



UNIVERSITAS DIPONEGORO

**SUPPORT VECTOR MACHINE (SVM) DAN NAIVE BAYES GUNA
MEMBANDINGKAN ANALISIS SENTIMEN HASIL PENERJEMAHAN
BAHASA MENGGUNAKAN PUSTAKA TEXTBLOB DAN VADER PADA
BAHASA PEMROGRAMAN PYTHON**

TUGAS AKHIR

**Diajukan sebagai salah satu syarat untuk memperoleh gelar
Sarjana Teknik**

MUHAMMAD IKHSAN

21120116120033

FAKULTAS TEKNIK

DEPARTEMEN TEKNIK KOMPUTER

SEMARANG

FEBRUARI 2020

HALAMAN PENGESAHAN

Tugas akhir ini diajukan oleh :

Nama : *Muhammad Ikhsan*
NIM : *21120116120033*
Departemen : *Teknik Komputer*
Judul Tugas Akhir : *Support Vector Machine (SVM) Dan Naive Bayes
Guna Membandingkan Analisis Sentimen Hasil
Penerjemahan Bahasa Menggunakan Pustaka
Textblob Dan Vader Pada Bahasa Pemrograman
Python*

Telah berhasil dipertahankan di hadapan Tim Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana Teknik pada Departemen Teknik Komputer, Fakultas Teknik, Universitas Diponegoro


TIM PENGUJI

Pembimbing I : *Agung Budi Prasetyo, ST, M.TI, Ph.D*
Pembimbing II : *Dania Eridani, ST, M.Eng*
Ketua Penguji : *Kurniawan Teguh Martono, ST, MT*
Anggota Penguji : *Ike Pertiwi Windasari, ST, MI*



Semarang, Maret 2020

Kepala Departemen Teknik Komputer



Dr. R. Rizal Isyanto, ST, MM, MT

NIP. 197007272000121001

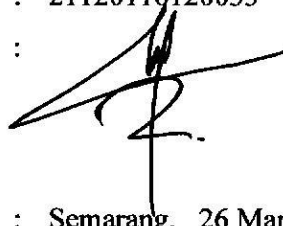
HALAMAN PERNYATAAN ORISINALITAS

Tugas Akhir ini adalah hasil karya saya sendiri, dan semua sumber baik yang dikutip maupun yang dirujuk telah saya nyatakan dengan benar.

Nama : Muhammad Ikhsan

NIM : 21120116120033

Tanda Tangan :

A handwritten signature in black ink, consisting of a stylized 'M' and 'I' followed by a horizontal line and a vertical line extending downwards.

Tanggal : Semarang, 26 Maret 2020

**HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS
AKHIR UNTUK KEPENTINGAN AKADEMIS**

Sebagai mahasiswa Universitas Diponegoro, saya yang bertanda tangan di bawah :

NAMA : MUHAMMAD IKHSAN
NIM : 21120116120033
DEPARTEMEN : TEKNIK KOMPUTER
FAKULTAS : TEKNIK
JENIS KARYA : TUGAS AKHIR

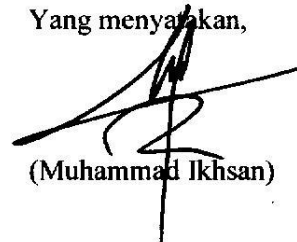
Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Diponegoro Hak Bebas Royalti Non-eksklusif (*Non-exclusive Royalty Free Right*) atas karya ilmiah saya berjudul : ***Support Vector Machine (SVM) Dan Naive Bayes Guna Membandingkan Analisis Sentimen Hasil Penerjemahan Bahasa Menggunakan Pustaka Textblob Dan Vader Pada Bahasa Pemrograman Python***. Dengan Hak Bebas Royalti/Non-eksklusif ini Universitas Diponegoro berhak menyimpan, mengalih-media/formatkan, merawat, dan mempublikasikan Tugas Akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Semarang

Pada tanggal : 26 Maret 2020

Yang menyatakan,



(Muhammad Ikhsan)

KATA PENGANTAR

Assalamualaikum Warahmatullahi Wabarakatuh

Alhamdulillah, puji syukur kehadiran Allah Subhanahu Wa Ta ala yang telah melimpahkan rahmat, taufik, dan hidayahnya sehingga penulis bisa menyelesaikan Tugas Akhir yang berjudul “ *Support Vector Machine (SVM) Dan Naive Bayes Guna Membandingkan Analisis Sentimen Hasil Penerjemahan Bahasa Menggunakan Pustaka *Textblob* Dan *Vader* Pada Bahasa Pemrograman *Python**” yang merupakan salah satu syarat yang harus dipenuhi untuk menyelesaikan pendidikan sarjana sesuai kurikulum di Departemen Teknik Komputer, Fakultas Teknik, Universitas Diponegoro dengan sebaik-baiknya dan tepat waktu.

Proses penyusunan laporan Tugas Akhir ini tidak lepas dari bantuan dan dukungan dari berbagai pihak. Oleh karena itu, penulis mengucapkan banyak terima kasih kepada :

1. Ayah, Ibu dan keluarga penulis yang selalu memberikan doa dan dukungan serta bimbingannya.
2. Bapak Dr. R. Rizal Isnanto, S.T., M.M., M.T. selaku Ketua Departemen Teknik Komputer Universitas Diponegoro.
3. Bapak Agung Budi Prasetijo, ST. M.IT, Ph.D dan Ibu Dania Eridani, ST, M.Eng. selaku dosen pembimbing Tugas Akhir.
4. Seluruh dosen, mahasiswa, staf, dan pegawai Teknik Komputer Undip.
5. Muhammad Rizky Nur Majiid dan Yogie Meysa Tama yang telah berbagi ilmu kepada penulis untuk menyelesaikan Tugas Akhir.
6. Siti Lina Suryaningsih dan Reza Kandriani yang telah membantu selama masa perkuliahan.
7. Laeli Chumaeroh, Aurel Firstanio dan Wulan Syaftira yang telah membantu pembuatan data uji yang digunakan.
8. Seluruh pihak yang telah membantu penulis hingga terselesaikannya Tugas Akhir ini.

Penulis berharap laporan Tugas Akhir ini dapat memberikan manfaat bagi masyarakat dan pengembang ilmu pengetahuan walaupun masih terdapat banyak kekurangan dan jauh dari kata sempurna. Akhir kata, penulis ucapkan terima kasih.

Semarang, 26 Maret 2020

A handwritten signature in black ink, consisting of several fluid, overlapping strokes that form a stylized representation of the name Muhammad Ikhsan.

Muhammad Ikhsan

DAFTAR ISI

HALAMAN PENGESAHAN.....	Error! Bookmark not defined.
HALAMAN PERNYATAAN ORISINALITAS...	Error! Bookmark not defined.
HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS	Error! Bookmark not defined.
KATA PENGANTAR	v
DAFTAR ISI.....	vii
DAFTAR TABEL.....	ix
DAFTAR GAMBAR	x
ABSTRAK	xi
BAB I PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah.....	2
1.3. Tujuan Penelitian	2
1.4. Batasan Masalah	2
1.5. Manfaat Penulisan.....	3
1.6. Sistematika Penulisan	3
BAB II KAJIAN PUSTAKA	5
2.1. Penelitian Terdahulu	5
2.2. <i>Text Mining</i>	6
2.3. <i>Natural Language Processing</i>	7
2.4. Analisis Sentimen	9
2.5. <i>Machine Learning</i>	9
2.6. <i>Term Frequency Inverse Document Frequency (TF-IDF)</i>	11
2.7. <i>Support Vector Machine (SVM)</i>	11
2.8. <i>Naive Bayes</i>	13
2.9. Ketepatan Klasifikasi	14

BAB III METODE PENELITIAN.....	15
3.1. Spesifikasi Perangkat Penelitian	15
3.2. Sumber Data.....	16
3.3. Struktur Data.....	16
3.4. Langkah Penelitian.....	18
BAB IV HASIL PENELITIAN DAN PEMBAHASAN	22
4.1. Pengumpulan Data	21
4.2. <i>Preprocessing</i>	26
4.3. Klasifikasi Menggunakan <i>Support Vector Machine</i>	29
4.4. Klasifikasi Menggunakan <i>Naive Bayes</i>	36
4.5. Rangkuman Dan Analisis	42
BAB V PENUTUP.....	44
5.1. Kesimpulan	44
5.2. Saran	45
DAFTAR PUSTAKA	46

DAFTAR TABEL

Tabel 2.1. Confusion Matrix	14
Tabel 3.1. Sampel data pada data latih.....	16
Tabel 3.2. Sampel data pada data uji.....	17
Tabel 4.1. Daftar Channel dan URL yang digunakan	21
Tabel 4.2. Variabel Hasil Keluaran dari <i>Scraping</i>	22
Tabel 4.3. Sampel data pada data uji.....	23
Tabel 4.4. Sampel data hasil penerjemahan bahasa pada data latih.....	24
Tabel 4.5. Sampel data pada data latih.....	25
Tabel 4.6. Contoh proses <i>case folding</i>	26
Tabel 4.7. Contoh proses <i>tokenization</i>	27
Tabel 4.8. Contoh proses <i>removal</i>	28
Tabel 4.9. Contoh proses Stemming	29
Tabel 4.10. Akurasi pelatihan model SVM <i>Kernel Linear</i>	30
Tabel 4.11. Perhitungan hasil klasifikasi menggunakan SVM <i>Kernel Linear</i>	32
Tabel 4.12. Akurasi pelatihan model SVM <i>Kernel RBF</i>	33
Tabel 4.13. Perhitungan hasil klasifikasi menggunakan SVM <i>Kernel RBF</i>	35
Tabel 4.14. Akurasi pelatihan model <i>Gaussian Naive Bayes</i>	36
Tabel 4.15. Perhitungan hasil klasifikasi menggunakan <i>Gaussian Naive Bayes</i> ..	38
Tabel 4.16. Akurasi pelatihan model <i>Bernoulli Naive Bayes</i>	39
Tabel 4.17. Perhitungan hasil klasifikasi menggunakan <i>Bernoulli Naive Bayes</i> ..	42
Tabel 4.18. Rangkuman akurasi pelatihan model	42
Tabel 4.19. Rangkuman akurasi pengujian model	43

DAFTAR GAMBAR

Gambar 2.1 Ilustrasi pencarian <i>hyperlane terbaik</i> dalam algoritma <i>Support Vector Machine</i>	11
Gambar 3.1. Visualisasi <i>wordcloud</i> data latih	17
Gambar 3.2. Visualisasi <i>wordcloud</i> data uji	18
Gambar 3.3. Diagram alir Tahapan Penelitian.....	20
Gambar 4.1 Visualisasi polaritas sentimen untuk data uji	23
Gambar 4.2. Visualisasi polaritas sentimen data latih	25
Gambar 4.3. Hasil pengujian model SVM <i>Linear</i> dengan data latih <i>textblob</i> terhadap (a) Sampel A; (b) Sampel B; (c) Sampel C.....	31
Gambar 4.4. Hasil pengujian model SVM <i>Linear</i> dengan data latih <i>vader</i> terhadap (a) Sampel A; (b) Sampel B; (c) Sampel C.....	32
Gambar 4.5. Hasil pengujian model SVM RBF dengan data latih <i>textblob</i> terhadap (a) Sampel A; (b) Sampel B; (c) Sampel C.....	34
Gambar 4.6. Hasil pengujian model SVM RBF dengan data latih <i>vader</i> terhadap (a) Sampel A; (b) Sampel B; (c) Sampel C	35
Gambar 4.7. Hasil pengujian model <i>Gaussian Naive Bayes</i> dengan data latih <i>textblob</i> terhadap (a) Sampel A; (b) Sampel B; (c) Sampel C	37
Gambar 4.8. Hasil pengujian model <i>Gaussian Naive Bayes</i> dengan data latih <i>vader</i> terhadap (a) Sampel A; (b) Sampel B; (c) Sampel C.....	38
Gambar 4.9. Hasil pengujian model <i>Bernoulli Naive Bayes</i> dengan data latih <i>textblob</i> terhadap (a) Sampel A; (b) Sampel B; (c) Sampel C	40
Gambar 4.10. Hasil pengujian model <i>Bernoulli Naive Bayes</i> dengan data latih <i>vader</i> terhadap (a) Sampel A; (b) Sampel B; (c) Sampel C.....	41

ABSTRAK

Kebutuhan analisis pada sebuah pendapat menjadi salah satu kegiatan penting yang dapat dilakukan untuk melakukan evaluasi terhadap sebuah kebijakan atau produk yang dikeluarkan. Kegiatan ini mengumpulkan pendapat beberapa orang untuk diklasifikasikan ke dalam kelas-kelas evaluasi yang diinginkan. Masalahnya untuk melakukan pengklasifikasian dibutuhkan beberapa data latih yang digunakan untuk melatih model mengenali bentuk kata atau kalimat yang ada pada masing-masing kelas.

Textblob dan *Vader* merupakan dua pustaka dalam bahasa pemrograman *Python* yang mampu mengklasifikasikan kata atau kalimat ke dalam beberapa kelas. Kedua pustaka ini mampu mengklasifikasi kata ke dalam tiga kelas berbeda yaitu kelas positif, negatif dan netral. Untuk melakukan perhitungan dan perbandingan hasil klasifikasi dari pustaka *textblob* dan *vader* menggunakan dua algoritma yaitu *Support Vector Machine* (SVM) dan *Naive Bayes*. Setiap algoritma menggunakan dua pendekatan berbeda sehingga total ada empat percobaan yang dilakukan dalam penelitian ini.

Dari empat kali percobaan yang dilakukan dalam penelitian ini, pustaka *vader* selalu mendapatkan nilai yang lebih baik dari pustaka *textblob* dalam semua aspek perhitungan. Perbandingan nilai tertinggi dari pustaka *textblob* dan *vader* adalah sebesar 0.618 dan 0.656 untuk perhitungan akurasi. Pada perhitungan presisi dan recall nilai tertinggi yang dihasilkan *textblob* dan *vader* sebesar 0.601 dan 0.648 pada perhitungan presisi dan sebesar 0.597 dan 0.617 pada perhitungan *recall*.

Kata Kunci : *Textblob, Vader, Support Vector Machine, Naive Bayes, Text Mining, Scraping, Youtube, Python.*

ABSTRACT

The Analysis in an opinion becomes one of the important activities that can be done to evaluate an issued policy and product. This activity collects opinions from several people to be classified into desired evaluation classes. The problem is to do classification, needs some of training data that used to train model to recognize words or sentences in each classes.

Textblob and vader is two library for python programming language that can classify words or sentences into several class. Both of them can classify of words into three different classes, namely positive, negative and neutral class. To calculate and compare the result of classification from textblob and vader use two algorithms that are Support Vector Machine (SVM) and Naive Bayes. Each algorithm uses two different Approaches so totally there are four experiments conducted in this study.

From the four experiments that conducted in this study, vader always gets better score than textblob in all aspect of calculation. The highest comparison of textblob and vader is 0.618 and 0.656 for accuracy calculations. In the calculation of precision and recall the highest value from textblob and vader is 0.601 and 0.648 in the precision calculation and 0.597 and 0.617 in the recall calculation.

Keywords : *Textblob, Vader, Support Vector Machine, Naive Bayes, Text Mining, Scraping, Youtube, Python.*