

Sejudol

Website Sentiment Analysis

Sentimen Judi



Keny Sade
cipeng0d@gmail.com

Menu

- █ Data
- █ Analisis Sentimen

Data dari Tabel Komentar

Berapa banyak tweet ingin di-crawl? Crawl Preprocessing Delete All

Total Data: 175

No	Tanggal	Data Mentah	Data Bersih	Label	Aksi
1	6/12/2025, 6:39:17 PM	@BANGVELL_ @akulezy Koordinatnya dkt lapas.. sepertinya pelaku menipu dr ldm lapas, apalagi qr's nya ke situs judol.. kemungkinan pelaku ad d ldm lapas	koordinatdktlapaslakutip <ul style="list-style-type: none">lapasqr'syasisitusjudolke mungkinylakulapas	negatif	Hapus
2	6/12/2025, 6:10:40 PM	Hati-hati! Modus penipuan judi online via telpon & Telegram lagi marak. Data ratusan juta penduduk Indonesia bocor, bikin nomor kita jadi target. Kenali triknya biar ga jadi korban! #JudiOnline #KeamananData	hatihatimodustipujudionl ineviatelponamptelegram marakdataratusnjutaduk indonesiabocorbiinnomo rtargetkenaltrikbiarkorba njudionlinekeamanandata	negatif	Hapus
3	6/12/2025, 5:45:23 PM	Hi guys jujur gua baru aja login X, dan gua melihat bahwa akun gua bernama Dipta...? Gua ga kenal ini siapa please, sumpah ini kayaknya akun admin judol deh. Bantu up!!!!	higuyssjujurgualoginkaligu aakunguanamadiptaguake nalpleaseumphayakak unadminjudoldehbantuup	negatif	Hapus
4	6/12/2025, 5:33:11 PM	@akulezy qr's judol itu, mau main tapi gak punya duit WKKWKW	qr'sjudolmainduitwkkwk	negatif	Hapus

Sentimen Judi



Keny Sade
cipeng0d@gmail.com

Menu

- █ Data
- █ Analisis Sentimen

Avg Confidence per Sentimen

Positif: **90.12%**

Negatif: **98.13%**

Netral: **94.35%**

Risk Flag per Sentimen

Positif: **44.44%** (8/18)

Negatif: **74.07%** (100/135)

Netral: **59.09%** (13/22)

Top Komentar Berdasarkan Confidence

Positif (paling yakin):

"@papantengah Kl aku sih trimakasih fans judol SE Indonesia raya.,berkat semangat mendukung yg sangat militan,,sehingga klub ini lsu punya pelatih dan pemain jempolan smua,,hi idup judoool"

— Confidence: 99.82%

Negatif (paling yakin):

"@lucentbeam12 Negara hancur udh mau salah pilih pemimpin tp yg dipekarain ttp homo mulu. Wapres lu noh follow akun judol, presiden lu noh perjahat HAM, RAJA AMPAT NOH HANCUR!"

— Confidence: 99.96%

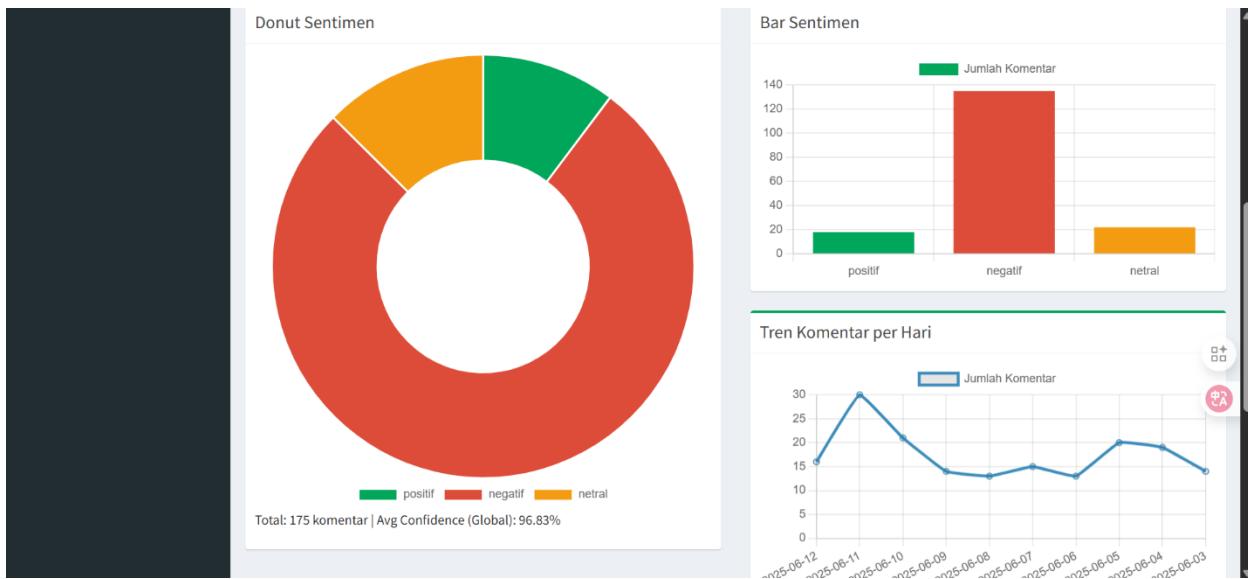
Donut Sentimen



Bar Sentimen



Jumlah Komentar: 140



- **Problem Statement**

Maraknya kecanduan judi online di kalangan remaja dan dewasa muda di Indonesia semakin mengkhawatirkan. Banyak kasus ditemukan melalui unggahan di media sosial, namun tidak ada sistem otomatis untuk mendeteksi atau memantau percakapan tersebut secara real-time. Hal ini menyulitkan lembaga pemerintah, keluarga, dan organisasi sosial untuk mengambil tindakan pencegahan sejak dini. Oleh karena itu, dibutuhkan sistem yang mampu mengenali gejala atau tanda percakapan terkait judi online dengan akurasi tinggi menggunakan AI.

- **Deskripsi Produk/Aplikasi**

Saya mengembangkan sebuah aplikasi web bernama **SEJUDOL** (Sentimen Judi Detector) yang dirancang untuk mengidentifikasi komentar-komentar di Twitter yang berpotensi mengandung unsur kecanduan judi online. Aplikasi ini menggunakan teknik pemrosesan bahasa alami (NLP) dan model klasifikasi sentimen untuk mendeteksi apakah sebuah komentar bersentimen positif, negatif, atau netral terhadap judi online, sehingga dapat membantu dalam pemantauan dan analisis risiko perilaku adiktif di media sosial.

Aplikasi ini akan mengklasifikasikan komentar ke dalam kategori positif, negatif, atau netral, serta mendeteksi kata-kata berisiko tinggi seperti “slot”, “pinjol”, “gacor”, “menang berapa”, dll.

Produk ini ditujukan untuk peneliti, pemerintah, dan aktivis sosial yang ingin memahami pola penyebaran konten judi online.

- **Fitur Utama dan Teknologi yang Digunakan**

1. Fitur Utama:

-  1. Modul Data

Fungsi Utama:

- Input jumlah tweet yang ingin di-crawl.
- Tombol "Crawl": Mengambil tweet dari sumber (kemungkinan dari Twitter API).
- Tombol "Preprocessing": Proses preprocessing mencakup pembersihan teks seperti menghapus tanda baca, angka, dan simbol, melakukan case folding, normalisasi kata, tokenisasi, stopword removal, dan stemming. Hasil dari proses ini disimpan dalam kolom "Data Bersih" atau "cleaning". Setelah itu, teks yang telah dibersihkan langsung diklasifikasikan menggunakan model (Naive Bayes, SVM, atau IndoBERT), dan hasil prediksi sentimen disimpan dalam kolom "Label" dengan kategori seperti "positif", "negatif", atau "netral".
- Tombol "Delete All": Menghapus seluruh data dalam tabel.

Tampilan Tabel Komentar:

- Kolom-kolom:
 - No
 - Tanggal
 - Data Mentah
 - Data Bersih (hasil preprocessing)
 - Label sentimen (positif / negatif / netral)
 - Aksi (hapus per komentar)

-  2. Modul Analisis Sentimen

- a. Statistik Confidence Rata-rata per Sentimen

Menampilkan rata-rata confidence dari model untuk masing-masing sentimen:

Positif: 90.12%

Negatif: 98.13%

Netral: 94.35%

- b. Risk Flag per Sentimen

Menampilkan jumlah dan persentase komentar yang masuk kategori "berisiko" per kelas sentimen. Misalnya: 74.07% komentar negatif masuk kategori berisiko.

c. Top Komentar Berdasarkan Confidence

- Menampilkan komentar paling representatif (dengan confidence tertinggi) dari masing-masing kelas:
- Positif paling yakin
- Negatif paling yakin

3. Visualisasi

- Donut Chart Sentimen: Proporsi sentimen dari semua data komentar (warna hijau = positif, merah = negatif, orange = netral).
- Bar Chart Sentimen: Jumlah komentar berdasarkan sentimen.
- Line Chart Tren Komentar per Hari: Pergerakan jumlah komentar dari waktu ke waktu.

2. Teknologi yang digunakan

Machine Learning & NLP (Natural Language Processing)

- Scikit-learn (joblib, tfidf_vectorizer, Naive Bayes, SVM pipeline)
- Sastrawi: Library bahasa Indonesia untuk stemming
- NLTK: Untuk stopwords Bahasa Indonesia
- HuggingFace Transformers API: Model indonesian-roberta-base-sentiment-classifier untuk inferensi sentimen

Python Core Libraries

- os, re: Untuk akses environment & regex
- asyncio: Untuk menangani asynchronous crawling

Eksternal Service Integration

- HuggingFace Inference API: Digunakan untuk klasifikasi sentimen berbasis transformer model
 - dotenv (python-dotenv): Mengelola variabel environment seperti API key dan DB credentials
-

Data Handling

- Pandas: Untuk manipulasi kamuskatabaku.xlsx menjadi dictionary kata baku
 - NumPy: Untuk operasi numerik & sigmoid konversi confidence SVM
-

Web Framework & API Handling

- Flask: Framework utama backend
 - Flask-CORS: Untuk mengizinkan cross-origin request (CORS)
 - Flask JSON API endpoints: CRUD operasi dan endpoint prediksi/preprocessing
-

Database

- PyMySQL: Koneksi ke database MySQL (akses dan update tabel comments)
-

Crawler (eksternal)

- crawl_and_return dari modul crawler: Fungsi async untuk mengambil tweet (diasumsikan dari modul buatan sendiri)
-

Keamanan

- Pemisahan API key dan DB credentials menggunakan .env (tidak ditulis langsung di source code)
-

File I/O

- kamuskatabaku.xlsx: File Excel untuk normalisasi kata tidak baku → baku
- .joblib files: Model dan vektorizer yang disimpan untuk inference

Frontend Web Interface

- PHP: Digunakan sebagai server-side rendering dan pengelolaan session
- HTML + JavaScript (Vanilla): Untuk tampilan antarmuka dan dynamic table
- Bootstrap: Untuk styling dan elemen UI (form, alert, tabel, dll.)
- JavaScript Fetch API: Untuk polling data komentar secara berkala dari API Flask
- AJAX-like Behavior: Menggunakan fetch dan DOM manipulation untuk live data update
- cURL (PHP): Untuk komunikasi dengan REST API Flask (crawl, preprocess, delete, dsb.)

• Cara Penggunaan Product

Pengguna dapat menggunakan SEJUDOL melalui antarmuka web yang intuitif dan mudah dipahami. Proses dimulai dengan melakukan crawling data dari Twitter, di mana pengguna menentukan jumlah tweet yang ingin diambil lalu menekan tombol "Crawl". Sistem akan secara otomatis mengambil data tweet sesuai jumlah yang ditentukan dan menyimpannya ke dalam database. Selanjutnya, pengguna dapat menekan tombol "Preprocessing" untuk melakukan pembersihan data, yang mencakup penghapusan simbol, normalisasi, stemming, dan tahap-tahap lain dalam pemrosesan teks. Proses ini juga sekaligus melakukan klasifikasi sentimen terhadap komentar menggunakan model machine learning, dan menentukan apakah komentar tersebut termasuk sentimen positif, negatif, atau netral.

SEJUDOL juga menyediakan opsi penghapusan data, baik secara individual melalui tombol "Hapus" di setiap baris, maupun secara menyeluruh dengan tombol "Delete All". Tabel pada antarmuka akan secara otomatis memperbarui dan menampilkan hasil dari crawling, preprocessing, serta klasifikasi secara real-time. Setiap entri komentar disajikan lengkap dengan tanggal, teks asli, hasil pembersihan, dan label sentimen.

Setelah proses klasifikasi selesai, pengguna dapat mengakses halaman Analisis Sentimen untuk melihat ringkasan statistik secara visual. Rangkuman ini meliputi rata-rata tingkat keyakinan (confidence score) dari model untuk masing-masing kategori sentimen—positif, negatif, dan netral. Selain itu, tersedia pula informasi mengenai risk flag, yaitu persentase komentar dalam tiap kategori sentimen yang dianggap memiliki potensi risiko terkait konten judi online. SEJUDOL juga menampilkan komentar-komentar dengan tingkat keyakinan tertinggi untuk sentimen positif dan negatif, disertai skor confidence-nya. Visualisasi data disajikan dalam bentuk donut chart untuk memperlihatkan distribusi proporsi sentimen, serta bar chart untuk menunjukkan jumlah komentar dalam masing-masing kategori.