

LAPORAN PROYEK

I. Introduction

Dalam zaman teknologi ini, berita dapat menyebar dengan begitu mudahnya. Namun hal yang tidak dapat dipungkiri adalah sulit untuk memilah informasi dan berita untuk dicerna dan dikonsumsi. Terlalu banyak berita yang tersebar mengakibatkan ketidakefisienan dalam mencari informasi. Oleh karena hal tersebut, munculah ide dalam kelompok kami untuk membuat sebuah algoritma untuk memilah dan mengelompokkan berita global berdasarkan sentimennya. Hal ini dapat memudahkan orang untuk memilah dan mencerna informasi dari berita. Juga mempermudah orang untuk memilih berita mana yang ingin dibaca.

Maka dari itu terpikirlah untuk merancang algoritma yang dapat membedakan sentiment dari berita, baik itu sentiment positif, negatif, maupun netral. Dengan melakukan Analisa sentiment seperti ini akan sangat mempermudah pembaca karena telah dikelompokkannya berita-berita berdasarkan sentimennya.

II. Method

Dalam menganalisa sentimen dari dataset yang ada yaitu berita global, kami menggunakan metode AFINN-lexicon, yang terbukti adalah salah satu metode dengan pendekatan terbaik dalam melakukan sentiment analisis. Namun tentu saja, sebelum melakukan Analisa sentiment, harus mem-pre-processing datasetnya terlebih dahulu. Proses Pre-Processing yang dilakukan adalah Case Folding, Tokenizing (menghilangkan leading space and trailing space, menghilangkan punctuation, menghilangkan simbol dan karakter spesial, dan memperjelas contractions dalam Bahasa Inggris, menghilangkan macrons dan karakter beraksen), Filtering (menghilangkan stopwords). Setelah itu barulah menggunakan AFINN-lexicon. Metode penggunaannya adalah dengan mendeteksi keberadaan kata positif maupun kata negatif dalam kalimat berita global. Dalam mendeteksi keberadaan kata positif dan negatif, AFINN-lexicon merubah penemuan kata tersebut ke dalam variabel Score. Score ini lah yang akan memperlihatkan dan membedakan berita global dengan sentimen positif ataupun negatif. Score ini dibatasi dengan angka, bila Score lebih besar dari 0 atau positif, artinya berita tersebut adalah berita dengan sentimen positif. Bila Score lebih kecil dari 0 atau negatif, maka berita tersebut akan dinilai sebagai berita dengan sentiment negatif. Bila Score adalah sama dengan 0, maka berita tersebut adalah berita netral.

III. Results and Discussion

Dari metode yang digunakan, yaitu AFINN-lexicon, berikut adalah hasilnya berdasarkan masing-masing tahapannya. Pada tahap Pre-Processing, dataset berhasil untuk dihurufkecilkan atau Case Folding. Tahap Tokenizing pun berhasil dilakukan dengan hilangnya leading space dan trailing space, punctuation, simbol dan karakter special, macrons dan karakter beraksen. Tahap Filtering pun berhasil dimana stopwords berhasil dihilangkan. Namun yang menjadi masalah adalah hanya memilih dataset berita global Bahasa Inggris yang belum bisa dilaksanakan. Jadi karena itu, dataset yang dilanjutkan ke tahapan selanjutnya masih mengandung beberapa data Bahasa diluar Bahasa Inggris. Pada tahap AFINN-lexicon, metode ini berhasil mengklasifikasi dataset berdasarkan sentimennya, dengan mencari dan menghitung kata-kata yang bersifat negatif maupun positif. Dari tahap tersebut,

diakumulasiannya lah jumlah kata-kata tersebut, baik kata positif maupun negatif. Akumulasi jumlah tersebut disimpan dalam variabel Score, sebagai penanda apakah berita global tersebut termasuk ke dalam berita dengan sentiment negatif atau sentiment positif atau juga sentiment netral.

Jika dalam tahap Pre-Processing bisa dipisahkan antara Bahasa Inggris dengan Bahasa lainnya, tentu hasilnya akan lebih akurat. Karena dengan adanya Bahasa selain Bahasa Inggris, tentu saja hasilnya tidak akan benar dalam menentukan apakah berita tersebut masuk ke dalam berita sentiment negatif, positif, ataupun netral. Namun hasil dari AFINN-lexicon pada dataset Bahasa Inggris cukup baik dan akurat. Karena pada dataset ada variabel yang berisi positif dan negatif. Kebanyakan dari dataset tersebut dengan hasil algoritma AFINN-lexicon akurat, namun ada beberapa yang berbeda dikarenakan adanya hasil sentiment netral jika Score adalah sama dengan 0 sedangkan pada dataset hanya ada data positif dan negatif.

IV. Conclusion

Hasil dari algoritma ini belum 100% akurat. Ada beberapa faktor yang mempengaruhi keakuratannya, yaitu adanya perbedaan hasil antara dataset dengan hasil algoritma dimana di dataset hanya ada 2 hasil yaitu positif dan negatif sedangkan pada hasil algoritma ada 3 hasil, yaitu positif, negatif, dan netral. Faktor lainnya adalah adanya Bahasa selain Bahasa Inggris yang tentu saja tidak bisa dianalisa sentiment menggunakan AFINN-lexicon. Faktor selanjutnya adalah karena metode AFINN-lexicon ini memproses data per kata, dan tiap kata memiliki nilai nya masing-masing. Dari data tersebut diakumulasi lah kata sentiment negatif dan positif menjadi sebuah variabel Score. Faktor-faktor inilah yang mempengaruhi akurasi dari metode ini. Meskipun tidak 100% akurat, tapi hasilnya cukup akurat dan memuaskan.