

BIDIRECTIONAL ENCODER REPRESENTATION FROM TRANSFORMERS DALAM MEMPREDIKSI MISSING TEXT USER EXPERIENCE PADA PLATFORM AIRLANE REVIEW

Muhammad Rafi Al Anbiya, 202231089
Dibimbing oleh Dine Tiara Kusuma ST, M.Kom.

ABSTRAK

Perkembangan platform digital telah meningkatkan jumlah ulasan pengguna (*user review*) pada industri penerbangan yang dapat dimanfaatkan untuk analisis pengalaman pengguna (*user experience*). Dalam *Natural Language Processing* (NLP), kemampuan model bahasa dalam memahami konteks dan memprediksi token yang disembunyikan menjadi indikator penting kualitas representasi bahasa. Penelitian ini bertujuan untuk menganalisis kemampuan model Bidirectional Encoder Representations from Transformers (BERT) dalam memprediksi teks yang di-mask (*missing text*) pada ulasan pengguna maskapai penerbangan melalui pendekatan Masked Language Modeling (MLM). Penelitian ini menggunakan metode eksperimental dengan model pra-latih *bert-base-uncased* yang dilakukan *fine-tuning* pada 600 data ulasan yang diperoleh melalui *web scraping* dari platform Airline Review (Skytrax). Data diproses melalui tahap *preprocessing*, tokenisasi, dan pembentukan data MLM, kemudian dibagi menjadi data latih dan validasi dengan rasio 80:20. Evaluasi dilakukan menggunakan *validation loss* dan metrik ROUGE-1. Hasil penelitian menunjukkan nilai *validation loss* sebesar 3.45 serta skor ROUGE-1 yang tinggi, yang mengindikasikan bahwa model mampu menghasilkan prediksi token yang relevan dan sesuai konteks kalimat. Secara keseluruhan, model BERT menunjukkan kemampuan yang memadai dalam merekonstruksi token yang di-mask pada domain ulasan maskapai penerbangan.

Kata Kunci: BERT, Masked Language Modeling, Missing Text, User Experience, Airline Review.

BIDIRECTIONAL ENCODER REPRESENTATION FROM TRANSFORMERS DALAM MEMPREDIKSI MISSING TEXT USER EXPERIENCE PADA PLATFORM AIRLANE REVIEW

Muhammad Rafi Al Anbiya, 202231089
Supervised by Dine Tiara Kusuma ST, M.Kom.

ABSTRACT

The growth of digital platforms has increased the volume of user reviews in the airline industry, which can be utilized for analyzing user experience. In the field of Natural Language Processing (NLP), a language model's ability to understand context and predict masked tokens serves as an important indicator of its representational capability. This study aims to analyze the performance of the Bidirectional Encoder Representations from Transformers (BERT) model in predicting masked text within airline user reviews using the Masked Language Modeling (MLM) approach. This research employs an experimental method using the pre-trained bert-base-uncased model, which is fine-tuned on a dataset of 600 airline reviews collected through web scraping from the Airline Review (Skytrax) platform. The data undergo preprocessing, tokenization, and MLM data construction before being divided into training and validation sets with an 80:20 ratio. Model performance is evaluated using validation loss and the ROUGE-1 metric. The evaluation results show a validation loss of 3.45 and a high ROUGE-1 score, indicating that the model is capable of generating contextually relevant token predictions. Overall, the BERT model demonstrates adequate performance in reconstructing masked tokens within airline review texts.

Keywords: BERT, Masked Language Modeling, Missing Text, User Experience, Airline Review.