

ABSTRACT

Local governments as public service providers are obliged to provide services for each citizen, including complaint services as a community response to the services provided. Through Diskominfo DIY, the DIY Regional Government provides the DIY e-Lapor application which is an electronic public complaint service. This application is used by the DIY Regional Government to obtain suggestions, criticisms, and reports from the community to improve the quality of public services.

The classification of complaints that are appropriate and by the problem, affects the speed of the government in responding and following up on the complaint. One of the problems in classification is data imbalance, where the distribution of data is not evenly distributed in each category and causes low classification performance. Data imbalance can be overcome by the oversampling method, which adds new data to the minority class until the number between classes becomes balanced. The process of adding data can be done with augmentation techniques. The augmentation technique used in this research is using SMOTE and Synonym Replacement techniques.

This research uses a dataset obtained from complaint data on the DIY e-Lapor application. This research aims to classify public complaint data according to the problem category. To find out how good the classification results are, classification performance is measured by calculating accuracy, precision, recall, and f1-score values.

The SVM classification model showed that the Synonym Replacement training data obtained an accuracy value of 85.19%, precision of 88.87%, recall of 80.51%, and f1-score of 83.75%. These results have a better value than using the original training data with an accuracy value of 78.43%, precision of 88.44%, recall of 68.95%, and f1-score of 74.50%, as well as SMOTE oversampling training data with an accuracy value of 79.89%, precision of 89.24%, recall of 71.15%, and f1-score of 76.74%.

Keywords: Public Complaint, Text Classification, Data Imbalance, Synonym Replacement, SMOTE, SVM.

INTISARI

Pemerintah daerah sebagai penyelenggara pelayanan publik wajib menyediakan layanan bagi setiap warganya, termasuk layanan pengaduan sebagai respon masyarakat terhadap layanan yang diberikan. Melalui Diskominfo DIY, Pemda DIY menyediakan aplikasi e-Lapor DIY sebagai layanan aduan masyarakat elektronik. Aplikasi ini digunakan Pemda DIY untuk mendapatkan saran, kritik, dan laporan dari masyarakat dalam rangka meningkatkan kualitas layanan publik di Daerah Istimewa Yogyakarta.

Klasifikasi aduan yang tepat dan sesuai dengan permasalahannya, mempengaruhi kecepatan pemerintah dalam merespon dan menindaklanjuti aduan tersebut. Salah satu permasalahan dalam klasifikasi adalah ketidakseimbangan data, dimana distribusi data tidak tersebar secara merata pada setiap kategori dan menyebabkan kinerja klasifikasi menjadi rendah. Ketidakseimbangan data dapat diatasi dengan menggunakan metode *oversampling*, yaitu menambahkan data baru pada kelas minoritas hingga jumlah antar kelas menjadi seimbang. Proses penambahan data tersebut dapat dilakukan dengan teknik augmentasi. Teknik augmentasi yang digunakan pada penelitian ini, adalah dengan menggunakan teknik SMOTE dan *Synonym Replacement*.

Penelitian ini menggunakan dataset yang diperoleh dari data aduan pada aplikasi e-Lapor DIY. Tujuan yang ingin dicapai dari penelitian ini adalah mengklasifikasikan data aduan masyarakat sesuai dengan kategori permasalahannya. Untuk mengetahui seberapa baik hasil klasifikasi tersebut, kinerja klasifikasi diukur dengan menghitung nilai akurasi, *precision*, *recall*, dan *f1-score*.

Menggunakan model klasifikasi SVM, hasil penelitian menunjukkan data latih *Synonym Replacement* memperoleh nilai akurasi sebesar 85,19%, *precision* 88,87%, *recall* 80,51%, dan *f1-score* 83,75%. Hasil tersebut memiliki nilai yang lebih baik dibandingkan dengan menggunakan data latih asli dengan nilai akurasi sebesar 78,43%, *precision* 88,44%, *recall* 68,95%, dan *f1-score* 74,50%, maupun data latih *oversampling* SMOTE dengan nilai akurasi sebesar 79,89%, *precision* 89,24%, *recall* 71,15%, dan *f1-score* 76,74%.

Kata Kunci : Aduan Masyarakat, Klasifikasi Teks, Ketidakseimbangan Data, *Synonym Replacement*, SMOTE, SVM.