

ABSTRACT

Public complaints using short message service (SMS) is one of the media used by the public in delivering suggestions, criticisms, and complaints about the government to improve the quality of public services. Classification of SMS content of public complaints precisely and accurately in accordance with the authority of the Office, greatly influences the speed and accuracy of the follow up of complaints by the government. The problem that is often found in classifications is the condition of data imbalance where the data is not evenly distributed in each class, this can result in a low classification performance. SMOTE is an oversampling technique to overcome data imbalances. This study discusses the application of SMOTE to overcome the problem of data imbalance in a limited dataset. The classification methods being compared are SVM, NB, KNN, LR, DT and RF.

The results showed variations in the value of the parameter $k_neighbors$ on SMOTE did not affect the performance of SVM and LR. This is because the SVM and LR methods use a one-vs-rest scheme. The performance of the KNN method becomes very low when applying SMOTE with the parameter $k_neighbors = 10$. There was a significant decrease in the value of precision, recall and f1-score respectively to 26.43%, 45.24% and 33.16%. RF method performance becomes very high by using the parameter $k_neighbors = 5$, can increase the value of precision, recall and f1-score respectively to 97.14%, 95.24% and 95.56%. To improve the performance of classification methods a combination of preprocessing, feature extraction and oversampling methods needs to be done.

Keywords : *Public Complaints, Short Message Service, Classification, Data Imbalance, SMOTE*

INTISARI

SMS pengaduan masyarakat adalah salah satu bentuk layanan elektronik yang digunakan pemerintah dalam mendapatkan saran, kritik, dan aduan dari masyarakat dalam rangka meningkatkan kualitas layanan publik. Klasifikasi konten SMS pengaduan masyarakat secara tepat dan akurat sesuai dengan kewenangan OPD, sangat mempengaruhi dalam kecepatan dan ketepatan tindak lanjut pengaduan oleh pemerintah. Salah satu masalah yang sering ditemukan dalam klasifikasi adalah kondisi ketidakseimbangan data dimana distribusi data tidak tersebar secara merata pada setiap kelas, hal ini dapat mengakibatkan kinerja klasifikasi menjadi rendah. SMOTE merupakan teknik *oversampling* untuk mengatasi ketidakseimbangan data. Penelitian ini bertujuan mengetahui pengaruh penerapan SMOTE untuk mengatasi masalah ketidakseimbangan data pada dataset yang terbatas. Metode klasifikasi yang dibandingkan adalah SVM, NB, KNN, LR, DT dan RF.

Hasil penelitian menunjukkan variasi nilai parameter $k_neighbors$ pada SMOTE tidak berpengaruh pada kinerja SVM dan LR. Hal ini disebabkan metode SVM dan LR menggunakan skema *one-vs-rest*. Kinerja metode KNN menjadi sangat rendah saat diterapkan SMOTE dengan parameter $k_neighbors=10$. Terjadi penurunan yang signifikan pada nilai *precision*, *recall* dan *f1-score* masing-masing menjadi 26,43%, 45,24% dan 33,16%. Kinerja metode RF menjadi sangat tinggi dengan menggunakan parameter $k_neighbors=5$, dapat menaikkan nilai *precision*, *recall* dan *f1-score* masing-masing menjadi 97,14%, 95,24% dan 95,56%. Untuk meningkatkan kinerja klasifikasi perlu dilakukan kombinasi metode *preprocessing*, *feature extraction* dan *oversampling* yang lain.

Kata kunci – Pengaduan Masyarakat, SMS, Klasifikasi, Ketidakseimbangan Data, SMOTE.