

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 finansial 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 finansial, 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 finansial. 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 finansial 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 dihuruf-kecilkan 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 finansial 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 finansial tersebut termasuk ke dalam berita dengan sentiment negatif atau sentiment positif atau juga sentiment netral. Hasil yang didapatkan adalah akurasi 59%.

Untuk memberikan perbandingan, maka dilakukan pula Analisa sentiment menggunakan metode non-filtering. Jadi ada satu tahap dalam Pre-Processing yang dihilangkan. Namun setelah dijalankan, ternyata hasilnya lebih baik dibandingkan dengan yang menggunakan metode Filtering. Meskipun algoritma yang melalui tahap Filtering hasilnya baik, namun tidaklah terlalu memuaskan.

Apabila tidak menggunakan metode AFINN-lexicon namun menggunakan metode lain seperti Naïve Bayes ataupun Deep Learning, hasilnya akan lebih memuaskan. Hal ini dikarenakan algoritma Naïve Bayes dan Deep Learning menganalisa data per kalimatnya, sedangkan AFINN-lexicon menganalisa per kata dalam kalimat data. Hal tersebut membuyarkan akurasi dari algoritma ini. Karena hal tersebut, tidak bisa atau sulit untuk mendeteksi kalimat sarkasme. Masalah lainnya yang muncul adalah sulit mendeteksi kalimat ambigu. 2 hal tersebut yang mempengaruhi faktor akurasi dari metode AFINN-lexicon.

IV. Conclusion

Menggunakan metode AFINN-lexicon membuahkan hasil yang cukup baik. Menghasilkan tingkat akurasi sebanyak 59%, meskipun cukup baik, namun tidak terlalu memuaskan. Apalagi setelah dibandingkan dengan metode tanpa tahap Filtering, tingkat akurasi lebih tinggi. Juga dengan digunakannya metode Naïve Bayes dan Deep Learning, hasil yang dihasilkan akan lebih baik.