

INTISARI

METODE SMOTE UNTUK DATA TIDAK SEIMBANG PADA ANALISIS SENTIMEN MENGGUNAKAN *SUPPORT VECTOR MACHINE* DAN *RANDOM FOREST*

(Studi Kasus: Komentar Video YouTube)

Oleh

Venessa Yumadila Syahra

19/445727/PA/19551

Media sosial tidak lagi asing bagi masyarakat, salah satunya YouTube yang merupakan platform video terbesar dengan berbagai macam konten. Pengguna bisa membagikan video untuk umum dan menggunakan ulasan dalam bentuk *likes*, *dislikes*, dan komentar untuk mengevaluasi konten video. Data komentar yang diambil dari sebuah media sosial umumnya tidak seimbang sehingga data perlu dilakukan penanganan dengan SMOTE dan Borderline-SMOTE agar data menjadi seimbang. Komentar merupakan ulasan berbentuk tekstual sehingga sulit menemukan informasi secara manual untuk dijadikan indikator evaluasi konten video. Oleh karena itu, diperlukan metode yang mampu mengklasifikasikan komentar secara otomatis. Analisis sentimen dilakukan untuk mengklasifikasikan komentar dengan metode *machine learning*, yaitu *Support Vector Machine* dan *Random Forest* ke dalam kelas sentimen positif, netral, dan negatif. Dalam penelitian ini analisis memberikan hasil bahwa metode *Support Vector Machine* dengan dataset yang sudah ditangani dengan SMOTE memberikan hasil tertinggi di setiap ukuran performa, yaitu akurasi sebesar 86.32%, presisi sebesar 86.67%, recall sebesar 86.32%, dan untuk f1-score sebesar 86.42%.

Kata kunci: analisis sentimen, *Support Vector Machine*, *Random Forest*, SMOTE, Borderline-SMOTE

ABSTRACT

SMOTE METHOD FOR UNBALANCED DATA IN ANALYSIS SENTIMENT USING SUPPORT VECTOR MACHINE AND RANDOM FOREST

(Case study: YouTube Video Comments)

By

Venessa Yumadila Syahra

19/445727/PA/19551

Social media is no stranger to society, and YouTube is one of the largest video platforms with a wide variety of content. Users can share videos publicly and use reviews in the form of likes, dislikes, and comments to evaluate video content. Comment data taken from social media is generally unbalanced, so the data needs to be handled with SMOTE and Borderline-SMOTE so that the data becomes balanced. Comments are textual reviews, so it is difficult to find information manually to be used as an indicator of video content evaluation. Therefore, a method is needed that is able to classify comments automatically. Sentiment analysis is performed to classify comments using machine learning methods, namely Support Vector machines and Random forests, into positive, neutral, and negative sentiment classes. In this study, the analysis provides results showing that the Support Vector Machine method with a dataset that has been handled with SMOTE provides the highest results in each performance measure, namely accuracy of 86.32%, precision of 86.67%, recall of 86.32%, and an f1-score as large as 86.42%.

Keywords: sentiment analysis, Support Vector Machine, Random Forest, SMOTE, Borderline-SMOTE