

# **Analisis Eksplorasi dan Prediksi Sentimen Terhadap Tweet Fufufafa Menggunakan Algoritma Naive Bayes**

Disusun dalam rangka memenuhi nilai UAS mata kuliah Sistem Temu Kembali Informasi

Dosen pengampu mata kuliah : Abu Salam, M.Kom



Oleh :

- ❖ Nama : FATU RAHMAT
- ❖ NIM : A11.2022.14831
- ❖ KELP : A11.4701~Sistem Temu Kembali Informasi

**PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS ILMU KOMPUTER  
UNIVERSITAS DIAN NUSWANTORO  
SEMARANG  
2025**

## **I. Ringkasan dan Permasalahan project**

### **Deskripsi:**

Kemunculan jejak digital akun kaskus fufufafa yang diyakini milik politikus sekaligus putra mantan Presiden Indonesia yakni Gibran Rakabuming Raka yang memuat postingan tidak pantas menjadi keresahan dan perdebatan di platform X atau Twitter. Oleh karena itu, saya melakukan analisis dan mengumpulkan data untuk diolah dan melatih model agar dapat memiliki luaran berupa prediksi teks baru yang bisa memperkirakan teks tersebut memiliki sentimen positif, negatif, atau netral. Naive Bayes dipilih karena sederhana, cepat memproses data teks, dan efektif untuk analisis sentimen.

### **Masalah :**

Masalah utama adalah ketidakseimbangan dalam data labeling dengan tiga kategori: netral, positif, dan negatif. Karena kasus ini terkait keresahan masyarakat, data cenderung didominasi oleh fitur negatif, yang menurunkan akurasi model. Untuk meningkatkan akurasi, perlu dilakukan improvisasi dengan menambahkan data relevan, khususnya terkait politik. Tujuan spesifik dari analisis ini meliputi:

- Membuat aplikasi prediksi sentimen terhadap teks inputan baru untuk menentukan apakah suatu teks memiliki sentimen negatif, netral, atau positif.
- Mengidentifikasi sentimen dominan dalam tweet Fufufafa.
- Mengamati tren topik yang sering muncul.

## **II. Tujuan yang akan dicapai + Model / Alur Penyelesaian (Buatkan dalam Bagan)**

### **Tujuan :**

- Membangun model klasifikasi sentimen yang akurat menggunakan algoritma Naive Bayes.
- Mendeploy model ke dalam aplikasi web menggunakan Streamlit untuk memprediksi sentimen dari input teks.

### **Alur :**

- a) Pengumpulan Dataset
- b) Preprocessing Data (Cleaning, Tokenization, Stopword Removal, Stemming)
- c) Feature Extraction (TF-IDF Vectorization)
- d) Pembuatan Model (Naive Bayes)
- e) Evaluasi Model (Akurasi, Recall, Precision, F1-Score)
- f) Deployment ke Streamlit

### III. Penjelasan Dataset, EDA dan Proses Features Dataset

#### Dataset:

Dataset diperoleh dari hasil scrapping dari platform X dan melakukan preprocessing serta labeling data, untuk meningkatkan performa dan akurasi model saya menambahkan dataset yang memiliki topik yang sama diambil dari kaggle dan disesuaikan dengan dataset awal.

<https://www.kaggle.com/datasets/waodeasriyani/sentiment1>

Dataset yang digunakan adalah clean\_dataset\_part01.csv diperoleh setelah preprocessing, yang berisi kolom:

tweet: Teks tweet yang akan dianalisis.

sentimen: Label sentimen (Negative: -1, Neutral: 0, Positive: 1).

#### Exploratory Data Analysis (EDA):

Distribusi sentimen:

- Negative: 26.8%
- Neutral: 21.6%
- Positive: 49.8%
- 

#### Proses Features Dataset:

Preprocessing:

- Membersihkan teks (menghapus tanda baca, stopwords, dll).
- Tokenization: Memecah teks menjadi kata-kata.

#### Feature Extraction:

- Menggunakan TfidfVectorizer untuk mengubah teks menjadi vektor numerik.

### IV. Proses Learning / Modeling

- Model yang digunakan: Naive Bayes (MultinomialNB).
- Proses pelatihan dilakukan dengan pembagian data 70% training dan 30% testing.
- Hyperparameter tuning: Optimasi alpha menggunakan GridSearchCV

### V. Performa Model

Accuracy Training set: 0.9954198473282443

Accuracy Test set: 0.6809015421115066

Recall Training set: 0.9954198473282443

Recall Test set: 0.6809015421115066

Precision Training set: 0.995445167158411

Precision Test set: 0.669481011149779

F1 Training set: 0.9954261877557283

F1 Test set: 0.6730521773875721

## **VI. Diskusi Hasil dan Kesimpulan**

Diskusi Hasil:

- Model Naive Bayes menunjukkan performa yang baik dengan akurasi 0.68.
- Visualisasi distribusi probabilitas menggunakan pie chart membantu pengguna memahami seberapa yakin model terhadap prediksinya.
- Ketidakseimbangan data memengaruhi hasil prediksi untuk kelas minoritas.
- Performa model cukup baik untuk data teks pendek seperti tweet.

Kesimpulan:

- Model Naive Bayes efektif untuk analisis sentimen pada dataset tweets.
- Penambahan data relevan dapat meningkatkan akurasi model.
- Aplikasi Streamlit yang dibuat memungkinkan pengguna untuk memprediksi sentimen dari teks secara interaktif.
- Deployment ke Streamlit Community Cloud memudahkan akses ke aplikasi secara online.