

## INTISARI

Penerapan gamifikasi dalam aplikasi pembelajaran digital telah menjadi instrumen vital untuk peningkatan motivasi dan keterlibatan siswa. Namun, evaluasi efektivitas aplikasi ini sering kali terkendala oleh metode konvensional seperti survei yang memiliki keterbatasan dalam skalabilitas, rentan terhadap bias subjektivitas, dan tidak mampu menangkap umpan balik secara *real-time*. Di sisi lain, volume besar ulasan pengguna di *platform* distribusi aplikasi menyediakan data autentik yang belum dimanfaatkan secara optimal. Permasalahan mendasar yang dihadapi saat ini adalah sulitnya mengidentifikasi persepsi pengguna secara akurat dari data teks tidak terstruktur yang kaya akan bahasa informal, serta minimnya validasi pedagogis dalam pendekatan komputasi terdahulu. Akibatnya, hasil analisis sering kali kurang relevan untuk dijadikan dasar evaluasi kualitas pembelajaran yang objektif.

Penelitian ini bertujuan untuk mengembangkan kerangka evaluasi berbasis data yang mampu memetakan topik permasalahan dan mengklasifikasikan sentimen pengguna secara presisi. Studi ini memanfaatkan dataset ulasan dari tiga aplikasi representatif (*Duolingo, Kahoot, Quizizz*) di *Google Play Store*. Metode yang diusulkan menggabungkan pendekatan *Educational Data Mining* (EDM) dengan teknik hibrida *Latent Dirichlet Allocation* (LDA) untuk pemodelan topik dan *Robustly Optimized BERT Pretraining Approach* (RoBERTa) untuk analisis sentimen. Kebaruan utama (*novelty*) penelitian ini terletak pada integrasi validasi *Subject Matter Expert* (SME) untuk memverifikasi relevansi pedagogis dari hasil klasifikasi mesin, menjembatani kesenjangan antara akurasi algoritmik dan interpretasi pendidikan.

Hasil penelitian menunjukkan bahwa model RoBERTa berhasil mencapai akurasi sebesar 94,17%, mengungguli model baseline tradisional dalam menangani kompleksitas bahasa ulasan. Analisis LDA berhasil mengidentifikasi 21 topik dominan yang mencakup aspek teknis hingga pengalaman belajar (*Learning Fun*), memberikan wawasan granular mengenai elemen gamifikasi yang paling berpengaruh.

Validasi pakar menghasilkan nilai kesepakatan *Fleiss' Kappa* sebesar 0,8499 (*Almost Perfect Agreement*), menegaskan bahwa model yang dikembangkan memiliki reliabilitas tinggi. Temuan ini mengimplikasikan bahwa integrasi analisis sentimen berbasis *Deep Learning* dan validasi pakar dapat menjadi solusi efektif bagi pengembang dan pendidik dalam merancang ekosistem pembelajaran digital yang lebih adaptif, responsif, dan mendukung keberhasilan akademik secara berkelanjutan.

**Kata kunci:** Gamifikasi, Analisis Sentimen, *Educational Data Mining*, LDA, RoBERTa, Validasi SME.

## ABSTRACT

The implementation of gamification in digital learning applications has emerged as a vital instrument to support student motivation and engagement. However, the evaluation of these applications' effectiveness is often constrained by conventional methods such as surveys, which are limited in scalability, susceptible to subjective bias, and unable to capture real-time feedback. Conversely, the vast volume of user reviews on application distribution platforms offers authentic data that remains underutilized. The fundamental challenge currently faced is the difficulty in accurately identifying user perceptions from unstructured text data laden with informal language, coupled with the scarcity of pedagogical validation in prior computational approaches. Consequently, analysis results are often insufficiently relevant to serve as a basis for objective evaluation of learning quality.

This research aims to develop a data-driven evaluation framework capable of mapping problem topics and classifying user sentiments with precision. This study utilizes a dataset of reviews from three representative applications (Duolingo, Kahoot, and Quizizz) on the Google Play Store. The proposed method integrates an Educational Data Mining (EDM) approach with a hybrid technique of Latent Dirichlet Allocation (LDA) for topic modeling and the Robustly Optimized BERT Pretraining Approach (RoBERTa) for sentiment analysis. The primary novelty of this research lies in the integration of Subject Matter Expert (SME) validation to verify the pedagogical relevance of machine classification results, thereby bridging the gap between algorithmic accuracy and educational interpretation.

The results indicate that the RoBERTa model achieved an accuracy of 94.17%, outperforming traditional baseline models in handling the linguistic complexity of reviews. LDA analysis successfully identified 21 dominant topics ranging from technical aspects to learning experiences (Learning Fun), providing granular insights into the most influential gamification elements. Expert validation yielded a Fleiss' Kappa agreement score of 0.8499 (Almost Perfect Agreement), confirming the high reliability of the developed model. These findings imply that the

integration of Deep Learning-based sentiment analysis and expert validation offers an effective solution for developers and educators to design digital learning ecosystems that are more adaptive, responsive, and supportive of sustainable academic success.

**Keywords** – Gamification, Sentiment Analysis, Educational Data Mining, LDA, RoBERTa, SME Validation.