

ABSTRAK

Depresi adalah kondisi kesehatan yang umum terjadi dan berdampak pada individu di seluruh dunia. Ciri-ciri khasnya meliputi perasaan sedih, putus asa, dan ketidaktertarikan pada aktivitas yang sebelumnya menyenangkan. Di sisi lain, deteksi depresi adalah tugas untuk menentukan apakah seseorang mengalami depresi atau tidak mengalami depresi dengan menggunakan beberapa fitur atau indikator. Indikator linguistik atau mendeteksi depresi dari tulisan teks digunakan dalam penelitian ini. Teks tersebut diambil dari data tweet X (Twitter). Mengingat tweet sering menggunakan singkatan dan bahasa gaul, serta kurangnya ketepatan tata bahasa, tweet sering kali menjadi tantangan tersendiri. Penggunaan frasa gaul dalam tweet dapat menyebabkan ketidaksesuaian kosakata. Machine Learning sering kali menemukan vocabulary mismatches dalam menangkap makna tulisan pengguna (dalam hal ini, digunakan untuk mendeteksi depresi) dan menghasilkan akurasi yang kurang optimal. Implementasi feature enrichment, dengan memanfaatkan user-based, content-based, dan LIWC feature, serta feature extraction dan feature expansion, akan menghasilkan peningkatan akurasi pada sistem pendeteksian depresi dengan menggunakan data tweet di X. Cara kerja feature enrichment dalam klasifikasi adalah dengan mengkombinasikan fitur user-based, content-based, dan LIWC. Metode yang diusulkan telah meningkatkan kinerja akurasi pada beberapa percobaan. Penggabungan baseline dengan fitur LIWC dapat meningkatkan akurasi sebesar 25% dan dengan fitur content-based dapat meningkatkan akurasi sebesar 18,75%. Di sisi lain, fitur user-based tidak dapat meningkatkan sistem dari baseline. Performa terbaik dari beberapa skenario pengujian didapatkan dengan kombinasi baseline (mengestrak data tweet menggunakan TFIDF) + fitur LIWC. Kombinasi fitur LIWC mampu memberikan performa akurasi sebesar 93.75%.

Kata kunci: depresi, deteksi depresi, vocabulary mismatches, feature enrichment, LIWC