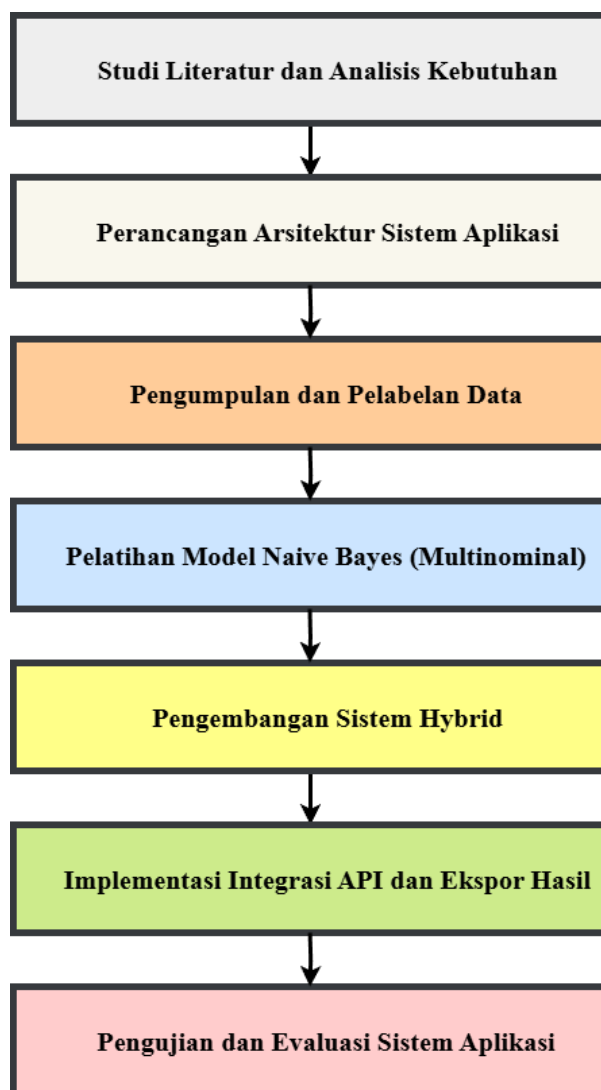


BAB III

METODE PENELITIAN

3.1 Prosedur Penelitian

Prosedur penelitian ini dirancang secara sistematis untuk memastikan pengembangan sistem deteksi komentar *spam* promosi judi online di YouTube dapat dilakukan secara terarah dan terukur. Tahapan penelitian terdiri atas tujuh langkah utama yang saling terkait dan berkelanjutan, mulai dari studi awal hingga evaluasi akhir sistem. Flowchart prosedur penelitian sebagai berikut.



Gambar 3.1 Prosedur Penelitian

1. Studi Literatur dan Analisis Kebutuhan

Tahap awal penelitian dilakukan dengan melakukan studi literatur terkait deteksi *spam*, klasifikasi teks, serta pendekatan *hybrid* dalam sistem informasi. Selain itu, dilakukan pula analisis kebutuhan fungsional dan non-fungsional dari sistem yang akan dikembangkan, termasuk target pengguna, fitur utama, dan alur interaksi sistem.

2. Perancangan Arsitektur Sistem Aplikasi

Pada tahap ini, arsitektur sistem dirancang berbasis web menggunakan teknologi Node.js dan Express. Komponen-komponen sistem seperti integrasi API, *rule-based* filter, preprocessing teks, klasifikasi *Naive bayes*, dan ekspor hasil ke Excel dirancang sedemikian rupa agar dapat bekerja secara terintegrasi. Diagram alur sistem dibuat untuk membantu proses pengembangan dan dokumentasi.

3. Pengumpulan dan Pelabelan Data

Sebanyak 10910 komentar dikumpulkan dari 17 video publik YouTube menggunakan skrip Python dan YouTube Data API. Komentar tersebut kemudian disimpan dalam file Excel (.xlsx) dan diberi label berdasarkan kemunculan kata kunci *spam* promosi judi online. Dataset ini selanjutnya dikonversi ke format JSON untuk digunakan dalam pelatihan model klasifikasi.

Berikut tabel data video yang komentarnya dijadikan pelatihan data:

Tabel 3. 1 Data Komentar dari Vidio

Nama Channel	Id Vidio	Judul Vidio	Spam	Bukan Spam	Total
MALAKA	W9HuTy_bioY	Satir Politik ala Kementerian Kegelapan	310	190	500
Deddy Corbuzier	YbTBIIoAz74	FERRY IRWANDI, OM SAMPAIKAN KE GIBRAN INI!!OM KAN BELAIN DIA!!	116	384	500
Raditya Dika	AQ7UGzn7QKE	Rahasia Sukses Film Jumbo	361	139	500
Kenyalang 96	o9zIzRWf41Q	Camping Hujan Deras Dapat Rusa 3 Hari di Hutan	333	167	500
ANTB OFFICIAL	2WIWuTfrGWs	GONDRONG DIUSIR	177	323	500
Tuah Kreasi	hAi1OWvYQo4	TITIK KUMPUL - AKHIRNYA ABDUR DISAMPERIN SALMA SALSABIL!!	348	152	500
Garasi Drift	ulfkHcmYeRE	Road Trip Berujung Mogok di Jalan IMX Semarang 2025 Pecah!!	399	101	500
MALAKA	qVgNdMvHvIQ	Kuliah S2Buat Apa?	311	189	500
guru gembul	FydVF Xu_k4	Eps 866 KEPALSUAN AKADEMIK DAN KARYA ILMIAH INDONESIA TERTINGGI KEDUA DI DUNIA. UGM APA KABAR?	180	320	500
MALAKA	0wwa-H0Rx0o	Influencer Micro vs Macro: Siapa yang Lebih Berpengaruh?	93	514	607
Fiersa Besari	Gs8TXj6rcbM	GUNUNG LEUSER - Atap Negeri Aceh #1	18	982	1000
HAHAHA TV	PuUOI5aXRT0	STANDUPFEST 2023 Extended Version dari The Founders Lucu Terus Nih	1	588	589
Ruangguru	dY_3acmFFRQ	Ruangguru Clash of Champions Episode 1 PARAH!!! HITUNG RATUSAN ANGKA SAMPE PUSING	0	1000	1000
Ruangguru	DSM-ZO1bIUY	Clash of Champions (COC) Season 2 - EPS. 1 Baru Mulai Udah CHAOS	9	991	1000

Deddy Corbuzier	PVfltgp9zeM	ORANG LAMA OPEJE NGELAWAK!!SULE MASIH KOCAK, INDAH KUS SAMPE NGAKAK WKWKWAKK!!😂😂 -VIDI	146	854	1000
Raditya Dika	B3e3c-HYw-U	Obrolan Ini Banyak Sensornya	162	513	675
Rian Arifinnn	lAX96HadzjU	Akhirnya Fajar Istri Dan Anaknya Datang Ke Podcast!	266	273	539
TOTAL			3230	7680	10910

4. Pelatihan Model *Naive bayes* (Multinomial)

Dataset berlabel kemudian digunakan untuk melatih algoritma klasifikasi teks berbasis *Naive bayes* (Multinomial) menggunakan library Natural di Node.js. Proses pelatihan dilakukan secara supervised, dan model yang dihasilkan disimpan dalam file JSON untuk digunakan dalam sistem produksi.

5. Pengembangan Sistem *Hybrid*

Sistem deteksi dikembangkan dengan memadukan dua pendekatan utama:

- Rule-Based Filter : Mendeteksi komentar *spam* eksplisit berdasarkan daftar kata kunci terlarang dari file blockedword.json dan beberapa aturan dalam normalization teks.
- Naive bayes* Classifier : Menangani komentar ambigu atau tidak terdeteksi oleh *rule-based* filter.

6. Implementasi Integrasi API dan Ekspor Hasil

Sistem dihubungkan dengan YouTube Data API v3 untuk mengambil komentar berdasarkan Video ID yang dimasukkan pengguna. Setelah diklasifikasikan, hasil deteksi disimpan dan diekspor ke dalam file Excel (.xlsx) menggunakan library ExcelJS. File hasil mencakup informasi seperti isi komentar, status *spam*, metode deteksi dan nama pengguna.

7. Pengujian dan Evaluasi Sistem Aplikasi

Sistem diuji coba menggunakan komentar dari channel pribadi milik peneliti. Komentar yang terdeteksi sebagai *spam* akan ditandai secara otomatis dari menggunakan sistem *hybrid*. Hasil prediksi dievaluasi menggunakan metrik klasifikasi seperti akurasi, presisi, dan recall untuk menilai performa sistem secara objektif.

3.2 Arsitektur Sistem

Arsitektur sistem dalam penelitian ini dirancang untuk mendeteksi komentar *spam* promosi judi online secara otomatis dan efisien dengan memanfaatkan pendekatan *hybrid*, yaitu kombinasi antara *rule-based filtering* dan klasifikasi teks berbasis machine learning menggunakan algoritma *Naive bayes*. Sistem ini dibangun berbasis web menggunakan Node.js dan Express, serta terintegrasi langsung dengan YouTube Data API v3 untuk mengakses komentar video secara real-time. Sebelum masuk ke sistem terlebih dahulu melakukan Autentikasi YouTube Data API dengan OAuth 2.0 Berikut adalah alur yang harus dilakukan jika autentikasi belum dilakukan sebelum sistem memulai:

1. credentials.json digunakan untuk inisialisasi client OAuth yang didapat dari aplikasi Youtube data API v3
2. Sistem cek apakah file token.json sudah ada:
 - a. Jika belum ada, user diarahkan ke browser untuk login dan konfirmasi izin dengan menggunakan perintah JavaScript.
 - b. Jika sudah ada, sistem langsung autentikasi otomatis
3. Sistem mendapatkan akses ke YouTube API (misalnya commentThreads.list , comments.setModerationStatus dan lain-lain)

4. Proses pemindaian komentar bisa dimulai

3.2.1 Komponen Utama Sistem

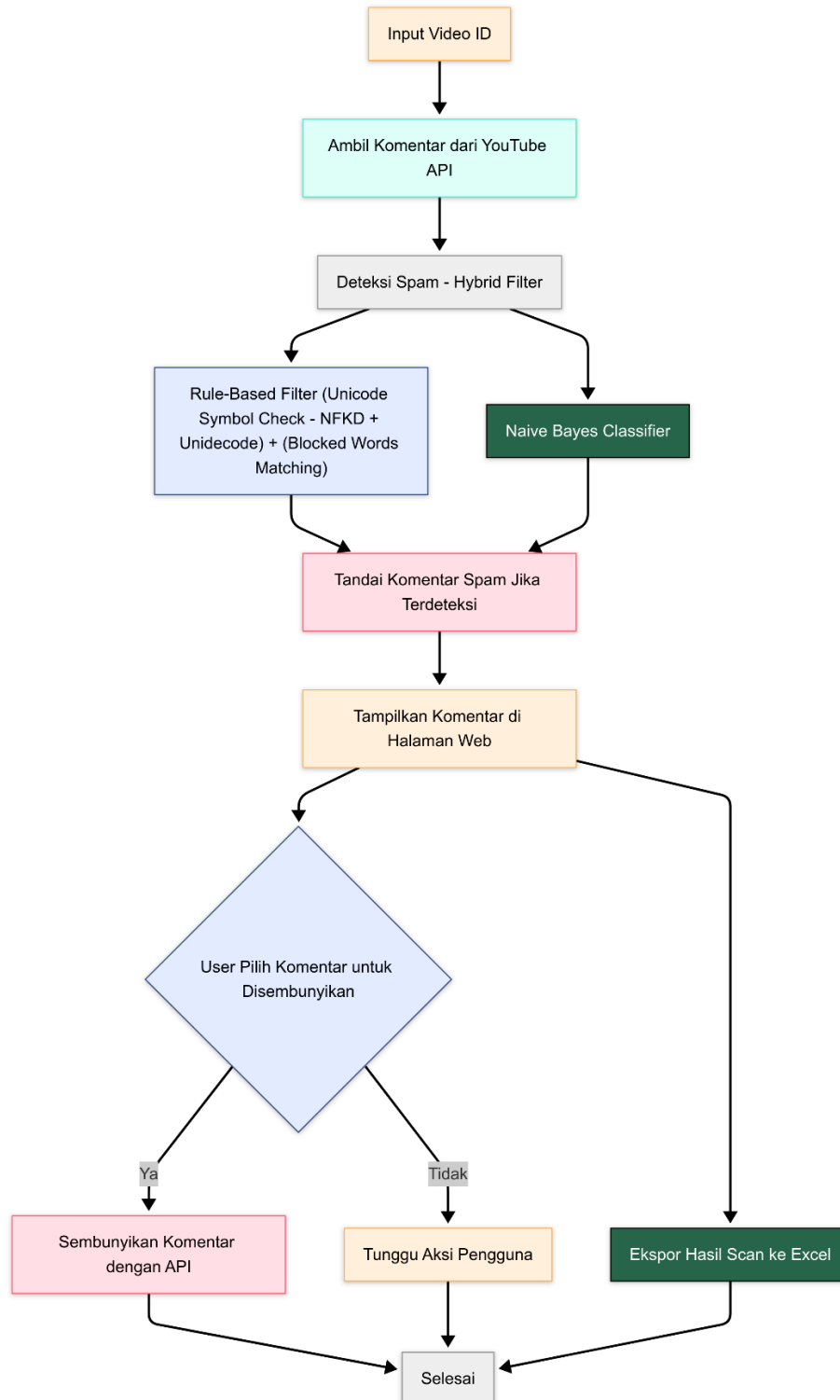
Sistem terdiri dari beberapa modul inti yang saling terhubung satu sama lain, berikut tabel yang menjelaskan komponen sistem :

Tabel 3.2 Komponen Utama Sistem

Komponen	Fungsi
YouTube Data API v3	Digunakan untuk mengambil dan menyembunyikan daftar komentar dari video YouTube menggunakan <i>endpoint comment Threads.list</i> .
<i>Preprocessing Engine</i>	Modul yang bertugas membersihkan teks komentar dengan cara normalisasi karakter, tokenisasi, penghapusan simbol, angka, dan stopwords agar siap untuk analisis lebih lanjut.
<i>Rule Based Filter</i>	Komponen yang mendeteksi komentar <i>spam</i> secara langsung berdasarkan daftar kata kunci terlarang, pola penggunaan huruf kapital berlebihan, atau karakter aneh yang sering muncul pada komentar <i>spam</i> .
<i>Naive bayes Classifier</i>	Model pembelajaran mesin (<i>supervised learning</i>) yang digunakan untuk mengklasifikasikan komentar ambigu setelah melewati tahap <i>rule-based filter</i> . Model ini dilatih menggunakan dataset komentar berlabel " <i>spam</i> " dan "bukan".
ExcelJS Exporter	Modul yang menyimpan hasil klasifikasi dalam format Excel untuk dokumentasi dan analisis lebih lanjut.

3.2.2 Alur Kerja Sistem Aplikasi Penghapus Komentar Spam

Alur kerja sistem ini terdiri dari beberapa tahapan, berikut alur kerja sistem



Gambal 3.2 Alur Kerja Sistem Aplikasi

1. Input Video ID

Pengguna memasukkan ID video YouTube channel Pribadi ke dalam form aplikasi. ID ini digunakan untuk mengakses komentar melalui YouTube Data API. Aplikasi mendukung input lebih dari satu ID (dipisahkan dengan koma).

2. Pengambilan Komentar via API

Setelah ID video dimasukkan, sistem menghubungi endpoint *commentThreads.list* dari YouTube Data API untuk mengambil komentar. Komentar yang diperoleh disimpan sementara untuk dianalisis.

3. Deteksi *Spam – Hybrid* Filter

Pada tahap ini, komentar yang telah diperoleh akan dianalisis oleh sistem deteksi spam menggunakan pendekatan hybrid, yaitu:

a. Rule-Based Filter

Terdiri dari dua jenis pemeriksaan:

1. Unicode Symbol Check: Sistem akan melakukan normalisasi teks menggunakan metode `NFKD` dan `unicodecode` untuk mendeteksi karakter mencurigakan seperti huruf penuh (**A**, **M**, **p**) atau simbol melingkar (**®**, **©**), yang biasanya digunakan spammer untuk menyamarkan kata kunci.
2. Blocked Words Matching: Sistem mencocokkan komentar dengan daftar kata terlarang (misalnya: "slot", "gacor", "jepe", "jp", "maxwin", dan nama-nama situs judi online. ").

b. Naive Bayes Classifier

Sistem pembelajaran mesin yang dilatih dengan dataset komentar spam dan bukan spam. Komentar akan diberi skor probabilitas dan diputuskan apakah tergolong spam atau bukan berdasarkan hasil klasifikasi.

4. Tandai Komentar Sebagai *Spam* Jika Terdeteksi

Jika komentar terdeteksi *spam* oleh salah satu atau kombinasi metode di atas, maka sistem akan menandai komentar tersebut sebagai *spam*. Namun, komentar belum langsung disembunyikan. Ini memberi kontrol lebih besar kepada pengguna..

5. Tampilkan Komentar di Halaman Web

Semua komentar (baik spam maupun bukan) ditampilkan pada tampilan web (halaman `index.ejs`). Komentar yang ditandai sebagai spam akan ditandai secara visual dan diberi checkbox agar dapat dipilih untuk disembunyikan.

6. User Pilih Komentar untuk Disembunyikan

Pengguna dapat mencentang komentar yang ingin disembunyikan. Ini mencegah sistem menyembunyikan komentar secara otomatis, dan memberi kesempatan pengguna untuk meninjau ulang komentar yang ditandai.

7. Sembunyikan Komentar dengan API

Jika pengguna menekan tombol "Sembunyikan Komentar Ditandai", sistem akan mengirim permintaan ke YouTube API untuk mengubah status komentar menjadi "rejected", sehingga tidak akan muncul di publik meskipun masih terlihat oleh pemilik channel.

8. Ekpor Hasil Scan ke Exel

Seluruh hasil analisis komentar, termasuk informasi video ID, nama pengguna, isi komentar, status spam, dan metode deteksi, akan diekspor ke file Excel (`hasil_scan.xlsx`). File ini bisa diunduh untuk dokumentasi atau laporan skripsi.

9. Selesai

Tahapan pemindaian selesai. Pengguna dapat mengulangi proses untuk video lain atau menutup aplikasi.

3.2.3 Pemilihan Pendekatan *Hybrid*

Pendekatan *hybrid* dipilih karena memiliki keunggulan dalam hal:

1. Kecepatan deteksi : *Rule-based* filter mampu mendeteksi komentar *spam* eksplisit secara instan.
2. Akurasi adaptif : Algoritma *Naive bayes* membantu mendeteksi komentar *spam* yang tidak terlalu jelas atau bersifat terselubung.
3. Fleksibilitas : Dengan pendekatan gabungan ini, sistem dapat menangani variasi pola *spam* yang berkembang di platform YouTube.

Model *Naive bayes* dilatih menggunakan dataset komentar berlabel yang telah diolah sebelumnya, dan disimpan dalam bentuk file json untuk digunakan secara langsung dalam sistem produksi.

3.2.4 Teknologi yang Digunakan

Berikut adalah tabel yang menjelaskan tentang modul, teknologi dan fungsi yang digunakan dalam penelitian ini:

Tabel 3.3 Teknologi yang Digunakan

Modul	Teknologi	Fungsi
Backend	Node.js + Express	Menjalankan server web dan API
Otentikasi	Google OAuth2	Akses YouTube Data API

API	YouTube Data API v3	Mengambil dan menyembunyikan komentar
Preprocessing	<i>natural,string-sanitizer</i>	Membersihkan dan memproses teks komentar
Machine Learning	<i>Natural (Bayes Classifier)</i>	Membaca model <i>Naive bayes</i> dan melakukan klasifikasi
Ekspor Data	ExcelJS	Menyimpan hasil ke dalam file Excel

3.3 Jenis dan Pendekatan Penelitian

Penelitian ini termasuk dalam kategori penelitian rekayasa perangkat lunak (software engineering research), di mana fokus utamanya adalah pada perancangan dan penerapan sebuah sistem aplikasi yang berfungsi untuk mendeteksi serta menyaring komentar *spam* promosi judi online secara otomatis pada platform YouTube. Penelitian ini tidak hanya bertujuan untuk menghasilkan suatu solusi teknologi, tetapi juga berperan mengevaluasi efektivitas pendekatan yang digunakan dalam menangani masalah *spam* teks di kolom komentar.

Sebagai bagian dari penelitian eksperimental, pendekatan kuantitatif digunakan untuk mengukur performa sistem berdasarkan *metrik* evaluasi seperti akurasi, presisi, dan *recall*. Data yang digunakan dalam penelitian ini bersumber dari komentar pengguna YouTube, yang dikumpulkan melalui YouTube Data API v3. Setelah melalui tahapan preprocessing dan filtering awal, data kemudian diklasifikasikan menggunakan algoritma *Naive bayes* sebagai bagian dari model machine learning yang dilatih secara *supervised*.

Selain itu, penelitian ini juga menerapkan pendekatan studi kasus, di mana sistem diimplementasikan dan diuji pada sejumlah video YouTube yang memiliki

komentar relevan dengan konten promosi ilegal seperti judi online. Hal ini memungkinkan analisis yang lebih spesifik terhadap pola komentar *spam* yang sering muncul serta respons sistem terhadap variasi data tersebut.

Pendekatan *hybrid* juga menjadi salah satu ciri khas dari penelitian ini, yaitu kombinasi antara *rule-based filtering* dan *machine learning*. Pendekatan pertama digunakan untuk membersihkan dan menyaring komentar yang jelas-jelas merupakan *spam* berdasarkan pola tertentu, sedangkan pendekatan kedua digunakan untuk mengklasifikasikan komentar yang sifatnya lebih kompleks dan tidak selalu dapat dideteksi hanya dengan aturan sederhana. Kombinasi kedua metode ini diharapkan dapat meningkatkan efisiensi dan ketepatan sistem dalam mendeteksi *spam* promosi judi online.

Berikut tabel yang merangkum pendekatan penelitian secara visual:

Tabel 3.4 Ringkasan Pendekatan Penelitian

Aspek	Penjelasan
Jenis Penelitian	Rekayasa Perangkat Lunak (Software Engineering Research)
Pendekatan	Kuantitatif, Eksperimental, Studi Kasus
Teknik Deteksi	<i>Hybrid: Rule-based Filtering + Klasifikasi Naive bayes</i>
Metode Pengumpulan Data	Scraping komentar menggunakan YouTube Data API v3
Data yang Digunakan	Komentar YouTube yang mengandung promosi judi online
Model Machine Learning	<i>Naive bayes</i> (Supervised Learning)
Evaluasi Kinerja	Akurasi, Presisi, Recall

3.4 Lokasi dan Waktu Penelitian

Penelitian ini dilaksanakan secara daring (*online*) karena seluruh proses pengumpulan data, pengembangan sistem serta pengujian dilakukan melalui akses internet dan platform digital. Lokasi teknis pelaksanaan berada pada lingkungan pengembangan lokal (*local development environment*), dengan menggunakan perangkat lunak seperti Node.js dan Express sebagai backend, serta Google Cloud Console untuk integrasi YouTube Data API v3.

Sumber data utama berupa komentar pengguna YouTube diperoleh melalui YouTube Data API, sehingga tidak memerlukan observasi langsung ke lokasi fisik. Seluruh tahapan pengolahan data dan implementasi sistem dilakukan pada perangkat komputer pribadi peneliti yang terhubung dengan jaringan internet. Penelitian ini direncanakan berlangsung dari April 2025 hingga Juni 2025, dengan tahapan kegiatan penelitian adalah sebagai berikut:

Tabel 3.5 Jadwal Penelitian

Waktu Kegiatan	Kegiatan
April 2025	Studi literatur, identifikasi masalah, analisis kebutuhan sistem, serta perancangan arsitektur sistem, pengambilan data komentar YouTube, <i>preprocessing</i> teks, serta implementasi <i>rule-based filtering</i> dan klasifikasi <i>Naive bayes</i> .
Mei 2025	Pengujian sistem dan evaluasi kinerja berdasarkan metrik akurasi, presisi, dan <i>recall</i> .
Juni 2025	Dokumentasi hasil, penyusunan laporan akhir, dan finalisasi naskah skripsi.

Dengan alokasi waktu yang terencana dan tahapan yang terstruktur, penelitian ini diharapkan dapat menghasilkan sistem deteksi *spam* promosi judi online yang efektif.

3.5 Metode Pengumpulan Data

Pengumpulan data dalam penelitian ini dilakukan menggunakan YouTube Data API v3. Data yang dikumpulkan berupa komentar pengguna dari sejumlah video YouTube, yang kemudian diproses dan digunakan untuk melatih serta menguji sistem deteksi komentar *spam* berbasis *hybrid*, yaitu kombinasi *rule-based filtering* dan *Naive bayes classifier*.

3.5.1 Pengambilan Komentar dari Vidio YouTube

Pengumpulan komentar dilakukan melalui eksekusi skrip *Python* yang terintegrasi dengan pustaka *googleapiclient.discovery* untuk mengakses *endpoint commentThreads.list* dari YouTube Data API v3. Dalam satu permintaan agar tidak dianggap *bot spam* oleh pihak YouTube, sistem hanya mengambil hingga 500 komentar per video dan proses ini diterapkan pada 11 video berbeda sehingga total komentar yang dikumpulkan sebanyak 5.107 komentar.

Untuk mempermudah eksekusi dan penyimpanan data, proses pengambilan komentar dilakukan melalui Google Colab. Data yang berhasil diambil meliputi informasi seperti:

1. nama pengguna (*author*),
2. isi komentar (*text*),
3. tanggal dan waktu komentar (*timestamp*).

Seluruh data tersebut disimpan ke dalam file Excel (.xlsx) menggunakan pustaka *pandas*.

3.5.2 Pelabelan Komentar *Spam*

Setelah data tersedia dalam format Excel, tahapan selanjutnya adalah pelabelan manual dengan Label “*spam*” atau “bukan” diberikan berdasarkan kemunculan kata kunci *spam* promosi judi online, seperti:

1. Komen Berisi kode seperti huruf atau angka
2. “slot”
3. “gacor”
4. “judi ”
5. dll.

Jika suatu komentar mengandung salah satu dari kata kunci tersebut, maka diberi label “*spam*”. Sebaliknya, jika tidak mengandung kata kunci tersebut maka komentar diberi label “bukan”.

3.5.3 Konversi Data Exel ke Format JSON

Setelah proses pelabelan selesai, file Excel kemudian dikonversi menjadi file berformat JSON untuk memenuhi kebutuhan pelatihan model machine learning. Format JSON dipilih karena kompatibel dengan pustaka *natural*, library yang digunakan dalam Node.js untuk implementasi algoritma *Naive bayes*.

Contoh struktur entri dalam file JSON:

```
1 {
2   "text": "isi komentar pengguna",
3   "label": "spam"
4 }
```

Gambal 3.3 Format Isi File JSON

3.5.4 Pelatihan Model Klasifikasi *Naive bayes*

File JSON yang berisi komentar berlabel kemudian digunakan sebagai dataset pelatihan (training set) untuk model klasifikasi teks berbasis algoritma *Naive bayes*. Model ini dilatih untuk mengenali pola-pola teks yang umum ditemukan pada komentar *spam* promosi judi online.

Model klasifikasi dilatih menggunakan dataset yang telah dibersihkan dan dilabeli secara otomatis. Pada dasarnya, algoritma *Naive bayes* bekerja dengan menghitung probabilitas suatu komentar termasuk kategori “*spam*” atau “bukan” berdasarkan frekuensi kemunculan kata-kata tertentu di dalamnya.

Secara teknis, model menggunakan pendekatan *Multinomial Naive bayes*, yang cocok untuk data yang direpresentasikan sebagai frekuensi kata dalam dokumen teks. Hal yang dilakukan selama pelatihan:

1. Model mempelajari pola-pola kata yang sering muncul dalam komentar *spam*.
2. Setiap kata diberi bobot berdasarkan kontribusinya dalam menentukan apakah suatu komentar bersifat *spam* atau bukan.
3. Hasil pelatihan disimpan dalam bentuk model yang dapat digunakan untuk memprediksi label komentar baru.

Setelah selesai dilatih, model disimpan dan siap digunakan untuk mendeteksi komentar baru secara real-time dalam sistem akhir.

3.5.5 Pengumpulan Data Uji dari Channel Pribadi

Sebagai bagian dari evaluasi sistem, data uji diperoleh dari komentar video yang diunggah di channel YouTube pribadi milik peneliti. Komentar tersebut

diambil secara langsung melalui YouTube Data API dan diproses oleh sistem *hybrid*: tahap awal berupa filtering berbasis aturan, diikuti dengan klasifikasi menggunakan model *Naive bayes* yang telah dilatih sebelumnya.

Hasil akhir berupa komentar yang dikategorikan sebagai *spam* atau bukan *spam* diekspor kembali ke dalam file Excel untuk analisis lebih lanjut dan evaluasi performa sistem.

3.6 Alat dan Bahan

Penelitian ini menggunakan kombinasi perangkat lunak dan perangkat keras untuk merancang, mengembangkan, serta menguji sistem deteksi komentar *spam* promosi judi online di platform YouTube. Berikut adalah rincian alat dan bahan yang digunakan dalam penelitian ini.

3.6.1 Perangkat Keras (Hardware)

Perangkat keras yang digunakan dalam penelitian ini adalah komputer pribadi dengan spesifikasi minimum sebagai berikut:

Tabel 3.6 Hardware yang Digunakan

Komponen	Spesifikasi
Prosesor	Intel Core i5 / setara
RAM	8 GB
Penyimpanan	SSD 256 GB
Sistem Operasi	Windows 10 / Linux Ubuntu
Koneksi Internet	Stabil untuk akses YouTube Data API

Perangkat tersebut digunakan untuk menjalankan server backend, melakukan pengujian sistem, serta integrasi dengan layanan eksternal seperti YouTube API.

3.6.2 Perangkat Lunak (Software) dan Library

Berikut adalah perangkat lunak dan library yang digunakan dalam pengembangan sistem:

Tabel 3.7 Perangkat Lunak dan Library

Perangkat Lunak / Library	Fungsi
Node.js	Platform runtime JavaScript yang digunakan untuk menjalankan kode backend di sisi server.
Express.js	Framework web minimalis yang digunakan untuk membuat server HTTP, routing, dan API endpoint secara efisien.
EJS	Template engine yang memungkinkan pembuatan tampilan HTML dinamis berdasarkan data dari server.
Google APIs Client Library	Library resmi dari Google untuk mengakses berbagai API Google, termasuk proses autentikasi OAuth2 dan interaksi dengan YouTube Data API v3.
YouTube Data API v3	API yang memungkinkan aplikasi mengambil data publik seperti komentar video, profil channel, dan informasi terkait YouTube.
Natural (NPM)	Library Natural Language Processing (NLP) untuk Node.js, digunakan untuk pelatihan dan klasifikasi komentar menggunakan algoritma Naive Bayes.
ExcelJS	Library Node.js untuk membuat dan menulis file Excel (.xlsx), digunakan untuk mengeksport hasil klasifikasi ke format spreadsheet.
dotenv	Library untuk membaca dan mengelola variabel lingkungan dari file .env, seperti API Key dan port server.
body-parser	Middleware Express yang digunakan untuk memproses data yang dikirim dari form HTML dan permintaan berformat JSON.
unidecode (NPM)	Library untuk mengubah karakter Unicode menjadi bentuk ASCII sederhananya, membantu normalisasi teks dalam deteksi simbol fancy Unicode.

fs (Node.js File System)	Modul bawaan Node.js untuk membaca, menulis, dan memanipulasi file lokal, digunakan dalam pemuatan model, konfigurasi, dan hasil scan.
path (Node.js Module)	Modul bawaan untuk menangani dan menyusun path file sistem operasi secara lintas platform.
Google Colab	Platform notebook interaktif dari Google yang digunakan untuk eksperimen awal, scraping komentar, dan persiapan dataset.
Python	Bahasa pemrograman yang digunakan pada tahap awal proyek untuk scraping komentar dan pelabelan data training.
pandas (Python)	Library Python untuk pengolahan dan analisis data tabular (Excel), digunakan untuk membersihkan dan memberi label pada data komentar.
openpyxl (Python)	Digunakan bersama pandas untuk membaca/menulis file Excel .xlsx pada tahap persiapan dataset di Python sebelum dilatih dengan Naive Bayes.
Node.js	Platform runtime JavaScript yang digunakan untuk menjalankan kode backend di sisi server.

Pemilihan teknologi seperti Node.js dan Express didasarkan pada kebutuhan sistem yang ringkas, responsif, dan mudah diintegrasikan dengan API pihak ketiga seperti YouTube Data API v3.

3.6.3 Dataset

Dataset yang digunakan dalam penelitian ini terdiri dari 5107 komentar YouTube yang dikumpulkan dari 11 video publik dengan estimasi 100 - 500 komentar per video. Data komentar disimpan dalam format Excel (.xlsx) dan kemudian dikonversi ke format JSON untuk pelatihan model klasifikasi.

Proses pelabelan awal dilakukan secara otomatis berdasarkan daftar kata kunci *spam* promosi judi online. Label yang digunakan adalah "*spam*" dan "*bukan*".

Dataset ini kemudian digunakan sebagai input pelatihan training set untuk model klasifikasi teks berbasis algoritma *Naive bayes* menggunakan library *natural*.

3.7 Metode Evaluasi

Evaluasi sistem dilakukan untuk mengukur efektivitas dan akurasi sistem deteksi komentar *spam* promosi judi online di YouTube. Metrik evaluasi yang digunakan meliputi:

1. Akurasi (Accuracy)
2. Presisi (Precision)
3. Recall (Sensitivitas)
4. F1-Score

Metrik-metrik ini dihitung berdasarkan confusion matrix, yaitu tabel yang menunjukkan hasil prediksi sistem dibandingkan dengan label aktual.

Tabel 3.8 Confusion Matrix

	Prediksi: <i>Spam</i>	Prediksi: Bukan <i>Spam</i>
Aktual: <i>Spam</i>	TP (True Positive): Komentar <i>spam</i> yang berhasil dikenali sebagai <i>spam</i> .	FN (False Negative): Komentar <i>spam</i> yang tidak terdeteksi (salah diklasifikasikan sebagai bukan <i>spam</i>).
Aktual: Bukan <i>Spam</i>	FP (False Positive): Komentar normal yang salah diklasifikasikan sebagai <i>spam</i> .	TN (True Negative): Komentar normal yang berhasil dikenali sebagai bukan <i>spam</i> .

Dari confusion matrix tersebut, dapat dihitung empat metrik utama sebagai berikut:

1. Akurasi (Accuracy)

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN}$$

Mengukur seberapa tepat sistem dalam memprediksi seluruh data (*spam* + bukan *spam*).

2. Presisi (Precision)

$$Precision = \frac{TP}{TP + FP}$$

Mengukur proporsi komentar yang benar-benar *spam* dari semua komentar yang diprediksi sebagai *spam*.

3. Recall (Sensitivity / True Positive Rate)

$$Recall = \frac{TP}{TP + FN}$$

Mengukur kemampuan sistem mendeteksi komentar *spam* yang sebenarnya.

4. F1-Score

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Rata-rata harmonis antara presisi dan recall, berguna ketika distribusi kelas tidak seimbang.

3.7.1 Contoh Penghitungan Evaluasi

Jika pada pengujian sistem dari 200 komentar diperoleh hasil sebagai berikut:

TP = 80, FN = 10, FP = 5 dan TN = 105

Maka perhitungannya:

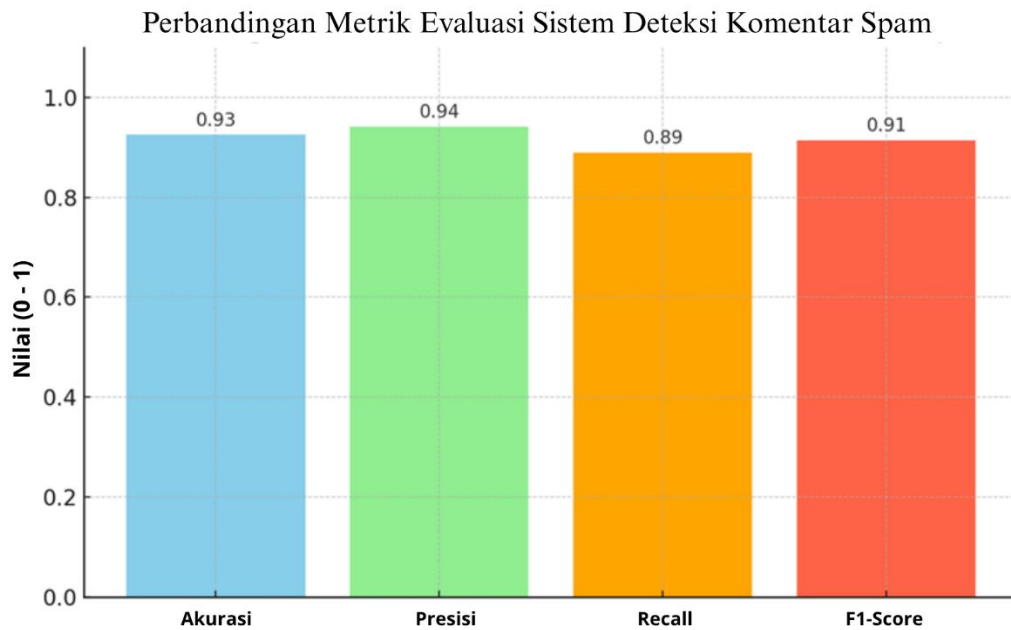
$$1. \text{ Akurasi} = (TP + TN) / (TP + FN + FP + TN) = (80 + 105) / 200 = 92.5\%$$

$$2. \text{ Presisi} = TP / (TP + FP) = 80 / (80 + 5) = 94.1\%$$

3. $\text{Recall} = \text{TP} / (\text{TP} + \text{FN}) = 80 / (80 + 10) = 88.9\%$

4. $\text{F1-Score} = 2 \times (\text{Presisi} \times \text{Recall}) / (\text{Presisi} + \text{Recall}) = 91.4\%$

Berikut adalah grafik visual perbandingan metrik evaluasi sistem, berdasarkan hasil uji contoh:



Gambal 3.4 Hasil Uji Contoh Metrik Evalua