

**KLASIFIKASI KOMENTAR PENGGUNA *INSTAGRAM*  
YANG TERINDIKASI *CYBERBULLYING*  
MENGUNAKAN ALGORITMA *NAÏVE BAYES***

**SKRIPSI**

**Oleh:**

**PRASETYO HANIEF AHMADA**

**201910225063**



**PROGRAM STUDI INFORMATIKA  
FAKULTAS ILMU KOMPUTER  
UNIVERSITAS BHAYANGKARA JAKARTA RAYA**

**2023**

## LEMBAR PERSETUJUAN PEMBIMBING

Judul Tugas Akhir : Klasifikasi Komentar Pengguna Instagram  
Yang Terindikasi Cyberbullying  
Menggunakan Algoritma Naïve Bayes

Nama Mahasiswa : Prasetyo Hanief Ahmada  
Nomor Induk Pokok : 201910225063  
Program Studi / Fakultas : Informatika / Ilmu Komputer  
Tanggal Lulus Ujian Skripsi :

Jakarta, 5 Agustus 2023

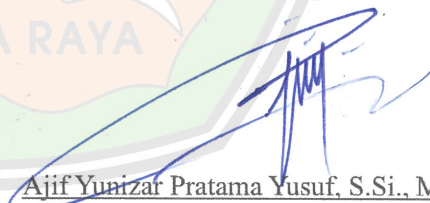
MENYETUJUI,

Pembimbing I

Pembimbing II

  
Prima Dina Atika, S.Kom., M.Kom.

NIDN: 0311037107

  
Ajif Yunizar Pratama Yusuf, S.Si., M.Eng

NIDN: 0328068603

## LEMBAR PENGESAHAN

Judul Skripsi : Klasifikasi Komentar Pengguna Instagram  
Yang Terindikasi Cyberbullying  
Menggunakan Algoritma Naïve Bayes

Nama Mahasiswa : Prasetyo Hanief Ahmada

Nomor Pokok Mahasiswa : 201910225063

Program Studi/Fakultas : Informatika / Ilmu Komputer

Tanggal Lulus Ujian Skripsi : 5 Agustus 2023

Jakarta, 5 Agustus 2023

Mengesahkan,

Ketua Tim Penguji : Abrar Hiswara, S.T., M.M., M.Kom. .....  
NIDN. 0324028101

Penguji (I) : Dwipa Handayani, S.Kom., M.M.S.I. .....  
NIDN. 0317078008

Penguji (II) : Prima Dina Atika, S.Kom., M.Kom. .....  
NIDN. 0311037107

Mengetahui,

Ketua

Dekan

Program Studi Informatika

Fakultas Ilmu Komputer

Ahmad Fathurrozi, S.E., M.M.S.I.

Dr. Dra. Tyastuti Sri Lestari, M.M.

NIP. 2012486

NIP. 1408206



UNIVERSITAS BHAYANGKARA JAKARTA RAYA  
FAKULTAS ILMU KOMPUTER  
PROGRAM STUDI INFORMATIKA

LEMBAR PERNYATAAN BUKAN PLAGIASI

Yang bertanda tangan dibawah ini :

Nama : Prasetyo Hanief Ahmada  
NPM : 201910225063  
Program Studi : Informatika  
Fakultas : Ilmu Komputer  
Judul Tugas Akhir : Klasifikasi Komentar Pengguna Instagram  
Yang Terindikasi Cyberbullying Menggunakan  
Algoritma Naïve Bayes

Dengan ini menyatakan bahwa hasil penulisan skripsi yang telah saya buat ini merupakan **hasil karya saya sendiri dan benar keasliannya**. Apabila dikemudian hari penulisan skripsi ini merupakan plagiat atau penjiplakan terhadap karya orang lain, maka saya bersedia mempertanggungjawabkan sekaligus bersedia menerima sanksi berdasarkan tata tertib di Universitas Bhayangkara Jakarta Raya.

Demikian pernyataan ini saya buat dalam keadaan sadar dan tidak dipaksakan dari pihak manapun.

Bekasi, 5 Agustus 2023

Penulis



Prasetyo Hanief Ahmada



## ABSTRAK

**Prasetyo Hanief Ahmada. 201910225063.** Klasifikasi Komentar Pengguna *Instagram* Yang Terindikasi *Cyberbullying* Menggunakan Algoritma *Naïve Bayes*. Bekasi: Fakultas Ilmu Komputer. Universitas Bhayangkara Jakarta Raya. 2023.

Instagram merupakan salah satu media sosial yang cukup populer. Instagram adalah sebuah aplikasi berbagi foto dan video yang memungkinkan para pengguna untuk mengambil foto atau video, menerapkan *filter* digital dan membagikannya. Dibalik semua fitur yang ditawarkan, *Instagram* juga memiliki dampak negatif, salah satunya adalah penindasan secara *online* atau *cyberbullying*. Tidak banyak pengguna *Instagram* yang menyadari bahwa ulasan atau komentar yang dilontarkan kepada seseorang ataupun kelompok merupakan tindakan *cyberbullying*. *Cyberbullying* tentunya sangat meresahkan karena dampak yang dapat ditimbulkan, oleh karena itu dapat dilakukan klasifikasi komentar dalam upaya untuk mengetahui apakah komentar tersebut mengandung unsur *bullying* atau tidak. Klasifikasi adalah sebuah proses menemukan kesamaan karakteristik dalam suatu kelompok atau kelas. Untuk mengklasifikasikan komentar digunakan algoritma *Naïve Bayes*. Dataset yang digunakan berisi 1024 dengan 578 komentar berkategori *bullying* dan 446 data komentar berkategori tidak *bullying*. Dataset dibagi menjadi 30% data uji dan 70% data latih. Berdasarkan pengujian yang telah dilakukan menghasilkan nilai *accuracy* 89%, *precision* 89%, *recall* 94% dan nilai *f1-score* 91%.

**Kata Kunci:** *Naïve Bayes*, *Cyberbullying*, Klasifikasi, *Confusion Matrix*.

## **ABSTRACT**

**Prasetyo Hanief Ahmada. 201910225063. Classification of Instagram User Comments Indicated of Cyberbullying Using the Naïve Bayes Algorithm. Bekasi: Faculty of Computer Science. Jakarta Bhayangkara University. 2023.**

Instagram is one of the most popular social media. Instagram is a photo and video sharing application that allows users to take photos or videos, apply digital filters and share them. Behind all the features offered, Instagram also has a negative impact, one of which is online bullying or cyberbullying. Not many Instagram users realize that a review or comment directed at a person or group is an act of cyberbullying. Cyberbullying is of course very troubling because of the impact that can be caused, therefore comments can be classified in an effort to find out whether these comments contain elements of bullying or not. Classification is a process of finding common characteristics in a group or class. To classify comments, the Naïve Bayes algorithm is used. The dataset used contains 1024 with 578 comments categorized as bullying and 446 data comments not categorized as bullying. The dataset is divided into 30% test data and 70% training data. Based on the tests that have been carried out, the accuracy is 89%, the precision is 89%, the recall is 94% and the f1-score is 91%.

**Keywords:** Naïve Bayes, Cyberbullying, Classification, Confusion Matrix.

## LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIK

Sebagai sivitas akademik Universitas Bhayangkara Jakarta Raya, saya yang bertanda tangan di bawah ini :

Nama : Prasetyo Hanief Ahmada  
NPM : 201910225063  
Program Studi : Informatika  
Fakultas : Ilmu Komputer  
Jenis Karya : Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Bhayangkara Jakarta Raya **Hak Bebas Royalti Non-Eksklusif (Non-Exclusive Royalty-Free Right)**, atas karya ilmiah saya yang berjudul :

**“KLASIFIKASI KOMENTAR PENGGUNA INSTAGRAM YANG TERINDIKASI CYBERBULLYING MENGGUNAKAN ALGORITMA NAÏVE BAYES”**

beserta perangkat yang ada (bila diperlukan). Dengan hak bebas royalti non-eksklusif ini, Universitas Bhayangkara Jakarta Raya berhak menyimpan, mengalihmediakan, mengelolanya dalam bentuk pangkalan data (*database*), mendistribusikannya dan mempublikasikannya di Internet atau media lain untuk kepentingan akademis tanpa perlu meminta ijin dari saya selama tetap mencantumkan nama saya sebagai penulis dan sebagai pemilik hak cipta.

Segala bentuk tuntutan hukum yang timbul atas pelanggaran hak cipta dalam karya ilmiah ini menjadi tanggung jawab saya pribadi

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Bekasi  
Pada tanggal : 5 Agustus 2023  
Yang Menyatakan



Prasetyo Hanief Ahmada



## KATA PENGANTAR

Puji syukur kehadirat Allah Swt yang telah memberikan rahmat, hidayah dan karunia-Nya, sehingga penulis dapat menyelesaikan skripsi yang berjudul **Klasifikasi Komentar Pengguna Instagram Yang Terindikasi Cyberbullying Menggunakan Algoritma Naïve Bayes** dengan baik.

Penulis sangat menyadari bahwa dalam penulisan skripsi ini banyak pihak yang telah membantu dan memberikan dukungan dalam menyelesaikan tugas akhir ini. Oleh sebab itu, penulis ingin mengucapkan terima kasih kepada:

1. Bapak Irjen Polisi (Purn) Prof. Dr. Drs. Bambang Karsono, S.H., M.M. Selaku Rektor Universitas Bhayangkara Jakarta Raya.
2. Ibu Dr. Dra. Tyastuti Sri Lestari, M.M. Selaku Dekan Fakultas Ilmu Komputer Universitas Bhayangkara Jakarta Raya.
3. Bapak Ahmad Fathurrozi, S.E., M.M.S.I. Selaku Kepala Program Studi Informatika Universitas Bhayangkara Jakarta Raya.
4. Ibu Prima Dina Atika, S.Kom., M.Kom. Selaku Dosen pembimbing satu dalam penyusunan skripsi yang telah memberikan bimbingan, arahan dan masukan yang sangat berharga dalam penulisan skripsi ini.
5. Bapak Ajif Yunizar Pratama Yusuf, S.Si., M.Eng. Selaku Dosen pembimbing dua dalam penyusunan skripsi yang telah memberikan bantuan, bimbingan, dan masukan yang sangat berharga dalam penulisan skripsi ini.
6. Ibu Rafika Sari, S.Si., M.Si. selaku dosen pembimbing akademik
7. Kepada Orang Tua yang selalu memberikan dukungan dan motivasi kepada penulis untuk menyelesaikan skripsi ini.
8. Teman-teman mahasiswa seperjuangan Program Studi Informatika Universitas Bhayangkara Jakarta Raya.
9. Teman-teman dan sahabat yang memberikan dukungan dan bantuan dalam penulisan skripsi.



Semoga Allah Swt selalu memberikan balasan yang terbaik kepada seluruh pihak yang telah membantu, membimbing dan memotivasi dalam penulisan skripsi ini. Penulis menyadari bahwa skripsi ini masih jauh dari sempurna dan masih banyak kekurangan. Oleh karena itu, penulis mengharapkan kritik dan saran yang membangun dari pembaca untuk perbaikan penelitian dimasa depan.

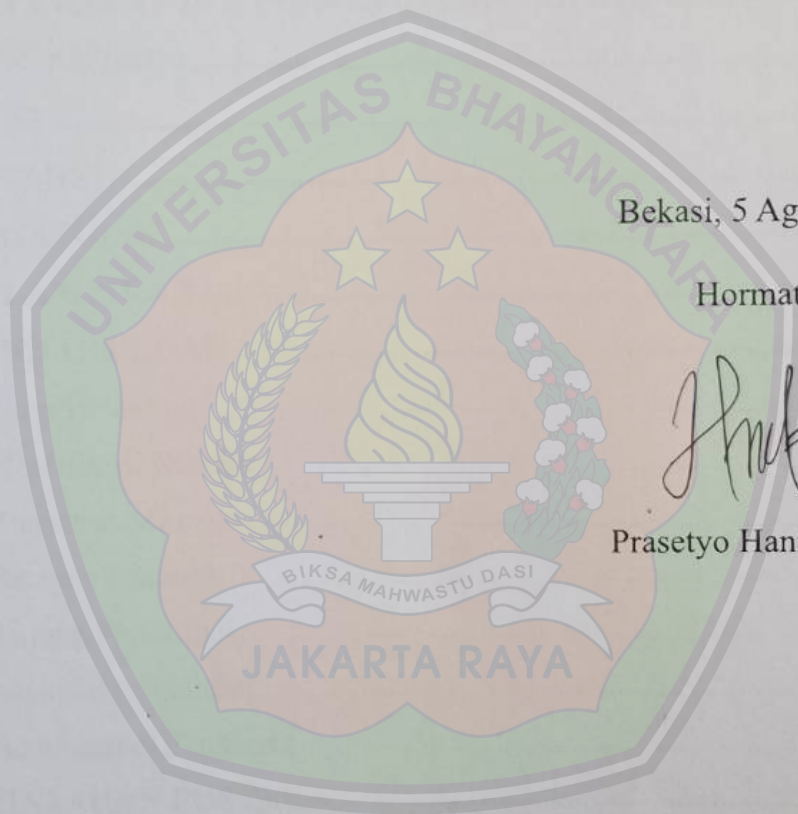
Akhir kata, penulis berharap skripsi ini dapat berguna serta memberikan manfaat kepada semua pembaca, khususnya bagi penulis pribadi.

Bekasi, 5 Agustus 2023

Hormat saya



Prasetyo Hanief Ahmada



## DAFTAR ISI

|                                                      |             |
|------------------------------------------------------|-------------|
| <b>LEMBAR PERSETUJUAN PEMBIMBING .....</b>           | <b>i</b>    |
| <b>LEMBAR PENGESAHAN .....</b>                       | <b>ii</b>   |
| <b>LEMBAR PERNYATAAN BUKAN PLAGIASI .....</b>        | <b>iii</b>  |
| <b>ABSTRAK .....</b>                                 | <b>iv</b>   |
| <b>ABSTRACT .....</b>                                | <b>v</b>    |
| <b>LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI .....</b> | <b>vi</b>   |
| <b>KATA PENGANTAR .....</b>                          | <b>vii</b>  |
| <b>DAFTAR ISI .....</b>                              | <b>ix</b>   |
| <b>DAFTAR TABEL .....</b>                            | <b>xi</b>   |
| <b>DAFTAR GAMBAR .....</b>                           | <b>xii</b>  |
| <b>DAFTAR LAMPIRAN .....</b>                         | <b>xiii</b> |
| <b>BAB I PENDAHULUAN .....</b>                       | <b>1</b>    |
| 1.1 Latar Belakang .....                             | 1           |
| 1.2 Identifikasi Masalah .....                       | 2           |
| 1.3 Rumusan Masalah .....                            | 3           |
| 1.4 Batasan Masalah .....                            | 3           |
| 1.5 Tujuan Penelitian .....                          | 3           |
| 1.6 Manfaat Penelitian .....                         | 3           |
| 1.7 Sistematika Penulisan .....                      | 4           |
| <b>BAB II TINJAUAN PUSTAKA .....</b>                 | <b>5</b>    |
| 2.1 Tinjauan Pustaka .....                           | 5           |
| 2.2 Landasan Teori .....                             | 8           |
| 2.2.1 Cyberbullying .....                            | 8           |
| 2.2.2 Komentar .....                                 | 8           |
| 2.2.3 Klasifikasi .....                              | 9           |
| 2.2.4 Text Classification .....                      | 9           |
| 2.2.5 Text Mining .....                              | 9           |
| 2.2.6 Algoritma Naïve Bayes .....                    | 10          |
| 2.2.7 Text Preprocessing .....                       | 10          |
| 2.2.7.1 Case Folding .....                           | 11          |
| 2.2.7.2 Tokenizing .....                             | 11          |

|                                                                        |           |
|------------------------------------------------------------------------|-----------|
| 2.2.7.3 <i>Stopword Removal</i> .....                                  | 12        |
| 2.2.7.4 <i>Stemming</i> .....                                          | 12        |
| 2.2.7.5 <i>Instagram</i> .....                                         | 13        |
| 2.2.8 <i>User</i> .....                                                | 14        |
| 2.2.9 <i>Confusion Matrix</i> .....                                    | 14        |
| 2.2.10 <i>Python</i> .....                                             | 16        |
| 2.2.11 <i>Term Frequency-Inverse Document Frequency (TF-IDF)</i> ..... | 17        |
| <b>BAB III METODOLOGI PENELITIAN .....</b>                             | <b>19</b> |
| 3.1 <i>Object Penelitian</i> .....                                     | 19        |
| 3.2 <i>Kerangka Penelitian</i> .....                                   | 19        |
| 3.3 <i>Analisis Kebutuhan Hardware dan Software</i> .....              | 22        |
| 3.3.1 <i>Spesifikasi Hardware</i> .....                                | 22        |
| 3.3.2 <i>Spesifikasi Software</i> .....                                | 22        |
| <b>BAB IV HASIL DAN PEMBAHASAN.....</b>                                | <b>23</b> |
| 4.1 <i>Hasil</i> .....                                                 | 23        |
| 4.2 <i>Pembahasan</i> .....                                            | 23        |
| 4.2.1 <i>Pengumpulan Data</i> .....                                    | 23        |
| 4.2.2 <i>Pelabelan Data</i> .....                                      | 25        |
| 4.2.3 <i>Preprocessing Data</i> .....                                  | 26        |
| 4.2.3.1 <i>Case Folding</i> .....                                      | 26        |
| 4.2.3.2 <i>Tokenizing</i> .....                                        | 27        |
| 4.2.3.3 <i>Stopword Removal</i> .....                                  | 28        |
| 4.2.3.4 <i>Stemming</i> .....                                          | 30        |
| 4.2.4 <i>Pembobotan Data</i> .....                                     | 31        |
| 4.2.5 <i>Implementasi Algoritma Naïve Bayes</i> .....                  | 40        |
| 4.2.6 <i>Evaluasi</i> .....                                            | 48        |
| 4.2.7 <i>Visualisasi Data</i> .....                                    | 50        |
| <b>BAB V PENUTUP .....</b>                                             | <b>52</b> |
| 5.1 <i>Kesimpulan</i> .....                                            | 52        |
| 5.2 <i>Saran</i> .....                                                 | 52        |
| <b>DAFTAR PUSTAKA.....</b>                                             | <b>53</b> |
| <b>LAMPIRAN.....</b>                                                   | <b>57</b> |



## DAFTAR TABEL

|                                                                   |    |
|-------------------------------------------------------------------|----|
| Tabel 2.1 Penelitian Terdahulu.....                               | 7  |
| Tabel 2. 2 Contoh <i>Case Folding</i> .....                       | 11 |
| Tabel 2. 3 Contoh <i>Tokenizing</i> .....                         | 11 |
| Tabel 2. 4 Contoh <i>Stopword Removal</i> .....                   | 12 |
| Tabel 2. 5 Contoh <i>Stemming</i> .....                           | 13 |
| Tabel 4. 1 Sampel Data Latih.....                                 | 25 |
| Tabel 4. 2 Sampel Data Uji .....                                  | 25 |
| Tabel 4. 3 Sampel data latih proses <i>case folding</i> .....     | 26 |
| Tabel 4. 4 Sampel data uji proses <i>case folding</i> .....       | 27 |
| Tabel 4. 5 Sampel data latih proses <i>tokenizing</i> .....       | 27 |
| Tabel 4. 6 Sampel data uji proses <i>tokenizing</i> .....         | 28 |
| Tabel 4. 7 Sampel data latih proses <i>stopword removal</i> ..... | 29 |
| Tabel 4. 8 Sampel data uji proses <i>stopword removal</i> .....   | 29 |
| Tabel 4. 9 Sampel data latih proses <i>stemming</i> .....         | 30 |
| Tabel 4. 10 Sampel data uji proses <i>stemming</i> .....          | 31 |
| Tabel 4. 11 Perhitungan TF dan DF .....                           | 32 |
| Tabel 4. 12 Perhitungan IDF .....                                 | 35 |
| Tabel 4. 13 Pembobotan TF-IDF.....                                | 38 |
| Tabel 4. 14 Sampel Data Latih.....                                | 41 |
| Tabel 4. 15 Sampel Data Uji .....                                 | 41 |
| Tabel 4. 16 Hasil uji kategori .....                              | 48 |
| Tabel 4. 17 Hasil <i>confusion matrix</i> data uji.....           | 48 |

## DAFTAR GAMBAR

|                                                                     |    |
|---------------------------------------------------------------------|----|
| Gambar 2.1 <i>Confusion Matrix</i> .....                            | 15 |
| Gambar 3. 1 Akun Instagram @raffinagita1717 .....                   | 19 |
| Gambar 3. 2 Kerangka Penelitian .....                               | 20 |
| Gambar 4. 1 <i>Website PhantomBuster</i> .....                      | 24 |
| Gambar 4. 2 Contoh Data Komentar.....                               | 24 |
| Gambar 4. 3 Hasil <i>confusion matrix naive bayes</i> data uji..... | 49 |
| Gambar 4. 4 <i>WordCloud</i> Kategori Bullying.....                 | 51 |
| Gambar 4. 5 <i>WordCloud</i> Kategori Tidak Bullying.....           | 51 |



## DAFTAR LAMPIRAN

|                                         |    |
|-----------------------------------------|----|
| Lampiran 1 Uji Plagiasi.....            | 58 |
| Lampiran 2 Biodata Mahasiswa.....       | 59 |
| Lampiran 3 Kartu Bimbingan Skripsi..... | 60 |
| Lampiran 4 Validasi Instrumen.....      | 62 |

