

ABSTRACT

Merdeka Learning Campus Merdeka (MBKM) program has been ongoing since 2020, and in the process, there have been changes and developments. Opinions from the public regarding the MBKM program have been expressed, one of which is through social media, Twitter. Sentiment analysis will be carried out to identify the views, feelings, or attitudes contained in the data regarding each MBKM category, which is carried out by applying clustering. By analyzing sentiment, one can explore whether opinions on MBKM tend to be positive or negative in each category with the application of classification. This study aims to analyze the sentiment on the implementation of MBKM among Twitter users using two approaches, namely clustering and classification. Through clustering, the tweet data will be grouped based on the similarity of sentiments, so that can identify sentiment patterns that appear in groups of tweet data related to MBKM. Meanwhile, using classification, will build a model to classify tweet sentiment data into positive or negative categories. Data that has gone through the text preprocessing stage will be grouped into three clusters using the K-Means algorithm, and then each cluster will be classified using the K-Nearest Neighbors algorithm. In the classification process, using the Random Oversampling technique to balance unbalanced data The ratio of test data to training data is 80:20. The results of cluster 0, which discussed apprenticeships, had an accuracy of 80.00%, then cluster 1, which discussed teaching campuses, had an accuracy of 83.91%, and cluster 2, which discussed independent studies, had an accuracy of 91.53%.

Keywords: Sentiment Analysis, K-Means, K-Nearest Neighbors, MBKM

ABSTRAK

Merdeka Belajar Kampus Merdeka (MBKM) telah dilaksanakan dari tahun 2020, dalam prosesnya telah terjadi perubahan dan perkembangan. Opini dari masyarakat mengenai program MBKM ini telah diungkapkan salah satunya melalui media sosial *Twitter*. Analisis sentimen akan dilakukan untuk mengidentifikasi pandangan, perasaan, atau sikap yang terkandung dalam data terkait setiap kategori MBKM yang dilakukan dengan penerapan *clustering*. Dengan menganalisis sentimen, dapat mengeksplorasi apakah opini terhadap MBKM cenderung positif atau negatif dalam masing-masing kategori dengan penerapan *classification*. Tujuan dari penelitian ini adalah untuk menganalisis sentimen pada implementasi MBKM dari pengguna *Twitter* dengan menggunakan dua pendekatan, yaitu *clustering* dan *classification*. Melalui *clustering*, data *tweet* akan dikelompokkan berdasarkan kesamaan sentimen, sehingga dapat mengidentifikasi pola-pola sentimen yang muncul dalam kelompok data *tweet* terkait MBKM. Sementara itu, menggunakan *classification*, akan membangun model untuk mengklasifikasikan sentimen data *tweet* ke dalam kategori positif atau negatif. Data yang sudah melalui tahapan *text preprocessing* akan dikelompokkan menjadi tiga *cluster* menggunakan algoritma *K-Means*, lalu setiap *cluster* akan di klasifikasi menggunakan algoritma *K-Nearest Neighbors*. Dalam proses klasifikasi memakai teknik *Random Oversampling* untuk menyeimbangkan data yang tidak seimbang. Perbandingan data uji dan data latih yaitu 80:20. Hasil cluster 0 yang membahas tentang magang mempunyai akurasi sebesar 80.00%, cluster 1 yang membahas tentang kampus mengajar memiliki akurasi sebesar 83,91%, dan cluster 2 yang membahas studi independent mempunyai akurasi sebesar 91.53%.

Kata kunci: Sentimen Analisis, *K-Means*, *K-Nearest Neighbors*, MBKM