

## **Rencana Judul Proyek**

### **Perbandingan Analisis Sentimen Aplikasi Lazada, Shopee dan Tokopedia**

#### **Ringkasan**

Proyek ini bertujuan untuk melihat perbandingan analisis sentimen dari review pengguna pada aplikasi Lazada, Shopee dan Tokopedia di Google Play Store. Masalah yang ingin dipecahkan adalah untuk melihat bagaimana perbandingan analisis sentiment pengguna terhadap ketiga aplikasi tersebut, sehingga dapat memberikan wawasan berharga bagi pengembang dan pemasaran. Manfaat yang diharapkan adalah kemampuan untuk mengidentifikasi kekuatan dan kelemahan aplikasi, tuntutan atau harapan yang diharapkan oleh pengguna aplikasi dan pembuatan model sentiment analysis berdasarkan ulasan pengguna serta meningkatkan pengalaman pengguna secara keseluruhan.

#### **Latar Belakang**

Dalam industri aplikasi, memahami feedback pengguna adalah kunci untuk pengembangan dan pemasaran yang efektif. Review dari pengguna di Google Play Store dapat memberikan wawasan langsung tentang pengalaman, kepuasan pengguna dan ekspektasi pengguna layanan. Namun, dengan volume review yang sangat besar, analisis manual menjadi tidak efisien. Sentiment analysis menggunakan teknik machine learning dapat membantu mengidentifikasi pola dan tren dalam data review, memberikan pandangan yang lebih mendalam tentang opini pengguna.

#### **Tujuan Proyek**

1. Perbandingan analisis sentimen dari review aplikasi Lazada, shopee dan tokopedia di Google Play Store.
2. Mengidentifikasi aspek-aspek yang paling sering dibicarakan pengguna dan sentimen yang terkait dengan masing-masing aplikasi.
3. Memberikan rekomendasi kepada pengembang aplikasi untuk perbaikan dan pengembangan fitur berdasarkan analisis sentimen.
4. Membuat model analisis sentiment.

#### **Metodologi**

##### **1. Pengumpulan Data:**

- Dataset berasal dari web scraping dari website Goole Play Store menggunakan library python google\_play\_scraper.
- Dataset akan terdiri dari teks review, rating, dan metadata terkait lainnya.

##### **2. Pemrosesan Data:**

- Pembersihan data: menghapus duplikat, menghilangkan karakter khusus, dan melakukan normalisasi teks.
- Tokenisasi dan lemmatization untuk mempersiapkan data teks untuk analisis lebih lanjut.
- Pembersihan spam review.

### 3. Analisis Sentimen:

- Menggunakan library nlp-id dan Vader lexicon untuk analisis awal sentimen.
- Model machine learning seperti Logistic Regression, Random Forest dan Naïve bayes.

### 4. Identifikasi Aspek:

- Menggunakan teknik NLP untuk mengidentifikasi aspek-aspek yang dibicarakan dalam review (misalnya, kecepatan, fitur, bug).
- Analisis sentimen pada masing-masing aspek untuk memahami sentimen terkait.

### 5. Evaluasi dan Penyempurnaan:

- Evaluasi model menggunakan metrik seperti akurasi, precision, recall, dan F1-score.
- Penyempurnaan model berdasarkan hasil evaluasi untuk meningkatkan performa.

### Teknologi yang Digunakan:

- Bahasa: Python 3.x
- Library: google\_play\_scraper, scikit-learn (<=1.5.1), nlp-id, vader lexicon, pysastrawi
- IDE: VS Code
- Streamlit

### Pemilihan Algoritma dan Model

#### 1. Model Baseline:

- Nlp-id, Vader lexicon, pysastrawi untuk analisis awal sentiment.

#### 2. Model Machine Learning:

- **Logistic Regression:** Sederhana dan efektif untuk klasifikasi teks.
- **Random Forest:** Memiliki kemampuan untuk menangani dataset besar dan memberikan hasil yang dapat diinterpretasikan.
- **Naïve Bayes :** algoritma pembelajaran mesin yang sering digunakan untuk tugas klasifikasi, terutama karena kesederhanaannya dan kinerja yang baik pada data berukuran besar.

### Proses Pelatihan Model dan Parameter:

- **Logistic Regression:** Grid search untuk hyperparameter seperti regularization strength.
- **Random Forest:** Penyetelan jumlah pohon dan kedalaman maksimum melalui grid search.
- **Naïve bayes :** Penyetelan parameter smoothing dengan grid search