

**PERBANDINGAN METODE *SUPPORT VECTOR MACHINE (SVM)*  
DAN *RANDOM FOREST (RF)* UNTUK DETEKSI *CYBERBULLYING*  
PADA SOSIAL MEDIA**

**SKRIPSI**

untuk memenuhi salah satu syarat mencapai derajat Sarjana  
Program Studi Informatika



diajukan oleh

**SAGIYAN DIMAS AGUSTI**

**17.11.1650**

Kepada

**PROGRAM SARJANA**

**PROGRAM STUDI INFORMATIKA**

**FAKULTAS ILMU KOMPUTER**

**UNIVERSITAS AMIKOM YOGYAKARTA**

**YOGYAKARTA**

**2022**

**PERBANDINGAN METODE *SUPPORT VECTOR MACHINE (SVM)*  
DAN *RANDOM FOREST (RF)* UNTUK DETEKSI *CYBERBULLYING*  
PADA SOSIAL MEDIA**

**SKRIPSI**

untuk memenuhi salah satu syarat mencapai derajat Sarjana  
Program Studi Informatika



diajukan oleh

**SAGIYAN DIMAS AGUSTI**

**17.11.1650**

Kepada

**PROGRAM SARJANA**

**PROGRAM STUDI INFORMATIKA**

**FAKULTAS ILMU KOMPUTER**

**UNIVERSITAS AMIKOM YOGYAKARTA**

**YOGYAKARTA**

**HALAMAN PERSETUJUAN**

**SKRIPSI**

**PERBANDINGAN METODE *SUPPORT VECTOR MACHINE (SVM)* DAN  
*RANDOM FOREST (RF)* UNTUK DETEKSI *CYBERBULLYING*  
PADA SOSIAL MEDIA**

yang disusun dan diajukan oleh

**SAGIYAN DIMAS AGUSTI**

**17.11.1650**

telah disetujui oleh Dosen Pembimbing Skripsi  
pada tanggal 9 Juni 2022

**Dosen Pembimbing,**

**Mulia Sulistiyono, M.Kom**

**NIK. 190302248**

**HALAMAN PENGESAHAN**

**SKRIPSI**

**PERBANDINGAN METODE *SUPPORT VECTOR MACHINE (SVM)* DAN  
*RANDOM FOREST (RF)* UNTUK DETEKSI *CYBERBULLYING* PADA SOSIAL  
MEDIA**

yang disusun dan diajukan oleh

**Sagiyan Dimas Agusti**

**17.11.1650**

Telah dipertahankan di depan Dewan Penguji  
pada tanggal 27 Juni 2022

**Susunan Dewan Penguji**

**Nama Penguji**

**Tanda Tangan**

**Ferry Wahyu Wibowo, S.Si, M.Cs**  
**NIK. 190302235**

**Andriyan Dwi Putra, M.Kom**  
**NIK. 190302270**

**Mulia Sulistiyono, M.Kom**  
**NIK. 190302248**

Skripsi ini telah diterima sebagai salah satu persyaratan  
untuk memperoleh gelar Sarjana Komputer  
Tanggal 1 Juli 2022

**DEKAN FAKULTAS ILMU KOMPUTER**

**Hanif Al Fatta, S.Kom., M.Kom.**  
**NIK. 190302096**

## HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Sagiyan Dimas Agusti  
NIM : 17.11.1650

Menyatakan bahwa Skripsi dengan judul berikut:

**Perbandingan Metode *Support Vector Machine (SVM)* Dan *Random Forest (RF)* Untuk Deteksi *Cyberbullying* Pada Sosial Media**

Dosen Pembimbing : Mulia Sulistiyono, M.Kom

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 23 Juni 2022

Yang Menyatakan,

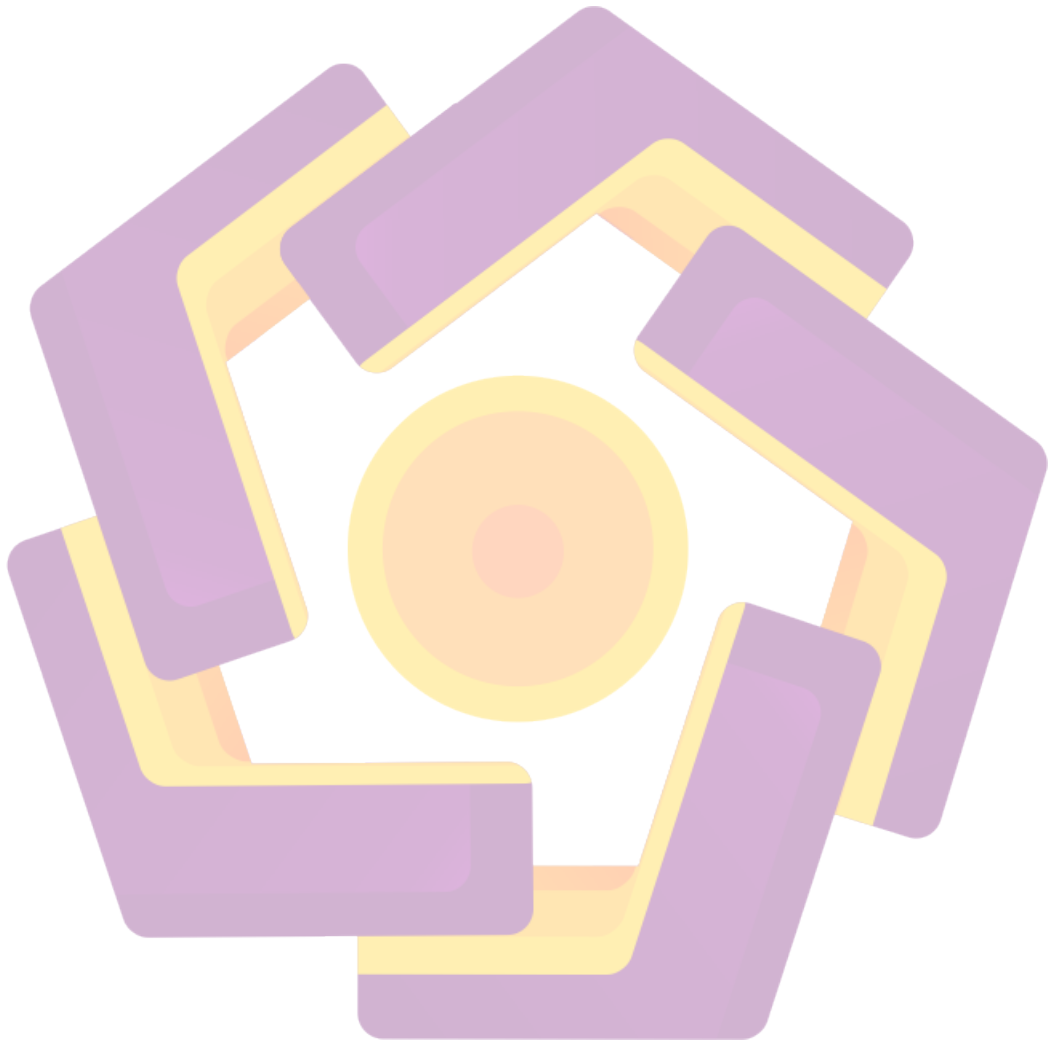


Sagiyan Dimas Agusti

## **MOTTO**

“Rezeki tidak mungkin tertukar, tidak pernah berkurang, dan tidak akan hilang”

(Penulis)



## KATA PENGANTAR

Puji dan syukur kehadiran Allah SWT yang telah memberikan rahmat dan hidayah-Nya, sehingga penulis dapat menyelesaikan Tugas Akhir / Skripsi dengan topic penelitian “Perbandingan Metode *Support Vector Machine (SVM)* Dan *Random Forest (RF)* Untuk Deteksi *Cyberbullying* Pada Sosial Media” ini sebagai salah satu syarat dalam menyelesaikan program S1 Informatika. Di Fakultas Ilmu Komputer Universitas Amikom Yogyakarta. Penulis mengucapkan terima kasih kepada pihak-pihak yang telah membantu dalam penyelesaian Tugas Akhir ini, terutama kepada:

1. Teristimewa kepada Bapak saya Hargiyanto dan almarhumah Ibu saya Saeni serta seluruh keluarga besar yang telah banyak memberikan dukungan moral dan material.
2. Bapak Prof. Dr. M. Suyanto, M.M selaku Rektor Universitas Amikom Yogyakarta
3. Ibu Windha Mega Pradnya D, M.Kom selaku ketua jurusan S1 Informatika Universitas Amikom Yogyakarta
4. Bapak Mulia Sulistiyono, M.Kom selaku Pembimbing Tugas Akhir Skripsi.
5. Seluruh Dosen dan Staff Jurusan Informatika Fakultas Ilmu Komputer Universitas Amikom Yogyakarta
6. Kepada *support system* yang menemani saya Vira Suci Ramayani.
7. Dan semua pihak yang telah mensupport saya yang tidak bisa saya sebutkan satu persatu.

Penulis menyadari bahwa dalam penulisan ini yang jauh dari sempurna, baik dari segi penyusunan, bahasa, ataupun penulisannya. Oleh karena itu penulis mengharapkan kritik dan saran yang sifatnya membangun.

Akhir kata penulis mengucapkan terima kasih dan semoga Tugas Akhir ini bermanfaat terutama untuk penulis sendiri, perusahaan serta bagi yang membaca.

Yogyakarta, 23 Juni 2022

Penulis

## DAFTAR ISI

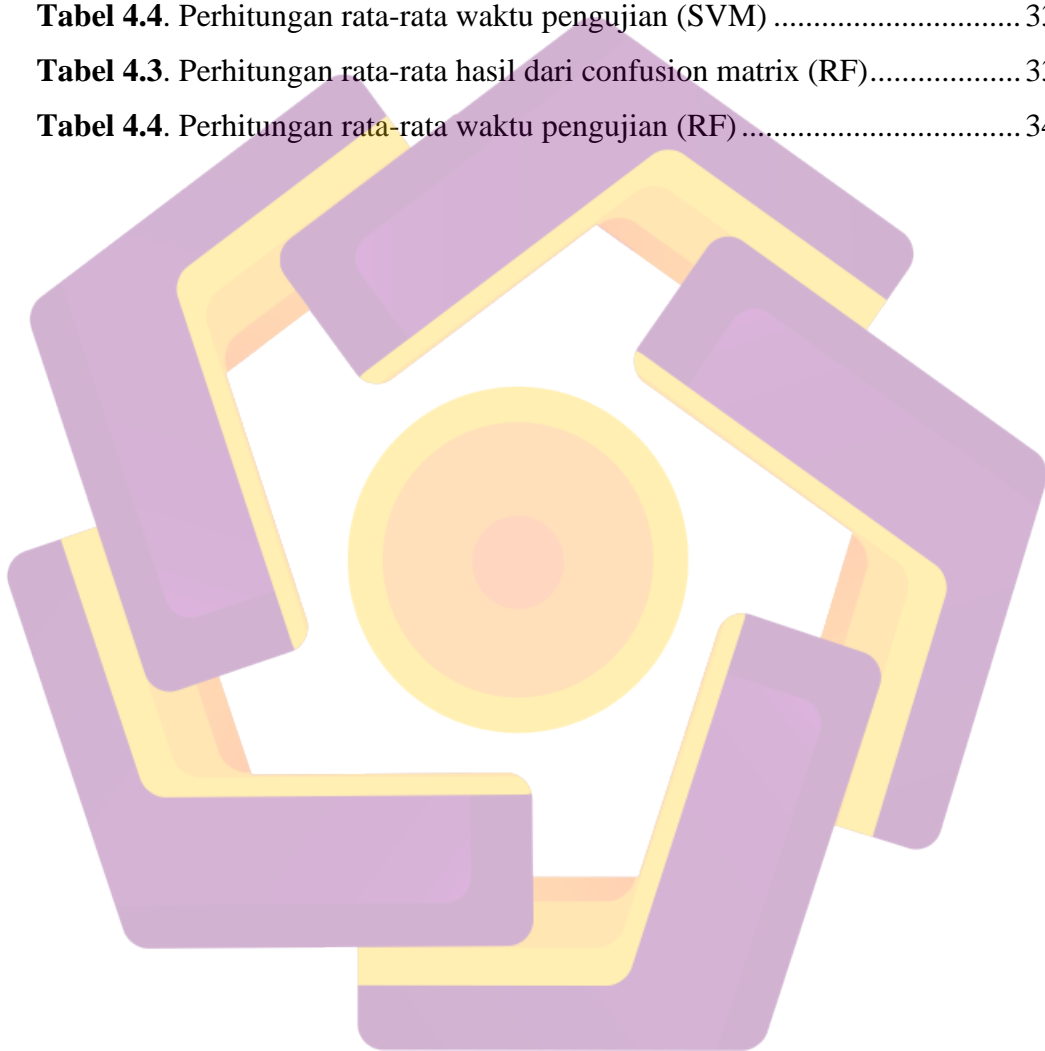
|   |             |
|---|-------------|
| <b>HALAMAN JUDUL .....</b>                      | <b>i</b>    |
| <b>HALAMAN PERSETUJUAN .....</b>                | <b>ii</b>   |
| <b>HALAMAN PENGESAHAN.....</b>                  | <b>iii</b>  |
| <b>HALAMAN PERNYATAAN KEASLIAN SKRIPSI.....</b> | <b>iv</b>   |
| <b>DAFTAR ISI.....</b>                          | <b>vi</b>   |
| <b>DAFTAR TABEL .....</b>                       | <b>viii</b> |
| <b>DAFTAR GAMBAR.....</b>                       | <b>ix</b>   |
| <b>DAFTAR LAMBANG DAN SINGKATAN .....</b>       | <b>x</b>    |
| <b>DAFTAR ISTILAH .....</b>                     | <b>xi</b>   |
| <b>INTISARI .....</b>                           | <b>xii</b>  |
| <b>ABSTRACT .....</b>                           | <b>xiii</b> |
| <b>BAB I PENDAHULUAN.....</b>                   | <b>1</b>    |
| 1.1 Latar Belakang.....                         | 1           |
| 1.2 Perumusan masalah .....                     | 3           |
| 1.3 Tujuan Penelitian.....                      | 3           |
| 1.4. Batasan Masalah.....                       | 3           |
| 1.5 Manfaat Penelitian.....                     | 3           |
| <b>BAB II TINJAUAN PUSTAKA.....</b>             | <b>5</b>    |
| 2.1 Literature Review .....                     | 5           |
| 2.2. Dasar Teori .....                          | 13          |
| 2.2.1. <i>Cyberbullying</i> .....               | 13          |
| 2.2.2. <i>Twitter</i> .....                     | 13          |
| 2.2.3. <i>Phyton</i> .....                      | 13          |
| 2.1.4. <i>Anaconda</i> .....                    | 13          |
| 2.1.5. <i>Tweepy</i> .....                      | 14          |
| 2.1.6. <i>NLTK</i> .....                        | 14          |
| 2.3. Sejarah Machine Learning.....              | 15          |



|                             |                                    |           |
|-----------------------------|------------------------------------|-----------|
| 2.3.1.                      | Support Vector Machine (SVM).....  | 15        |
| 2.3.2.                      | Random Forest (RF).....            | 17        |
| 2.3.2.                      | Evaluasi.....                      | 18        |
| <b>BAB III</b>              | <b>METODOLOGI PENELITIAN .....</b> | <b>20</b> |
| 3.1                         | Deskripsi Singkat Objek.....       | 20        |
| 3.2.                        | Alat dan Bahan Penelitian .....    | 20        |
| 3.3.                        | Langkah Penelitian .....           | 20        |
| 3.3.1.                      | Pengumpulan Data .....             | 22        |
| 3.3.2.                      | Input Data Hasil Crawling .....    | 22        |
| 3.3.3.                      | Pemrosesan Data .....              | 22        |
| 3.3.4.                      | Pembobotan Term Frekuensi .....    | 23        |
| 3.3.5.                      | Proses Klasifikasi.....            | 23        |
| 3.3.6.                      | Validasi .....                     | 25        |
| <b>BAB IV</b>               | <b>HASIL DAN PEMBAHASAN .....</b>  | <b>26</b> |
| 4.1.                        | Implementasi .....                 | 26        |
| 4.1.1.                      | <i>Mining Data</i> .....           | 26        |
| 4.1.2.                      | <i>Prosessing NLTK</i> .....       | 26        |
| 4.2.                        | Pengujian .....                    | 29        |
| 4.2.1.                      | Proses klasifikasi.....            | 29        |
| 4.3.                        | Hasil dan Analisis .....           | 32        |
| 4.3.1.                      | Perhitungan Hasil.....             | 32        |
| 4.3.2.                      | Analisa Hasil .....                | 34        |
| <b>BAB V</b>                | <b>KESIMPULAN DAN SARAN .....</b>  | <b>35</b> |
| 5.1.                        | Kesimpulan.....                    | 35        |
| 5.2.                        | Saran .....                        | 35        |
| <b>DAFTAR PUSTAKA .....</b> |                                    | <b>36</b> |

## DAFTAR TABEL

|  |    |
|--|----|
| <b>Tabel 2.1.</b> Matrix perbandingan penelitian .....                         | 6  |
| <b>Tabel 2.2.</b> Confusion Matrix .....                                       | 18 |
| <b>Tabel 4.3.</b> Perhitungan rata-rata hasil dari confusion matrix (SVM)..... | 32 |
| <b>Tabel 4.4.</b> Perhitungan rata-rata waktu pengujian (SVM) .....            | 33 |
| <b>Tabel 4.3.</b> Perhitungan rata-rata hasil dari confusion matrix (RF).....  | 33 |
| <b>Tabel 4.4.</b> Perhitungan rata-rata waktu pengujian (RF).....              | 34 |

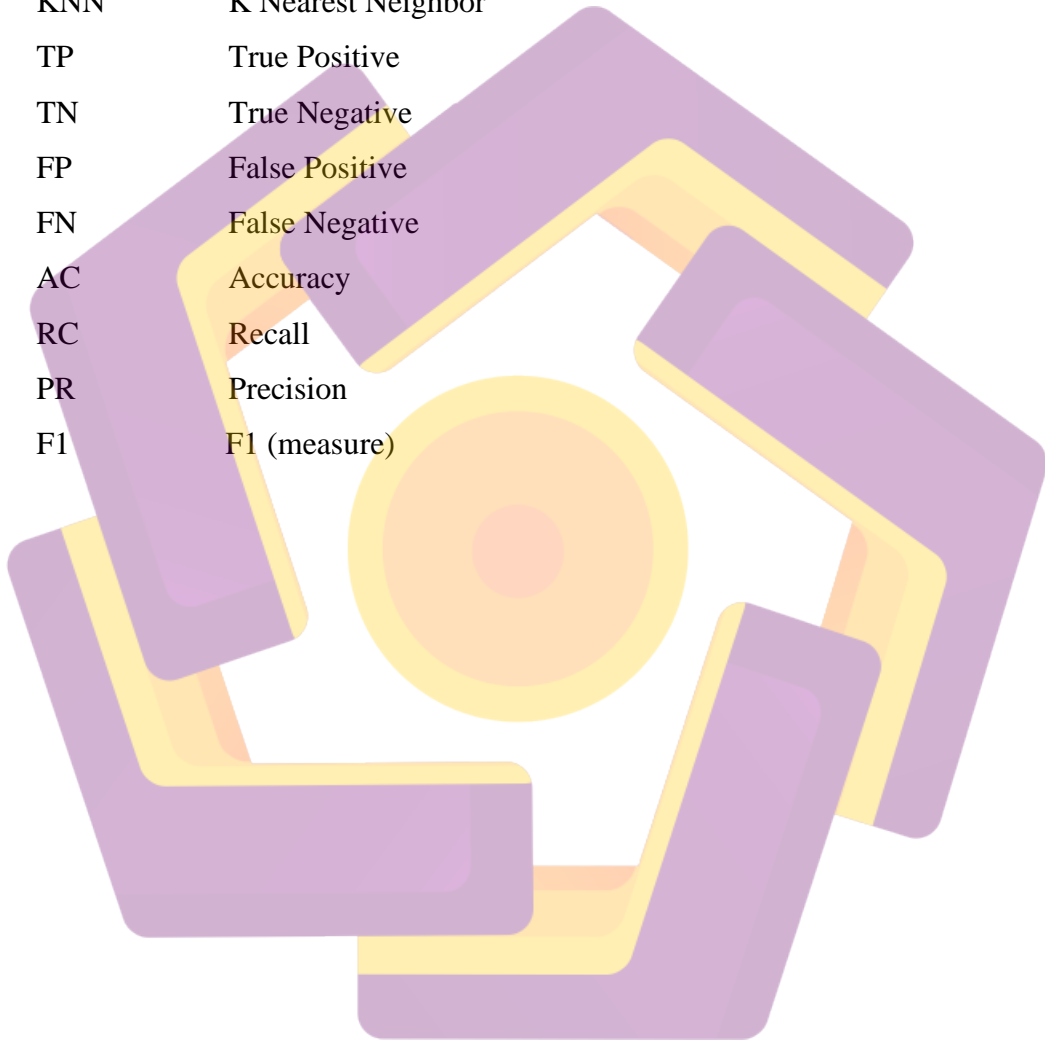


## DAFTAR GAMBAR

|   |    |
|---|----|
| Gambar 2.1. Source code Tweepy .....  | 14 |
| Gambar 2.2. Source code NLTK .....  | 14 |
| Gambar 2.3. Ilustrasi Support Vector Machine .....                                  | 16 |
| Gambar 2.4. Ilustrasi Random Forest (Struktur pohon pada metode Chart).....         | 17 |
| Gambar 3.1. Diagram Alur Penelitian .....   | 21 |
| Gambar 3.2. Klasifikasi SVM dan RF .....  | 24 |
| Gambar 4.1. Hasil dari crawling data .....  | 26 |
| Gambar 4.2. Proses <i>Load data crawling</i> .....                                  | 27 |
| Gambar 4.3. Hasil proses <i>case folding</i> .....                                  | 27 |
| Gambar 4.4. Hasil proses <i>tokenizing</i> .....                                    | 27 |
| Gambar 4.5. Hasil proses <i>stopword removal</i> .....                              | 28 |
| Gambar 4.6. Proses <i>Normalisasi</i> .....   | 28 |
| Gambar 4.7. Hasil setelah <i>normalisasi</i> dan <i>stemming</i> .....              | 28 |
| Gambar 4.8. Alur proses klasifikasi.....  | 29 |
| Gambar 4.9. Grafik hasil uji <i>accuracy, recall, precision, dan f1 (SVM)</i> ..... | 30 |
| Gambar 4.10. Grafik hasil waktu uji 5 kali SVM.....                                 | 30 |
| Gambar 4.11. Contoh klasifikasi bullying .....                                      | 31 |
| Gambar 4.12. Grafik hasil uji <i>accuracy, recall, precision, dan f1 (RF)</i> ..... | 31 |
| Gambar 4.13. Grafik hasil waktu uji 5 kali RF.....                                  | 31 |

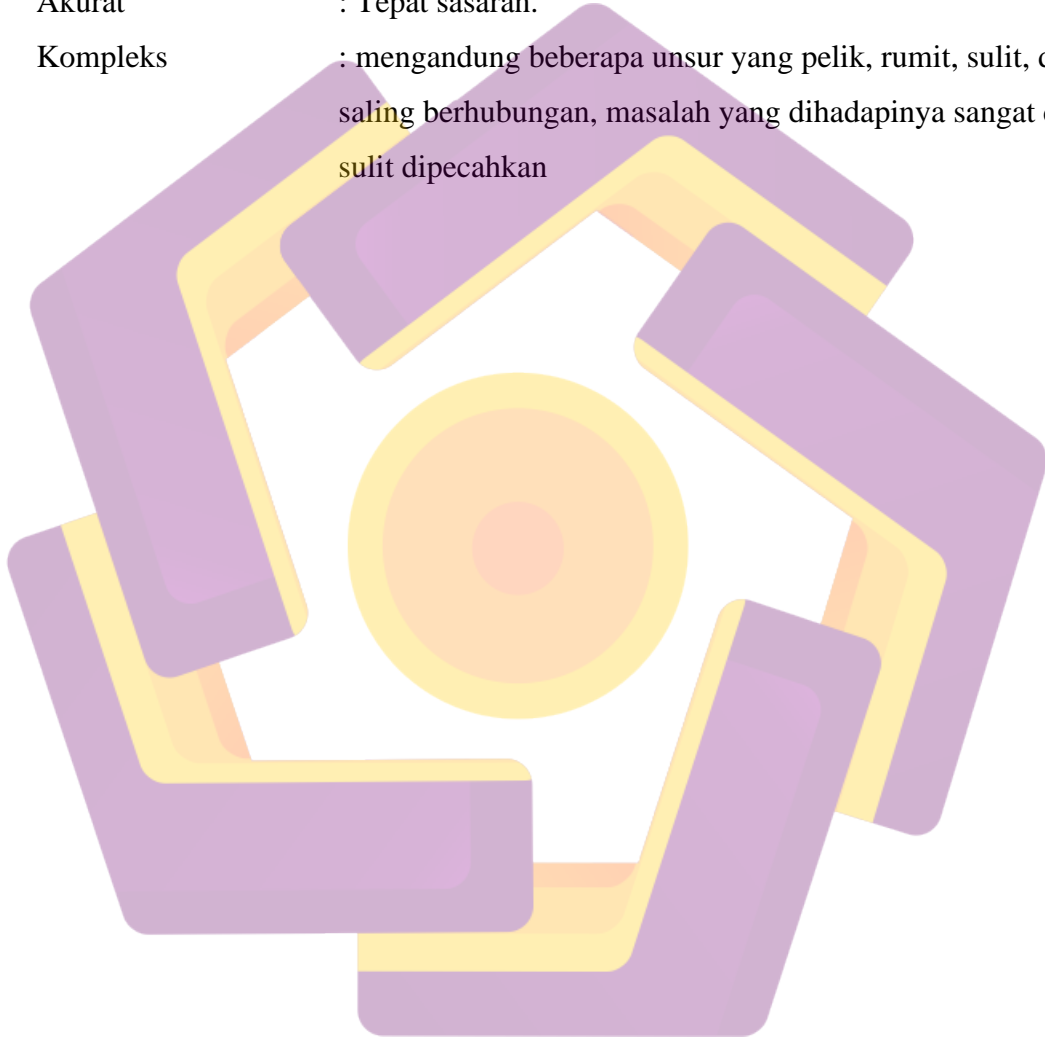
## DAFTAR LAMBANG DAN SINGKATAN

|     |                         |
|-----|-------------------------|
| RF  | Random Forest           |
| SVM | Support Vector Machines |
| KNN | K Nearest Neighbor      |
| TP  | True Positive           |
| TN  | True Negative           |
| FP  | False Positive          |
| FN  | False Negative          |
| AC  | Accuracy                |
| RC  | Recall                  |
| PR  | Precision               |
| F1  | F1 (measure)            |



## DAFTAR ISTILAH

|            |  |
|------------|--|
| Alay       | : Bahasa yang digunakan oleh sedikit orang yang hanya tahu pada kalangan tersebut.   |
| Signifikan | : Perbedaan yang sangat kecil ataupun sangat berbeda.  |
| Akurat     | : Tepat sasaran.   |
| Kompleks   | : mengandung beberapa unsur yang pelik, rumit, sulit, dan saling berhubungan, masalah yang dihadapinya sangat dan sulit dipecahkan |



## INTISARI

Dalam perkembangan zaman yang begitu pesat tentu akan berdampak pula dalam perkembangan teknologi internet dengan data di Indonesia pada akhir 2019 berjumlah kurang lebih 196 juta jiwa terhubung ke internet dan sosial media, sehingga memungkinkan terjadi kejahatan didalamnya seperti Cyberbullying dengan data yang tertera 51,2% dari jumlah penduduk yang menggunakan internet pernah terkena dampak bullying mengingat dalam suatu negara presentase ini cukup besar, dan dengan ini berkenaan pula sedang tren dalam penggunaan (machine learning) beberapa waktu terakhir.

Terdapat dua jenis metode pembelajaran machine learning yang dapat digunakan analisis sentiment yaitu supervised learning dan unsupervised learning, dalam penelitian akan dilakukan perbandingan dari dua metode yaitu SVM dan RF dalam pengujian analisis sentimen tokoh masyarakat Presiden Jokowi, dengan jumlah data tweet sebanyak 10.000 data yang di crawling dengan menggunakan library tweepy kemudian di simpan dengan ekstensi (.csv).

Dalam pengujian model dari dua metode dengan lima kali data pengujian didapati rata-rata keakurasian RF lebih unggul dari SVM dengan 90% tingkat keakurasian dan waktu proses hanya 16.44 detik dengan SVM hanya 77% keakurasian serta waktu proses 17.90 detik. Dengan demikian kesimpulan dari penelitian metode RF yang dinilai lebih akurat ketimbang SVM.

**Kata kunci:** *Random forest, Support Vector Machine, Python, Cyberbullying.*

## ABSTRACT

*In the rapid development of the era, it will certainly have an impact on the development of internet technology with data in Indonesia at the end of 2019 totaling approximately 196 million people connected to the internet and social media, making it possible for crimes to occur in them such as Cyberbullying with 51.2% of the data listed. The number of people who use the internet has been affected by bullying, considering that in a country this percentage is quite large, and with this regard, there is also a trend in use (machine learning) in recent times.*

*There are two types of machine learning learning methods that can be used for sentiment analysis, namely supervised learning and unsupervised learning, in this study a comparison will be made of the two methods, namely SVM and RF in testing the sentiment analysis of President Jokowi's public figures, with a total of 10,000 tweet data crawled. by using the tweepy library then saved with the extension (.csv).*

*In testing the model from two methods with five times the test data, it was found that the average accuracy of RF is superior to SVM with 90% accuracy rate and processing time is only 16.44 seconds with SVM only 77% accuracy and processing time is 17.90 seconds. Thus, the conclusion from the research of the RF method is that it is considered more accurate than SVM.*

**Keyword: Random forest, Support Vector Machine, Python, Cyberbullying**

