

BAB 5 PENUTUP

5.1 Kesimpulan

Klasifikasi teks yang dilakukan dengan menggunakan data yang berasal dari *review* dari pembeli pada *marketplace* yang ada di Indonesia. Dengan total 1000 data, dimana 618 data merupakan kelas produk, 214 data kelas pengiriman, dan 168 data kelas pelayanan. Data tersebut akan melalui *preprocessing* untuk dibersihkan serta dapat menghilangkan *noise* yang ada. Pada *preprocessing* terdapat 6 tahap yang dilalui, yaitu *case folding*, *cleansing*, normalisasi, tokenisasi, *filtering*, dan *stemming*. Setelah data sudah melalui *preprocessing* data akan dihitung bobotnya dengan menggunakan metode *tf-idf*. Jika bobot setiap kata pada data sudah diketahui, data akan dibagi menjadi *data training* dan *data testing*. Kemudian model klasifikasi akan dengan *k-nearest neighbor*, *random forest*, dan *logistic regression* dibangun dengan menggunakan *data training* serta diuji dengan *data testing*. Dan model klasifikasi dievaluasi untuk mengetahui kinerjanya dengan menggunakan *confusion matrix*. Dari penelitian ini didapat kesimpulan sebagai berikut.

1. Algoritma *k-nearest neighbor*, *random forest*, dan *logistic regression* dapat diimplementasikan untuk mengklasifikasi data teks *review* pada *marketplace*.
2. Nilai persentase pembagian data berpengaruh pada kinerja klasifikasi menggunakan, *k-nearest neighbor*, *random forest*, dan *logistic regression*. Dalam penelitian ini perbandingan *data train* 80% : *data test* 20% lebih baik dibandingkan dengan *data train* 50% : *data test* 50% dan *data train* 20% : *data test* 80%.
3. Pada penelitian ini algoritma *random forest* dan *logistic regression* mendapat hasil yang cukup memuaskan, karena mendapat akurasi diatas 80%. Sedangkan algoritma *k-nearest neighbor* mendapat hasil yang belum memuaskan, karena hasil akurasi yang didapat masih kurang dari 80%.
4. Pada penelitian ini kinerja algoritma *random forest* cenderung lebih tinggi dibandingkan dengan algoritma *k-nearest neighbor*, dan *logistic regression*. Dengan kinerja tertinggi dari *random forest* yaitu akurasi sebesar 81%, *average*

precision 76%, *average recall* 72%, dan *average f1-score* 73%, sedangkan akurasi tertinggi pada *k-nearest neighbor* 76% serta *logistic regression* 80%.

Dari hal tersebut dapat diketahui algoritma *random forest* mendapat hasil terbaik dibandingkan dengan *k-nearest neighbor* dan *logistic regression*, hal tersebut sudah sesuai dengan penelitian-penelitian terdahulu, yang memberi informasi bahwa *random forest* lebih baik dari *logistic regression*, dan *logistic regression* lebih baik dari *k-nearest neighbor*. Sehingga algoritma *random forest* merupakan algoritma terbaik pada penelitian ini yang dapat diimplementasikan untuk mengklasifikasi kategori *review* pada *marketplace*.

5.2 Saran

Masih terdapat banyak kekurangan dari penelitian ini, sehingga ada beberapa saran untuk peneliti-peneliti pada masa yang akan datang yang mungkin dapat membuat lebih baik lagi. Berikut beberapa saran yang dapat digunakan untuk penelitian pada masa yang akan datang.

1. Menambah kelas seperti “produk dan pengiriman”, “produk dan pelayanan”, dan sebagainya, karena banyak data yang *review* yang dapat masuk ke dalam 2 atau 3 kelas sekaligus.
2. Membuat kamus kata-kata yang digunakan untuk normalisasi menjadi lebih banyak lagi, karena pada penelitian ini kamus kata-kata yang dipakai masih terlalu dikit dan belum mencakup semua *slang word* yang ada pada data.
3. Mencoba metode atau *library* lain pada tahap *filtering* untuk menghapus *stopword*, karena masih banyak kata *stopword* yang belum terhapus pada saat *preprocessing*.
4. Menambahkan proses *balancing data*, karena data yang digunakan tidak *balance* pada setiap kelas yang ada.

Selain itu penelitian ini juga dapat menjadi saran kepada para *developer marketplace* dalam mengembangkan penelitian ini kedepannya. Karena penelitian ini cukup berguna bagi *marketplace* yang ada, karena dapat membantu para pengguna *marketplace* (pembeli maupun penjual), yaitu untuk memudahkan pengguna dalam membedakan kategori *review* yang diberikan para pembeli. Sehingga penelitian ini dapat diimplementasikan pada *marketplace* yang ada.