

**IMPLEMENTASI METODE KLASIFIKASI NAÏVE BAYES  
UNTUK MENDETEKSI SPAM DI DISCORD**

**SKRIPSI**

untuk memenuhi salah satu syarat mencapai derajat Sarjana  
Program Studi Informatika



diajukan oleh  
**IKHWAN TAUFIQ FARHAN**  
**18.11.2408**

Kepada  
**PROGRAM SARJANA**  
**PROGRAM STUDI INFORMATIKA**  
**FAKULTAS ILMU KOMPUTER**  
**UNIVERSITAS AMIKOM YOGYAKARTA**  
**YOGYAKARTA**  
**2022**

**IMPLEMENTASI METODE KLASIFIKASI NAÏVE BAYES  
UNTUK MENDETEKSI SPAM DI DISCORD**

**SKRIPSI**

untuk memenuhi salah satu syarat mencapai derajat Sarjana  
Program Studi Informatika



diajukan oleh

**IKHWAN TAUFIQ FARHAN**

**18.11.2408**

Kepada

**PROGRAM SARJANA**

**PROGRAM STUDI INFORMATIKA**

**FAKULTAS ILMU KOMPUTER**

**UNIVERSITAS AMIKOM YOGYAKARTA**

**YOGYAKARTA**

**HALAMAN PERSETUJUAN**

**SKRIPSI**

**IMPLEMENTASI METODE KLASIFIKASI NAÏVE BAYES  
UNTUK MENDETEKSI SPAM DI DISCORD**

yang disusun dan diajukan oleh

**Ikhwan Taufiq Farhan**

**18.11.2408**

telah disetujui oleh Dosen Pembimbing Skripsi  
pada tanggal 22 Agustus 2022

**Dosen Pembimbing.**

**Dwi Nurani, M.Kom**

**NIK. 190302236**

**HALAMAN PENGESAHAN**  
**SKRIPSI**  
**IMPLEMENTASI METODE KLASIFIKASI NAÏVE BAYES**  
**UNTUK MENDETEKSI SPAM DI DISCORD**

yang disusun dan diajukan oleh

**Ikhwan Taufiq Farhan**

**18.11.2408**

Telah dipertahankan di depan Dewan Penguji  
pada tanggal 22 Agustus 2022

**Susunan Dewan Penguji**

**Nama Penguji**

**Tanda Tangan**

**Andika Agus Slameto, M.Kom**

**NIK. 190302109**

**Alnul Yaqln, M. Kom**

**NIK. 190302255**

**Dwi Nurani, M.Kom**

**NIK. 190302236**

Skripsi ini telah diterima sebagai salah satu persyaratan  
untuk memperoleh gelar Sarjana Komputer  
Tanggal 22 Agustus 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 : Ikhwan Taufiq Farhan

NIM : 18.11.2408

Menyatakan bahwa Skripsi dengan judul berikut:

### IMPLEMENTASI METODE KLASIFIKASI NAÏVE BAYES UNTUK MENDETEKSI SPAM DI DISCORD

Dosen Pembimbing : Dwi Nurani, MKom

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, 13 September 2022

Yang Menyatakan,



Ikhwan Taufiq Farhan

## HALAMAN PERSEMBAHAN

Pertama-tama puji syukur saya panjatkan pada Allah SWT atas terselesaikannya Skripsi ini dengan baik dan lancar. Dan Skripsi ini saya persembahkan untuk:

- Ibuku Makinem M.P.d tercinta, yang selalu mendukung penulis baik dalam materi, motivasi, doa dan kasih sayang yang tulus.
- Keluarga besar saya yang telah senantiasa telah membantu menyelesaikan Skripsi ini
- Ibu Dwi Nurani M.Kom selaku dosen pembimbing terimakasih atas bimbingannya dan motivasi-motivasinya yang akan selalu saya ingat dan sebagai pembelajaran.
- Orang spesial didalam hidup saya yaitu Ayu Septika Izilia, terima kasih atas support,motivasi agar saya segera menyelesaikan skripsi ini.
- Teman-teman saya yang telah membantu dalam menyelesaikan skripsi yaitu Silmi,Ridho,Pares dan lainnya.
- *Last but not least, I wanna thank me, I wanna thank me for believing in me, I wanna thank me for doing all this hard work, I wanna thank me for having no days off, I wanna thank me for never quitting, I wanna thank me for always being a giver And tryna give more than I recieve, I wanna thank me for tryna do more right than wrong. I wanna thank me for just being me at all times.*

## KATA PENGANTAR

Assalamualaikum Warahmatulahi Wabarakatuh

Segala puji bagi Allah yang telah memberikan saya kemudahan sehingga saya dapat menyelesaikan skripsi yang berjudul “Implementasi Metode Klasifikasi Naïve Bayes Untuk Mendeteksi Spam di Discord” dengan sebaik mungkin. Sholawat dan salam semoga terlimpah curahkan kepada baginda tercinta kita yakni Nabi Muhammad SAW.

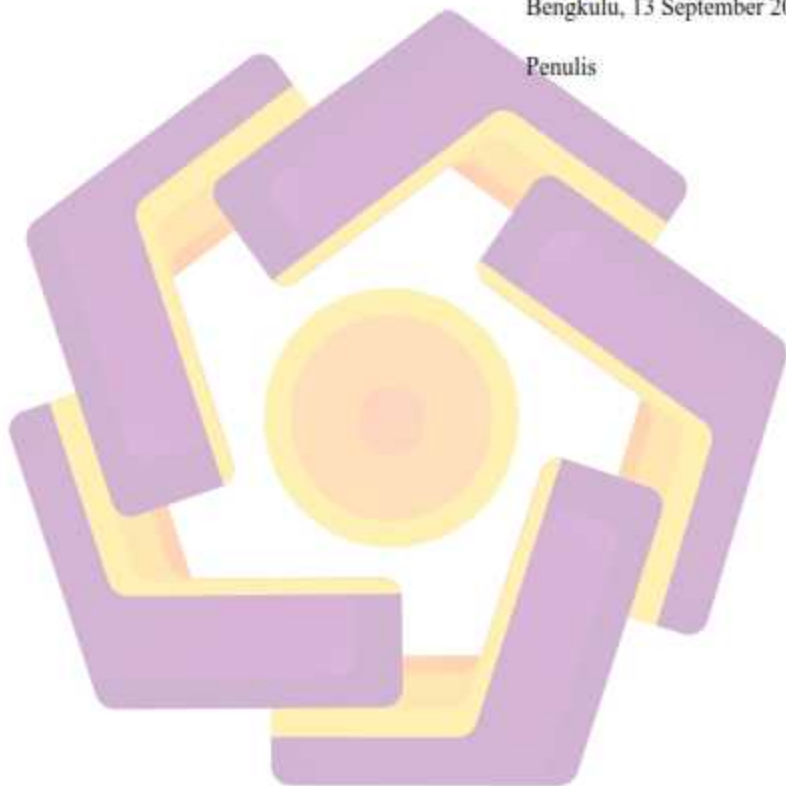
Penulis menyadari terselenggaranya penulisan skripsi ini tak lepas dari bimbingan dan bantuan dari berbagai pihak. Dalam kesempatan ini pula saya mengucapkan terima kasih kepada:

1. Bapak Prof. Dr. M. Suyanto, MM selaku Rektor Universitas Amikom Yogyakarta, yang telah memberikan fasilitas dalam menimba ilmu pengetahuan di Universitas Amikom Yogyakarta.
2. Bapak Hanif Al Fatta, S.Kom., M.Kom. selaku Dekan Fakultas Ilmu Komputer di Universitas Amikom Yogyakarta
3. Ibu Windha Mega Pradnya D, M.Kom. selaku Ketua Program Studi Informatika di Universitas Amikom Yogyakarta.
4. Ibu Dwi Nurani, M.Kom. selaku Dosen Pembimbing yang telah meluangkan waktu serta dengan penuh kesabaran telah memberikan bimbingan dalam penyusunan Skripsi.
5. Bapak Andika Agus Slameto, M.Kom selaku Dosen Penguji satu yang telah banyak memberikan bimbingan, kritik dan sarannya untuk perbaikan skripsi ini.
6. Bapak Ainul Yaqin, M. Kom selaku Dosen Penguji dua yang telah banyak memberikan masukan dan bimbingan dalam skripsi ini
7. Ibunda tercinta yang dengan penuh kesabaran dan pengorbanannya selalu memberikan dorongan, bantuan material maupun non material agar penulis dapat menyelesaikan studi.
8. Semua pihak yang tidak bisa penulis sebutkan satu persatu, terimakasih atas bantuan dan dukungannya.

Penulis menyadari Skripsi ini masih jauh dari sempurna, karena hal tersebut tidak lepas dari kelemahan dan keterbatasan penulis. Penulis berharap agar Skripsi ini berguna sebagai tambahan ilmu pengetahuan serta dapat memberikan manfaat bagi semua pihak dan dijadikan implikasi selanjutnya bagi mahasiswa.

Bengkulu, 13 September 2022

Penulis





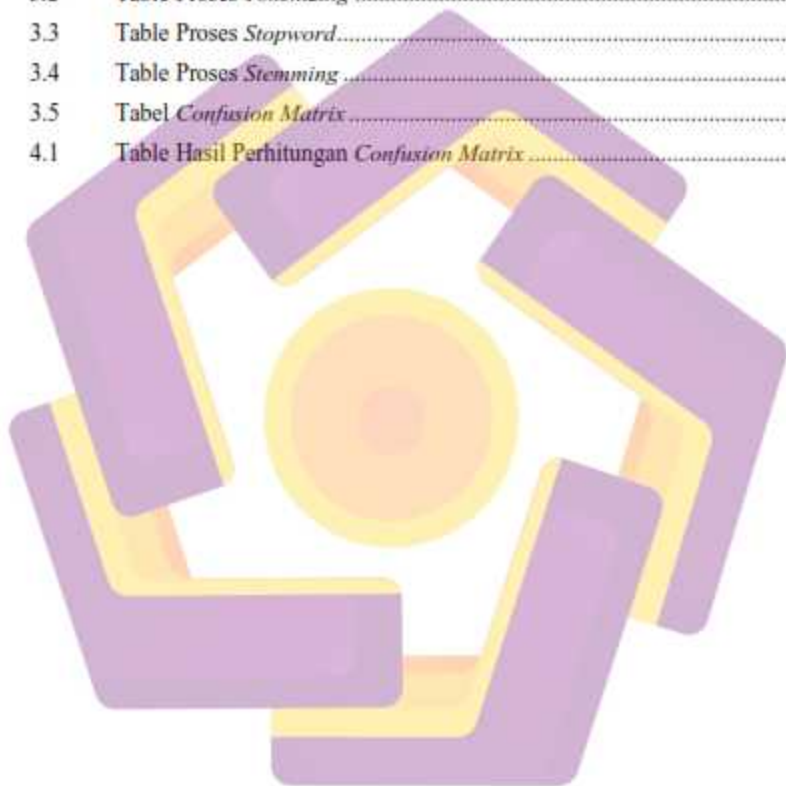
## DAFTAR ISI

HALAMAN JUDUL.....	i
HALAMAN PERSETUJUAN.....	ii
HALAMAN PENGESAHAN.....	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI.....	iv
HALAMAN PERSEMBAHAN.....	v
KATA PENGANTAR.....	vi
DAFTAR ISI.....	viii
DAFTAR TABEL.....	x
DAFTAR GAMBAR.....	xi
DAFTAR LAMPIRAN.....	xii
INTISARI.....	xiii
ABSTRACT.....	xiv
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Perumusan masalah.....	2
1.3 Tujuan Penelitian.....	3
1.4 Batasan Masalah.....	3
1.5 Manfaat Penelitian.....	3
BAB II TINJAUAN PUSTAKA.....	4
2.1 Literature Review.....	4
2.2 Discord.....	6
2.2.1 Daftar Istilah Discord.....	6
2.3 Spam.....	8
2.4 Klasifikasi.....	9
2.5 <i>Text Mining</i> .....	9
2.5.1 <i>Case Folding</i> .....	10
2.5.2 <i>Tokenizing</i> .....	10
2.5.3 <i>Filtering</i> atau <i>Stopword</i> .....	11
2.5.4 <i>Stemming</i> .....	11
2.6 <i>Term Frequency Inverse Document Frequency</i> .....	11

2.7	<i>Naïve Bayes Classifier</i> .....	12
2.8	Metode Pengujian.....	13
<b>BAB III ANALISIS DAN PERANCANGAN</b> .....		16
3.1	Diagram Alur Penelitian.....	16
3.2	Analisa Kebutuhan Sistem .....	17
3.2.1.	Analisis Kebutuhan Fungsional .....	17
3.2.2.	Analisis Kebutuhan Non-Fungsional .....	17
3.3	Pengumpulan Data .....	18
3.4	Teks Preprocessing.....	18
3.4.1	Case Folding.....	19
3.4.2	Tokenizing.....	19
3.4.3	Stopword Removal.....	19
3.4.4	Stemming .....	20
3.5	Pembobotan TF-IDF.....	20
3.6	Proses Klasifikasi .....	20
3.7	Pengujian Model.....	21
<b>BAB IV HASIL DAN PEMBAHASAN</b> .....		23
4.1	Implementasi .....	23
4.1.1	Impelementasi Pengumpulan data.....	23
4.1.2	Pelabelan Data.....	26
4.1.3	Implementasi Text Preprocessing .....	27
4.1.4	Impelementasi Pembobotan TF-IDF.....	28
4.1.5	Naïve Bayes .....	28
4.2	Pengujian .....	29
<b>BAB V KESIMPULAN DAN SARAN</b> .....		31
5.1	Kesimpulan.....	31
5.2	Saran.....	31
DAFTAR PUSTAKA .....		32
DAFTAR LAMPIRAN.....		34

## DAFTAR TABEL

2.1	Tabel Perbandingan Rujukan Penelitian.....	6
2.2	Table contoh confusion matrix .....	13
3.1	Table Proses <i>Case Folding</i> .....	19
3.2	Table Proses <i>Tokenizing</i> .....	19
3.3	Table Proses <i>Stopword</i> .....	20
3.4	Table Proses <i>Stemming</i> .....	20
3.5	Table <i>Confusion Matrix</i> .....	22
4.1	Table Hasil Perhitungan <i>Confusion Matrix</i> .....	30

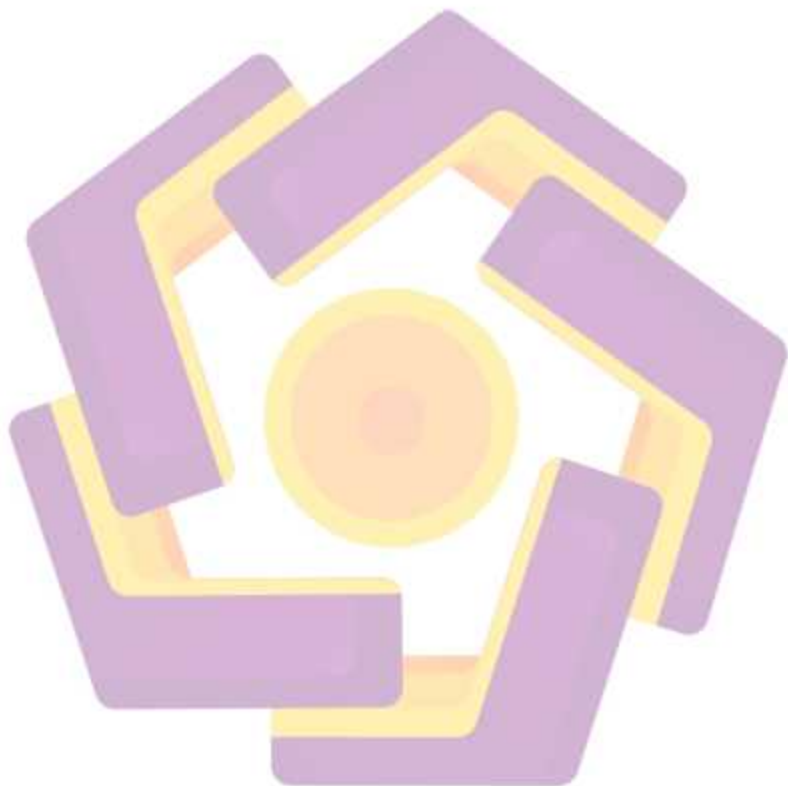


## DAFTAR GAMBAR

3.1	Gambar Alur Penelitian .....	16
3.2	Gambar Proses <i>Preprocessing</i> .....	18
3.3	Gambar Alur TF-IDF .....	20
3.4	Gambar Proses klasifikasi .....	21
4.1	Gambar Aplikasi Discord Chat Exporter .....	24
4.2	Gambar Pemilihan Server .....	24
4.3	Gambar Pemilihan channel .....	25
4.4	Gambar Proses Mendownload Data .....	25
4.5	Gambar Contoh Hasil Pengumpulan Data .....	26
4.6	Gambar Contoh Hasil Pelabelan Data .....	27
4.7	Gambar Proses <i>Preprocessing</i> .....	27
4.8	Gambar Hasil <i>Preprocessing</i> .....	27
4.9	Gambar <i>Script</i> Pembobotan kata .....	28
4.10	Gambar Total Data dan Kata .....	28
4.11	Gambar <i>Script Split Data</i> .....	29
4.12	Gambar <i>Pipeline Model</i> .....	29
4.13	Gambar <i>Script</i> dan Hasil <i>Confusion Matrix</i> .....	30

## DAFTAR LAMPIRAN

1. Lampiran Wordcloud Kelas Spam..... 34



## INTISARI

Dunia informasi saat ini seakan tidak bisa terlepas dari teknologi. Penggunaan teknologi oleh masyarakat menjadikan dunia teknologi semakin lama semakin canggih. Salah satunya dalam bidang teknologi komunikasi seperti adanya smartphone dan internet, membuat manusia semakin meningkatkan cara komunikasinya misalnya melalui media sosial. Kehadiran media sosial memberikan kemudahan bagi manusia untuk berkomunikasi dan bersosialisasi. Salah satu platform media sosial tersebut adalah aplikasi Discord. Discord adalah aplikasi chatting, dan platform distribusi digital yang dirancang untuk menciptakan komunitas yang dapat berjumlah hingga 250.000 pengguna dalam satu server. Didalam sebuah komunitas yang besar terkadang ada salah satu anggota yang membuat kerusuhan misalnya dengan spam baik menggunakan bot atau dengan mengirim spam berulang kali dalam waktu yang bersamaan. Spam dapat berwujud banyak hal, misalnya email spam, iklan spam, *click spam*, *link spam*, berita spam, dan tulisan (komentar) spam. Penelitian ini dilakukan bertujuan untuk mengklasifikasikan pesan spam menggunakan metode algoritma *naive bayes* dan menguji tingkat akurasi dari algoritma *naive bayes*. Data yang diperoleh dan digunakan pada penelitian ini yaitu bersumber dari pesan pada server Discord

**Kata kunci:** Media Sosial, Spam, Discord, Naive Bayes, Klasifikasi

## ABSTRACT

The world of information today seems to be inseparable from technology. The use of technology by society makes the world of technology increasingly sophisticated. One of them is in the field of communication technology such as the existence of smartphones and the internet, making humans increasingly improve their way of communication, for example through social media. The presence of social media makes it easy for humans to communicate and socialize. One such social media platform is the Discord application. Discord is a chat application, and digital distribution platform designed to create a community of up to 250,000 users on a single server. In a large community sometimes there is one member who makes a riot for example by spamming either using bots or by sending spam repeatedly at the same time. Spam can take the form of many things, such as spam emails, spam ads, click spam, spam links, spam news, and spam posts (comments). This research aims to classify spam messages using the naive bayes algorithm and to test the accuracy of the naive bayes algorithm. The data obtained and used in this study are sourced from messages on the Discord server.

**Keyword: Media Social, Spam, Discord, Naive Bayes, Classification**