

**RANCANG BANGUN APLIKASI PEMBERSIH KOMENTAR
SPAM PROMOSI JUDI ONLINE DI YOUTUBE DENGAN
PENDEKATAN *HYBRID: RULE-BASED FILTER* DAN
KLASIFIKASI *NAIVE BAYES*.**

SKRIPSI

Untuk Memenuhi Persyaratan Memperoleh Gelar Sarjana (S1) Pada
Program Studi Teknologi Informasi Fakultas Sains dan Teknologi
Universitas Labuhanbatu



OLEH :

**MUHAMMAD DIMAS RIVALDO
2108100083**

PROGRAM STUDI TEKNOLOGI INFORMASI

FAKULTAS SAINS DAN TEKNOLOGI

UNIVERSITAS LABUHANBATU

RANTAUPRAPAT

2025

PERNYATAAN

Judul Skripsi : RANCANG BANGUN APLIKASI PEMBERSIH KOMENTAR
SPAM PROMOSI JUDI ONLINE DI YOUTUBE DENGAN
PENDEKATAN *HYBRID*: RULE-BASED FILTER DAN
KLASIFIKASI *NAIVE BAYES*.

Nama : Muhammad Dimas Rivaldo

NPM : 2108100083

Prodi : Teknologi Informasi

Dengan ini, saya menyatakan bahwa skripsi ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana pada Program Studi Teknologi Informasi, Fakultas Sains dan Teknologi, Universitas Labuhanbatu, merupakan hasil karya saya sendiri. Segala kutipan dan referensi yang digunakan dalam penulisan ini telah dicantumkan sesuai dengan ketentuan yang berlaku. Apabila di kemudian hari terbukti bahwa sebagian atau seluruh isi skripsi ini bukan merupakan hasil karya asli saya atau mengandung unsur plagiarisme, maka saya bersedia menerima sanksi berupa pencabutan gelar akademik serta sanksi lain sesuai dengan peraturan perundang-undangan yang berlaku.

Rantauprapat, 19 Agustus 2025

Yang Membuat Pernyataan



MUHAMMAD DIMAS RIVALDO

NPM : 2108100083

LEMBAR PERSETUJUAN SKRIPSI

Judul Skripsi : RANCANG BANGUN APLIKASI PEMBERSIH KOMENTAR
SPAM PROMOSI JUDI ONLINE DI YOUTUBE DENGAN
PENDEKATAN *HYBRID*: RULE-BASED FILTER DAN
KLASIFIKASI *NAIVE BAYES*.

Nama : Muhammad Dimas Rivaldo

NPM : 2108100083

Prodi : Teknologi Informasi

Disetujui pada tanggal : 19 Agustus 2025

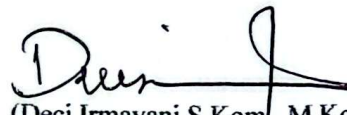
Pembimbing I

Pembimbing II



(Sahat Rarulian Sitorus, S.T., M.Kom.)

NIDN : 0124018703



(Deci Irmayani S.Kom., M.Kom.)

NIDN : 0127058602

LEMBAR PERSETUJUAN SKRIPSI

Judul Skripsi : RANCANG BANGUN APLIKASI PEMBERSIH KOMENTAR
SPAM PROMOSI JUDI ONLINE DI YOUTUBE DENGAN
PENDEKATAN *HYBRID*: RULE-BASED FILTER DAN
KLASIFIKASI *NAIVE BAYES*.

Nama : Muhammad Dimas Rivaldo
NPM : 2108100083
Prodi : Teknologi Informasi

Telah Diuji Dan Dinyatakan Lulus Dalam Ujian Sarjana
Pada Tanggal 19 Agustus 2025.

TIM PENGUJI Pembimbing I

Nama : Sahat Parulian Sitorus, S.T., M.Kom.
NIDN : 0124018703

Tanda Tangan



Pembimbing II

Nama : Deci Irmayani S.Kom., M.Kom.
NIDN : 0124018703



Penguji

Nama : Dr. Iwan Purnama, S.Kom., M.Kom
NIDN : 0112029202

Rantauprapat, 19 Agustus 2025.
Dekan
Fakultas Sains dan Teknologi



(Dr. Iwan Purnama, S.Kom., M.Kom)
NIDN : 0112029202

Ka. Program Studi
Teknologi Informasi



(Rahmadani Pane, S.Kom., M.Kom.)
NIDN : 0113028702

ABSTRAK

Perkembangan platform Youtube sebagai tempat berbagi video dan juga merupakan tempat yang menjanjikan untuk memperoleh penghasilan ini membawa dampak besar dalam komunikasi digital. Pengguna bisa dengan bebas mengakses video yang ingin ditonton, kolom komentar untuk tempat diskusi maupun mengapresiasi. Namun, kebebasan ini juga disertai dengan maraknya komentar *spam*, khususnya yang bermuatan promosi situs judi online. Komentar semacam ini tidak hanya mengganggu pengalaman pengguna, tetapi juga dapat merusak reputasi kanal dan menyalahi kebijakan platform. Sebagai salah satu cara untuk menghambat situs-situs judi online berpromosi penelitian ini bertujuan untuk merancang dan membangun sebuah aplikasi pembersih komentar *spam* bermuatan judi online di YouTube dengan pendekatan *hybrid*, yaitu menggabungkan metode *rule-based filter* dan algoritma klasifikasi *Naive bayes*.

Metode *rule-based* digunakan untuk menyaring komentar berdasarkan daftar kata kunci sensitif yang umum digunakan, simbol-simbol yang mirip dengan huruf alphabet, juga emoji-emoji khusus yang menunjukkan promosi situs judi online. Sementara itu, pendekatan *Naive bayes* dimanfaatkan untuk mengenali pola linguistik yang tidak secara eksplisit mencantumkan kata kunci, namun memiliki kecenderungan sebagai *spam*. Aplikasi ini dikembangkan menggunakan teknologi Node.js dengan antarmuka berbasis web (EJS), serta diintegrasikan dengan YouTube Data API untuk mengambil dan mengelola komentar secara otomatis.

Hasil pengujian menunjukkan bahwa pendekatan *hybrid* ini mampu meningkatkan akurasi deteksi *spam* dibandingkan penggunaan metode tunggal atau menyembunyikan *spam* komen satu per satu. Selain itu, aplikasi juga dilengkapi fitur ekspor data hasil deteksi ke dalam format Excel sebagai dokumentasi analisis. Dengan demikian, aplikasi ini dapat menjadi solusi praktis dan adaptif dalam membantu konten kreator menjaga kualitas interaksi di kanal YouTube mereka dari komentar *spam* bermuatan situs judi online.

Kata kunci: YouTube, *spam* komentar, judi online, *rule-based filter*, *Naive bayes*, klasifikasi teks, aplikasi.

ABSTRACT

The development of the YouTube platform as a place to share videos and also a promising place to earn income has had a big impact on digital communication. . Users can freely access the videos they want to watch, the comment column for discussion or appreciation. However, this freedom is also accompanied by the rise of *spam* comments, especially those containing promotions for online gambling sites. Such comments not only disrupt the user experience, but can also damage the channel's reputation and violate platform policies. As one way to prevent online gambling sites from promoting, this study aims to design and build an application to clean online gambling *spam* comments on YouTube with a *hybrid* approach, namely combining the *rule-based* filter method and the *Naive bayes* classification algorithm.

The *rule-based* method is used to filter comments based on a list of commonly used sensitive keywords, symbols that resemble letters of the alphabet, and special emojis that indicate online gambling site promotions. Meanwhile, the *Naive bayes* approach is used to recognize linguistic patterns that do not explicitly include keywords but have a tendency to be *spam*. This application is developed using Node.js technology with a web-based interface (EJS), and is integrated with the YouTube Data API to automatically retrieve and manage comments.

The test results show that this *hybrid* approach is able to improve the accuracy of *spam* detection compared to using a single method or hiding *spam* comments one by one. In addition, the application is also equipped with a feature to export detection data into Excel format as analysis documentation. Thus, this application can be a practical and adaptive solution in helping content creators maintain the quality of interaction on their YouTube channels from *spam* comments containing online gambling sites.

Keywords: YouTube, comment *spam*, online gambling, *rule-based* filter, *Naive bayes*, text classification, application.

KATA PENGANTAR

Terlebih dahulu, penulis mengucapkan Alhamdulillahirobbil'alamin, Puji serta Syukur Kehadirat Allah SWT. yang selalu memberikan Rahmat dan HidayahNya sehingga penulisan tugas akhir ini dapat diselesaikan. Dan Shalawat beserta salam teruntuk Baginda Nabi Muhammad SAW. yang telah membawa manusia dari alam kebodohan kepada alam yang penuh dengan ilmu pengetahuan.

Skripsi ini disusun sebagai langkah awal dalam penelitian yang bertujuan untuk membangun aplikasi pembersih komentar *spam* promosi judi online menggunakan pendeteksi otomatis *Rule-based* filter dan klasifikasi *Naïve bayes*. Teknologi ini diharapkan dapat membersihkan situs-situs judi online untuk berpromosi.

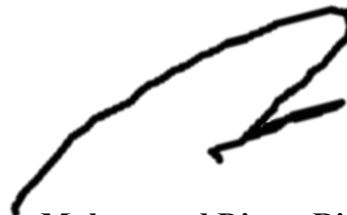
Selama proses penyusunan Skripsi ini, penulis telah menerima bimbingan dan dukungan dari berbagai pihak. Oleh karena itu, penulis mengucapkan terima kasih yang sebesar-besarnya kepada:

1. Bapak Halomoan Nasution, S.H., M.H selaku ketua Yayasan Universitas Labuhanbatu
2. Bapak Assoc, Prof. Ade Parlaungan Nasution, SE., M.Si., Ph.D. selaku Rektor Universitas Labuhanbatu.
3. Bapak Dr. Iwan Purnama, S.Kom., M.Kom. selaku Dekan Fakultas Sains dan Teknologi , Universitas Labuhanbatu.
4. Ibu Rahmadani Pane, S.Kom., M.Kom. selaku Ketua Program Studi Teknologi Informasi, Fakultas Sains dan Teknologi, Universitas Labuhanbatu.
5. Bapak Sahat Parulian Sitorus, S.T., M.Kom. selaku Dosen Pembimbing I.
6. Ibu Deci Irmayani S.Kom., M.Kom. selaku Dosen Pembimbing II.
7. Orang tua, Keluarga serta teman-teman yang selalu memberikan dukungan dan semangat.

Penulis menyadari bahwa skripsi ini masih jauh dari kesempurnaan. Oleh karena itu, penulis sangat mengharapkan kritik dan saran yang membangun demi penyempurnaan penelitian ini.

Semoga skripsi ini dapat menjadi langkah awal yang baik untuk pelaksanaan penelitian dan memberikan kontribusi nyata bagi dunia akademik serta industri.

Rantauprapat, 19 Agustus 2025



Muhammad Dimas Rivaldo
2108100083

DAFTAR ISI

ABSTRAK	i
ABSTRACT	ii
KATA PENGANTAR.....	iii
DAFTAR ISI	iv
DAFTAR TABEL.....	viii
BAB I PENDAHULUAN	1
1.1 Latar Belakang Masalah.....	1
1.2 Perumusan Masalah.....	5
1.3 Batasan Masalah.....	5
1.4 Tujuan Penelitian.....	6
1.5 Manfaat Penelitian.....	7
BAB II LANDASAN TEORI.....	8
2.1 YouTube Sebagai Media Sosial.....	8
2.2 Komentar <i>Spam</i> di YouTube.....	10
2.3 Judi Online Sebagai Ancaman Nasional	12
2.3.1 Ancaman Terhadap Kesehatan Mental.....	12
2.3.2 Dampak Sosial	14
2.3.3 Dampak Ekonomi	15
2.3.4 Ancaman Kejahatan Siber.....	16
2.3.5 Ancaman Hukum	18
2.3.6 Modus Judi Online	19
2.4 Deteksi <i>Spam</i> Secara Otomatis	21
2.4.1 Pengertian <i>Spam</i>	22
2.4.2 Cara Kerja Sistem Deteksi <i>Spam</i>	22
2.4.3 Faktor Utama dalam Klasifikasi <i>Spam</i>	23
2.4.4 Kelebihan dan Kekurangan Pendekatan Otomatis.....	25
2.5 Sistem Berbasis Aturan (<i>Rule-based Filter</i>)	26
2.5.1 Daftar Kata Terlarang (Bloked Words)	26
2.5.2 <i>Normalization Check</i> dalam <i>Rule-based Filter</i>	26
2.6 Klasifikasi Teks dan <i>Naive bayes</i>	29

2.6.1	Prinsip Kerja <i>Naive bayes</i> dalam Klasifikasi Teks	29
2.6.2	Jenis-Jenis <i>Naive bayes</i>	30
2.6.3	Proses Pelatihan dan Pengklasifikasian	31
2.6.4	Kelebihan dan Kelemahan <i>Naive bayes</i>	33
2.6.5	Relevansi <i>Naive bayes</i> dengan Penelitian Ini	33
2.7	Pendekatan <i>Hybrid</i> dalam Deteksi <i>Spam</i>	34
2.7.1	Implementasi Umum Pendekatan <i>Hybrid</i>	35
2.7.2	Keunggulan Pendekatan <i>Hybrid</i>	36
2.7.3	Tantangan Pendekatan <i>Hybrid</i>	36
2.7.4	Relevansi Pendekatan <i>Hybrid</i> dengan Penelitian Ini	37
2.8	Integrasi YouTube API	37
2.8.1	Tujuan Integrasi API dalam Sistem Deteksi <i>Spam</i>	38
2.8.2	Jenis API yang Digunakan	39
2.8.3	Tahapan Integrasi YouTube API	39
2.8.4	Biaya Penggunaan YouTube Data API.....	40
2.8.5	Manfaat Integrasi API	43
2.8.6	Tantangan Integrasi API.....	44
2.9	Ekspor Data Menggunakan ExcelJS	44
2.9.1	Tujuan Ekspor Data ke Excel.....	45
2.9.2	Struktur Data yang Diekspor.....	45
2.9.3	Langkah-Langkah Penggunaan ExcelJS.....	45
2.9.4	Keunggulan Menggunakan ExcelJS	46
2.9.5	Kendala dan Tantangan ExelJS	47
2.10	Penelitian sebelumnya.....	47
BAB III	METODE PENELITIAN	50
3.1	Prosedur Penelitian.....	50
3.2	Arsitektur Sistem.....	54
3.2.1	Komponen Utama Sistem	55
3.2.2	Alur Kerja Sistem Aplikasi Penghapus Komentar Spam.....	56
3.2.3	Pemilihan Pendekatan <i>Hybrid</i>	59
3.2.4	Teknologi yang Digunakan	59
3.3	Jenis dan Pendekatan Penelitian.....	60

3.4	Lokasi dan Waktu Penelitian	62
3.5	Metode Pengumpulan Data	63
3.5.1	Pengambilan Komentar dari Vidio YouTube	63
3.5.2	Pelabelan Komentar <i>Spam</i>	64
3.5.3	Konversi Data Exel ke Format JSON	64
3.5.4	Pelatihan Model Klasifikasi <i>Naive bayes</i>	65
3.5.5	Pengumpulan Data Uji dari Channel Pribadi.....	65
3.6	Alat dan Bahan	66
3.6.1	Perangkat Keras (Hardware).....	66
3.6.2	Perangkat Lunak (Software) dan Library.....	67
3.6.3	Dataset.....	68
3.7	Metode Evaluasi	69
3.7.1	Contoh Penghitungan Evaluasi	70
BAB IV	HASIL DAN PEMBAHASAN	72
4.1	Implementasi Sistem	72
4.1.1	Instalasi dan Konfigurasi Sistem.....	73
4.1.2	Lingkungan Pengembangan.....	75
4.1.3	Arsitektur Sistem	76
4.1.4	Implementasi Modul Frontend Web.....	79
4.1.5	Implementasi Modul Beckend	81
4.1.6	Implementasi YouTube Data API v3.....	81
4.1.7	Implementasi Rule-Based Filter	84
4.1.8	Implementasi Klasifikasi Komentar dengan Naive Bayes.....	85
4.1.9	Implementasi Ekspor Hasil Scan ke Excel	86
4.2	Pengujian Sistem	88
4.2.1	Pengujian Ambil Komentar dari Vidio Youtube	89
4.2.2	Pengujian Deteksi Spam dengan Blocked Words	89
4.2.3	Pengujian Deteksi fancy text/symbol aneh.....	90
4.2.4	Pengujian Klasifikasi Naïve Bayes	90
4.2.5	Pengujian Hide Komentar spam	90
4.2.6	Pengujian Ekpor Hasil Ke Excel.....	92
4.3	Pembahasan Hasil Pengujian	92

4.3.1	Efektivitas Rule-Based Filter	92
4.3.2	Akurasi Klasifikasi Naïve Bayes	93
4.3.3	Keunggulan Pendekatan Hybrid	94
4.3.4	Efektivitas Fitur Penyembunyian Komentar.....	94
4.3.5	Evaluasi Fitur Ekspor ke Excel.....	94
4.4	Simpulan Implementasi dan Pengujian Sistem	94
BAB V	KESIMPULAN DAN SARAN	96
5.1	Kesimpulan.....	96
5.2	Saran.....	97
	DAFTAR PUSTAKA	98
	LAMPIRAN-LAMPIRAN.....	100
	Lampiran 1. Lampiran isi kodingan file Index.js.....	100
	Lampiran 2. Lampiran isi kodingan file Index.ejs	105
	Lampiran 3. Lampiran isi kodingan file GetToken.js	107
	Lampiran 4. Lampiran isi kodingan file NaiveBayes.js.....	108
	Lampiran 5. Lampiran isi kodingan file test_naive_model.js.....	109

DAFTAR TABEL

Tabel 2.1 Fitur YouTube.....	9
Tabel 2.2 Jenis-Jenis <i>Spam</i>	10
Tabel 2.3 Kesehatan Mental.....	13
Tabel 2.4 Dampak Sosial.....	14
Tabel 2.5 Dampak Ekonomi.....	15
Tabel 2.6 Ancaman Kejahatan Siber	16
Tabel 2.7 Ancaman Hukum.....	18
Tabel 2.8 Modus Judi Online	19
Tabel 2.9 Kelebihan dan Kekurangan Pendekatan Otomatis	25
Tabel 2.10 Langkah Utama Normalisasi Teks.....	27
Tabel 2.11 Kelebihan dan kelemahan Naive bayes	33
Tabel 2.12 Integrasi YouTube API	38
Tabel 2.13 Biaya (unit) YouTube Data API v3.....	41
Tabel 2.14 Penelitian Sebelumnya	47
Tabel 3. 1 Data Komentar dari Vidio	52
Tabel 3.2 Komponen Utama Sistem.....	55
Tabel 3.3 Teknologi yang Digunakan.....	59
Tabel 3.4 Ringkasan Pendekatan Penelitian.....	61
Tabel 3.5 Jadwal Penelitian.....	62
Tabel 3.6 Hardware yang Digunakan.....	66
Tabel 3.7 Perangkat Lunak dan Library	67
Tabel 3.8 Confusion Matrix	69
Tabel 4.1 Spesifikasi Lingkungan Pengembangan	76
Tabel 4.2 Biaya Kuota YouTube API	83

DAFTAR GAMBAR

Gambar 2.1 Proses Deteksi <i>Spam</i> Otomatis.....	23
Gambar 2. 2 Normalisasi Teks	28
Gambar 2.3 YouTube Data API v3.....	39
Gambal 3.1 Prosedur Penelitian.....	50
Gambal 3.2 Alur Kerja Sistem Aplikasi.....	56
Gambal 3 3 Format Isi File JSON.....	64
Gambal 3. 4 Hasil Uji Contoh Metrik Evaluasi	71
Gambar 4.1 Struktur Folder	72
Gambar 4.2 Kodingan File"GetToken.js"	74
Gambar 4.3 Kode Akses Token.....	75
Gambar 4.4 Tampilan Terminal GetToken.js	75
Gambar 4.5 Diagram Arsitektur Sistem	77
Gambar 4.6 Tampilan Form Input.....	79
Gambar 4.7 Tampilan Hasil Pemindai Komentar	80
Gambar 4. 8 Inisialisasi Server dan Pustaka	81
Gambar 4.9 Proses Autentikasi OAuth 2.0	82
Gambar 4.10 Pengambilan Komentar dengan commentThreads.list.....	82
Gambar 4.11 Penyembunyian Komentar dengan comments.setModerationStatus83	
Gambar 4.12 Isi File blockedword.json	84
Gambar 4.13 Deteksi Unicode dan Fancy Text.....	85
Gambar 4.14 Contoh Pelabelan Data Untuk Pelatihan	85
Gambar 4.15 Isi data yang sudah dilatih Naive Bayes.....	86
Gambar 4.16 Pembuatan struktur file ekspor.....	87
Gambar 4.17 Hasil Pengujian Ambil Komentar dari Vidio Youtube	89
Gambar 4.18 Sebelum komentar spam disembunyikan.....	91
Gambar 4.19 Sesudah komentar spam disembunyikan.....	91
Gambar 4.20 Hasil ekspor ke excel.....	92