SENTIMENT CLASSIFICATION OF IMDB MOVIE REVIEWS USING IBM GRANITE

Aisya Mufidah Najwa

Hacktiv8 Capstone Project

DATASET

Sumber:
[IMDB Dataset of 50K Movie Reviews]

Deskripsi: Dataset untuk klasifikasi sentimen biner (positif/negatif) dengan 50.000 ulasan film.

	review	sentiment
0	One of the other reviewers has mentioned that ...	positive
1	A wonderful little production. The...	positive
2	I thought this was a wonderful way to spend ti...	positive
3	Basically there's a family where a little boy ...	negative
4	Petter Mattei's "Love in the Time of Money" is...	positive
...	...	...
49995	I thought this movie did a down right good job...	positive
49996	Bad plot, bad dialogue, bad acting, idiotic di...	negative
49997	I am a Catholic taught in parochial elementary...	negative
49998	I'm going to have to disagree with the previou...	negative
49999	No one expects the Star Trek movies to be high...	negative

50000 rows × 2 columns

PROJECT OVERVIEW

Proyek ini bertujuan untuk mengklasifikasikan sentimen (positif atau negatif) dari ulasan film dalam dataset IMDB menggunakan model AI IBM Granite 3.3-8B Instruct, sebuah Large Language Model (LLM). Hasil klasifikasi dibandingkan dengan label asli untuk mengukur akurasi dan performa model.

```python
from langchain_community.llms import Replicate

llm = Replicate(
    model="ibm-granite/granite-3.3-8b-instruct"
)
```

ANALYSIS PROCESS

1. Pengambilan dan Pemrosesan Data

- Dataset dimuat menggunakan pandas.
- Sampel acak sebanyak 100 ulasan diambil dari total 50.000 data untuk evaluasi.

2. Penggunaan IBM Granite LLM

- Model diakses melalui LangChain yang memungkinkan integrasi langsung dengan LLM.
- Setiap ulasan diklasifikasikan menggunakan prompt eksplisit berbasis teks.

Prompt yang digunakan:

```
def classify_sentiment(text):
    prompt = f"""
You are a sentiment classifier.

Classify the sentiment of the following movie review as either:
- positive
- negative

Respond with ONLY one word: either "positive" or "negative". No explanation.
No punctuation.

Review: {text}

Sentiment:
"""
    try:
        result = llm.invoke(prompt)
        return result.strip().lower()
    except Exception as e:
        return f"error: {str(e)}"
```

ANALYSIS PROCESS

3. Pembersihan Output

- Output model disaring agar hanya menyisakan hasil "positive" atau "negative".
- Pembersihan dilakukan menggunakan aturan berbasis regular expression (regex).

4. Evaluasi Akurasi

- Perbandingan antara label sentiment (asli) dan sentiment_ibm (hasil prediksi).
- Akurasi model pada sampel data mencapai 93%.

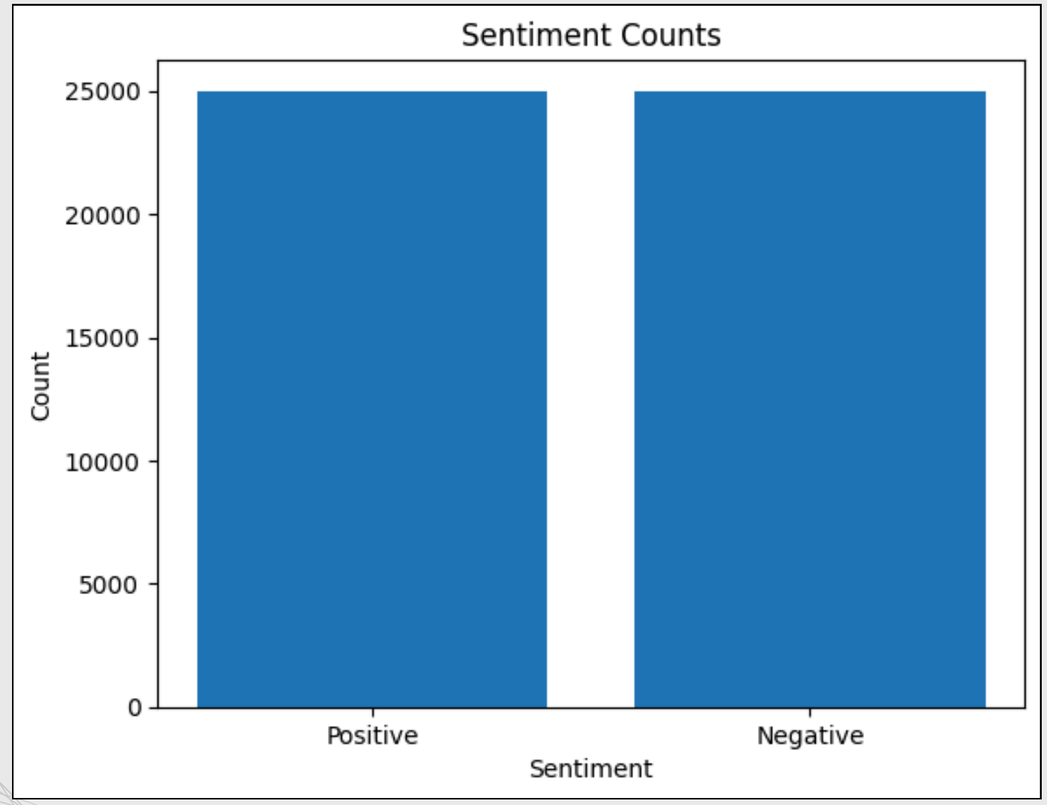
5. Visualisasi

Visualisasi data dilakukan dengan bantuan LangChain `pandas_dataframe_agent`, yaitu agent berbasis LLM yang dapat menjalankan instruksi analitik langsung terhadap `DataFrame`. Visualisasi yang dilakukan:

- Bar chart distribusi sentimen.
- Confusion matrix untuk mengevaluasi performa klasifikasi.
- Perbandingan visual sentimen asli dan hasil model.

RESULT

DISTRIBUSI SENTIMEN ASLI



Prompt:

```
agent.run("Tampilkan bar chart jumlah sentimen positif dan negatif di kolom sentiment")
```

Dataset IMDB terdiri dari 50.000 ulasan dengan jumlah seimbang:

- 25.000 positive
- 25.000 negative

Artinya model tidak bias dari awal (data balance) dan evaluasi model bisa lebih adil.

RESULT

AKURASI KLASIFIKASI

Setelah 100 sampel acak ulasan film dari dataset IMDB diklasifikasikan menggunakan IBM Granite LLM, hasil output model dibersihkan untuk memastikan hanya respons "positive" atau "negative" yang valid.

Hasil akurasi model pada 100 sampel: **93%**

Prompt:

```
agent2.run("Hitung akurasi prediksi sentimen_ibm terhadap label sentiment.")
```

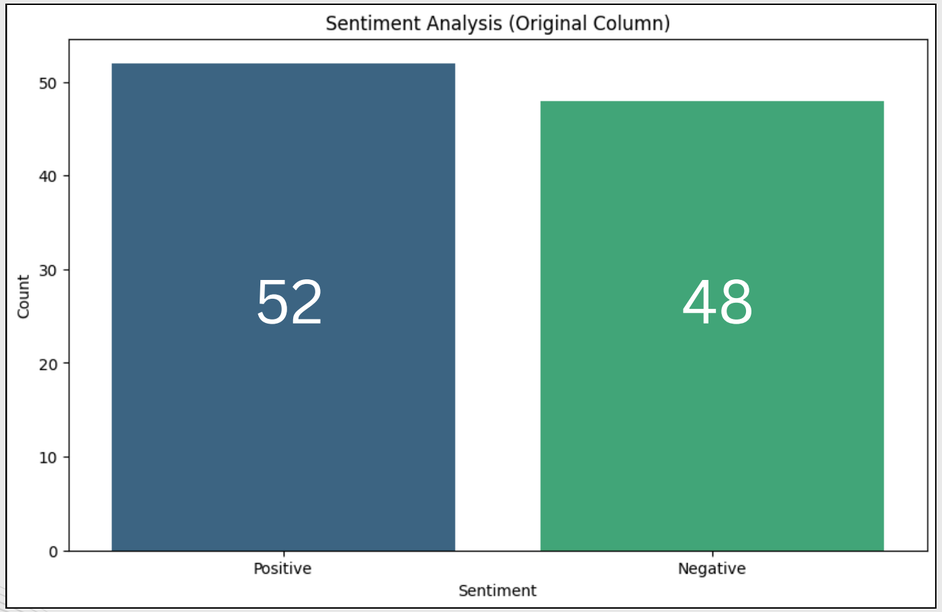
RESULT

PERBANDINGAN DISTRIBUSI PREDIKSI MODEL

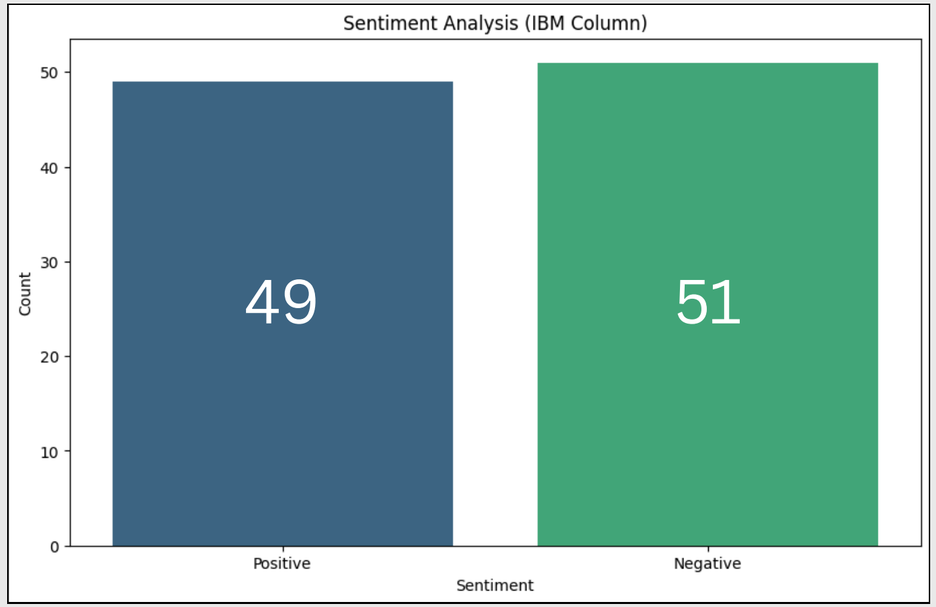
Prompt:

```
agent2.run(
    "Tampilkan perbandingan jumlah sentimen positif dan negatif antara kolom 'sentiment' dan 'sentiment_ibm' dalam bentuk bar chart. "
)
```

Data Asli (Sampel)

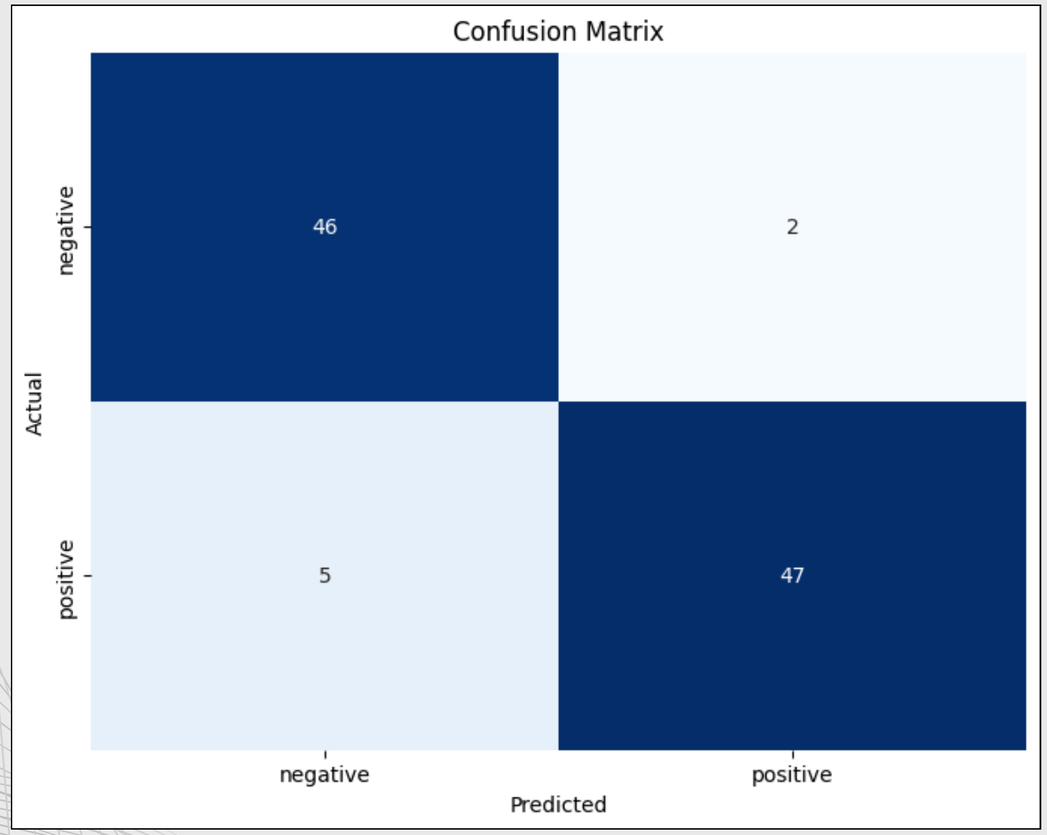


Data Prediksi



RESULT

CONFUSION MATRIX



	Predicted Negative	Predicted Positive
Actual Negative	46	2
Actual Positive	5	47

Prompt:

```
agent2.run("Buat confusion matrix sebagai visualisasi antara kolom
'sentiment' dan 'sentiment_ibm'. Gunakan matplotlib/seaborn, dan tampilkan
plot-nya dengan label positive dan negative.")
```

- Model sangat akurat dalam mengklasifikasi kedua kelas.
- Precision tinggi (96%) untuk positive → model cenderung hati-hati memberikan label positif.
- Recall tinggi (96%) untuk negative → model jarang melewatkan review yang negatif.

INSIGHT AND FINDINGS

- Distribusi Seimbang: Dataset asli memiliki distribusi seimbang antara sentimen positif dan negatif, cocok untuk evaluasi objektif.
- Akurasi Tinggi: Model IBM Granite berhasil mengklasifikasikan sentimen dengan akurasi 93% pada sampel data.
- Presisi dan recall tinggi untuk kedua label.
- Model cenderung sedikit lebih hati-hati dalam memprediksi positif dibanding negatif (Recall pos = 0.90 vs neg = 0.96).
- Model cenderung stabil dan tidak terlalu agresif, terbukti dari F1-score yang seimbang (0.93 untuk kedua kelas).
- Output Perlu Pembersihan: Meskipun prompt disusun dengan spesifik, model kadang tetap menghasilkan format seperti `<positive>` atau kalimat tambahan. Oleh karena itu, perlu proses pembersihan tambahan.
- Performa Stabil: Tidak ditemukan kecenderungan bias ekstrem terhadap salah satu kelas.

CONCLUSION

Model IBM Granite menunjukkan performa yang sangat baik dalam tugas klasifikasi sentimen terhadap dataset IMDB pada 100 sampel acak. Dengan akurasi 93%, serta presisi dan recall yang tinggi pada kedua label (positif dan negatif), model ini terbukti mampu melakukan analisis sentimen secara andal. Hasil evaluasi menunjukkan bahwa model bersifat seimbang dan tidak bias terhadap salah satu kelas, dengan F1-score yang merata. Namun, masih ditemukan beberapa output yang memerlukan proses pembersihan tambahan akibat ketidakkonsistenan format hasil dari model.

RECOMMENDATION

- Model IBM Granite cocok digunakan untuk analisis ulasan pelanggan dalam industri film, e-commerce, dan layanan konsumen.
- Karena hasil klasifikasi kadang tidak konsisten secara format, disarankan untuk menambahkan tahap *post-processing* untuk membersihkan output model sebelum dianalisis lebih lanjut.
- Untuk mendapatkan gambaran performa yang lebih representatif, evaluasi sebaiknya diperluas pada sampel yang lebih besar dan lebih bervariasi, termasuk data dari sumber lain seperti media sosial.
- Model ini dapat diintegrasikan ke dalam pipeline analisis opini secara otomatis untuk membantu pengambilan keputusan berbasis sentimen pelanggan.

AI SUPPORT EXPLANATION

Dalam proyek ini, Artificial Intelligence (AI) digunakan untuk melakukan klasifikasi sentimen menggunakan model IBM Granite 3.3-8B Instruct melalui integrasi LangChain.

1. Penyusunan Prompt Klasifikasi
   - Setiap ulasan dikirim ke model dengan instruksi eksplisit untuk mengklasifikasikan sebagai “positive” atau “negative”.
   - Prompt disusun agar model hanya menghasilkan satu kata sebagai output.

2. Eksekusi dengan LangChain Agent
   - Digunakan `create_pandas_dataframe_agent()` untuk membuat agen yang dapat berinteraksi langsung dengan DataFrame.
   - Agen digunakan untuk menjalankan analisis seperti visualisasi, evaluasi akurasi, pembuatan confusion matrix, hingga insight otomatis.

AI BENEFITS

- Akurasi tinggi: Model mampu mencapai akurasi 93% pada sampel data.
- Efisiensi waktu: Proses klasifikasi dan analisis dilakukan secara otomatis dan cepat.
- Fleksibilitas: Model dapat dijalankan melalui perintah natural language untuk tugas-tugas analitik seperti klasifikasi, visualisasi, dan rekomendasi.
- Lebih sederhana dibanding pendekatan NLP tradisional: Dengan menggunakan LLM seperti IBM Granite, proses klasifikasi tidak memerlukan tahapan-tahapan manual seperti preprocessing teks, tokenisasi, ekstraksi fitur, dan pelatihan model secara eksplisit. Cukup dengan menyusun prompt yang tepat, hasil klasifikasi bisa langsung diperoleh dari model, sehingga mempermudah dan mempercepat workflow analisis.

THANK YOU