

LAPORAN UAS — DATA MINING / NLP

Pengembangan Chatbot Layanan Pegadaian (TF-IDF + Naive Bayes)

Nama	: (ISI NAMA ANDA)
Kelas	: (ISI KELAS ANDA)
Reguler	: (ISI REGULER)
Dosen	: Dr. Ir. Agung Budi Susanto, MM
Tahun Akademik	: SMT Ganjil 2025/2026

1. Deskripsi Proyek

Chatbot ini dibuat untuk menjawab pertanyaan pelanggan seputar layanan Pegadaian, seperti syarat gadai, gadai emas, tenor/jatuh tempo, perpanjangan, tabungan emas, pembayaran online, serta kanal keluhan. Model mampu mengenali kemiripan input pengguna walaupun tidak persis sama dengan data latih.

2. Dataset

Dataset disimpan dalam format JSON dengan minimal 20 intent (tag). Setiap intent memiliki beberapa contoh kalimat pertanyaan (patterns) dan satu atau lebih jawaban (responses).

No	Tag (Intent)	Jumlah Patterns
1	salam	6
2	jam_operasional	4
3	lokasi_cabang	4
4	syarat_gadai	4
5	gadai_emas	4
6	gadai_hp	4
7	gadai_bpkb	4
8	simulasi_pinjaman	4
9	biaya_admin	4
10	bunga_sewa_modal	4
11	jatuh_tempo	4
12	perpanjang_gadai	4
13	tebus_barang	4
14	cicilan_emas	4
15	tabungan_emas_cetak	4
16	pembayaran_online	4
17	cek_transaksi	4
18	keluhan	4
19	keamanan_data	4
20	terima_kasih	4

3. Preprocessing

Tahapan preprocessing yang digunakan: (1) case folding (lowercase), (2) menghapus tanda baca/karakter non alfanumerik, (3) normalisasi spasi. Tahap ini membantu menstabilkan fitur teks sebelum vektorisasi.

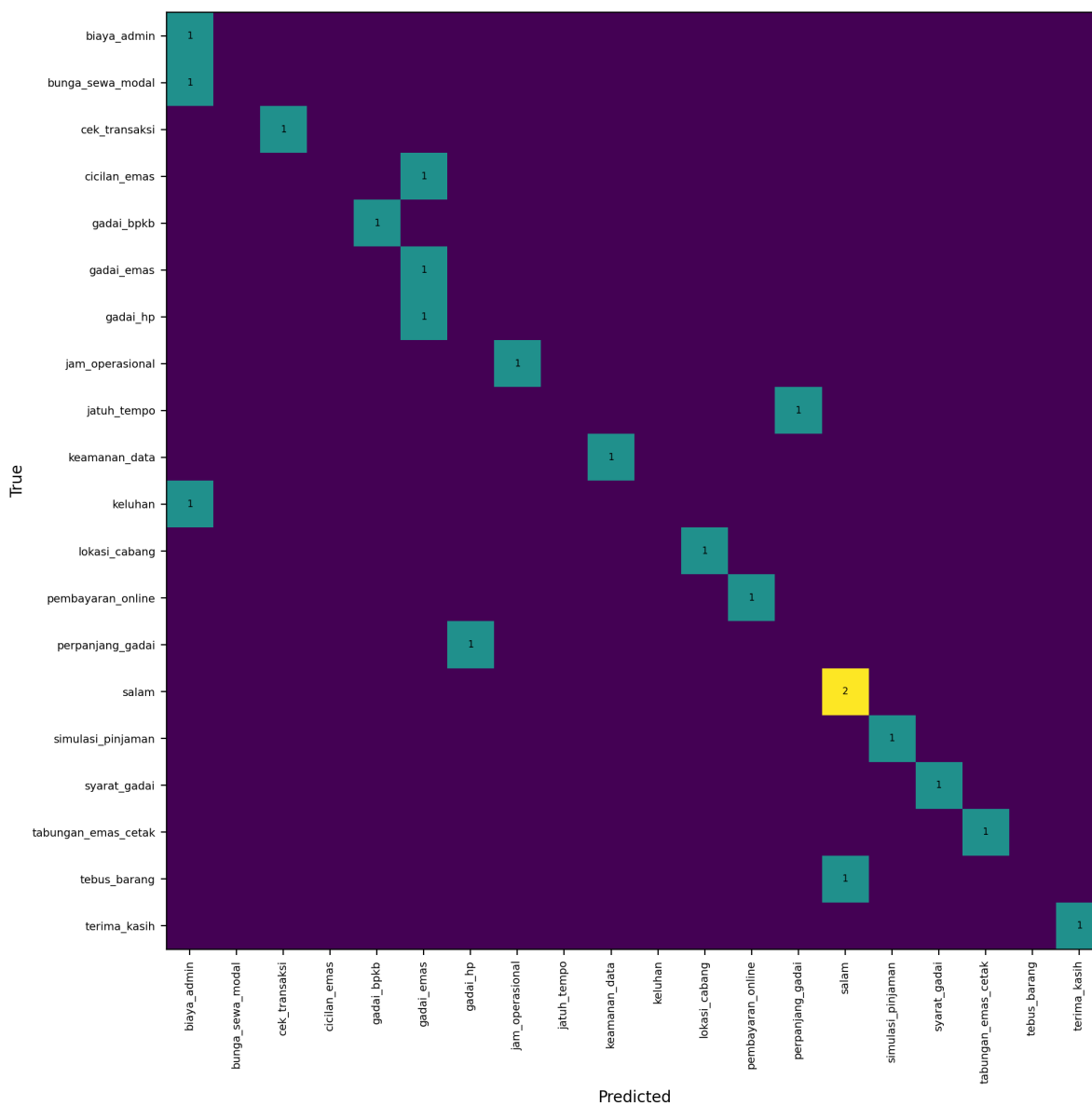
4. Algoritma (Supervised Learning)

Representasi teks menggunakan TF-IDF (ngram 1–2). Klasifikasi intent menggunakan Multinomial Naive Bayes. Hasil prediksi intent digunakan untuk memilih jawaban (response) yang sesuai dari dataset.

5. Evaluasi Akurasi

Evaluasi dilakukan dengan pembagian data (hold-out) 75% latih dan 25% uji (random_state=42, stratify label). Pada percobaan ini, akurasi hold-out sebesar 66.67%. Karena dataset relatif kecil, akurasi dapat meningkat jika variasi kalimat pada patterns diperbanyak.

Gambar 1. Confusion Matrix (Holdout Test)



6. Cara Menjalankan

a) Install dependensi:

```
pip install -r requirements.txt
```

b) Jalankan chatbot terminal:

```
python chatbot/chatbot.py
```

c) Jalankan dashboard Streamlit:

```
streamlit run dashboard/app.py
```

7. Kesimpulan

Pendekatan TF-IDF + Naive Bayes cukup efektif untuk chatbot berbasis intent dengan dataset kecil. Chatbot dapat menangkap maksud pertanyaan yang mirip (tidak harus persis sama). Peningkatan kualitas dapat dilakukan dengan menambah jumlah data, memperbanyak variasi patterns, serta melakukan tuning threshold confidence.

Tanda Tangan: _____