Sentiment Analysis for Streaming Application

This project will analyze review streaming applications that is widely used by the Indonesian people. The streaming applications are We TV, Netflix, Viu, Vidio, and Disney+ Hotstar. In detail, the objective of this project:

1. Discovering the number of downloads and ratings from users to each application globally
2. Discovering and classifying the opinions from users about each application
3. Finding positive and negative keywords in each application to be considered by streaming application companies in improving their application quality

Dataset

Data used in this project is applications data, and review data given by users from Google Play Store and Apple Store. Data extracted using Python library such as google-play-scraper and app-store-scraper.

Feature extracted from applications data:

1. The number of downloads
2. Ratings globally
3. The number of each ratings (rating 1 to 5)
4. The number of users which gives rating
5. The number of reviews

review\_id','date','app\_name','version','source','language','country','content','score','sentiment\_score','sentiment\_label'

Feature extracted from reviews data in Google Play Store:

1. ReviewId: unique identifier of reviews
2. At: date of review created
3. Content: the reviews from user
4. Score: the score from user

Feature extracted from reviews data in Apple Store:

1. Date: date of review
2. Rating:
3. Title
4. Review

Data yang digunakan pada projek ini merupakan data jumlah unduhan, rating, dan ulasan yang diberikan oleh pengguna aplikasi streaming yang berasal dari Google Play Store dan Apple Store. Data diambil menggunakan library Python google-play-scraper dan app-scraper.

Fitur yang diambil dari data aplikasi pada Google Play Store yaitu:

Fitur yang diambil dari data ulasan pada Google Play Store yaitu:

Fitur yang diambil dari data ulasan pada Apple Store yaitu:

ETL Process

Data yang telah diextract dari Google Play Store dan Apple Store berupa semi-structured data. Data tersebut disatukan dan ditransform menjadi structured menggunakan Pandas dan disimpan ke dalam Postgresql agar dapat diakses dan digunakan untuk analisis di waktu masa depan.

Sentiment Analysis

Proses analisis sentiment pada projek ini menggunakan TextBlob dan Google Translator API untuk mengklasifikasikan ulasan ke dalam kelas positif, netral, maupun negatif.

EDA