

INTISARI

Pada layanan pengaduan UPIK, Pemerintah Kota Yogyakarta masih mengelola laporan pengaduan masyarakat secara parsial dan tidak terkoordinir dengan maksimal sehingga masih terjadi redundansi data laporan pengaduan maupun terdapat laporan pengaduan yang tidak tertangani oleh Pemerintah Kota Yogyakarta. Salah satu akibat dari redundansi laporan adalah laporan menumpuk pada halaman *platform* sehingga mengurangi peran masyarakat sebagai pengawal transparansi dari Pemerintah Kota Yogyakarta. Oleh karena itu, untuk meningkatkan peran masyarakat dan memaksimalkan pemrosesan laporan maka dirancang suatu model atau pendekatan berbasis *Natural Language Processing* untuk memilah laporan-laporan yang berpotensi untuk terduplikat. Dengan adanya model tersebut, laporan pengaduan yang masuk dapat terorganisir dengan rapi dan mengurangi redundansi laporan yang masuk.

Model yang dikembangkan akan membandingkan setiap laporan pengaduan yang masuk dengan laporan yang telah masuk dan memeriksa kemiripan antara kedua laporan tersebut dibantu dengan data kategori laporan dan lokasi dari permasalahan terjadi. Terdapat tiga proses utama yaitu data *pre-processing*, pembobotan kata, dan pengujian kemiripan teks. Pada proses pertama yaitu data *pre-processing* akan dilakukan *case folding* untuk mengubah huruf kapital menjadi huruf kecil dan menghilangkan tanda baca. Kemudian untuk pembobotan kata menggunakan model pra-pelatihan IndoBERT dan data perhitungan yang dihasilkan akan dimasukkan kedalam rumus menggunakan metode *cosine similarity* dalam pengujian kemiripan teks untuk menghasilkan nilai kemiripan.

Kata kunci : Laporan Pengaduan, *Natural Language Processing*, Pemerintah Kota Yogyakarta, Redundansi Laporan, UPIK

ABSTRACT

In the UPIK complaint service, the Yogyakarta City Government still manages public complaint reports partially and is not optimally coordinated so that there is still redundancy of complaint report data and there are complaints reports that are not handled by the Yogyakarta City Government. One of the consequences of report redundancy is that reports accumulate on the platform page thereby reducing the role of the community as guardians of transparency from the Yogyakarta City Government. Therefore, to increase the role of the community and maximize report processing, a model or approach based on Natural Language Processing is designed to sort out reports that have the potential to be duplicated. With this model, incoming complaint reports can be neatly organized and reduce redundancy of incoming reports.

The model developed will compare each incoming complaint report with the incoming report and check the similarities between the two reports assisted with report category data and the location of the problem. There are three main processes, namely data pre-processing, word weighting, and text similarity testing. In the first process, namely pre-processing data, case folding will be carried out to convert capital letters to lowercase letters and remove punctuation marks. Then for word weighting using the IndoBERT pre-training model and the resulting calculation data will be entered into the formula using the cosine similarity method in testing the similarity of the text to produce similarity values.

Keywords: Complaint Report, Natural Language Processing, Yogyakarta City Government, Report Redundancy, UPIK