

**IMPLEMENTASI NAÏVE BAYES CLASSIFIER UNTUK  
PENDETEKSI CYBERBULLYING PADA  
KOMENTAR DI INSTAGRAM**

**SKRIPSI**



disusun oleh

**Abinda Dwi Nur Ahmad**

**15.11.9185**

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

**IMPLEMENTASI NAÏVE BAYES CLASSIFIER UNTUK  
PENDETEKSI CYBERBULLYING PADA  
KOMENTAR DI INSTAGRAM**

**SKRIPSI**

untuk memenuhi sebagian persyaratan  
mencapai gelar Sarjana  
pada Program Studi Informatika



disusun oleh

**Abinda Dwi Nur Ahmad**

**15.11.9185**

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

**PERSETUJUAN**

**SKRIPSI**

**IMPLEMENTASI NAÏVE BAYES CLASSIFIER UNTUK  
PENDETEKSI CYBERBULLYING PADA  
KOMENTAR DI INSTAGRAM**

yang dipersiapkan dan disusun oleh

**Abinda Dwi Nur Ahmad**

**15.11.9185**

telah disetujui oleh Dosen Pembimbing Skripsi  
pada tanggal 28 November 2019

**Dosen Pembimbing,**



**Sri Ngudi Wahyuni, S.T.,M.Kom.**

**NIK. 190302060**

**PENGESAHAN**  
**SKRIPSI**  
**IMPLEMENTASI NAÏVE BAYES CLASSIFIER UNTUK**  
**PENDETEKSI CYBERBULLYING PADA**  
**KOMENTAR DI INSTAGRAM**

yang dipersiapkan dan disusun oleh  
**Abinda Dwi Nur Ahmad**  
**15.11.9185**  
telah dipertahankan di depan Dewan Penguji  
pada tanggal 18 November 2019

**Susunan Dewan Penguji**

**Nama Penguji**

**Erni Seniwati, M.Cs**  
**NIK. 190302231**

**Wiwi Widayani, M.Kom**  
**NIK. 190302272**

**Sumarni Adi, S.Kom, M.Cs**  
**NIK. 190302256**


**Tanda Tangan**



Skripsi ini telah diterima sebagai salah satu persyaratan  
untuk memperoleh gelar Sarjana Komputer  
Tanggal 30 November 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 tanggungjawab saya pribadi.

Yogyakarta, 7 Desember 2019

METERAI  
TEMPEL

5C4D6AHF155816281

6000  
ENAM RIBURUPIAH

Abinda Dwi Nur Ahmad

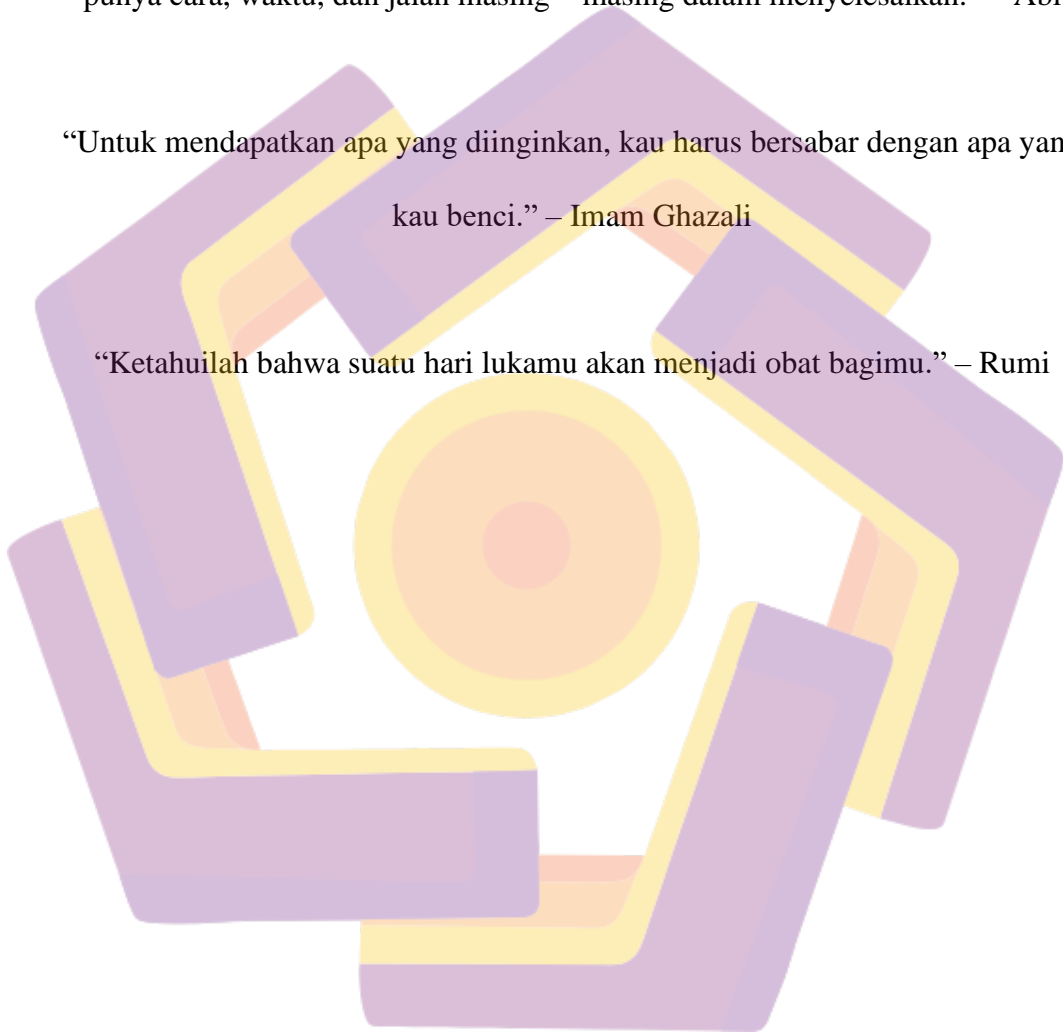
NIM. 15.11.9185

## MOTTO

“Skripsi adalah proses melawan diri sendiri. Keterlambatan dalam proses bukanlah sebuah kehinaan. Ketepatan dalam proses tidak selalu prestasi. Semua punya cara, waktu, dan jalan masing – masing dalam menyelesaikan.” – Abi

“Untuk mendapatkan apa yang diinginkan, kau harus bersabar dengan apa yang kau benci.” – Imam Ghazali

“Ketahuilah bahwa suatu hari lukamu akan menjadi obat bagimu.” – Rumi



## PERSEMBAHAN

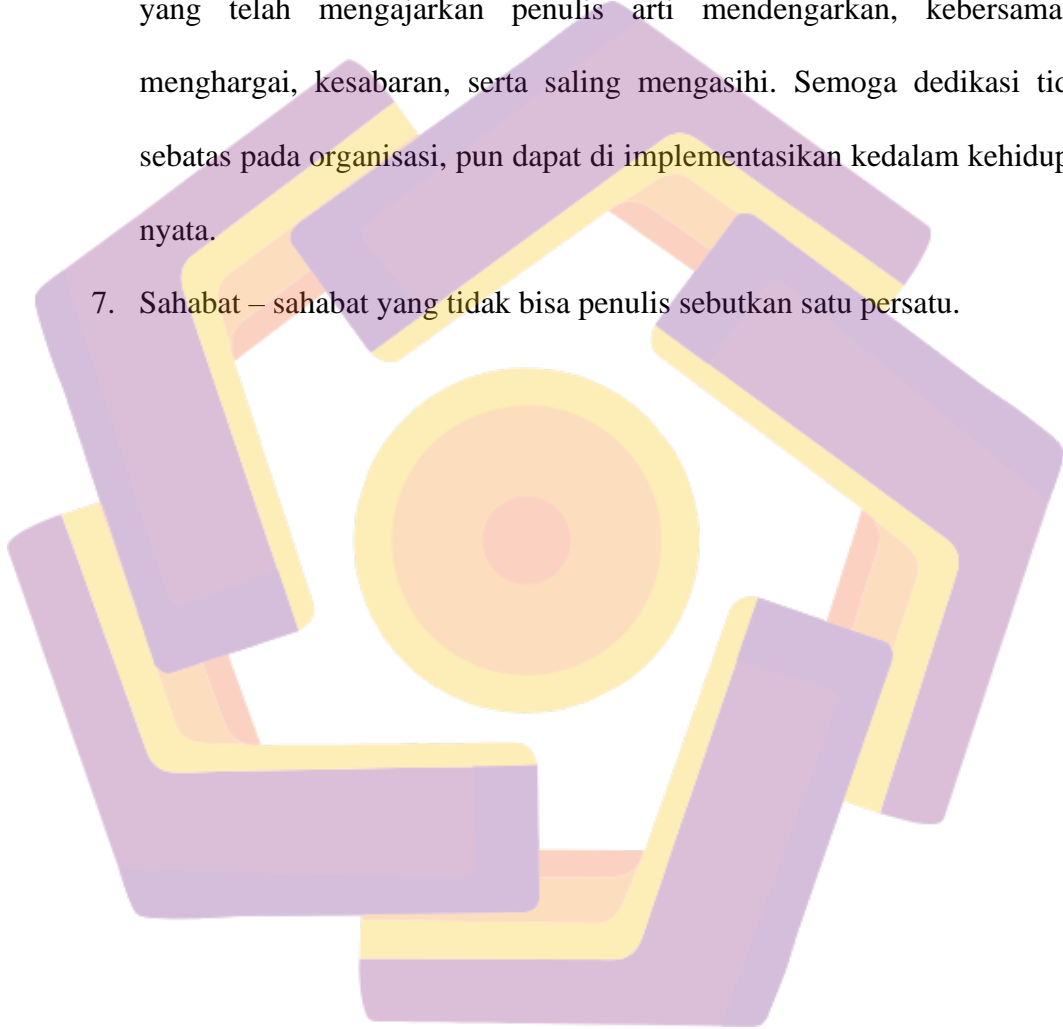
Segala puji bagi Allah ﷻ yang telah memberikan rahmat, hidayah serta karunia kepada seluruh makhluk-Nya tanpa terkecuali. Sholawat serta salam senantiasa kita panjatkan kepada junjungan nabi agung kita Nabi Muhammad ﷺ yang kita nantikan syafaatnya di Yaumul Qiyamah.

Dengan memanjatkan puji syukur atas kehadiran Allah ﷻ serta nikmat yang telah diberikan –Nya dan dukungan maupun doa dari orang tercinta, penulis dapat menyelesaikan laporan skripsi yang berjudul **“Implementasi Naïve Bayes Classifier Untuk Pendeteksi Cyberbullying Pada Komentar di Instagram”**.

Tidak lupa penulis mempersembahkan karya tulis ini kepada:

1. Kedua orang tua yang penulis sayangi, Ibu Dwi Hardiningsih dan Bapak Erdy Sumanto serta adik – kakak tersayang. Tanpa doa dan dukungannya, penulis tidak bisa mencapai sejauh ini. Do'a anakmu untuk beliau serta keluarga, semoga Allah memberikan umur yang berkah, kesehatan serta selalu memuliakan dimanapun berada. Aamiin.
2. Kekasih, Dian Indriastuti sebagai partner yang selalu memberikan dukungan dan pembelajaran disaat sedang berjuang bersama – sama.
3. Ibu Sri Ngudi Wahyuni, S.T.,M.Kom. selaku dosen pembimbing yang telah memberikan motivasi dan banyak pelajaran serta membimbing dengan penuh kesabaran, sehingga penulis dapat merampungkan skripsi ini.
4. Teman – teman satu kontrakan “Kondeman” yang menjadi keluarga kedua di perantuan yang selalu berbagi cerita dan dukungan selama kurang lebih 3 tahun bersama.

5. Teman – teman Informatika 10 angkatan 2015 yang telah berbagi kisah selama kurang lebih 3.5 tahun. Semoga kelak persahabatan kita di kelas sampai menua.
6. Teman organisasi IEEE ASB, AEC periode 2016/2017, BEM 2017/2018 yang telah mengajarkan penulis arti mendengarkan, kebersamaan, menghargai, kesabaran, serta saling mengasihi. Semoga dedikasi tidak sebatas pada organisasi, pun dapat di implementasikan kedalam kehidupan nyata.
7. Sahabat – sahabat yang tidak bisa penulis sebutkan satu persatu.





## KATA PENGANTAR

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

*Assalamu'alaikum Warahmatullahi Wabarakaatuh*

Alhamdulillah rabbil 'alamin segala puji bagi Allah ﷻ yang telah memberikan rahmat serta karunia kepada seluruh makhluk di alam semesta. Shalawat serta salam senantiasa kita lantukan kepada junjungan kita nabi akhir zaman Nabi Muhammad ﷺ yang telah membimbing dari zaman kegelapan menuju zaman yang lebih terang serta yang kita nantikan syafa'atnya di hari kiamat. Penulis ucapkan syukur kehadiran Allah ﷻ karena atas izin-Nya laporan ini dapat diselesaikan.

Laporan ini disusun sebagai salah satu syarat kelulusan program studi S1 – Informatika Universitas AMIKOM Yogyakarta. Pada penyusunan laporan ini, penulis mendapat bantuan dari berbagai pihak. Oleh karena itu, Penulis ingin mengucapkan terimakasih kepada beberapa pihak yang telah membantu dalam penulisan laporan skripsi ini. Penulis mengucapkan terima kasih kepada:

1. Kedua orang tua tercinta yang telah melahirkan penulis ke dunia, Ibu Dwi Hardiningsih dan Bapak Erdy Sumanto.
2. Bapak Prof. Dr. M. Suyanto, M.M selaku Rektor Universitas AMIKOM Yogyakarta.
3. Ibu Krisnawati, S.Si., M.T. selaku dekan Fakultas Ilmu Komputer Universitas AMIKOM Yogyakarta.

4. Bapak Sudarmawan, M.T selaku ketua jurusan S1-Informatika Universitas AMIKOM Yogyakarta.
5. Ibu Sri Ngudi Wahyuni, S.T.,M.Kom. selaku dosen pembimbing yang telah memberikan arahan, bimbingan serta motivasi sehingga skripsi ini dapat di selesai dengan baik.
6. Bapak Ibu Dosen penguji yang telah memberikan kritik dan saran pada laporan skripsi ini sehingga penelitian kedepan dapat lebih baik.
7. Kekasihku Dian Indriastuti yang selalu memberikan dukungan, motivasi serta semangat tak henti – henti dalam proses mengerjakan skripsi ini hingga selesai.
8. Beta Priyoko yang telah meluangkan waktu berdiskusi, serta membantu penulis selama proses pengerjaan skripsi.
9. Teman – teman Informatika 10 angkatan 2015, teman – teman organisasi, serta sahabat yang tidak bisa disebutkan satu persatu.
10. Para penulis sumber bacaan, jurnal, dan semacamnya yang penulis jadikan referensi serta pembuat *tools* yang mempermudah penulis dalam menyelesaikan laporan ini.

Penulis menyadari dalam penulisan laporan ini tentu masih ada kekurangan. Oleh karena itu, penulis juga menerima saran atau masukan sebagai bahan penelitian kedepan.

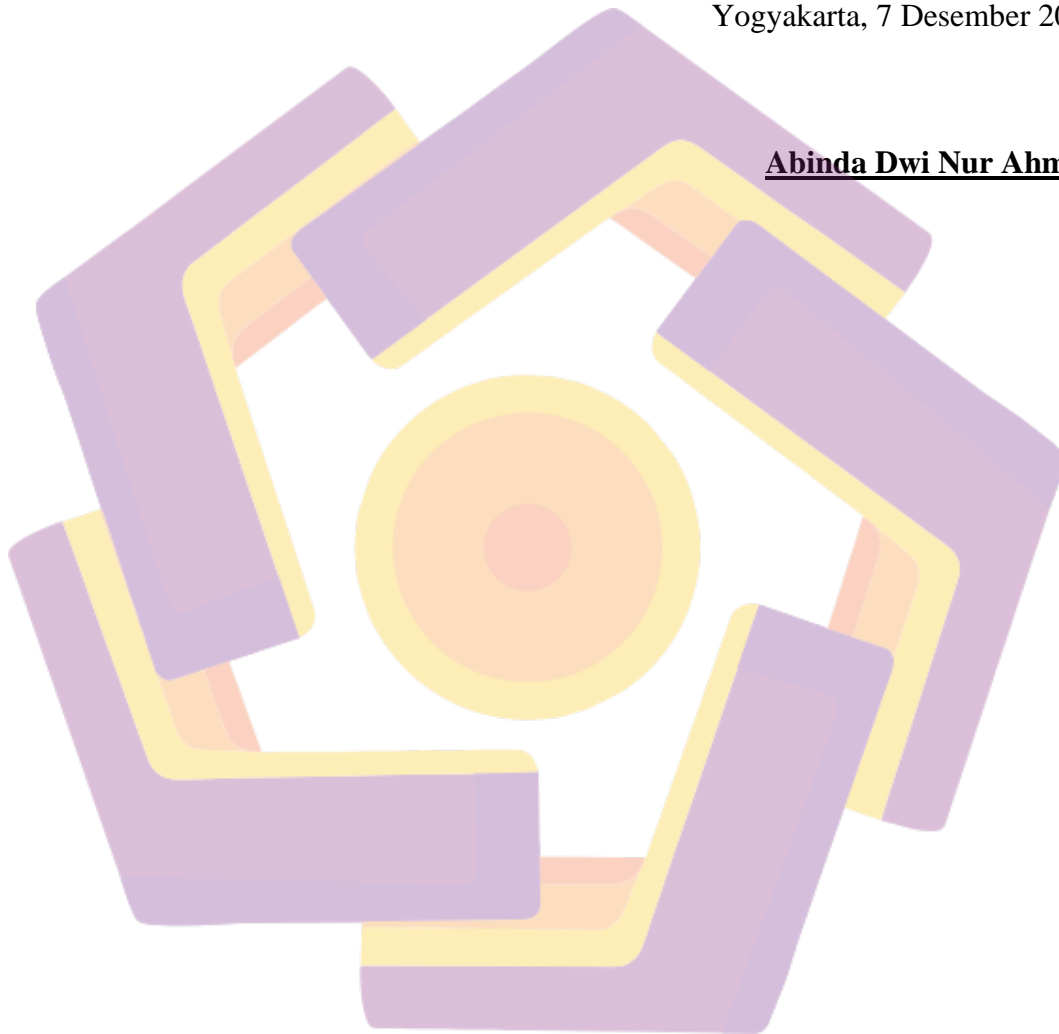
Akhir kata, penulis berharap laporan skripsi ini dapat memberikan kontribusi kepada para pembaca sebagai bahan kajian untuk mahasiswa dalam melakukan

penelitian selanjutnya baik di Universitas AMIKOM Yogyakarta maupun secara luas.

*Wassalamu'alaikum Warohmatullahi Wabarokatuh*

Yogyakarta, 7 Desember 2019

**Abinda Dwi Nur Ahmad**

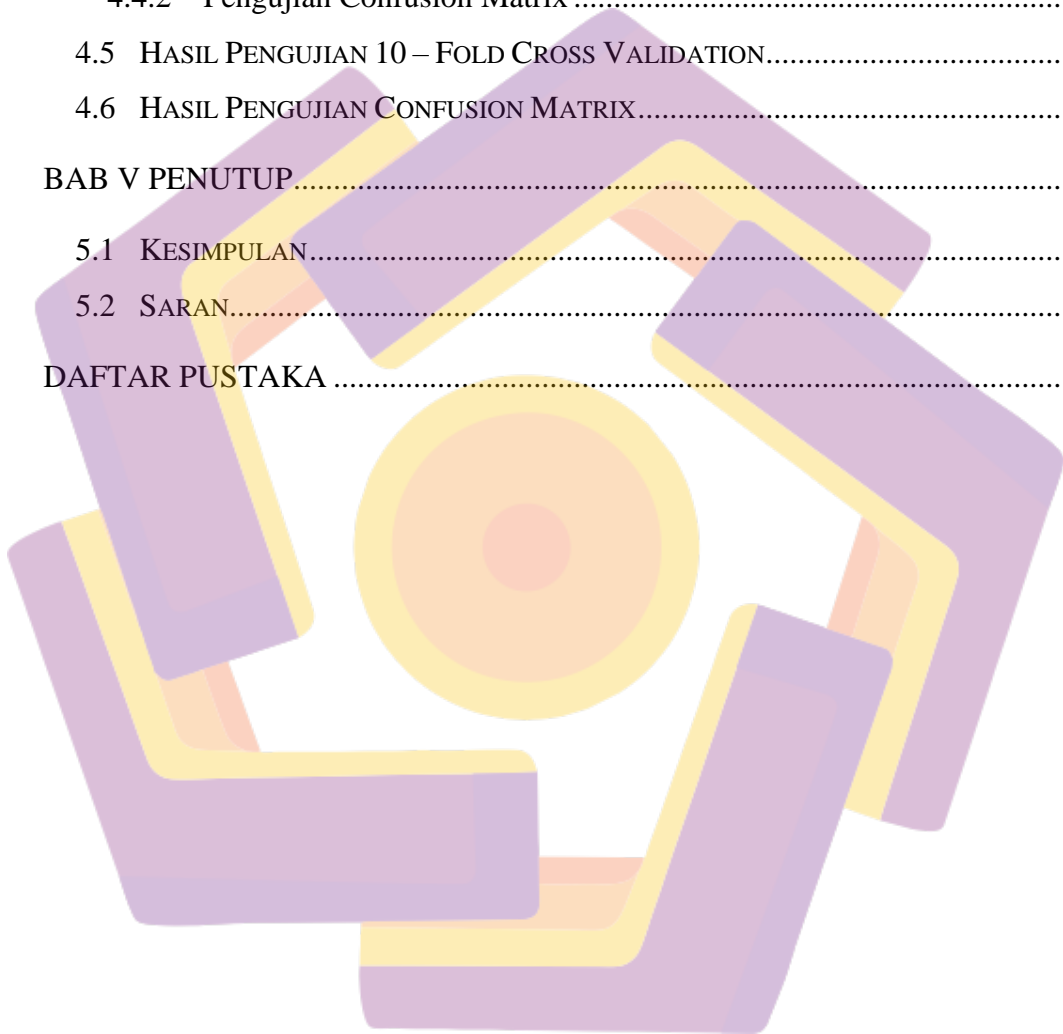


## DAFTAR ISI

JUDUL .....	I
PERSETUJUAN .....	II
PENGESAHAN .....	III
PERNYATAAN.....	IV
MOTTO .....	V
PERSEMBAHAN.....	VI
KATA PENGANTAR .....	VIII
DAFTAR ISI.....	X
DAFTAR TABEL.....	XIII
DAFTAR GAMBAR .....	XIV
INTISARI.....	XVI
ABSTRACT.....	XVII
BAB I PENDAHULUAN.....	1
1.1 LATAR BELAKANG.....	1
1.2 RUMUSAN MASALAH .....	3
1.3 BATASAN MASALAH .....	4
1.4 MAKSUD DAN TUJUAN PENELITIAN .....	5
1.5 METODE PENELITIAN .....	5
1.6 SISTEMATIKA PENULISAN .....	7
BAB II LANDASAN TEORI.....	9
2.1 KAJIAN PUSTAKA.....	9
2.2 DASAR TEORI.....	13
2.2.1 Analisis Kebutuhan Sistem .....	13
2.2.2 Text Mining.....	14

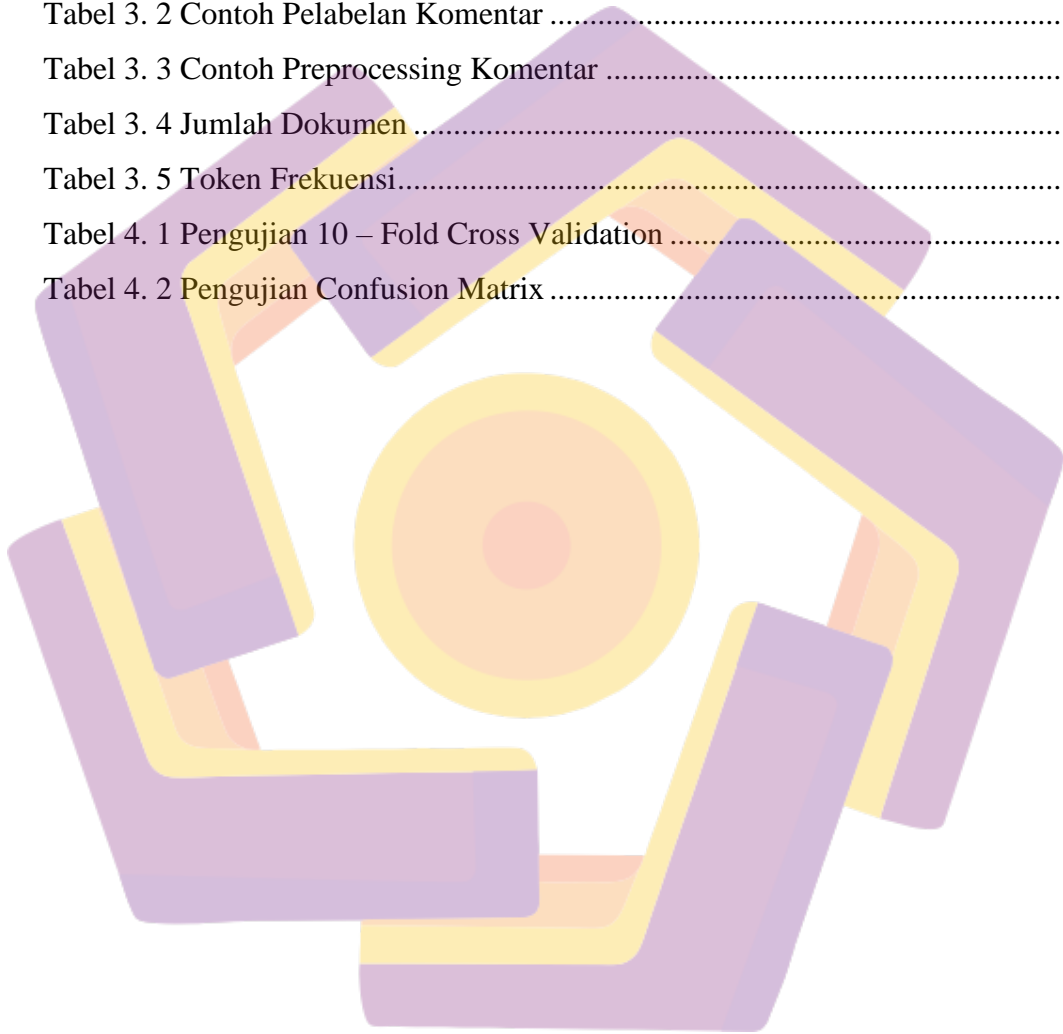
2.2.3	Cyberbullying.....	16
2.2.4	Cyberbullying di Instagram.....	18
2.2.5	Klasifikasi .....	20
2.2.6	Naïve Bayes Classifier .....	21
2.2.7	K - Fold Cross Validation .....	24
2.2.8	Confusion Matrix .....	25
2.2.9	Flowchart Diagram.....	27
<b>BAB III ANALISIS DAN PERANCANGAN .....</b>		<b>29</b>
3.1	<b>ANALISIS MASALAH .....</b>	<b>29</b>
3.1.1	Identifikasi Masalah.....	29
3.1.2	Analisis Diagram Fishbone.....	30
3.1.3	Hasil Analisis .....	32
3.2	<b>ANALISIS KEBUTUHAN PENELITIAN.....</b>	<b>35</b>
3.2.1	Analisis Kebutuhan Fungsional .....	35
3.2.2	Analisis Kebutuhan Non Fungsional .....	35
3.2.3	Analisis Kebutuhan Data.....	36
3.2.4	Pelabelan Data.....	37
3.3	<b>ANALISIS PERANCANGAN SISTEM.....</b>	<b>38</b>
3.3.1	Preprocessing .....	38
3.3.2	Klasifikasi .....	44
3.3.3	Perancangan Antarmuka (Mockup).....	49
<b>BAB IV IMPLEMENTASI DAN PEMBAHASAN .....</b>		<b>53</b>
4.1	<b>PENGUMPULAN DATASET TRAINING.....</b>	<b>53</b>
4.2	<b>IMPLEMENTASI ANTAR MUKA .....</b>	<b>54</b>
4.2.1	Implementasi Halaman Utama.....	54
4.2.2	Implementasi Halaman Training.....	55
4.2.3	Implementasi Halaman Klasifikasi .....	56
4.2.4	Implementasi Halaman Evaluasi.....	58
4.3	<b>PEMBAHASAN SOURCE CODE.....</b>	<b>58</b>
4.3.1	Source Code Scrapping Data .....	58

4.3.2	Source Code Training .....	60
4.3.3	Source Code Preprocessing.....	60
4.3.4	Source Code Klasifikasi.....	61
4.4	PENGUJIAN MODEL .....	62
4.4.1	Pengujian 10 – Fold Cross Validation .....	63
4.4.2	Pengujian Confusion Matrix .....	63
4.5	HASIL PENGUJIAN 10 – FOLD CROSS VALIDATION.....	64
4.6	HASIL PENGUJIAN CONFUSION MATRIX.....	70
BAB V PENUTUP.....		72
5.1	KESIMPULAN.....	72
5.2	SARAN.....	72
DAFTAR PUSTAKA .....		74



## DAFTAR TABEL

Tabel 2. 1 Ringkasan Penelitian.....	12
Tabel 2. 2 Confusion Matrix .....	25
Tabel 3. 1 Contoh Komentar.....	36
Tabel 3. 2 Contoh Pelabelan Komentar .....	38
Tabel 3. 3 Contoh Preprocessing Komentar .....	41
Tabel 3. 4 Jumlah Dokumen .....	46
Tabel 3. 5 Token Frekuensi.....	46
Tabel 4. 1 Pengujian 10 – Fold Cross Validation .....	70
Tabel 4. 2 Pengujian Confusion Matrix .....	71

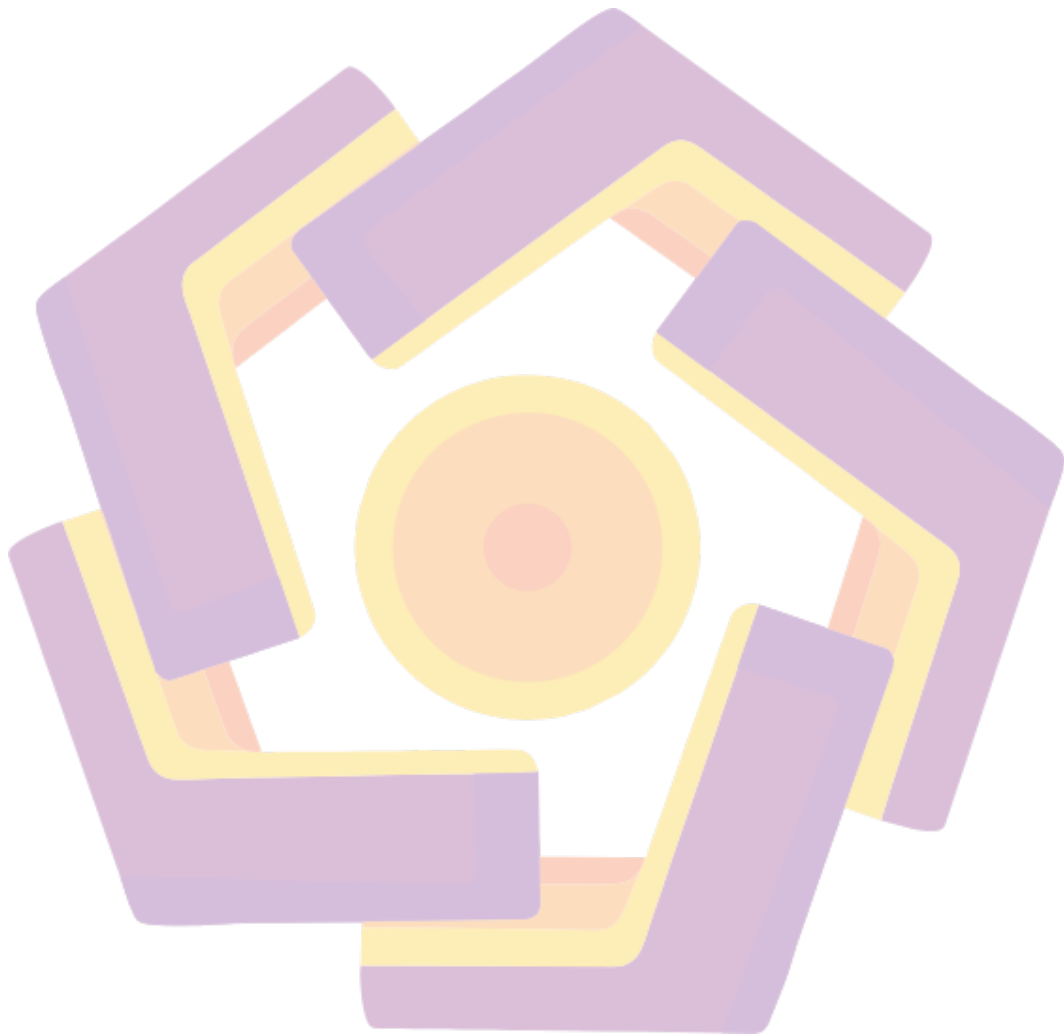


## DAFTAR GAMBAR

Gambar 2. 1 Menunjukkan arsitektur text mining .....	14
Gambar 2. 1 Ilustrasi 10 - fold cross validation .....	25
Gambar 3. 1 Diagram Fishbone .....	31
Gambar 3. 2 Flowchart Training.....	33
Gambar 3. 3 Flowchart Testing.....	34
Gambar 3. 4 Flowchart Preprocessing .....	39
Gambar 3. 5 Flowchart Klasifikasi Naïve Bayes Classifier.....	44
Gambar 3. 6 Halaman Utama.....	50
Gambar 3. 7 Halaman Training.....	50
Gambar 3. 8 Halaman Klasifikasi .....	51
Gambar 3. 9 Halaman Evaluasi Model .....	52
Gambar 4. 1 Komentar Cyberbullying.....	54
Gambar 4. 2 Komentar Not Cyberbullying.....	54
Gambar 4. 3 Halaman Utama.....	55
Gambar 4. 4 Halaman Training.....	56
Gambar 4. 5 Training Result.....	56
Gambar 4. 6 Halaman Klasifikasi .....	57
Gambar 4. 7 Hasil Klasifikasi .....	57
Gambar 4. 8 Halaman Evaluasi.....	58
Gambar 4. 9 Source Code Scraping .....	59
Gambar 4. 10 Source Code Training .....	60
Gambar 4. 11 Source Code Preprocessing .....	61
Gambar 4. 12 Source Code Klasifikasi .....	62
Gambar 4. 13 Hasil Pengujian Fold 1 .....	64
Gambar 4. 14 Hasil Pengujian Fold 2 .....	65
Gambar 4. 15 Hasil Pengujian Fold 3 .....	65
Gambar 4. 16 Hasil Pengujian Fold 4 .....	66
Gambar 4. 17 Hasil Pengujian Fold 5 .....	66
Gambar 4. 18 Hasil Pengujian Fold 6 .....	67



Gambar 4. 19 Hasil Pengujian Fold 7 ..... 68  
Gambar 4. 20 Hasil Pengujian Fold 8 ..... 68  
Gambar 4. 21 Hasil Pengujian Fold 9 ..... 69  
Gambar 4. 22 Hasil Pengujian Fold 10 ..... 69



## INTISARI

Dewasa ini setiap orang di seluruh dunia begitu lekat dengan media sosial. Instagram merupakan *platform* jejaring sosial populer yang hampir semua kalangan menggunakan pada zaman sekarang. Media sosial Instagram digunakan mulai dari anak – anak, remaja hingga orang dewasa untuk menunjukkan eksistensinya. Pada umumnya pengguna akan mengunggah foto, video, dan caption sebagai keterangan. Terlepas dari hal tersebut, terdapat permasalahan yang marak terjadi yakni *cyberbullying* dan sering dilakukan pengguna khususnya pada kolom komentar yang dapat dibaca oleh pengguna lain. Setidaknya menurut survei yang dilakukan Ditch the Label pada tahun 2017 sebanyak 42% korban *cyberbullying* paling banyak terjadi pada Instagram. Bahaya *cyberbullying* tentu meresahkan banyak pihak dikarenakan dampak yang ditimbulkan. Meskipun tidak terlihat secara langsung, namun efek dari *cyberbullying* sangat berpengaruh pada psikis korban.

Algoritma Naïve Bayes Classifier merupakan salah satu teknik yang dapat digunakan untuk membangun model klasifikasi dengan probabilitas dan statistik. Terdapat 2 tahapan pada proses klasifikasi yakni *training* dan *testing*. Dari permasalahan yang terjadi, peneliti memberikan solusi dengan membuat pendeteksi *cyberbullying* pada komentar di Instagram. Data yang diambil menggunakan metode *scrapping* pada komentar Instagram dari tokoh publik ataupun pengguna yang memiliki banyak *follower*. Untuk membersihkan data komentar dilakukan *preprocessing* seperti *casefolding*, *tokenizing*, *stopword removal*, dan *stemming*. Melalui penerapan algoritma Naïve Bayes Classifier diharapkan dapat diambil tindakan yang bijak terhadap komentar yang mengandung *cyberbullying* di Instagram.

Berdasarkan hasil eksperimen klasifikasi teks menggunakan Naïve Bayes Classifier dengan skenario 10 – *Fold Cross Validation* menghasilkan rata – rata akurasi sebesar 72%. Tingkat akurasi tertinggi dihasilkan pada *fold* 3 dengan akurasi sebesar 93%, akurasi terendah pada *fold* 6 yaitu 61%. Sedangkan pengujian menggunakan *Confusion Matrix* didapatkan akurasi sebesar 78%, presisi 88%, recall 74%, *error rate* 22%.

**Kata Kunci:** Instagram, *Cyberbullying*, *Naïve Bayes Classifier*, 10 – *fold cross validation*, *Confusion Matrix*

## ABSTRACT

Nowadays everyone in the whole world is so attached to social media. Instagram is a popular social networking platform that almost all people use today. Instagram social media is used starting from children, teenagers to adults to show their existence. In general, users will upload photos, videos, and captions as information. Apart from this, there is a widespread problem that is cyberbullying and is often done by users, especially in the comments column that can be read by other users. At least according to a survey conducted by Ditch the Label in 2017, as many as 42% of cyberbullying victims occurred most on Instagram. The danger of cyberbullying is certainly troubling for many parties because of the impact it causes. Although not seen directly, but the effect of cyberbullying is very influential on the victim's psychic.

Naïve Bayes Classifier algorithm is one technique that can be used to build classification models with probabilities and statistics. There are 2 stages in the classification process namely training and testing. Of the problems that occur, researchers provide a solution by making cyberbullying detectors in comments on Instagram. Data retrieved using the scrapping method on Instagram comments from public figures or users who have many followers. To clear the comment data, preprocessing is done such as case folding, tokenizing, stopword removal, and stemming. Through the application of the Naïve Bayes Classifier algorithm it is hoped that wise action can be taken against comments that contain cyberbullying on Instagram.

Based on the results of text classification experiments using Naïve Bayes Classifier with a 10-Fold Cross Validation scenario produces an average accuracy of 72%. The highest accuracy level is produced on fold 3 with an accuracy of 93%, the lowest accuracy on fold 6 is 61%. While testing using the Confusion Matrix obtained 78% accuracy, 88% precision, 74% recall, 22% error rate.

**Keyword:** Instagram, Cyberbullying, Naïve Bayes Classifier, 10 – fold cross validation, Confusion Matrix