

KLASIFIKASI SENTIMEN ULASAN GAME BUS SIMULATOR INDONESIA DENGAN PENDEKATAN TF-IDF DAN NAÏVE BAYES

ABSTRAK

Bus Simulator Indonesia (BUSSID) merupakan salah satu produk lokal unggulan dalam industri hiburan digital, khususnya *game* berbasis perangkat *mobile*. Penelitian ini bertujuan mengklasifikasikan sentimen ulasan pengguna menjadi kategori positif, negatif, dan netral menggunakan metode *Term Frequency-Inverse Document Frequency* (TF-IDF) dan algoritma *Naïve Bayes*. Data diperoleh dari ekstraksi data pada Google Play Store. Tahap *preprocessing data* meliputi *cleaning*, *case folding*, normalisasi, *tokenizing*, dan *stemming*. Pelabelan data menggunakan kamus *Indonesia Sentiment Lexicon* untuk menentukan kategori sentimen awal. Proses klasifikasi memanfaatkan pembobotan TF-IDF untuk menentukan bobot kata dalam korpus. Hasil penelitian menunjukkan bahwa pendekatan TF-IDF yang dikombinasikan dengan algoritma *Naïve Bayes* efektif dalam mengklasifikasikan sentimen ulasan *game* BUSSID. Pada skenario data terbaik (70:30), model mencapai akurasi sebesar 89,92%. Performa model pada kelas positif memperoleh *precision* 82%; *recall* 98%; dan *f1-score* 89%. Kelas negatif memiliki *precision* 98%; *recall* 90%; dan *f1-score* 94%, sementara kelas netral menunjukkan *precision* 95%; *recall* 73%; dan *f1-score* 83%.

COMMENTARY SENTIMENT CLASSIFICATION OF INDONESIAN BUS SIMULATOR GAME USING TF-IDF AND NAÏVE BAYES APPROACHES

ABSTRACT

Bus Simulator Indonesia (BUSSID) is one of the leading local products in the digital entertainment industry, particularly in mobile-based games. This study aims to classify user review sentiments into positive, negative, and neutral categories using the Term Frequency–Inverse Document Frequency (TF-IDF) method and the Naïve Bayes algorithm. The data were obtained through extraction from the Google Play Store. The data preprocessing stages include cleaning, case folding, normalization, tokenizing, and stemming. Data labeling was performed using the Indonesian Sentiment Lexicon dictionary to determine the initial sentiment categories. The classification process utilizes TF-IDF weighting to determine the importance of words within the corpus. The results of the study indicate that the TF-IDF approach combined with the Naïve Bayes algorithm is effective in classifying sentiment in BUSSID game reviews. In the best data scenario 70:30 split), the model achieved an accuracy of 89.92%. The model's performance for the positive class obtained a precision of 82%, a recall of 98%, and an F1-score of 89%. The negative class achieved a precision of 98%, recall of 90%, and an F1-score of 94%, while the neutral class showed a precision of 95%, recall of 73%, and an F1-score of 83%.