

ABSTRAK

Aplikasi Klasifikasi SMS dengan Algoritma Logistic Regression Berbasis Web dengan Menggunakan Flask Framework

Jenis SMS spam adalah jenis pesan teks yang tidak diinginkan atau tidak diminta yang dikirim ke ponsel pengguna, seringkali untuk tujuan komersial. Untuk mengetahui bagaimana sms spam dapat diklasifikasikan, penulis membuat penelitian dengan menggunakan machine learning yang diintegrasikan ke dalam aplikasi web. Aplikasi web dipilih karena hanya dengan memasukan pesan sms pada *form*, hasil klasifikasi dapat langsung diketahui tanpa perlu mengunduh aplikasi tambahan seperti pada *platform mobile*, yang tentunya akan memakan ruang memori lebih. Algoritma *machine learning* yang digunakan adalah *Logistic Regression*. *Logistic Regression* dipilih karena memiliki performa cukup baik dalam melakukan klasifikasi dengan dataset label 1 dan 0 menurut penelitian-penelitian yang telah dilakukan sebelumnya. Sebelum melakukan *training* pada model, dataset sms yang telah diperoleh dilakukan proses *preprocessing* menggunakan *tokenizer*, *stop words* dan TFIDF untuk mengubah data teks menjadi angka, sehingga *machine learning* dapat memprosesnya. Dataset yang telah melalui proses *preprocessing*, dilakukan pengujian pada parameter *Logistic Regression* yaitu *test size* dan *solver*. Hasil pengujian menghasilkan, untuk *test size* dengan nilai akurasi tertinggi adalah *test size* 0.4 dengan nilai akurasi 1 untuk *data train*, dan 0.97 untuk *data test*, sedangkan untuk *solver* yang menghasilkan nilai akurasi tertinggi adalah *lbfgs*, dan *liblinear*. Masing-masing *solver* menghasilkan nilai akurasi 1 untuk data train, dan 0.97 untuk data test. Nilai akurasi pada model *machine learning* yang telah dibuat menggunakan *logistic regression* dengan parameter *test size* 0.4, dan *solver lbfgs* menghasilkan 0.97 pada jenis *scoring Accuracy*. Model *Machine learning* dapat mendeteksi pesan mana yang spam dan tidak spam adalah dengan menganalisa pola dari kata sms tersebut, apakah pola kata tersebut mirip dengan pola kata dari sms yang ada di dataset yang sebelumnya telah diberi label 0 (spam) dan 1 (tidak spam). Model machine learning diimplementasikan ke dalam aplikasi web menggunakan *framework* dari Python yaitu Flask.

Kata kunci: SMS, Spam, *Machine Learning*, *Logistic Regression*, *Test Size*, *Solver*, Flask, Web.