

**PERBANDINGAN METODE PEMBOBOTAN TF-IDF DAN TF-RF
DALAM PERFORMA ALGORITMA SUPPORT VECTOR
MACHINE PADA ANALISIS SENTIMEN**

SKRIPSI



**disusun oleh
Rizki Arif Priyanto Nugroho
15.11.9304**

**PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2020**

**PERBANDINGAN METODE PEMBOBOTAN TF-IDF DAN TF-RF
DALAM PERFORMA ALGORITMA SUPPORT VECTOR
MACHINE PADA ANALISIS SENTIMEN
SKRIPSI**

untuk memenuhi sebagian persyaratan
mencapai gelar Sarjana
pada Program Studi Informatika



**disusun oleh
Rizki Arif Priyanto Nugroho
15.11.9304**

**PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2020**

PERSETUJUAN

SKRIPSI

PERBANDINGAN METODE PEMBOBOTAN TF-IDF DAN TF-RF DALAM PERFORMA ALGORITMA SUPPORT VECTOR MACHINE PADA ANALISIS SENTIMEN

yang dipersiapkan dan disusun oleh

Rizki Arif Priyanto Nugroho

15.11.9304

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 21 Februari 2019

Dosen Pembimbing,

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

NIK. 190302108

PENGESAHAN
SKRIPSI
PERBANDINGAN METODE PEMBOBOTAN TF-IDF
DAN TF-RF DALAM PERFORMA ALGORITMA
SUPPORT VECTOR MACHINE PADA ANALISIS
SENTIMEN

yang dipersiapkan dan disusun oleh
Rizki Arif Priyanto Nugroho
15.11.9304
telah dipertahankan di depan Dewan Penguji
pada tanggal 10 Februari 2020

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Rizqi Sukma Kharisma, M.Kom
NIK. 190302057

Erni Seniwati, S.Kom., M.Cs
NIK. 190302254

Mardhiya Hayaty, S.T., M.Kom
NIK. 190302185

Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 20 Februari 2020

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 tanggungjawab saya pribadi.

Yogyakarta, 20 Februari 2020

Rizki Arif Priyanto Nugroho
NIM. 15.11.9304

PERSEMBAHAN

Alhamdulillah segala puji syukur atas kehadiran 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 penelitian ini.

Skripsi ini saya persembahkan untuk kampus saya, Universitas AMIKOM Yogyakarta.

Untuk kedua orang tua saya yang telah susah payah mendukung saya tanpa kenal lelah.

Untuk ibu Mardhiya, yang sudah dengan sabar membimbing saya. Saya ucapkan terimakasih sedalam-dalamnya.

Untuk para pembaca, saya ucapkan terima kasih sebesar-besarnya. Semoga menjadi amal jariyah atas ilmu yang saya bagikan.

Dan untuk semua pihak yang telah banyak membantu dalam penyusunan skripsi ini yang tidak dapat saya sebutkan satu persatu.

MOTTO

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

” *Nothing last forever we can change the future.* ”

- Alucard Mobile Legends

” *Not knowing oneself, that's the worst* ”

- Chou Mobile Legends

” *Love other's as love yourself* ”

- Estes Mobile Legends

” *There is no shame in losing one battle. But you must take revenge by any means necessary!* ”

- Owl Sekiro: Shadows Die Twice

” *Sometimes, you can't think everything out. You have to listen to your heart. 'And, Heroic? Human? Those are just things people say after fact... It's not what the people say afterward... What's important is being true to oneself* ”

- Zidane Final fantasy IX

KATA PENGANTAR

Puji syukur penulis haturkan atas kehadiran Allah *Subhanahu wa Ta'ala* yang telah memberikan rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan skripsi dengan judul “Perbandingan Metode Support Vector Machine dan K-Nearest Neighbor Dalam Analisis Sentimen Twitter Terhadap Presiden Amerika Serikat Donald Trump” dengan sebaik-baiknya. Tidak lupa shalawat serta salam penulis haturkan kepada junjungan kita, nabi agung Nabi Muhammad *Shallallahu 'alaihi wa Sallam* yang telah membawa cahaya ilmu pengetahuan ke tengah-tengah kita.

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

1. Bapak Prof. Dr. M. Suyanto selaku Rektor Universitas AMIKOM Yogyakarta
2. Bapak Sudarmawan, S.T., M.T. selaku Ketua Program Studi S1 Informatika serta Wali Dosen penulis.
3. Ibu Krisnawati, S.Si., M.T. selaku Dekan Fakultas Ilmu Komputer.
4. Ibu Mardhiya Hayaty, S.T., M.Kom. selaku pembimbing yang telah dengan sabar membimbing dan banyak memberikan kontribusi bagi penulis dalam pembuatan skripsi ini.
5. Para Dosen dan Staff Universitas AMIKOM Yogyakarta yang telah banyak memberi bantuan informasi, pengalaman dan kontribusi lainnya selama penulis berkuliah di kampus ini hingga terselesaikannya skripsi.
6. Orang tua dan saudara penulis yang telah memberikan dukungan terbaiknya selama berkuliah.
7. Makhluk-makhluk Kontrakan yang sudah menemani dan bersenang-senang bersama penulis selama berkuliah.
8. Penghuni Group Ninja Saga Indonesia yang selalu menjadi tempat berkumpulnya para sadboy dan sadgirl.

9. Keluarga besar Gamebrott yang telah menjadi tempat mengais rezeki serta menjadi tempat curhat selama pengerjaan skripsi ini.

Penulis menyadari dengan betul bahawa dalam pembuatan skripsi ini masih banyak sekali kekurangan dan kelemahan. Oleh karena itu penulis berharap untuk semua pihak yang telah membaca dan memahami penelitian ini untuk dapat menyampaikan kritik dan saran yang membangun sehingga dapat menambah kesempurnaan skripsi ini.

Akhir kata, semoga skripsi ini dapat bermanfaat bagi pihak terkait dan pembaca pada umumnya.

Yogyakarta, 20 Februari 2020

Rizki Arif Priyanto Nugroho
NIM. 15.11.9304

DAFTAR ISI

| | |
|---|----------|
| COVER | i |
| PERSETUJUAN | iii |
| PENGESAHAN | iv |
| PERNYATAAN | v |
| PERSEMBAHAN | vi |
| MOTTO | vii |
| DAFTAR ISI | viii |
| DAFTAR GAMBAR | xii |
| DAFTAR TABEL | xiii |
| INTISARI | xiv |
| <i>ABSTRACT</i> | xv |
| BAB I PENDAHULUAN | 1 |
| 1.1 Latar Belakang | 1 |
| 1.2 Rumusan Masalah | 2 |
| 1.3 Batasan Masalah | 2 |
| 1.4 Maksud dan Tujuan Penelitian | 3 |
| 1.5 Manfaat Penelitian | 3 |
| 1.6 Metode Penelitian | 3 |
| 1.7 Sistematika Penulisan | 5 |
| BAB II LANDASAN TEORI | 7 |
| 2.1 Tinjauan Pustaka | 7 |
| 2.2 Landasan Teori | 10 |
| 2.2.1 Text Mining | 10 |
| 2.2.2 Analisis Sentimen | 10 |
| 2.2.3 Preprocessing Data | 10 |
| 2.2.4 Klasifikasi | 11 |
| 2.2.5 Term Frequency – Inverse Document Frequency | 12 |
| 2.2.6 Term Frequency- Relevant Frequency | 13 |

| | | |
|--|--|-----------|
| 2.2.7 | Support Vector Machine | 14 |
| 2.2.8 | Evaluasi..... | 20 |
| 2.2.9 | Validasi | 21 |
| BAB III METODE PENELITIAN | | 23 |
| 3.1 | Alat dan Bahan..... | 23 |
| 3.1.1 | Alat Penelitian..... | 23 |
| 3.1.2 | Bahan Penelitian | 24 |
| 3.2 | Alur Penelitian | 24 |
| 3.2.1 | Dataset..... | 26 |
| 3.2.2 | Pembobotan..... | 26 |
| 3.2.3 | Klasifikasi | 27 |
| 3.2.4 | Evaluasi..... | 28 |
| 3.2.5 | Validasi | 29 |
| BAB IV HASIL DAN PEMBAHASAN | | 30 |
| 4.1 | Eksperimen | 30 |
| 4.1.1 | Pembagian Data | 30 |
| 4.1.2 | Pembobotan..... | 31 |
| 4.1.3 | Pelatihan dan Pengujian Menggunakan SVM..... | 38 |
| 4.1.4 | Evaluasi..... | 47 |
| 4.1.5 | Validasi | 48 |
| 4.2 | Hasil dan Pembahasan | 50 |
| 4.2.1 | Hasil Evaluasi | 50 |
| 4.2.2 | Hasil Validasi..... | 53 |
| 4.2.3 | Analisis Hasil..... | 55 |
| BAB V KESIMPULAN | | 59 |
| 5.1 | Kesimpulan..... | 59 |
| 5.2 | Saran | 60 |
| DAFTAR PUSTAKA | | 61 |

DAFTAR GAMBAR

| | |
|--|----|
| Gambar 2.1 (a) Pencarian Hyperplane (b) Hyperplane Terbaik | 14 |
| Gambar 2.2 Maximum hyperplane dengan support vector | 15 |
| Gambar 2.3 Confusion Matrix | 20 |
| Gambar 2.4 Ilustrasi K-fold Cross Validation | 22 |
| Gambar 3.1 Diagram alir tahapan eksperimen | 25 |
| Gambar 3.2 Diagram alir tahapan SVM | 28 |
| Gambar 4.1 Implementasi pembagian data | 30 |
| Gambar 4.2 Implementasi TF-IDF | 31 |
| Gambar 4.3 Implementasi TF-RF | 36 |
| Gambar 4.4 Pelatihan dan pengujian menggunakan SVM | 38 |
| Gambar 4.5 Evaluasi menggunakan Confusion Matrix | 48 |
| Gambar 4.6 Validasi menggunakan K-fold cross validation | 49 |
| Gambar 4.7 Grafik Garis Hasil pembobotan TF-IDF | 51 |
| Gambar 4.8 Grafik Garis Hasil pembobotan TF-RF | 52 |
| Gambar 4.9 Grafik Garis K-Fold CV pada pembobotan TF-IDF | 54 |
| Gambar 4.10 Grafik Garis K-Fold CV pada pembobotan TF-IDF | 55 |
| Gambar 4.11 Grafik perbandingan akurasi Confusin Matrik | 56 |
| Gambar 4.12 Perbandingan Akurasi menggunakan K-Fold Cross validation | 57 |
| Gambar 4.13 Grafik Batang perbandingan rata-rata akurasi K-Fold CV | 57 |

DAFTAR TABEL

| | |
|---|----|
| Tabel 3.1 Spesifikasi Perangkat keras dan Perangkat Lunak..... | 23 |
| Tabel 3.2 Detail Dataset Penelitian..... | 24 |
| Tabel 4.1 Contoh proses pembobotan Term Frequency | 33 |
| Tabel 4.2 Contoh proses pembobotan Inverse Document Frequency..... | 34 |
| Tabel 4.3 Contoh proses pembobotan menggunakan TF-IDF..... | 35 |
| Tabel 4.4 Contoh proses pembobotan Relevance Frequency | 36 |
| Tabel 3.3 Contoh proses pembobotan menggunakan TF-RF | 37 |
| Tabel 4.5 Data latih dalam format SVM..... | 39 |
| Tabel 4.6 Hasil kernelisasi matrix xTx | 45 |
| Tabel 4.7 Hasil Evaluasi pembobotan TF-IDF | 50 |
| Tabel 4.8 Hasil evaluasi pembobotan TF-RF | 51 |
| Tabel 4.9 Hasil validasi Tf-IDF Menggunakan K-fold Cross Validation..... | 53 |
| Tabel 4.10 Hasil Validasi TF-RF Menggunakan K-fold Cross Validation | 54 |

INTISARI

Pada analisis sentimen terdiri dari beberapa proses, yaitu *preprocessing*, pembobotan fitur, klasifikasi dan evaluasi. Proses metode pembobotan ternyata proses penting yang dapat mempengaruhi performa dari algoritma klasifikasi yang diterapkan.

Berberapa penelitian sebelumnya melakukan penelitian terkait fitur pembobotan untuk mengetahui pengaruhnya terhadap performa algoritma klasifikasi. Maka dari itu, penelitian ini akan meneliti pengaruh metode pembobotan terhadap performa algoritma klasifikasi.

Pada penelitian ini, peneliti tidak melakukan proses *preprocessing* karena menggunakan dataset milik peneliti sebelumnya. Metode pembobotan yang digunakan adalah TF-IDF dan TF-RF. Metode pembobotan TF-IDF adalah salah satu metode pembobotan yang populer digunakan dalam membobotkan kata. Sedangkan TF-RF adalah metode pembobotan penyempurnaan dari metode pembobotan sebelumnya yang masih tradisional. Setelah melewati pembobotan kata, kemudian diklasifikasikan menggunakan algoritma *Support Vector Machine* yang dikombinasikan dengan tokenisasi fitur N-Gram. Hasil dari penelitian ini akan dievaluasi menggunakan *Confusion Matrix* dan *K-fold Cross Validation*.

Hasil dari penelitian ini menunjukkan akurasi dari kedua metode pembobotan tersebut tak memiliki perbedaan yang signifikan. Kesimpulan dari penelitian ini menunjukkan jika akurasi dari TF-IDF dan TF-RF tidak memiliki perbedaan yang signifikan dalam dokumen berbahasa Indonesia.

Kata Kunci: analisis sentimen, metode pembobotan, TF-IDF, TF-RF, support vector machine

ABSTRACT

Sentiment analysis consists of several processes, namely pre-processing, feature weighting, classification, and evaluation. The weighting method process turns out to be an important process that can affect the performance of the classification algorithm applied.

Several previous studies conducted research related to the weighting feature to determine its effect on the performance of classification algorithms. Therefore, this study will examine the effect of weighting methods on the performance of classification algorithms.

In this study, researchers did not do the pre-processing because they used the dataset of the previous researchers. The weighting methods used are TF-IDF and TF-RF. TF-IDF weighting method is one of the popular weighting methods used in weighting words. While TF-RF is a method of weighting the refinement of the traditional TF-IDF. After passing the word weighting, then classified using the Support Vector Machine algorithm in combination with the tokenization of the N-Gram feature.

The results experiment will be evaluated using Confusion Matrix and K-fold Cross-Validation method. The results of this experiment show the accuracy of the two weighting methods does not have a significant difference. The conclusion of this experiment shows that the accuracy of TF-IDF and TF-RF does not have a significant difference in Indonesian language documents.

Keyword : *sentiment analysis, term weight method, TF-IDF, TF-RF, support vector machin*