

ABSTRAK

Revolusi Industri 4.0 telah mendorong transformasi signifikan dalam cara manusia berinteraksi dan menyebarkan informasi, termasuk melalui media sosial seperti Twitter. Platform ini menjadi sumber data yang kaya untuk menangkap opini publik secara real-time terhadap berbagai kebijakan, termasuk program Merdeka Belajar Kampus Merdeka (MBKM). MBKM merupakan inisiatif pemerintah Indonesia yang bertujuan meningkatkan fleksibilitas pembelajaran mahasiswa agar lebih adaptif terhadap dunia kerja. Namun, implementasinya menimbulkan berbagai respons dari masyarakat. Sebagian pihak menyambut positif program ini karena memberikan kesempatan bagi mahasiswa untuk memperoleh pengalaman belajar di luar kampus, seperti magang, proyek di desa, pertukaran pelajar, dan riset. Di sisi lain, muncul pula kritik terkait kesiapan infrastruktur, beban administratif, ketimpangan akses antar perguruan tinggi, serta kejelasan konversi nilai dan jaminan mutu pembelajaran di luar kampus. Penelitian ini bertujuan untuk menganalisis dan membandingkan performa algoritma K-Nearest Neighbor (KNN) dan Logistic Regression dalam klasifikasi sentimen publik terhadap program MBKM berdasarkan data Twitter berbahasa Indonesia. Data penelitian diperoleh melalui proses scraping Twitter sebanyak 1.018 tweet dengan dua kelas sentimen, yaitu positif dan negatif. Tahapan preprocessing yang dilakukan meliputi case folding, cleaning, normalization, tokenizing, stopword removal, stemming, pembobotan TF-IDF, serta penerapan SMOTE untuk mengatasi ketidakseimbangan data. Evaluasi model dilakukan menggunakan metrik accuracy, precision, recall, dan F1-score. Hasil penelitian menunjukkan bahwa metode Logistic Regression dengan penerapan SMOTE pada skenario 4 memberikan performa terbaik dibandingkan metode lainnya, dengan nilai accuracy sebesar 79,90%. Sementara itu, metode KNN tanpa SMOTE pada skenario 1 memperoleh accuracy sebesar 73,53%, sedangkan KNN dengan SMOTE pada skenario 2 memperoleh accuracy sebesar 72,55%. Hasil tersebut menunjukkan bahwa Logistic Regression lebih stabil dan efektif dibandingkan KNN dalam mengklasifikasikan sentimen terkait program MBKM. Selain itu, penerapan SMOTE terbukti mampu meningkatkan performa model Logistic Regression dalam menangani ketidakseimbangan data. Penelitian ini diharapkan dapat menjadi rujukan metodologis dalam penerapan machine learning untuk analisis sentimen berbahasa Indonesia serta memberikan masukan berbasis data bagi evaluasi kebijakan program MBKM.

Kata Kunci: Merdeka Belajar Kampus Merdeka, Analisis Sentimen, KNN, Logistic Regression, Media Sosial.

ABSTRACT

The Industrial Revolution 4.0 has driven significant transformations in the way humans interact and disseminate information, including through social media like Twitter. This platform has become a rich data source for capturing real-time public opinion on various policies, including the Merdeka Belajar Kampus Merdeka (MBKM) program. MBKM is an Indonesian government initiative aimed at increasing student learning flexibility to better adapt to the world of work. However, its implementation has drawn mixed responses from the public. Some welcomed the program because it provided opportunities for students to gain off-campus learning experiences, such as internships, village projects, student exchanges, and research. On the other hand, criticisms have also emerged regarding infrastructure readiness, administrative burdens, inequality of access between universities, and clarity of grade conversion and quality assurance for off-campus learning. This study aims to analyze and compare the performance of the K-Nearest Neighbor (KNN) and Logistic Regression algorithms in classifying public sentiment toward the MBKM program based on Indonesian-language Twitter data. The research data was obtained through Twitter scraping of 1,018 tweets with two sentiment classes: positive and negative. The preprocessing steps included case folding, cleaning, normalization, tokenizing, stopword removal, stemming, TF-IDF weighting, and the application of SMOTE to address data imbalance. Model evaluation was conducted using accuracy, precision, recall, and F1-score metrics. The results showed that the Logistic Regression method with SMOTE in scenario 4 performed better than the other methods, with an accuracy of 79.90%. Meanwhile, the KNN method without SMOTE in scenario 1 achieved an accuracy of 73.53%, while KNN with SMOTE in scenario 2 achieved an accuracy of 72.55%. These results indicate that Logistic Regression is more stable and effective than KNN in classifying sentiment related to the MBKM program. Furthermore, the application of SMOTE has been shown to improve the performance of the Logistic Regression model in addressing data imbalance. This research is expected to serve as a methodological reference in the application of machine learning for Indonesian-language sentiment analysis and provide data-driven input for MBKM program policy evaluation.

Keywords: Independent Learning Independent Campus, Sentiment Analysis, KNN, Logistic Regression, Social Media.