PENGARUH PENGGUNAAN STEMMING DAN LEMMATIZATION TERHADAP AKURASI ANALISIS SENTIMEN

SKRIPSI

disusun oleh

Fadel Maulana Ichsan

16.11.0587

PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2020

PENGARUH PENGGUNAAN STEMMING DAN LEMMATIZATION

TERHADAP AKURASI ANALISIS SENTIMEN

SKRIPSI

untuk memenuhi sebagian persyaratan
mencapai gelar Sarjana S1
pada jurusan Informatika

disusun oleh

Fadel Maulana Ichsan

16.11.0587

PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2020

PERSETUJUAN

SKRIPSI

PENGARUH PENGGUNAAN STEMMING DAN LEMMATIZATION TERHADAP AKURASI ANALISIS SENTIMEN

yang disusun oleh

Fadel Maulana Ichsan

16.11.0587

telah disetujuai oleh Dosen Pembimbing Skripsi

pada tanggal 4 Juli 2020

Dosen Pembimbing,

Mardhiya Hayaty. S.T, M.Kom
NIK. 190302108

PENGESAHAN

SKRIPSI

PENGARUH PENGGUNAAN STEMMING DAN LEMMATIZATION TERHADAP AKURASI ANALISIS SENTIMEN

yang disusun oleh

Fadel Maulana Ichsan

16.11.0587

telah dipertahankan di depan Dewan Penguji
pada tanggal 20 Juli 2020

Susunan Dewan Penguji

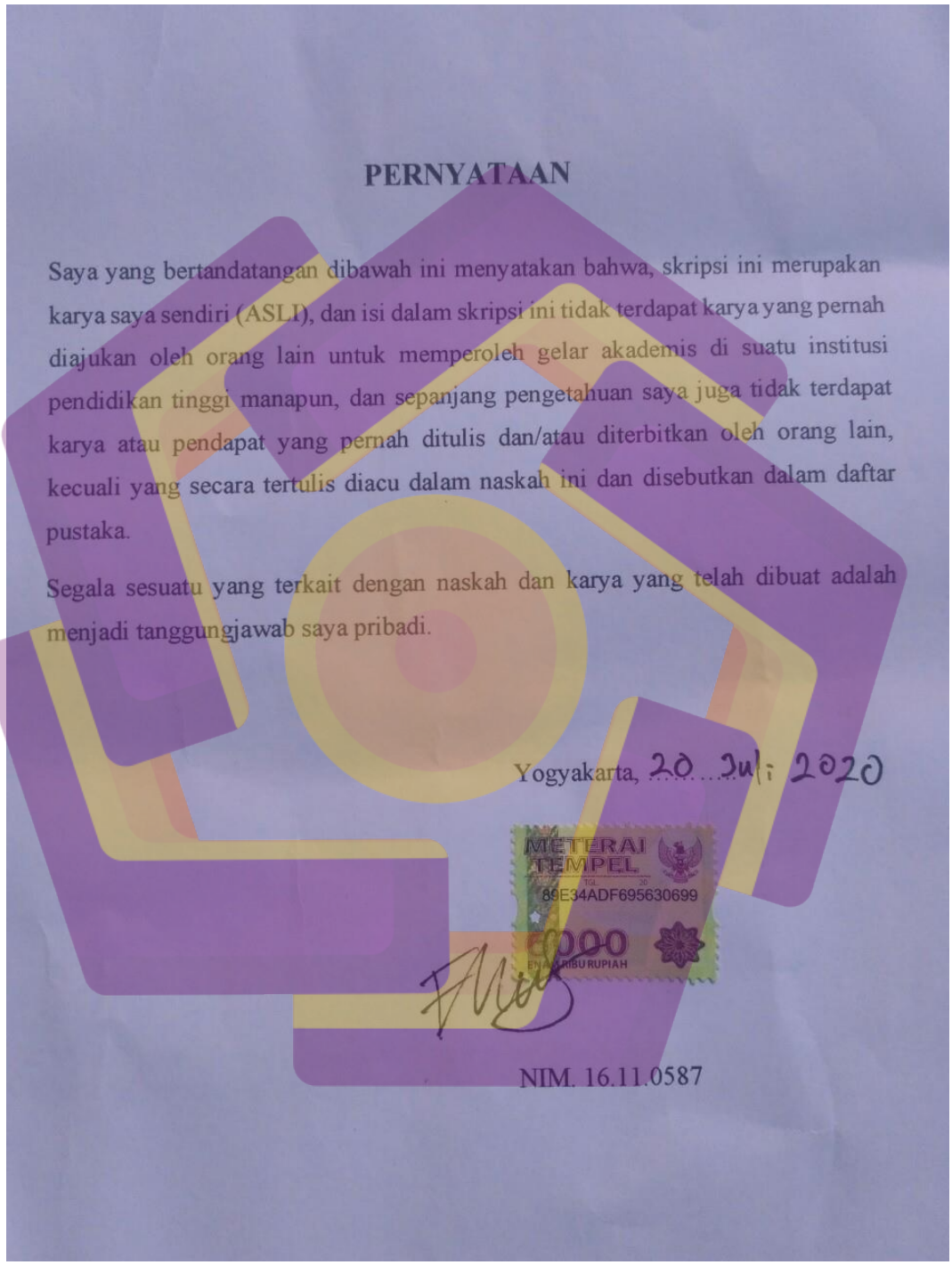
Nama Penguji	Tanda Tangan
Ainul Yaqin, M.Kom **NIK. 190302255**	_____
Majid Rahardi, S.Kom., M.Eng. **NIK. 190302393**	_____
Mardhiya Hayaty. S.T., M.Kom **NIK. 190302108**	_____

Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal .....

DEKAN FAKULTAS ILMU KOMPUTER

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

PERNYATAAN



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 tanggungjawab saya pribadi.

Yogyakarta, 20 Juli 2020

METERAI TEMPEL 89E34ADF695630699 6000 ENAM RIBU RUPIAH

NIM. 16.11.0587

MOTTO

“Bermimpilah setinggi langit, karena bila kau jatuh. Kau akan jatuh diantara bintang-bintang”

-Ir Soekarno

"Ask not what your country can do for you, but what you can do for your country."

-John F. Kennedy

“Life isn’t fair, get used to it”

-Patrick Star

“The only person you are destined to become, is the person you decide to be”

-Ralph Waldo Emerson

PERSEMBAHAN

Puji Syukur saya panjatkan kepada Allah SWT, karena berkat rahmat dan karunia-Nya sehingga saya dapat menyelesaikan skripsi ini. Untuk itu skripsi ini saya persembahkan untuk:

1. Kedua orang tua yang selalu memberikan motivasi, dorongan baik berupa materi maupun non materi. Berkat doa dari orang tua saya lah skripsi ini dapat terselesaikan.
2. Ibu Mardhiya Hayaty selaku dosen pembimbing yang senantiasa sabar dalam membimbing saya dan memberikan masukan serta saran terhadap skripsi yang saya kerjakan agar dapat terselesaikan dengan baik.
3. Seluruh Dosen Universitas Amikom Yogyakarta yang telah banyak memberikan ilmu pengetahuan serta informasi kepada saya yang sangat berguna dalam penyusunan skripsi ini.
4. Keluarga besar yang rutin menanyakan “kapan lulus”, terimakasih untuk support dan motivasinya sehingga saya bisa menyelesaikan skripsi ini.
5. Teman-teman grup “ Daily New Jokes” yang berisikan sobat sambat yang senantiasa menjadi ice breaker kala jenuh mengerjakan skripsi.
6. Teman-teman dari S1 IF-09 Universitas Amikom Yogyakarta yang selalu mendukung saya dari semester 1 hingga sekarang.
7. Creator dan pengguna Google, Github, Stackoverflow, Mas-mas youtuber tutorial dari India yang telah banyak berjasa membantu saya dalam pemecahan masalah skripsi ini.

KATA PENGANTAR

Puji dan syukur saya panjatkan kepada Tuhan Yang Maha Esa yang telah memberikan rahmat, hidayah dan kekuatan sehingga saya dapat menyelesaikan skripsi yang berjudul Pengaruh Penggunaan Stemming Dan Lemmatization Terhadap Akurasi Analisis Sentimen.

Skripsi ini saya buat guna menyelesaikan studi jenjang Strata Satu (S1) pada program studi Informatika fakultas Ilmu Komputer Universitas AMIKOM Yogyakarta. Selain itu juga merupakan suatu bukti bahwa mahasiswa telah menyelesaikan kuliah jenjang program strata 1 dan untuk memperoleh gelar Sarjana Komputer.

Dengan selesainya skripsi ini, Maka pada kesempatan ini saya mengucapkan terima kasih kepada :

1. Bapak Prof. Dr. M. Suyanto, MM. selaku Rektor Universitas AMIKOM Yogyakarta.
2. Bapak Sudarmawan, MT. selaku Ketua Program Studi S1 Informatika.
3. Ibu Mardhiya Hayaty, S.T, M.Kom. selaku Dosen Pembimbing, yang telah memberikan pengarahan yang sangat membantu dalam proses pembuatan skripsi ini.

Yogyakarta, 5 Juli 2020

Fadel Maulana Ichsan

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

INTISARI

Analisis sentimen adalah studi komputasional dari opini-opini orang dengan menggunakan ilmu *text mining*. Analisis sentimen akan mengelompokkan teks yang ada dalam kalimat atau dokumen untuk mengetahui pendapat yang dikemukakan dalam kalimat atau dokumen tersebut, bisa berupa sentimen positif, maupun negatif. Sebelum dapat mengelompokkan sentimen menjadi positif maupun negatif, data teks akan melalui tahap *preprocessing* terlebih dahulu. *Preprocessing* berguna untuk membersihkan data dari *noise*, serta mempermudah *classifier* untuk melakukan klasifikasi. Salah satu tahapan *preprocessing* data adalah *stemming* dan *lemmatization*. *Stemming* dan *lemmatization* adalah salah satu tahap *preprocessing* untuk mengubah kata berimbuhan menjadi kata dasar. Walaupun memiliki fungsi yang sama namun terdapat perbedaan dalam implementasi *stemming* dan *lemmatization*. Pada *stemming* pengubahan hanya dilakukan dengan memotong/ menghapus imbuhan kata. Sedangkan *lemmatization* memiliki alur yang lebih kompleks dengan melibatkan kamus bahasa untuk mencari kata dasar (*root*).

Pada penelitian ini dilakukan komparasi antara *stemming* dengan *lemmatization* untuk kasus analisis sentimen. Dataset akan diklasifikasi dengan algoritma *support vector machine*, kemudian dievaluasi dengan *confusion matrix* dan divalidasi menggunakan metode *k-fold cross validation* untuk melihat akurasi dari masing-masing preprocessing. Dan hasil akurasi keduanya akan dikomparasi menggunakan uji statistik untuk melihat apakah perbedaan diantara keduanya signifikan atau tidak.

Setelah dilakukan evaluasi dengan *confusion matrix*, *preprocessing* dengan *stemming* menghasilkan akurasi sebesar 85% sedangkan *lemmatization* menghasilkan akurasi 84%. Namun setelah dilakukan uji signifikansi, ternyata perbedaan dari keduanya tidak signifikan.

Kata-kunci: analisis sentimen, stemming, lemmatization, uji signifikan

ABSTRACT

Sentiment analysis is a computational study of people's opinions using text mining science. Sentiment analysis will classify the text in a sentence or document to find out the opinions expressed in the sentence or document, it can be in the form of positive or negative sentiments. Before being able to classify sentiments into positive or negative, text data will go through the preprocessing stage first. Preprocessing is useful for cleaning data from noise, as well as making it easier for classifiers to classify. One of the stages of data preprocessing is stemming and lemmatization. Stemming and lemmatization are preprocessing steps to convert affix words into basic words. Even though they have the same function, there are differences in the implementation of stemming and lemmatization. In stemming, changes are only done by cutting / removing affixes of words. Meanwhile, lemmatization has a more complex flow by involving a language dictionary to look for the root word.

In this study, a comparison between stemming and lemmatization was carried out for sentiment analysis cases. The dataset will be classified with a support vector machine algorithm, then evaluated with a confusion matrix and validated using the k-fold cross validation method to see the accuracy of each preprocessing. And the results of the accuracy of the two will be compared using statistical tests to see whether the difference between the two is significant or not.

After evaluating with confusion matrix preprocessing with stemming it produces an accuracy of 85% while lemmatization produces an accuracy of 84%. However, after the significance test was carried out, it turned out that the difference between the two was not significant.

Keywords: *sentiment analysis, stemming, lemmatization, significant test*