



Hak cipta dan penggunaan kembali:

Lisensi ini mengizinkan setiap orang untuk menggubah, 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 5

SIMPULAN DAN SARAN

5.1 Simpulan

Penelitian ini berhasil menganalisis sentimen kandidat Presiden Amerika Serikat 2016 berdasarkan data *Twitter*. Melalui *labelling* oleh narasumber, diketahui bahwa Bernie Sanders merupakan kandidat favorit di *Twitter*, sedangkan Hillary Clinton dan Donald Trump cenderung dibicarakan secara negatif. Popularitas kandidat di *Twitter* cenderung tidak merata dan terdapat sejumlah kandidat yang kurang dibicarakan, seperti Jim Gilmore dan Mike Huckabee.

Model *Naive Bayes* dilatih menggunakan data *training* yang diproses secara sederhana. Model ini memiliki akurasi 95,8% dan memprediksi Ted Cruz sebagai kandidat presiden yang unggul dari Partai Republik serta Bernie Sanders sebagai kandidat yang unggul dari Partai Demokrat. Namun, model hanya mencapai akurasi 26,7% ketika digunakan untuk memprediksi hasil *polling* pada *RealClearPolitics.com*. Akurasi prediksi model terhadap hasil *polling* diduga dipengaruhi kebaruan data yang digunakan untuk pelatihan model.

Penelitian ini juga menganalisis hubungan antara jumlah *tweet* positif serta popularitas kandidat di *Twitter* dengan hasil *polling*. Kedua faktor tersebut tidak

dapat digunakan untuk menebak hasil *polling* dengan cukup akurat. Hal ini mendukung argumen bahwa *Twitter* tidak dapat digunakan untuk memprediksi hasil *polling* (Gayo-Avello et al., 2011).

5.2 Saran

Meskipun cukup akurat, model *Naive Bayes* yang dibangun tidak mampu memprediksi hasil *polling* dengan cukup akurat. Hal ini dapat disebabkan oleh beberapa hal: (1) metode *preprocessing* yang kurang tepat, (2) kesalahan *labelling* oleh narasumber, (3) prediksi menggunakan *Naive Bayes* kurang tepat, dan (4) *Twitter* tidak dapat digunakan untuk memprediksi hasil *polling*.

Penelitian ini menggunakan metode *preprocessing* sederhana, yaitu hanya menyingkirkan URL dan gambar tanpa menghilangkan kata apapun dalam teks, agar tidak terjadi pergeseran makna *tweet* ketika akan diberi *label* oleh narasumber. Penelitian selanjutnya dapat bereksperimen dengan melakukan tahap *preprocessing* kedua, yaitu membersihkan kata yang kurang signifikan terhadap makna *tweet*, seperti kata depan dan kata ganti, untuk mengurangi kompleksitas penghitungan oleh model.

Narasumber yang dipilih dalam penelitian ini bukan penduduk asli Amerika Serikat sehingga dapat terjadi perbedaan pemahaman pada isi *tweet*. Selain itu, terdapat istilah asing yang digunakan pada *tweet*. Pada penelitian selanjutnya, dibutuhkan pemilihan narasumber yang lebih tepat untuk mengantisipasi adanya

kesalahan pada proses *labelling*.

Penggunaan algoritma *supervised* seperti *Naive Bayes* membutuhkan *data set* yang akurat untuk memperoleh hasil maksimal. Pada penelitian selanjutnya, model dapat dibangun dengan algoritma *supervised* lain, misalnya *Support Vector Machine* (SVM). Penggunaan algoritma *unsupervised* untuk melakukan prediksi juga dapat dilakukan karena tidak membutuhkan proses *labelling*.

Selain itu, terdapat kemungkinan bahwa *Twitter* tidak dapat digunakan untuk memprediksi hasil *polling*. Untuk mengurangi *bias* dan menghindari *spam* pada data *Twitter*, penelitian selanjutnya dapat membatasi pengambilan data menggunakan prinsip *one man one vote*, yaitu hanya mengambil satu *tweet* per pengguna. Pembatasan *tweet* yang digunakan berdasarkan lokasi geografis serta perkiraan usia pengguna *Twitter* juga dapat dilakukan.

UMMN