



UNIVERSITAS
GADJAH MADA

Sentiment Analysis

Natural Language Processing Assignment

Kelompok



UNIVERSITAS GADJAH MADA

M Rifat Bagus A 20/456371/TK/50501

Nismara Chandra H 20/456374/TK/50504

- Saat ini antusiasme masyarakat terhadap hiburan perfilman sedang dalam trend yang terus meningkat.
- Tingkat pencarian dan review rating film oleh masyarakat maupun stakeholder dunia perfilman juga turut meningkat.
- Saat ini, lazimnya kita dapat melihat rating film dengan melihat komentar dari user lain atau dengan parameter 'bintang'
- Metode 'bintang' dapat dengan cepat menampilkan penilaian kumulatif dari semua review namun metode komentar tidak
- Diperlukan suatu cara untuk “merangkum” komentar pengguna untuk menentukan apakah suatu film disukai atau tidak

Deskripsi Aplikasi



UNIVERSITAS GADJAH MADA

- Aplikasi yang dikembangkan berbasis google colab dan bahasa pemrograman python3
- Aplikasi dapat “membaca” setiap review pengguna lalu dapat mengkategorisasikan komentar tersebut
- Sementara aplikasi hanya dapat mengkategorikan dalam kategori review positif atau review negatif
- Aplikasi ini dapat “merangkum” review penonton terhadap suatu film, lalu menyimpulkan apakah film tersebut disukai atau tidak disukai

Basic Data Preprocessing



UNIVERSITAS GADJAH MADA

- Menghilangkan HTML tag
- Menghilangkan Karakter dalam square bracket
- Menghilangkan stopwords
- Menghilangkan url

Library yang Digunakan



UNIVERSITAS GADJAH MADA

- Simple Transformer
- Panda
- Numpy
- NLTK

Sumber Dataset (Training, Validation & Testing)



UNIVERSITAS GADJAH MADA

kaggle.com/datasets/columbine/imdb-dataset-sentiment-analysis-in-csv-format

Data berupa kumpulan review terhadap film-film yang berasal dari laman IMDB

Penjelasan Garis Besar Kode



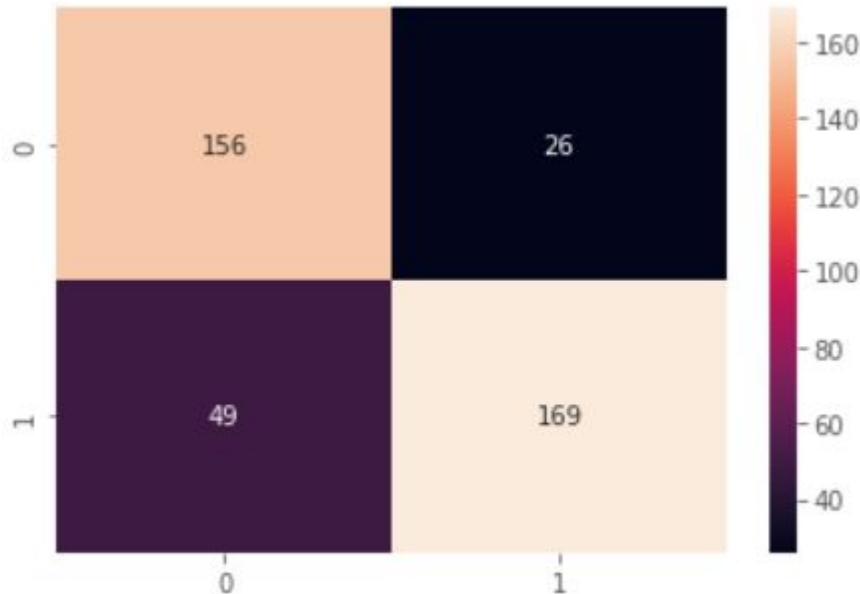
UNIVERSITAS GADJAH MADA

Evaluasi Model



UNIVERSITAS GADJAH MADA

Hasil evaluasi model setelah train menggunakan dataset eval.



Positive :

- Precision :76%
- Recall :85.7%

Negative :

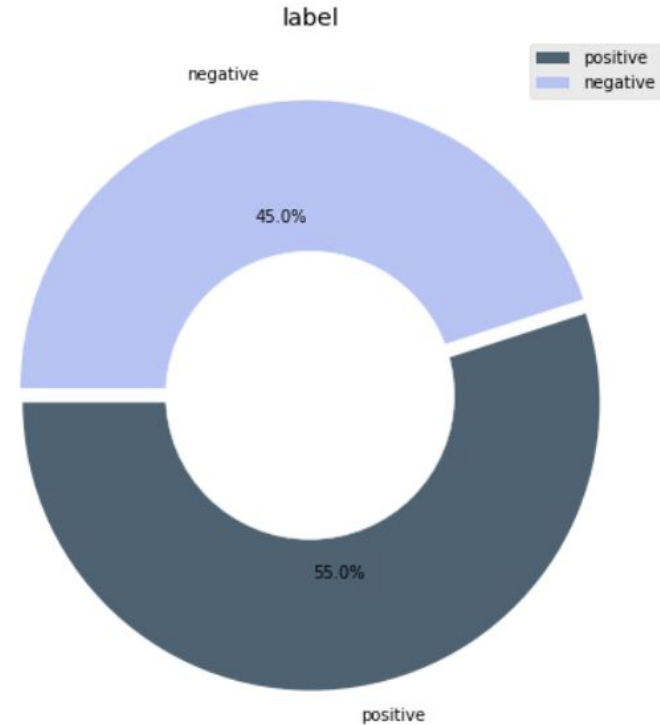
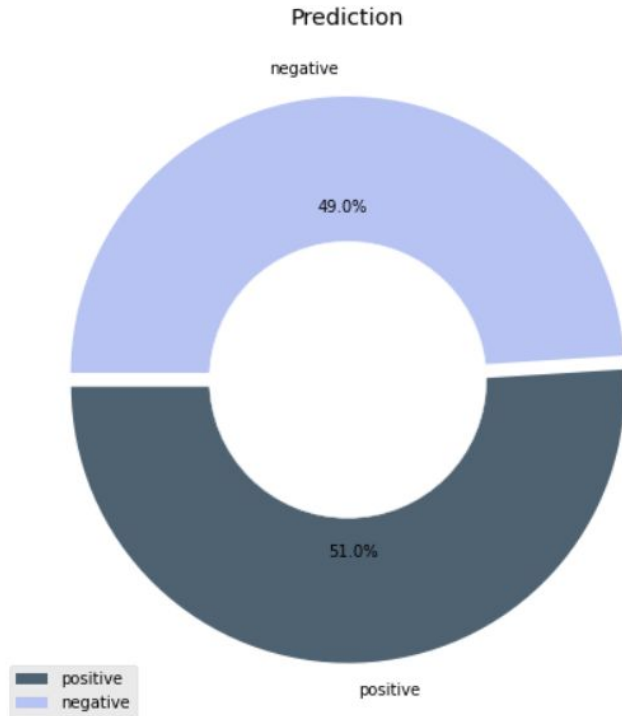
- Positive :86.67%
- Recall :77.54%

Model Accuracy : 81.25%

Hasil Prediksi pada TestDataset



UNIVERSITAS GADJAH MADA



Kesimpulan



UNIVERSITAS GADJAH MADA

Evaluation test pada model memberikan hasil precision dan recall diatas 75% bahkan menyentuh angka diatas 85%. Angka-angka tersebut menunjukkan bahwa model tidak under-predicting maupun over-predicting.