ABSTRAK

Ulasan pelanggan merupakan salah satu komponen penting dalam ekosistem e-commerce yang berperan dalam membantu konsumen mengambil keputusan pembelian. Namun, meningkatnya jumlah ulasan palsu (fake reviews) menjadi permasalahan serius karena dapat menyesatkan konsumen serta menurunkan tingkat kepercayaan terhadap platform digital. Oleh karena itu, diperlukan metode deteksi yang tidak hanya memiliki performa tinggi pada data latih, tetapi juga mampu melakukan generalisasi dengan baik pada data dari distribusi yang berbeda.

Penelitian ini bertujuan untuk mengevaluasi pengaruh teknik sampling pada proses generasi teks, khususnya Top-k dan Top-p, terhadap performa serta kemampuan generalisasi model berbasis transformer, yaitu RoBERTa, dalam mendeteksi ulasan palsu pada skenario cross-dataset. Dataset sumber yang digunakan adalah dataset Salminen yang berisi ulasan asli dan ulasan palsu berbasis Top-k. Sementara itu, dataset pembanding dibentuk melalui generasi ulasan palsu menggunakan model GPT-2 dengan teknik Top-p. Dataset Ott digunakan sebagai data target dalam pengujian cross-dataset untuk menguji kemampuan generalisasi model terhadap distribusi data yang berbeda.

Tahapan penelitian meliputi proses generasi teks sintetis, preprocessing ringan, pembagian data dengan rasio 80:20 untuk data latih dan validasi, tokenisasi menggunakan tokenizer RoBERTa, serta proses fine-tuning model roberta-base. Evaluasi model dilakukan menggunakan berbagai metrik kinerja, yaitu accuracy, precision, recall, F1-score, confusion matrix, serta persentase accuracy yang dipertahankan untuk mengukur tingkat penurunan performa pada skenario cross-dataset.

Hasil pengujian intra-dataset menunjukkan bahwa model RoBERTa dengan teknik Top-k memiliki performa yang lebih baik, dengan nilai accuracy sebesar 0,956844 dan F1-score sebesar 0,958428, dibandingkan dengan model Top-p yang memperoleh accuracy sebesar 0,920490 dan F1-score sebesar 0,925742. Namun, pada pengujian cross-dataset, model RoBERTa Top-p menunjukkan kemampuan generalisasi yang lebih baik dengan accuracy sebesar 0,701250 dan F1-score sebesar 0,641679, lebih tinggi dibandingkan model Top-k yang memperoleh accuracy sebesar 0,648125 dan F1-score sebesar 0,491418. Selain itu, model Top-p mampu mempertahankan accuracy sebesar 76,18%, lebih tinggi dibandingkan Top-k yang hanya sebesar 67,74%.

Berdasarkan hasil tersebut, dapat disimpulkan bahwa teknik Top-k lebih efektif dalam kondisi distribusi data yang sama (intra-dataset), sedangkan teknik Top-p lebih unggul dalam meningkatkan kemampuan generalisasi model terhadap perubahan distribusi data (cross-dataset).

Kata kunci: ulasan palsu, e-commerce, RoBERTa, GPT-2, Top-k, Top-p, cross-dataset, generalisasi model, klasifikasi teks

ABSTRACT

Customer reviews are a crucial component of the e-commerce ecosystem, as they assist consumers in making purchasing decisions. However, the increasing presence of fake reviews has become a significant issue, as it can mislead consumers and reduce trust in digital platforms. Therefore, it is necessary to develop detection methods that not only achieve high performance on training data but also demonstrate strong generalization capabilities across different data distributions.

This study aims to evaluate the effect of text generation sampling techniques, specifically Top-k and Top-p, on the performance and generalization ability of a transformer-based model, RoBERTa, for fake review detection in a cross-dataset scenario. The source dataset used in this research is the Salminen dataset, which consists of genuine reviews and Top-k-based fake reviews. Meanwhile, the comparison dataset is generated using GPT-2 with the Top-p sampling technique. The Ott dataset is employed as the target dataset for cross-dataset evaluation to assess the model's robustness against distribution shifts.

The research methodology includes synthetic text generation, light preprocessing, data splitting with an 80:20 ratio for training and validation, tokenization using the RoBERTa tokenizer, and fine-tuning of the roberta-base model. Model evaluation is conducted using multiple performance metrics, including accuracy, precision, recall, F1-score, confusion matrix, and retained accuracy percentage to measure performance degradation in cross-dataset scenarios.

The intra-dataset evaluation results indicate that the RoBERTa model with Top-k sampling achieves superior performance, with an accuracy of 0.956844 and an F1-score of 0.958428, compared to the Top-p model, which achieves an accuracy of 0.920490 and an F1-score of 0.925742. However, in the cross-dataset evaluation, the RoBERTa Top-p model demonstrates better generalization capability, achieving an accuracy of 0.701250 and an F1-score of 0.641679, outperforming the Top-k model, which achieves an accuracy of 0.648125 and an F1-score of 0.491418. Furthermore, the Top-p model retains a higher accuracy of 76.18%, compared to 67.74% for Top-k.

These findings suggest that Top-k is more effective under the same data distribution (intra-dataset), while Top-p is more robust in improving model generalization under distribution shifts (cross-dataset scenarios).

Keywords: fake reviews, e-commerce, RoBERTa, GPT-2, Top-k, Top-p, cross-dataset, model generalization, text classification