

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 WKWKWK	qr'sjudolmainduitwkWKWK	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 Isa 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%

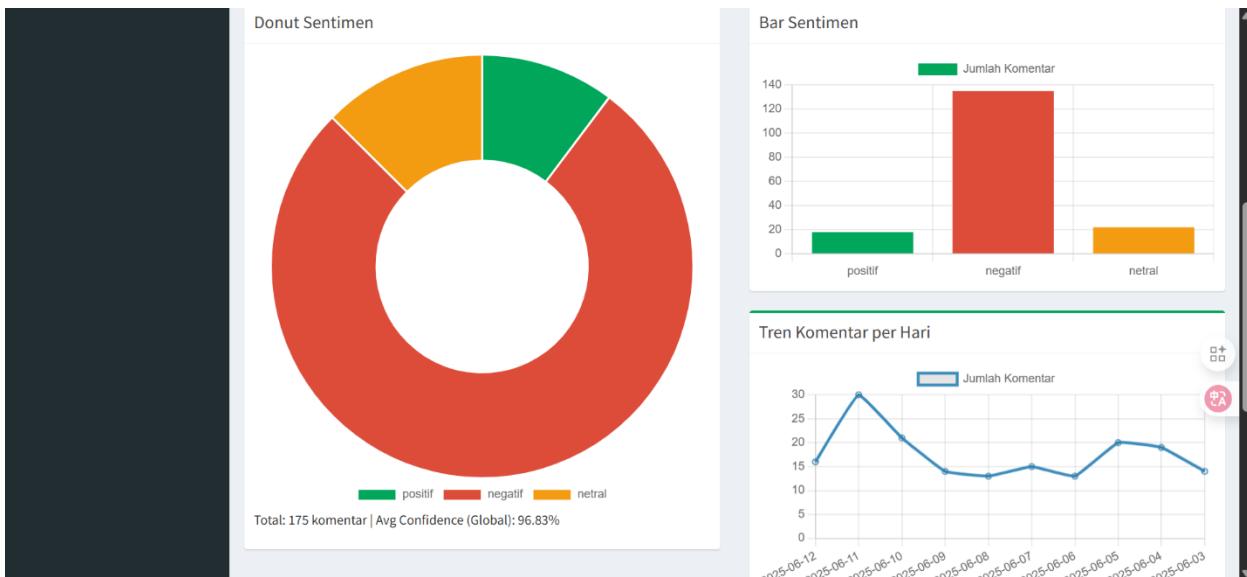
Donut Sentimen



Bar Sentimen



Jumlah Komentar: 140



- **Problem Statement**

The rise of online gambling addiction among teenagers and young adults in Indonesia is increasingly worrying. Many cases are discovered through social media posts, but there is no automated system to detect or monitor such conversations in real-time. This makes it difficult for government agencies, families, and social organizations to take precautions early. Therefore, a system that is able to recognize symptoms or signs of conversation related to online gambling with high accuracy using AI is needed.

- **Product/Application Description**

I developed a web application called SEJUDOL (Sentiment Gambling Detector) which is designed to identify comments on Twitter that have the potential to contain elements of online gambling addiction. The app uses natural language processing (NLP) techniques and sentiment classification models to detect whether a comment is positive, negative, or neutral to online gambling, so it can help in monitoring and analyzing the risk of addictive behavior on social media.

The app will classify comments into positive, negative, or neutral categories, as well as detect high-risk words such as "slots", "loans", "gacor", "how much do you win", etc.

This product is intended for researchers, governments, and social activists who want to understand the patterns of online gambling content spread.

- **Key Features and Technologies Used**

1. Key Features:

-  1. Data Module

Main Functions:

- Input the number of tweets you want to crawl.
- "Crawl" button: Retrieves tweets from the source (possibly from the Twitter API).
- "Preprocessing" button: The preprocessing process includes text cleanup such as removing punctuation, numbers, and symbols, performing case folding, word normalization, tokenization, stopword removal, and stemming. The results of this process are stored in the "Clean Data" or "cleaning" column. After that, the cleaned text is instantly classified using a model (Naive Bayes, SVM, or IndoBERT), and the results of the sentiment prediction are stored in a "Label" column with categories such as "positive", "negative", or "neutral".
- "Delete All" button: Deletes all data in the table.

Comment Table View:

- Columns:
- Yes
- Date
- Raw Data
- Net Data (preprocessing results)
- Sentiment labels (positive/negative/neutral)
- Actions (delete per comment)

-  Sec. 2. Sentiment Analysis Module

- a. Average Confidence Statistics per Sentiment

Displays the average confidence of the model for each sentiment:

Positive: 90.12%

Negative: 98.13%

Neutral: 94.35%

- b. Risk Flag per Sentiment

Displays the number and percentage of comments that are categorized as "at risk" per sentiment class. For example: 74.07% of negative comments are in the risk category.

c. Top Comments Based on Confidence

- Display the most representative (with the highest confidence) comments from each class:
- The most confident positive
- The most confident negatives

Sec. 3. Visualization

- Donut Sentiment Chart: The sentiment proportion of all comment data (green = positive, red = negative, orange = neutral).
- Sentiment Chart Bar: The number of comments based on sentiment.
- Comment Trend Chart Line per Day: The movement of the number of comments over time.

2. Technology used

Machine Learning & NLP (Natural Language Processing)

- Scikit-learn (joblib, tfidf_vectorizer, Naive Bayes, SVM pipeline)
- Sastrawi: Indonesian library for voting
- NLTK: For Indonesian stopwords
- HuggingFace Transformers API: A Sentiment-Classifier model for sentiment inference

Python Core Libraries

- os, re: For environment & regex access
- asyncio: To handle asynchronous crawling

External Service Integration

- **HuggingFace Inference API:** Used for model-based sentiment classification
 - **dotenv (python-dotenv):** Manage environment variables such as API keys and DB credentials
-

Data Handling

- **Pandas:** For manipulation kamuskatabaku.xlsx become a dictionary of standard words
 - **NumPy:** For numerical & sigmoid conversion operations SVM confidence
-

Web Framework & API Handling

- **Flask:** The backend main framework
 - **Flask-CORS:** To allow cross-origin requests (CORS)
 - **Flask JSON API endpoints:** CRUD operation and prediction/preprocessing endpoint
-

Database

- **PyMySQL:** Connection to MySQL database (access and update comments table)
-

Crawler (external)

- **crawl_and_return of the crawler module:** The `async` function to retrieve tweets (assumed from the homemade module)
-

Security

- **Separation of API keys and DB credentials using .env** (not written directly in the source code)

File I/O

- kamuskatabaku.xlsx: Excel file for non-standard → default word normalization
- .joblib files: Models and vectorizers stored for inference

Frontend Web Interface

- PHP: Used as server-side rendering and session management
- HTML + JavaScript (Vanilla): For interface and dynamic table display
- Bootstrap: For styling and UI elements (forms, alerts, tables, etc.)
- JavaScript Fetch API: To poll comment data periodically from the Flask API
- AJAX-like Behavior: Use fetch and DOM manipulation for live data updates
- cURL (PHP): For communication with the REST API Flask (crawl, preprocess, delete, etc.)

● How to Use the Product

Users can use SEJUDOL through an intuitive and easy-to-understand web interface. The process begins by crawling data from Twitter, where users determine the number of tweets they want to retrieve and then press the "Crawl" button. The system will automatically retrieve the tweet data according to the specified amount and store it in the database. Next, users can press the "Preprocessing" button to perform data cleansing, which includes symbol removal, normalization, stemming, and other stages in text processing. This process also classifies sentiment towards comments using machine learning models, and determines whether the comments include positive, negative, or neutral sentiments.

SEJUDOL also provides data deletion options, either individually through the "Delete" button in each row, or thoroughly with the "Delete All" button. The tables on the interface will automatically update and display the results of crawling, preprocessing, and classification in real-time. Each comment entry is presented complete with dates, original text, cleanup results, and sentiment labels.

Once the classification process is complete, users can access the Sentiment Analysis page to view a summary of the statistics visually. This summary includes the average confidence score of the model for each sentiment category—positive, negative, and neutral. In addition, information is also available about risk flags, which are the percentage of comments in each sentiment category that are considered to have potential risks related to online gambling content. SEJUDOL also displays the comments with the highest confidence level for positive and negative sentiments, along with their confidence scores. Data visualization is presented in the form of a donut chart to show the distribution of sentiment proportions, as well as a bar chart to show the number of comments in each category.