



Hak cipta dan penggunaan kembali:

Lisensi ini mengizinkan setiap orang untuk mengubah, memperbaiki, dan membuat ciptaan turunan bukan untuk kepentingan komersial, selama anda mencantumkan nama penulis dan melisensikan ciptaan turunan dengan syarat yang serupa dengan ciptaan asli.

Copyright and reuse:

This license lets you remix, tweak, and build upon work non-commercially, as long as you credit the origin creator and license it on your new creations under the identical terms.

BAB V

SIMPULAN DAN SARAN

5.1 Simpulan

Berdasarkan hasil dari implementasi dan hasil uji coba sistem, simpulan dari penelitian adalah sebagai berikut.

1. Implementasi *word embedding* untuk klasifikasi sentimen *user feedback* menggunakan Random Forest Classifier telah selesai dibuat. FastText *pre-trained* model yang digunakan merupakan buatan dari beberapa *dataset user feedback* yang didapatkan dari ulasan Tokopedia dan BYU di Playstore.
2. Hasil dari uji coba menunjukkan bahwa *dataset AppReview* dari penelitian Pamungkas (2017) yang sebanyak 553 baris data masih cenderung memberikan hasil performa yang rendah karena ukuran *dataset* masih tergolong kecil, apalagi jika label yang digunakan lebih dari 2. Pada penelitian ini, performa paling ideal diperoleh ketika menggunakan *dataset* BYU (200 dimensi) sebagai *embedding dataset* dengan rasio 80:20. Hasil yang diperoleh yaitu akurasi sebesar 70.27%, *precision* sebesar 80%, *recall* sebesar 54%, dan F1 sebesar 54%.

5.2 Saran

Berdasarkan penelitian yang dilakukan, terdapat beberapa saran untuk pengembangan lanjutan, yaitu sebagai berikut.

1. Pengembangan selanjutnya diharapkan dapat menyempurnakan proses klasifikasi dengan melakukan *upsampling* atau *downsampling* yang

lebih bervariasi. Diharapkan dengan itu, performa yang dihasilkan bisa meningkat dan lebih memuaskan.

2. Memperbesar *dataset* untuk klasifikasi. Diharapkan dengan ukuran *dataset* yang lebih besar dan *balance* hasil yang diperoleh lebih optimal.
3. Mencoba jenis *word embedding* yang lain seperti Word2Vec dan GloVe sebagai *feature extraction*. Dengan bantuan *library* dari *gensim* hal tersebut sangat dimudahkan dan diharapkan memberikan hasil performa yang lebih memuaskan.
4. Mencoba metode *cross validation* yang lain, misalnya melakukan *cross validation* dengan bantuan *library* *RandomizedSearchCV*. Diharapkan dengan *library* tersebut hasil *tuning hyperparameter* bisa lebih bervariasi dan memberikan hasil performa terbaik.
5. Mengkombinasikan algoritma klasifikasi dengan *synonym extraction*. Memang benar bahwa *word embedding* mampu mendeteksi kemiripan kata secara semantik dan sintatik, namun *word embedding* hanya mengukur berdasarkan jarak antar vektor yang hasilnya belum tentu sinonim seperti yang manusia pikirkan pada umumnya. Diharapkan dengan hal ini, hasil performa yang dihasilkan bisa lebih akurat dan mengurangi misklasifikasi, karena kata-kata yang memiliki kemiripan makna dapat diklasifikasikan sebagai satu kelompok.