

**APLIKASI FILTER PESAN MENGGUNAKAN METODE N-
GRAM BERBASIS TELEGRAM BOT**

SKRIPSI

Digunakan Sebagai Syarat Maju Ujian Diploma IV
Politeknik Negeri Malang

Oleh:

WAHYU HIDAYAH NIM. 1741720190



**PROGRAM STUDI TEKNIK INFORMATIKA
JURUSAN TEKNOLOGI INFORMASI
POLITEKNIK NEGERI MALANG
JULI 2021**



HALAMAN PENGESAHAN





APLIKASI FILTER PESAN MENGGUNAKAN METODE N-GRAMS BERBASIS TELEGRAM BOT

Disusun oleh:

WAHYU HIDAYAH **NIM. 1741720190**

Skripsi ini telah diuji pada 21 Juli 2021

Disetujui oleh:

1. Pembimbing I : Gunawan Budi Prasetyo, ST., MMT., Ph.D.
NIP. 197704242008121001 
2. Pembimbing II : Wilda Imama Sabilla, S.Kom., M.Kom
NIP. 199208292019032023 
3. Penguji I : Imam Fahrur Rozi, ST., MT.
NIP. 198406102008121004 
4. Penguji II : Rokhimatul Wakhidah, S.Pd., M.T.
NIP. 198903192019032013 

Mengetahui,

Ketua Jurusan
Teknologi Informasi

Rudy Ariyanto, S.T., M.CS.
NIP. 1971111101999031002

Ketua Program Studi
Teknik Informatika



Imam Fahrur Rozi, ST., MT.
NIP. 198406102008121004

PERNYATAAN

Dengan ini saya menyatakan bahwa pada Skripsi ini tidak terdapat karya, baik seluruh maupun sebagian, yang sudah pernah diajukan untuk memperoleh gelar akademik di Perguruan Tinggi manapun, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis disitasi dalam naskah ini serta disebutkan dalam daftar sitasi/pustaka.

Malang, 21 Juli 2021



Wahyu Hidayah

ABSTRAK

Hidayah., Wahyu. “Aplikasi Filter Pesan Menggunakan Metode *N-Gram* berbasis *Telegram Bot*”. Pembimbing: (1) Gunawan Budi Prasetyo, S.T., MMT., Ph.D., (2) Wilda Imama Sabilla, S.Kom., M.Kom.

Skripsi, Program Studi Teknik Informatika, Jurusan Teknologi Informasi, Politeknik Negeri Malang, 2021.

Masyarakat sekarang hampir mustahil tidak tersentuh dengan media seperti media sosial dan *instant messaging*. Media sosial membuka peluang bagi setiap orang yang terlibat, mulai dari anak kecil hingga orang tua untuk mengekspresikan pendapatnya secara bebas, maka peningkatan berbagai konsekuensinya juga harus diperhatikan. Tidak sedikit orang bertutur kata buruk saat berkomunikasi di media sosial yang dapat memberikan dampak negatif bagi semua kalangan. Salah satu akibat dari dampak negatif tersebut adalah munculnya kasus *cyberbullying* dan *miss communication* melalui media chat atau *instant messaging*. Dalam upaya mengurangi dampak negatif dari permasalahan di atas, maka dibangun sebuah sistem *filter* pesan dengan menggunakan *n-gram* berbasis *telegram bot*. Pada penelitian ini, model bahasa *n-gram* yang digunakan berupa *bigram* dan *trigram* yang digunakan untuk memecah sebuah kata menjadi beberapa potongan *n-gram* sesuai dengan model bahasanya. Agar model bahasa *n-gram* ini bisa menerima pesan dari *group chat*, diterapkan model bahasa *n-gram* ini ke dalam *telegram bot*. Penelitian ini menggunakan 1000 data *tweet* yang dibagi menjadi 70% data pelatihan dan 30% data pengujian. Tingkat akurasi tertinggi yang didapat dari kedua model bahasa *n-gram* sebesar 91% untuk skenario *bigram* dengan *threshold* 30% dan 89% untuk skenario *trigram* dengan *threshold* 60%.

Kata Kunci: pesan, *N-Gram*, Telegram Bot

ABSTRACT

Hidayah., Wahyu. “*Message Filter Application Using N-Gram Method based on Telegram Bot*”. Supervisors: (1) **Gunawan Budi Prasetyo, S.T., MMT., Ph.D.**, (2) **Wilda Imama Sabilla, S.Kom., M.Kom.**

Thesis, Informatics Engineering Study Program, Department of Information Technology, State Polytechnic of Malang, 2021.

Today's society is almost impossible not to be touched by media such as social media and instant messaging. Social media opens up opportunities for everyone involved, from young children to the elderly to express their opinions freely, so the increased consequences must also be considered. However, a few people speak bad words when communicating in social media which can have a negative impact on all circles. One of the consequences of this negative impact is the emergence of cases of cyberbullying and miss communication through chat media or instant messaging. In an effort to reduce the negative impact of the above problems, a message filter system was built using an n-gram based telegram bot. In this study, the n-gram language model used were bigram and trigram, which were used to break a word into several n-gram pieces according to the language model. To make this n-gram language model able to receive messages from group chats, this n-gram language model was applied to the telegram bot. This study used 1000 tweet data which was divided into 70% training data and 30% test data. The highest level of accuracy obtained from the two n-gram language models was 91% for the bigram scenario with a 30% threshold and 89% for the trigram scenario with a 60% threshold.

Keywords: *message, N-Gram, Telegram Bot*

KATA PENGANTAR

Puji Syukur kami panjatkan kehadirat Allah SWT/Tuhan YME atas segala rahmat dan hidayah-Nya penulis dapat menyelesaikan skripsi dengan judul “APLIKASI FILTER PESAN MENGGUNAKAN METODE N-GRAM BERBASIS TELEGRAM BOT”. Skripsi ini penulis susun sebagai persyaratan untuk menyelesaikan studi program Diploma IV Program Studi Teknik Informatika, Jurusan Teknologi Informasi, Politeknik Negeri Malang.

Kami menyadari bahwasannya dengan tanpa adanya dukungan dan kerja sama dari berbagai pihak, kegiatan laporan akhir ini tidak akan dapat berjalan baik. Untuk itu, kami ingin menyampaikan rasa terima kasih kepada:

1. Allah SWT yang senantiasa memberikan rahmat dan hidayah-Nya sehingga penulis dapat menyelesaikan skripsi ini dengan lancar.
2. Kedua almarhum orang tua saya yang sebelumnya selalu memberikan dukungan moral, materil, motivasi dan doa agar saya dimudahkan dalam menghadapi kehidupan ini termasuk skripsi ini dan menjadi anak yang sholeh serta dapat membahagiakan mereka di alam sana.
3. Bapak Gunawan Budi Prasetyo, S.T., MMT., Ph.D., selaku Pembimbing 1 yang telah memberi bimbingan, arahan, dan meluangkan waktu kepada penulis dalam menyusun skripsi ini dengan baik.
4. Ibu Wilda Imama Sabilla, S.Kom., M.Kom., selaku Pembimbing 2 yang telah memberi bimbingan, arahan, dan meluangkan waktu kepada penulis dalam menyusun skripsi ini dengan baik.
5. Bapak Rudy Ariyanto, ST., M.Cs., selaku Ketua Jurusan Teknologi Informasi
6. Bapak Imam Fahrur Rozi, ST., MT., selaku Ketua Program Studi DIV Teknik Informatika
7. Dan seluruh pihak yang telah membantu dan mendukung lancarnya pembuatan Laporan Akhir dari awal hingga akhir yang tidak dapat kami sebutkan satu persatu.

Penulis menyadari bahwa dalam penyusunan laporan akhir ini, masih banyak terdapat kekurangan dan kelemahan yang dimiliki penulis baik itu sistematika penulisan maupun penggunaan bahasa. Untuk itu penulis mengharapkan saran dan

kritik dari berbagai pihak yang bersifat membangun demi penyempurnaan laporan ini. Semoga laporan ini berguna bagi pembaca secara umum dan penulis secara khusus. Akhir kata, penulis ucapkan banyak terima kasih.

Malang, 21 Juli 2021

Penulis

DAFTAR ISI

HALAMAN PENGESAHAN.....	ii
PERNYATAAN.....	iii
ABSTRAK	iv
<i>ABSTRACT</i>	v
KATA PENGANTAR	vi
DAFTAR ISI.....	vii
DAFTAR GAMBAR	ix
DAFTAR TABEL.....	x
DAFTAR LAMPIRAN.....	xi
BAB I. PENDAHULUAN	12
1.1 Latar Belakang.....	12
1.2 Rumusan Masalah.....	13
1.3 Batasan Masalah.....	13
1.4 Tujuan.....	14
1.5 Manfaat.....	14
BAB II. LANDASAN TEORI	15
2.1 Studi Literatur.....	15
2.2 Dasar Teori	17
2.2.1. Pesan	17
2.2.2. Kata Baku & Tidak Baku.....	17
2.2.3. <i>Python</i>	18
2.2.4. <i>Text Mining</i>	18
2.2.5. Telegram Bot.....	19
2.2.6. Telegram <i>Bot API</i>	19
2.2.7. <i>N-gram</i>	20
2.2.8. <i>Confusion Matrix</i>	20
BAB III. METODOLOGI PENELITIAN	22
3.1 Waktu dan Tempat Penelitian.....	22
3.2 Teknik Pengumpulan Data	22
3.3. Teknik Pengolahan Data.....	22
3.4. Uji Coba Sistem.....	26
BAB IV. ANALISIS DAN PERANCANGAN SISTEM.....	28
4.1 Deskripsi Sistem.....	28
4.2 Analisa Aktor.....	28
4.3 Analisa Kebutuhan Fungsional.....	28
4.4 Analisa Kebutuhan Non Fungsional.....	29
4.4.1. Kebutuhan Perangkat Lunak	29

4.4.2. Kebutuhan Perangkat Keras	29
4.5 Perancangan Sistem.....	29
4.5.1. Diagram <i>Use Case</i>	29
4.5.2. Arsitektur Sistem.....	31
4.5.3. <i>Activity Diagram</i> Filter Pesan	32
4.5.4. Database	34
4.5.4.1. Struktur Tabel Database.....	34
4.5.5. Desain <i>Interface</i>	35
BAB V. IMPLEMENTASI DAN PENGUJIAN.....	42
5.1 Implementasi Sistem.....	42
5.1.1 <i>Pre-Processing Text</i>	42
5.1.2 Pemecahan Bentuk <i>N-Gram</i>	42
5.1.3 Pemodelan <i>N-Gram</i> dan Perhitungan Probabilitas Kata.....	42
5.1.4 Perhitungan Persamaan <i>N-Gram</i>	43
5.2 Implementasi <i>Database</i>	43
5.2.1 Implementasi Tabel abusive.....	43
5.2.2 Implementasi Tabel slangword	43
5.2.3 Implementasi Tabel dataset.....	43
5.3 Implementasi Desain <i>Interface</i>	44
5.3.1 Implementasi Tampilan Halaman Utama.....	44
5.3.2 Implementasi Tampilan Halaman Kamus Kata Kasar	45
5.3.3 Implementasi Tampilan Halaman Kamus Kata Alay.....	45
5.3.4 Implementasi Tampilan Halaman Dataset <i>Tweet</i>	46
5.3.5 Implementasi Tampilan Halaman <i>Testing</i>	46
5.3.6 Implementasi Tampilan Halaman <i>Result Testing</i>	47
5.4 Implementasi Bot pada <i>Group Chat</i>	47
5.5 Hasil Pengujian Sistem.....	48
5.4.1 Pengujian Fungsionalitas	48
5.4.2 Pengujian Metode.....	55
BAB VI. HASIL DAN PEMBAHASAN	59
6.1 Hasil.....	59
6.2 Pembahasan	59
BAB VII. KESIMPULAN DAN SARAN	61
7.1 Kesimpulan.....	61
7.2 Saran	61
DAFTAR PUSTAKA	62
LAMPIRAN – LAMPIRAN	64

DAFTAR GAMBAR

Gambar 3.1 Skenario pengolahan data.....	23
Gambar 3.2 <i>Text-Preprocessing</i>	23
Gambar 4.1 <i>Use Case</i> Sistem.....	30
Gambar 4.2 Arsitektur Diagram.....	31
Gambar 4.3 <i>Activity Diagram</i> Filter Pesan	32
Gambar 4.4 Proses pada kalkulasi persamaan <i>N-Gram</i>	33
Gambar 4.5 Tampilan Halaman Utama	36
Gambar 4.6 Tampilan Halaman Kamus Kata Kasar (<i>Abusive</i>).....	37
Gambar 4.7 Tampilan Halaman Kamus Kata Alay (<i>Slangword</i>)	38
Gambar 4.8 Tampilan Halaman Dataset	39
Gambar 4.9 Tampilan Halaman <i>Testing</i>	40
Gambar 4.10 Tampilan Halaman <i>Result Testing</i>	41
Gambar 5.1 Implementasi Tabel abusive.....	43
Gambar 5.2 Implementasi Tabel slangword	43
Gambar 5.3 Implementasi Tabel dataset.....	44
Gambar 5.4 Implementasi Tabel message	44
Gambar 5.5 Implementasi Tampilan Admin Halaman Utama.....	44
Gambar 5.6 Implementasi Tampilan Admin Halaman Kamus Kata Kasar	45
Gambar 5.7 Implementasi Tampilan Admin Halaman Kamus Kata Alay.....	45
Gambar 5.8 Implementasi Tampilan Admin Halaman Dataset <i>Tweet</i>	46
Gambar 5.9 Implementasi Tampilan Admin Halaman <i>Testing</i>	46
Gambar 5.10 Implementasi Tampilan Admin Halaman <i>Result Testing</i>	47
Gambar 5.11 Pilih Group Chat.....	47
Gambar 5.12 Undang Bot ke Group Chat.....	47
Gambar 5.13 Masuk ke Konfigurasi Admin Group.....	48
Gambar 5.14 Menambah Admin Group.....	48

DAFTAR TABEL

Tabel 2.1 Studi Literatur	15
Tabel 2.2 Contoh <i>N-gram</i> dari kata “PALSU”	20
Tabel 2.3 Tabel <i>Confusion Matrix</i>	20
Tabel 3.1 Proses <i>Case Folding</i>	23
Tabel 3.2 Proses <i>Tokenizing</i>	24
Tabel 3.3 Proses <i>Replacing</i>	24
Tabel 3.4 Proses <i>Filtering</i>	24
Tabel 3.5 Proses <i>Stemming</i>	25
Tabel 3.6 <i>Confusion Matrix</i> pada Uji Coba Sistem	27
Tabel 4.1 Kebutuhan Fungsional	28
Tabel 4.2 Kebutuhan Perangkat Lunak.....	29
Tabel 4.3 Kebutuhan Perangkat Keras.....	29
Tabel 4.4 Deskripsi Tabel abusive	34
Tabel 4.5 Deskripsi Tabel slangword.....	34
Tabel 4.6 Deskripsi Tabel dataset	35
Tabel 4.7 Deskripsi Tabel message.....	35
Tabel 5.1 Pengujian Fungsional Halaman Utama.....	48
Tabel 5.2 Pengujian Fungsional Halaman Kamus Kata Kasar (<i>Abusive</i>).....	49
Tabel 5.3 Pengujian Fungsional Halaman Kamus Kata Alay (<i>Slangword</i>).....	50
Tabel 5.4 Pengujian Fungsional Halaman Dataset <i>Tweet</i>	52
Tabel 5.5 Pengujian Fungsional Halaman <i>Testing</i>	52
Tabel 5.6 Pengujian Fungsional Halaman <i>Result Testing</i>	53
Tabel 5.7 Pengujian Fungsional Bot Telegram dalam <i>Private Chat</i>	54
Tabel 5.8 Pengujian Fungsional Bot Telegram dalam <i>Group Chat</i>	54
Tabel 5.9 Pengujian Skenario <i>Bigram</i>	55
Tabel 5.10 Pengujian Skenario <i>Trigram</i>	56
Tabel 5.11 Hasil Prediksi Skenario <i>Bigram Threshold 30%</i>	56
Tabel 5.12 Hasil <i>Confusion Matrix</i> dari pengujian Tabel 5.11	57
Tabel 5.13 Penentuan <i>True Positif, True Negatif, False Positif, dan False Negatif</i> dari Uji Coba Sistem.....	57
Tabel 5.14 Hasil Perhitungan <i>Precision, Recall, F1-Score, Accuracy</i> dari Uji Coba Sistem	57

DAFTAR LAMPIRAN

Lampiran 1. <i>Source Code Preprocessing Text</i>	64
Lampiran 2. <i>Source Code</i> Pemecahan Bentuk <i>N-Gram</i>	65
Lampiran 3. <i>Source Code</i> Pemodelan <i>N-Gram</i> Perhitungan Probabilitas Kata...	66
Lampiran 4. <i>Source Code</i> Perhitungan Persamaan <i>N-Gram</i>	66
Lampiran 5. Contoh Hasil Uji Coba Bot pada Group Chat	67
Lampiran 6. Form Verifikasi Abstrak dan Tata Cara Penulisan Laporan Skripsi	68