

**PENERAPAN METODE NAÏVE BAYES DAN TF-IDF UNTUK
IDENTIFIKASI CYBERBULLYING PADA KOMENTAR INSTAGRAM**

SKRIPSI



MUHAMMAD FARKHAN IMFROZIN FEBRIAN PUTRA

NIM. 1955202002

PROGRAM STUDI TEKNIK INFORMATIKA

FAKULTAS SAINS DAN TEKNOLOGI

UNIVERSITAS ISLAM RADEN RAHMAT MALANG

2024

LEMBAR PERSETUJUAN SKRIPSI

Judul : Penerapan Metode *Naïve Bayes* dan TF-IDF untuk
Identifikasi *Cyberbullying* pada Komentar Instagram

Penyusun : Muhammad Farkhan Imfrozin Febrian Putra

NIM 1955202002

Telah diperiksa dan disetujui untuk diuji pada tanggal 14 juni 2024

Disetujui oleh :

Pembimbing I,



Priska Choirina, M.Tr.T
NIDN. 0729119301

Pembimbing II,



Pangestuti Prima Darajat, S.Si., M.Si
NIDN. 0710089201

UNIVERSITAS ISLAM
RADEN RAHMAT

LEMBAR PENGESAHAN SKRIPSI

Judul : Penerapan Metode *Naïve Bayes* dan TF-IDF untuk
Identifikasi *Cyberbullying* pada Komentar Instagram

Penyusun : Muhammad Farkhan Imfrozin Febrian Putra

NIM 1955202002

Skrripsi oleh Muhammad Farkhan Imfrozin Febrian Putra ini telah dipertahankan
di depan dewan penguji pada Tanggal 14 Juni 2024.

Disetujui oleh :

Pembimbing I,



Priska Choirina, M.Tr.T
NIDN. 0729119301

Pembimbing II,



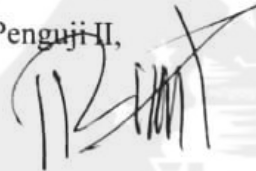
Pangestuti Prima Darajat, M.Si
NIDN. 0710089201

Penguji I,



Urnika M. Jannah, S.Kom., M.Pd
NIDN. 0722078905

Penguji II,



Rachmad Imam Tarecha, M.Kom
NIDN. 2111510051

Mengesahkan,
Dekan Fakultas SAINTEK



Dr. Zinnal Abidin, M.Si
NIDN. 0704018804

Mengetahui,
Ketua Program Studi



Urnika M. Jannah, S.Kom., M.Pd
NIDN. 0722078905

KATA PENGANTAR

Puji syukur kepada Allah SWT berkat Rahmat, Hidayah, dan Karunia-Nya kepada kita semua sehingga penulis dapat menyelesaikan skripsi dengan judul *"PENERAPAN METODE NAIVE BAYES DAN TF-IDF UNTUK IDENTIFIKASI CYBERBULLYING PADA KOMENTAR INSTAGRAM"*. Shalawat serta salam senantiasa tercurahkan kepada junjungan kita Nabi besar Muhammad SAW yang mengantarkan kita semua dari zaman kegelapan menuju zaman yang terang benderang ini. Penyusunan skripsi ini disusun sebagai salah satu syarat untuk mengerjakan skripsi pada program Strata-1 di Jurusan Teknik Informatika, Fakultas Sains dan Teknologi, Universitas Islam Raden Rahmat Malang.

Penulis menyadari dalam penyusunan proposal skripsi ini tidak akan selesai tanpa bantuan dari berbagai pihak baik secara moril maupun materil. Oleh karena itu pada kesempatan ini penulis ingin mengucapkan terima kasih kepada :

1. Allah SWT yang memberikan rezeki untuk menempuh pendidikan dan penyusunan proposal skripsi ini.
2. Bapak H. Imron Rosyadi Hamid, M.Si., Selaku Rektor Universitas Islam Raden Rahmat (UNIRA) Malang.
3. Bapak Dr. Zainal Abidin, M.Si., Selaku Dekan Fakultas Sains dan Teknologi Universitas Islam Raden Rahmat (UNIRA) Malang.
4. Ibu Urnika Mudhifatul Jannah, S.Kom., M.Pd., selaku Ketua Program Studi Teknik Informatika Universitas Islam Raden Rahmat (UNIRA) Malang.
5. Bapak Rachmad Imam Tarecha, M.Kom., Selaku Dosen Wali yang telah mengamati proses belajar saya dari awal menempuh Perguruan Tinggi hingga akhir Perguruan Tinggi.
6. Ibu Priska Choirina, M.Tr.T. selaku Dosen Pembimbing pertama yang telah meluangkan waktu dan tenaga untuk mengarahkan dan membimbing.
7. Ibu Pangestuti Prima Darajat, M.Si. selaku Dosen Pembimbing kedua yang telah meluangkan waktu dan tenaga untuk mengarahkan dan membimbing.
8. Seluruh Dosen Fakultas Sains dan Teknologi khususnya Program Studi Teknik Informatika yang telah memberikan ilmu untuk mendidik dan

mengajar selama penulis menempuh Pendidikan di Program Studi Teknik Informatika.

9. Seluruh Staff Universitas Islam Raden Rahmat (UNIRA) Malang
10. Orang tua yang tercinta senantiasa memberikan doa dan dukunan agar tercapai maksud dan tujuan penulis.
11. Nafisa Lutfi Zahro selaku *support system* sekaligus kekasih, terima kasih telah kebersamaan penulis dalam hari hari yang tidak mudah dalam menyelesaikan skripsi.
12. Serta teman-teman yang sudah membantu proses penyusunan skripsi dan memotivasi serta meluangkan waktu

Penulis menyadari bahwa skripsi ini tidak luput dari berbagai kekurangan dan jauh dari kesempurnaan, namun penulis mengharapkan saran serta masukan bahkan kritik yang membangun demi kesempurnaan dan perbaikannya sehingga akhirnya proposal skripsi ini dapat memberikan manfaat bagi bidang pendidikan dan penerapan di lapangan khususnya dalam bidang teknik informatika serta bisa dikembangkan lagi lebih lanjut.

Malang, 14 Juni 2024

Penyusun

Muhammad Farkhan

UNIVERSITAS ISLAM
RADEN RAHMAT

PERNYATAAN KEASILAN TULISAN

Saya yang bertanda tangan dibawah ini :

Nama : Muhammad Farkhan Imfrozin Febrian Putra

NIM : 1955202002

Program Studi : Fakultas Sains dan Teknologi, Universitas Islam Raden

Rahmat Malang

Menyatakan dengan sesungguhnya bahwa skripsi yang saya tulis ini benar benar tulisan saya, dan bukan merupakan plagiasi baik Sebagian atau seluruhnya.

Apabila diselanjutnya hari terbukti skripsi ini hasil plagiasi, baik sebagian atau seluruhnya, maka saya bersedia menerima sanksi atas perbuatan saya tersebut dengan ketentuan yang berlaku.

Malang 14 Juni 2023

Yang membuat pernyataan,



Muhammad Farkhan Imfrozin F.P
NIM 1955202002

UNIVERSITAS ISLAM
RADEN RAHMAT

DAFTAR ISI

LEMBAR PERSETUJUAN SKRIPSI.....	i
KATA PENGANTAR	iii
DAFTAR ISI.....	vi
DAFTAR GAMBAR	viii
DAFTAR TABEL	ix
Abstrak.....	ix
Abstract.....	xi
BAB I PENDAHULUAN	1
1.1 Latar belakang.....	1
1.2 Rumusan Masalah.....	3
1.3 Tujuan Penelitian	4
1.4 Batasan Masalah	4
1.5 Manfaat penelitian	4
BAB II TINJAUAN PUSTAKA	6
2.1 <i>Bullying</i>	6
2.1.1 <i>Cyberbullying</i>	7
2.2 Python	8
2.3 <i>Text Mining</i>	8
2.4 Modul Sastrawi	9
2.5 Teks <i>Preprocessing</i>	10
2.5.1 <i>Case Folding</i>	10
2.5.2 <i>Tokenizing</i>	11
2.5.3 <i>Stopword Removal</i>	11
2.5.4 <i>Stemming</i>	12
2.6 <i>Term Weighting</i> TF-IDF	13
2.6.1 Hitung kata yang muncul di semua dokumen (TF).....	14
2.6.2 Normalisasi <i>Term Frequency</i> (TF).....	14
2.6.3 Hitung kata yang muncul di semua dokumen (DF)	16
2.6.4 <i>Inverse Document Frequency</i> (IDF)	17
2.6.5 Mengkalikan TF dan IDF (TF-IDF).....	18
2.7 <i>Multinomial Naïve Bayes Classifier</i>	19
2.8 <i>Confusion Matrix</i>	26
2.9 <i>Flask Python Framework</i>	26
2.10 Penelitian Terkait.....	26

2.11	Metode Penelitian <i>SDLC Waterfall</i>	28
2.12	Kerangka Berpikir.....	28
BAB III METODE PENELITIAN		30
3.1	Metode Pembangunan Sistem <i>SDLC Waterfall</i>	30
3.1.1	Analisis kebutuhan.....	30
3.1.2	Desain sistem.	30
3.1.3	Implementasi.....	32
3.1.4	Testing.....	32
3.2	Pengumpulan Data	32
3.3.2	<i>Tokenizing</i>	34
3.3.3	<i>Stopword removal</i>	34
3.3.4	<i>Stemming</i>	36
3.4	<i>Term weighting</i>	37
3.5	<i>Klasifikasi Naïve Bayes</i>	41
3.6	Implementasi.....	43
3.7	<i>Testing</i> atau pengujian	43
3.8	Kerangka Operasional.....	44
BAB IV HASIL DAN PEMBAHASAN		46
4.1	Pengumpulan Data	46
4.2	Inisialisasi Data.....	47
4.3	Preprocessing	48
4.4	<i>Term Weighting</i>	56
4.5	Klasifikasi Naïve Bayes	57
4.6	Metode Cross validation	60
4.7	Model <i>Complexity</i>	61
4.8	<i>Deploying</i> menggunakan flask.....	62
BAB V KESIMPULAN DAN SARAN.....		66
5.1	Kesimpulan	66
5.2	Saran	67
DAFTAR PUSTAKA		68
LAMPIRAN.....		72

DAFTAR GAMBAR

Gambar 2. 1 SDLC Waterfall	28
Gambar 2. 2 Kerangka berpikir.....	29
Gambar 3. 1 Desain Sistem.....	31
Gambar 3. 2 Tahapan Preprocessing.....	33
Gambar 3. 3 Flowchart Case Folding	33
Gambar 3. 4 Flowchart Tokenizing	34
Gambar 3. 5 Flowchart stopword removal.....	35
Gambar 3. 6 Flowchart stemming.....	36
Gambar 3. 7 Diagram alir menghitung TF.....	37
Gambar 3. 8 Flowchart TF	38
Gambar 3. 9 Diagram alir DF	39
Gambar 3. 10 Diagram alir IDF	40
Gambar 3. 11 Diagram alir mengalikan TF dan IDF	41
Gambar 3. 12 flowchart naive bayes.....	42
Gambar 3. 13 Kerangka operasional.....	44
Gambar 4. 1 Kode Case folding.....	49
Gambar 4. 2 Hasil Case Folding	49
Gambar 4. 3 Kode tokenizing	51
Gambar 4. 4 Hasil tokenizing	51
Gambar 4. 5 Kode Stopword Removal	53
Gambar 4. 6 Hasil Stopword removal.....	53
Gambar 4. 7 Kode program stemming.....	54
Gambar 4. 8 Hasil stemming	55
Gambar 4. 9 Kode program TF-IDF	56
Gambar 4. 10 Hasil TF-IDF menggunakan Python	57
Gambar 4. 11 Pemodelan Naïve Bayes.....	58
Gambar 4. 12 Model confusion matrix	59
Gambar 4. 13 Kode program Cross Validation.....	60
Gambar 4. 14 framework flask	63
Gambar 4. 15 Tampilan UI pada web.....	64
Gambar 4. 16 Tampilan UI pada web II	64

DAFTAR TABEL

Tabel 2. 1 Contoh kalimat yang akan dihitung	14
Tabel 2. 2 Hitung kata yang muncul	14
Tabel 2. 3 Normalisasi term frekuensi (TF).....	15
Tabel 2. 4 Sebelum perhitungan	16
Tabel 2. 5 Setelah perhitungan.....	17
Tabel 2. 6 Mencari IDF	18
Tabel 2. 7 Sebelum TF.IDF dihitung	18
Tabel 2. 8 TF.IDF sesudah dihitung	19
Tabel 2. 9 Hasil metode TF-IDF	19
Tabel 2. 10 Tabel keputusan main atau tidak.....	20
Tabel 2. 11 Penelitian terkait	27
Tabel 4. 1 Beberapa data sebelum preprocessing	46
Tabel 4. 2 Hasil pemberian label	47
Tabel 4. 3 Contoh sebelum dan sesudah case folding.....	50
Tabel 4. 4 Struktur data setelah case folding	50
Tabel 4. 5 Sebelum dan sesudah Tokenizing	52
Tabel 4. 6 Sampel data Tokennizing.....	52
Tabel 4. 7 Contoh sebelum dan setelah dilakukan Stopword Removal	53
Tabel 4. 8 Struktur data setelah stopwords removal.....	54
Tabel 4. 9 Sebelum dan setelah dilakukan stemming	55
Tabel 4. 10 Struktur data setelah stemming	55
Tabel 4. 11 hasil akurasi Pembagian data latih dan data uji	58
Tabel 4. 12 Hasil cross validation dari ketiga proporsi.....	61
Tabel 4. 13 Cross validation tiap fold	61

Abstrak

Penggunaan Internet mengalami peningkatan yang pesat hal ini berbanding lurus dengan pengguna sosial media. Dalam hal ini pengguna instagram sering menghadapi berbagai opini terhadap postingan yang diunggah. *Cyberbullying*

merupakan salah satu bentuk pelecehan yang dilakukan oleh para penggunanya. *Cyberbullying* memiliki dampak negatif yang berpengaruh terhadap mentalitas korban sehingga menimbulkan trauma pada diri korban. Pada penelitian ini penulis akan mengidentifikasi komentar instagram yang dapat dikategorikan sebagai tindakan *cyberbullying* dengan penerapan algoritma TF-IDF dan *naïve bayes*. Penelitian ini menggunakan metode penelitian *Waterfall* untuk model pembangunan yang berurutan dan sistematis dengan diawali tahap analisa diikuti desain sistem kemudian implementasi dan tahap pegujian. Dilakukan *text mining* yang mengimplementasikan pustaka python agar dapat melakukan *preprocessing*, serta kombinasi TF-IDF dan *Naïve Bayes*. Untuk menguji keakuratan sebuah model maka dilakukan metode pengujian *Cross Validation* dan pengujian akurasi, pada pengujian ini dilakukan tiga pembagian data latih dan data uji sebesar 60:40, 70:30, dan 80:20 pengujian ini menghasilkan masing masing 79%, 98% dan 80%, lalu pemodelan tersebut akan diterapkan pada platfrom berbasis web menggunakan *flask framework* untuk menginputkan kalimat dan akan menampilkan hasilnya.

Kata kunci : *Naïve bayes*, TF-IDF, *cyberbullying*, Instagram



UNIVERSITAS ISLAM
RADEN RAHMAT

Abstract

Internet usage has increased rapidly, this is directly proportional to social media users. In this case Instagram users often face various opinions on uploaded posts. Cyberbullying is one form of harassment committed by its users. Cyberbullying has a negative impact that affects the mentality of the victim, causing trauma to the victim. In this research, the author will identify Instagram comments that can be categorized as cyberbullying by applying the TF-IDF algorithm and naïve bayes. This research uses the Waterfall research method for a sequential and systematic development model starting with the analysis stage followed by system design then implementation and testing stages. Text mining is carried out which implements the python library in order to do preprocessing, as well as a combination of TF-IDF and Naïve Bayes. To test the accuracy of a model, the Cross Validation testing method and accuracy testing are carried out, in this test three divisions of training data and test data of 60:40, 70:30, and 80:20 this test produces 79%, 98% and 80% respectively, then the modeling will be applied to a web-based platform using the flask framework to input sentences and will display the results. Translated with DeepL.com (free version)

Keywords : Naïve Bayes, TF-IDF, Cyberbullying, Instagram



UNIVERSITAS ISLAM
RADEN RAHMAT

BAB I

PENDAHULUAN

1.1 Latar belakang

Penggunaan Internet merupakan hal yang tidak dapat dilepaskan dari kehidupan manusia. internet memudahkan manusia untuk berinteraksi serta mendapat informasi dari berbagai belahan dunia. Rahman (2021) berpendapat bahwa internet merupakan jaringan global yang menghubungkan beribu-ribu, bahkan berjuta-juta jaringan komputer dari berbagai belahan bumi untuk mengirim dan memperoleh informasi ataupun mentransfer data. Selain itu Alifah (2024) mengemukakan bahwa internet adalah salah satu teknologi yang paling penting bagi setiap orang dalam kehidupan sehari-hari dan merupakan sebuah teknologi yang terus berkembang, menurut dua pakar di atas, dapat disimpulkan bahwa internet adalah teknologi yang terus berkembang berupa jaringan global yang menghubungkan berbagai informasi hingga data dan berperan penting dalam kehidupan manusia.

Pertumbuhan Jumlah pengguna internet di Indonesia mengalami peningkatan yang signifikan setiap tahun. Informasi dari Asosiasi Pengguna Jasa Internet Indonesia (APJII) menunjukkan bahwa dalam periode tahun 2019 hingga kuartal II/2020, jumlah pengguna internet mencapai 196,7 juta orang. Terdapat peningkatan sebanyak 23,5 juta orang, atau sekitar 8,9%, dibandingkan dengan angka pada tahun 2018. Sebagian besar pengguna internet berasal dari provinsi Jawa Barat, dengan jumlah mencapai 35,1 juta orang. Disusul oleh Jawa Tengah yang mencatat 26,5 juta orang, dan Jawa Timur dengan jumlah 23,4 juta orang. Jumlah pengguna internet di Sumatera Utara mencapai 11,7 juta orang, sementara di Banten mencapai 9,98 juta orang. Jakarta juga memiliki sekitar 8,9 juta orang pengguna internet. Survei yang dilakukan oleh APJII menggunakan kuesioner dan wawancara terhadap 7.000 sampel, dengan tingkat toleransi kesalahan sebesar 1,27%.

Dibalik peningkatan pengguna internet yang cukup pesat, maka pengguna sosial media akan semakin banyak, sosial media dapat diartikan sebagai suatu platform yang menyuguhkan fasilitas untuk melakukan aktivitas sosial, mulai dari komunikasi hingga informasi. Media sosial merupakan bentuk dari perkembangan

internet salah satunya adalah Instagram, Media sosial yang sering digunakan dalam kehidupan sehari-hari. Instagram adalah salah satu platform jejaring sosial di mana pengguna dapat berbagi foto dan video untuk mempublikasikan karya ataupun gaya hidup pengguna atau pemilik akun instagram.

Pengguna Instagram berasal dari berbagai latar belakang, termasuk pejabat, selebriti, dan orang biasa. Maraknya media sosial pada waktu itu pengguna instagram sering menghadapi dengan berbagai opini terhadap postingan yang diunggah bahkan sering terjadi sebuah hujatan atau *hate speech*, hal tersebut merupakan pergeseran fungsi utama dari media sosial yang berupa dampak negatif sesuai dengan pendapat yang di kemukakan oleh Maulana, F. A., dan Ernawati, I. (2020) saat ini banyak oknum pengguna media sosial yang melakukan penyalahgunaan untuk melakukan penghinaan ataupun pencemaran nama baik, hal tersebut masuk kategori *cyberbullying*. Meskipun media sosial memudahkan kita untuk berbagi informasi, *cyberbullying* merupakan salah satu bentuk pelecehan yang dilakukan oleh para penggunanya.

Penindasan yang terjadi melalui sarana digital seperti ponsel, laptop, dan tablet dikenal sebagai *cyberbullying*. Penindasan siber dapat terjadi melalui aplikasi media sosial, forum, atau permainan di mana pengguna dapat membaca, berpartisipasi, atau bertukar konten. Mengirim, memposting, atau menyebarkan konten negatif, menyakitkan, menyesatkan, atau berbahaya tentang orang lain merupakan perundungan siber. Ada sebuah hubungan yang berkesinambungan dan signifikan antara perilaku provokator *cyberbullying* dan perilaku korban *cyberbullying* yang mana semakin reaktif perilaku provokator *cyberbullying* maka semakin reaktif pula perilaku korban *cyberbullying*. Hal ini membuktikan betapa besarnya pengaruh *cyberbullying* terhadap kehidupan sosial (Khaira et al., 2020).

Maka dari itu penulis tertarik untuk melakukan penelitian guna mengidentifikasi komentar instagram yang dapat dikategorikan sebagai tindakan *cyberbullying*, agar pengguna sosial media terutama instagram dapat menyaring komentar yang tidak layak seperti *hate speech*. Penelitian ini akan menggunakan algoritma *Naïve Bayes* karena penulis menggunakan data latih dalam jumlah yang kecil pada proses klasifikasi data. Beberapa penelitian sebelumnya yang

menggunakan pengklasifikasi *Naïve Bayes* ini diantaranya adalah penelitian yang dilakukan oleh Khadafi dkk (2022) yang bertujuan untuk menguji akurasi dari pengelompokan ujaran *cyberbullying* terhadap akun sosial media Instagram (Badan Penyelenggara Jaminan Sosial) BPJS. Penelitian tersebut mengkombinasikan metode *Naïve Bayes* dengan (*term frequency–inverse document frequency*) TF-IDF mendapatkan akurasi 80% dan 512 data terkumpul. Pada penelitian lain yang diusut oleh Syarif D.R dkk (2019) dalam mengidentifikasi *cyberbullying* pada komentar instagram menggunakan metode *Naïve Bayes* menggunakan ekstraksi fitur (*term frequency–inverse document frequency*) TF-IDF menghasilkan akurasi 97% dengan jumlah data sebanyak 360 data. Dengan menggunakan algoritma *Naïve Bayes*, akan dibangun sebuah aplikasi pendeteksi *cyberbullying* terhadap media sosial Instagram untuk menentukan apakah sebuah komentar memiliki *cyberbullying* atau tidak..

Pada penelitian ini akan dibangun sebuah sistem indentifikasi *cyberbullying* menggunakan pengolahan dataset *cyberbullying* kometar instagram menggunakan bahasa pemrograman *python*, dengan menjalankan metode-metdoe *preprocessing* *casefolding*, *tokenizing*, normalisasi, *stopword removal*, dan *stemming*. Membutuhkan *library* dari *python* untuk melakukan pendekatan dengan melatih *classifier* menggunakan algoritma TF-IDF dan *Naïve Bayes* pada *dataset* yang telah dilabelkan secara manual. Metode *naïve bayes* dikombinasikan dengan metode ekstraksi fitur (*term frequency–inverse document frequency*) TF-IDF. Berfungsi untuk mencari representasi nilai dari tiap-tiap dokumen dari suatu kumpulan data training (training set), dimana nantinya dibentuk suatu vektor antara dokumen dengan kata (Grootendorst, M., 2022). Sehingga dapat mengklasifikasikan komentar kedalam kelas *Cyberbullying* dan *Noncyberbullying*. Data komentar Instagram ini akan bermanfaat bagi para pengembang untuk membuat sistem yang dapat memprediksi apakah sebuah pesan mengandung *cyberbullying* atau tidak sebagai sistem peringatan dini, dengan harapan dapat meminimalisir jumlah kejadian perundungan di dunia maya.

1.2 Rumusan Masalah

Berdasarkan pada latar belakang penelitian yang telah dipaparkan, dapat di temukan rumusan masalah sebagai berikut

1. Bagaimana merancang dan membangun sistem identifikasi *cyberbullying* dengan algoritma *Naïve Bayes* dan TF-IDF yang diterapkan pada metode pengembangan *Waterfall*?
2. Bagaimana akurasi dari penerapan kombinasi algoritma TF-IDF dan *Naïve Bayes* pada identifikasi *cyberbullying*?

1.3 Tujuan Penelitian

Tujuan dari penelitian ini diuraikan dalam urutan sebagai berikut :

1. Merancang dan mengimplementasikan metode *Waterfall* serta kombinasi TF-IDF dan *Naïve Bayes* untuk membangun sistem pendeteksi *cyberbullying*.
2. Mengetahui akurasi dari penerapan algoritma *Naïve Bayes* dan TF-IDF, yang dapat menentukan probabilitas kelas komentar, yang berpotensi untuk melakukan *cyberbullying* dan *Noncyberbullying*.

1.4 Batasan Masalah

Agar penelitian ini lebih terfokus, ada beberapa batasan masalah ditetapkan untuk memfokuskan penelitian ini, di antaranya :

1. Menggunakan data pelatihan *cyberbullying* Instagram yang diambil dari Kaggle.com.
2. Data yang digunakan hanya berfokus pada kalimat berbahasa Indonesia.
3. Menggunakan *library* yang tersedia pada *python*.
4. Penggunaan *stopword* removal menggunakan *stoplist* yang telah tersedia pada *library* python
5. Menggunakan dua kategori sentimen positif dan negatif
6. Menggunakan model berbasis *Naïve Bayes* dan tidak membandingkan dengan algoritma lain.

1.5 Manfaat penelitian

Manfaat yang diperoleh dari penelitian ini adalah, tersedianya sebuah aplikasi pendeteksi dan dataset mengenai kata atau gabungan kata yang dapat digunakan untuk melakukan *bullying* pada sosial media instagram. Dataset dan aplikasi pendeteksi ini suatu saat dapat berguna bagi pengembang perangkat lunak untuk

membuat sistem atau aplikasi yang mampu mendeteksi terhadap terjadinya *cyberbullying* dengan harapan dapat mengurangi angka terjadinya *bullying* di media sosial khususnya instagram. Adapun dampak positif bagi semua pihak yang berkaitan diantaranya sebagai berikut:

1. Bagi Universitas.

Universitas Islam Raden Rahmat sebagai acuan dalam mengelompokkan sebuah kritik dan saran yang mengandung unsur *cyberbullying* pada forum daring. Selain itu, semua institusi kampus dapat menilai kemampuan mahasiswa dalam menggunakan informasi yang telah diperoleh selama ini.

2. Bagi Penulis

Menerapkan informasi atau metode penelitian yang telah dipelajari selama masa perkuliahan untuk digunakan dalam menganalisis isu-isu terkini dan mencari sebuah solusi.

3. Bagi pembaca

Memberikan pengetahuan dan ilmu baru mengenai klasifikasi menggunakan *naïve bayes*, selain itu bisa menjadikan penelitian ini sebagai referensi apabila ingin membuat sebuah sistem informasi yang serupa



UNIVERSITAS ISLAM
RADEN RAHMAT