

Abstrak

Depresi merupakan penyakit mental yang umum ditemukan di dunia dan diasosiasikan menjadi penyebab utama disabilitas. Layanan pengobatan penyakit mental di dunia yang belum memadai memberikan motivasi untuk dilakukan upaya pencegahan. Penelitian di bidang kesehatan mental yang masih kekurangan data kuantitatif akibat kompleksitas penyakit mental menjadikan media sosial sebagai sumber data yang berpotensi dalam menciptakan upaya pencegahan penyakit mental khususnya depresi. Pada penelitian ini, *sentiment analysis* digunakan untuk mendeteksi perilaku depresi pada media sosial Reddit. Analisis kata ganti orang dan pola kata yang mengikuti dilakukan untuk melihat pola bahasa pada data depresi. Berikutnya, beberapa eksperimen dilakukan untuk mendapatkan performa model terbaik, seperti perbandingan jenis *preprocessing*, perbandingan *feature selection* *Information Gain* dan *Categorical Proportional Difference*, dan pencarian parameter *smoothing* terbaik dalam Multinomial Naïve Bayes. Hasil penelitian menunjukkan bahwa terdapat peningkatan performa model akibat dari *stopword removal* dan reduksi kata dengan *stemming*. Pada perbandingan *feature selection*, *Information Gain* menghasilkan subset fitur terbaik sebanyak 40% dari total fitur dan berhasil memberikan peningkatan akurasi terbaik sebanyak 5.59% menjadi 87.82% dan peningkatan f1-score sebanyak 3.91% menjadi 89.74%. Terakhir, tuning parameter *smoothing* dengan nilai alpha sebesar 0.1 pada Multinomial Naïve Bayes menghasilkan peningkatan akurasi terbaik sebanyak 2.54% menjadi 84.77% dan peningkatan f1-score sebanyak 1.56% menjadi 87.39%.

Kata kunci : Reddit, Sentiment Analysis, Information Gain, Categorical Proportional Difference, Multinomial Naïve Bayes