

INTISARI

ANALISIS SENTIMEN TERHADAP *REVIEW* DRAMA DARI TWITTER MENGUNAKAN METODE *K-NEAREST NEIGHBOR* (K-NN) DENGAN PENANGANAN DATA *IMBALANCED*

Oleh

Anisah Cahyani

17/409516/PA/17823

Media sosial memiliki peran sangat penting untuk komunikasi secara online antar satu orang dengan orang lain. Salah satu media sosial yang populer adalah Twitter, dengan Twitter masyarakat dapat menyatakan pendapat atau opini secara bebas terkait suatu topik, menjual suatu barang, dan hal lainnya. Salah satu pendapat atau opini yang sering dibicarakan masyarakat di Twitter adalah serial drama yang sedang populer. Banyaknya pengguna dan *review* dari penonton drama di Twitter menyebabkan banyak data yang terkumpul, sehingga sulit untuk melihat gambaran sentimen secara jelas. Oleh karena itu, pada penelitian ini akan menganalisis Twitter berbahasa Indonesia yang memiliki opini atau sentimen tentang serial drama, agar data yang tersedia dapat dimanfaatkan untuk keperluan-keperluan terkait *review* drama. Pada penelitian ini dilakukan klasifikasi sentimen menggunakan *K-Nearest Neighbor* serta penanganan data *imbalanced* dengan *random oversampling*, *random undersampling*, dan *random over-under sampling*. Penanganan data *imbalanced* dilakukan agar prediksi klasifikasi tidak cenderung pada kelas mayoritas dan mengabaikan kelas minoritas. Dari analisis yang dilakukan, didapatkan kesimpulan bahwa metode klasifikasi terbaik untuk *review* drama dari Twitter adalah metode K-NN dengan *random oversampling* ($k=1$) karena dapat meningkatkan nilai akurasi, sensitivitas, dan *f-measure*, serta memiliki nilai spesifisitas yang cukup baik.

Kata Kunci : analisis sentimen, *K-Nearest Neighbor*, data *imbalanced*, *random oversampling*, *random undersampling*, *random over-under sampling*.

ABSTRACT

**SENTIMENT ANALYSIS OF THE DRAMA REVIEW FROM TWITTER
USING K-NEAREST NEIGHBOR (K-NN) METHOD WITH
IMBALANCED DATA HANDLING**

By

Anisah Cahyani

17/409516/PA/17823

Social media has a significant role in online communication between one person and another. One of the popular social media is Twitter, with Twitter can express freely opinions related to a topic, sell an item, and other things. One of the opinions often talks about on Twitter is the currently popular drama series. A large number of users and reviews from drama viewers on Twitter means that a lot of data is collected, making it difficult to see a clear picture of the sentiment. Therefore, this study will analyze Twitter Indonesian-language has opinions or sentiment about drama series so that the available data can be used for purposes related to drama review. In this study conducted sentiment classification using K-Nearest Neighbor and imbalanced data handling with random oversampling, random undersampling, and random over-under sampling. Handle imbalanced data so that the classification results not tend to predict the majority class and ignore the minority class. From the analysis, can be concluded that the best classification method for a review drama from Twitter is K-NN with random oversampling ($k = 1$) because it can increase the value of accuracy, sensitivity, f-measure, and the value of specificity is good.

Keywords : sentiment analysis, K-Nearest Neighbor, imbalanced data, random

oversampling, random undersampling, random over-under sampling.