

## ABSTRACT

Ease of use of the internet and social media to communicate increasingly gives freedom to their users. The unrestricted freedom can cause users to attack someone or an organization using hate speech. Systems that can limit users are needed, so that users can't use hate speech. Systems that apply the classification method are believed can solve that problem. However, hate speech in the form of text is very difficult to find, this makes the distribution of training data for the classification system becomes unbalanced or called imbalanced data. The classification system using imbalanced data will adversely affect the performance of the classification. There are several methods to overcome the problem of imbalanced data, one of them with undersampling method. This research comparing several undersampling methods for hate speech classification using the TF-IDF weighting method and SVM classification method. Undersampling methods in this study are *Random Undersampling* (RUS), *Cluster Centroid* (CC), *NearMiss* versi 2 (NM2), *Instance Hardness Threshold* (IHT), *Condensed Nearest Neighbour* (CNN), *Edited Nearest Neighbour* (ENN), *Repeated Edited Nearest Neighbour* (RENN), *All-kNN* (AllKNN), and *One Sided Selection* (OSS). In addition, this research also presents when the undersampling method is implemented to obtain optimal results. The results of this research shows that the undersampling method: *Instances Hardness Threshold* implemented when weighting text data has been weighted to TF-IDF, proven to balance the dataset with class ratio is 1:1 and has the value of classification performance evaluation that is superior compared to other undersampling methods with Index of balanced Accuracy is 0.92.

**Keywords :** Text Classification, Imbalanced Data, Undersampling, TF-IDF, SVM.

## INTISARI

Kemudahan penggunaan internet dan media sosial untuk berkomunikasi semakin memberikan kebebasan kepada penggunanya. Kebebasan pengguna tersebut yang tidak terkontrol, dapat memudahkan penggunanya untuk menyerang seseorang atau sebuah organisasi dengan ujaran kebencian. Sistem yang dapat membatasi penggunanya sangat dibutuhkan, agar penggunanya tidak dapat menggunakan ujaran kebencian. Sistem yang menerapkan metode klasifikasi diyakini dapat mengatasi masalah tersebut. Akan tetapi, ujaran kebencian berbentuk teks sangat susah ditemukan, hal ini membuat distribusi data latih untuk sistem klasifikasi tersebut menjadi tidak seimbang atau disebut *imbalanced data*. Sistem klasifikasi menggunakan *imbalanced data* akan berdampak buruk terhadap kinerja klasifikasi tersebut. Terdapat beberapa metode untuk mengatasi masalah *imbalanced data*, salah satunya dengan metode *undersampling*. Penelitian ini membandingkan beberapa metode *undersampling* untuk klasifikasi teks ujaran kebencian dengan menggunakan metode pembobotan TF-IDF dan metode klasifikasi SVM, metode-metode tersebut yaitu *Random Undersampling* (RUS), *Cluster Centroid* (CC), *NearMiss* versi 2 (NM2), *Instance Hardness Threshold* (IHT), *Condensed Nearest Neighbour* (CNN), *Edited Nearest Neighbour* (ENN), *Repeated Edited Nearest Neighbour* (RENN), *All-kNN* (AllKNN), dan metode *One Sided Selection* (OSS). Selain itu, penelitian ini juga menyajikan kapan metode *undersampling* tersebut diimplementasikan untuk mendapatkan hasil yang optimal. Hasil penelitian ini menunjukkan bahwa metode *undersampling : Instance Hardness Threshold* yang diimplementasikan ketika pembobotan data teks telah berbentuk TF-IDF, terbukti dapat menyeimbangkan *dataset* dengan rasio perbandingan kelas 1:1 (IR=1) dan mempunyai nilai evaluasi kinerja klasifikasi yang lebih unggul dibandingkan metode *undersampling* lainnya dengan nilai evaluasi *Index of Balanced Accuracy* 0,92.

**Kata kunci** – Klasifikasi Teks, Imbalanced Data, Undersampling, TF-IDF, SVM.