

INTISARI

Pemodelan Topik Menggunakan Kombinasi *Latent Dirichlet Allocation* dan BERT *Embedding* (Studi Kasus: Ulasan Aplikasi Mobile Anteraja)

Oleh

Balqis Ima Khariyah

21/479860/PA/20819

Perkembangan teknologi digital kini telah mengubah kebiasaan belanja masyarakat menjadi serba *online*. Hal itu meningkatkan pertumbuhan industri *e-commerce* dan permintaan terhadap layanan pengiriman berbasis aplikasi, seperti Anteraja. Namun, pada tahun 2023 Anteraja mengalami penurunan pendapatan seiring dengan aktivitas belanja masyarakat yang kembali normal setelah Covid-19. Untuk mengetahui penyebab penurunan tersebut, diperlukan adanya evaluasi secara internal. Salah satu sumber data yang relevan digunakan adalah ulasan pengguna aplikasi Anteraja yang tersedia di Google Play Store. Penelitian ini bertujuan untuk mengidentifikasi topik-topik utama dalam ulasan pengguna Anteraja. Sebanyak 6451 ulasan dikumpulkan dari Google Play Store sejak Januari 2023 hingga November 2024. Pemodelan dilakukan dengan mengombinasikan algoritma *Latent Dirichlet Allocation* (LDA) dan representasi kalimat *Sentence-BERT* (SBERT), dilanjutkan dengan reduksi dimensi menggunakan UMAP dan klusterisasi dengan algoritma *KMeans*. Model terbaik adalah kombinasi antara LDA dan SBERT yang menghasilkan *Silhouette Coefficient* sebesar 0,7342 dengan tiga kluster optimal. Hasil interpretasi topik menunjukkan bahwa pengguna menyoroti tiga aspek utama, yaitu: layanan pelanggan dan masalah teknis, kecepatan serta efisiensi pengiriman, dan pengalaman positif pelanggan.

ABSTRACT

Hybrid Topic Modeling Using Latent Dirichlet Allocation and BERT Embedding (Case Study: Anteraja Mobile App Reviews)

By

Balqis Ima Khariyah

21/479860/PA/20819

The growth of digital technology has changed people's shopping habits to become more online. This trend has helped the e-commerce industry grow and increased the demand for delivery services that use mobile apps, such as Anteraja. However, in 2023, Anteraja's revenue dropped as people returned to normal shopping activities after the Covid-19 pandemic. To understand the reason for this decline, an internal evaluation is needed. One useful source of data is user reviews of the Anteraja app on the Google Play Store. This study aims to find the main topics in user reviews of Anteraja. A total of 6451 reviews were collected from the Google Play Store between January 2023 and November 2024. The analysis combined the Latent Dirichlet Allocation (LDA) algorithm with sentence representations from Sentence-BERT (SBERT), followed by dimensionality reduction using UMAP and clustering using the KMeans algorithm. The best model is the combination of LDA and SBERT which produces a Silhouette Coefficient of 0.7342 with three optimal clusters. The results showed that users mostly talked about three main things: customer service and technical problems, delivery speed and efficiency, and positive experiences with the service.