populer untuk klasifikasi teks. Namun, penelitian sebelumnya memberikan hasil yang berbeda mengenai pengaruh kedua metode tersebut terhadap akurasi pada klasifikasi sentimen.

Oleh karena itu, penelitian ini melakukan investigasi lebih lanjut tentang pengaruh *stemming* dan penghapusan *stopword* pada analisis sentimen bahasa Indonesia. Selain itu, penelitian ini mengusulkan dua kondisi preprocessing yang menggunakan *stemming* dan penghapusan *stopword* dan tanpa menggunakan keduanya. *Support Vector Machine* digunakan untuk algoritma klasifikasi dan TF-IDF sebagai skema pembobotan. Hasilnya dievaluasi menggunakan *confusion matrix* dan kemudian divalidasi menggunakan metode *k-fold cross validation*.

Hasil eksperimen menunjukkan bahwa semua akurasi tidak meningkat dan cenderung menurun ketika melakukan skenario *stemming* dan penghapusan *stopword*. Karya ini menyimpulkan bahwa penerapan teknik *stemming* dan penghapusan *stopword* tidak berpengaruh signifikan terdahap akurasi analisis sentimen pada dokumen teks bahasa Indonesia.

Kata Kunci: *stemming*, penghapusan *stopword*, *preprocessing*, analisis sentimen, *svm*

ABSTRACT

Preprocessing is an essential task for sentiment analysis since textual information carries a lot of noisy and unstructured data. Both stemming and stopword removal are pretty popular preprocessing techniques for text classification. However, the prior research gives different results concerning the influence of both methods toward accuracy on sentiment classification.

Therefore, this research conducts further investigations about the effect of stemming and stopword removal on Indonesian language sentiment analysis. Furthermore, we propose two preprocessing conditions which are with using both stemming and stopword removal and without using both. Support Vector Machine was used for the classification algorithm and TF-IDF as a weighting scheme. The result was evaluated using the confusion matrix and then validated using k-fold cross-validation methods.

The experiments result show that all accuracy did not improve and tends to decrease when performing stemming or stopword removal scenarios. This work concludes that the application of stemming and stopword removal technique does not significantly affect the accuracy of sentiment analysis in Indonesian text documents.

Keywords: *stemming, stopword removal, preprocessing, sentiment analysis, svm*