

ABSTRAK

Organisasi Kesehatan Dunia (WHO) pada tahun 2023 melaporkan bahwa lebih dari 700.000 orang meninggal akibat bunuh diri setiap tahun. Proses bunuh diri dimulai dengan ideasi bunuh diri dan kemudian berkembang menjadi percobaan bunuh diri. Saat ini banyak orang aktif di beberapa jaringan media sosial populer seperti Reddit untuk berbagi kehidupan sehari-hari dan pengalaman, baik atau buruk. Bunuh diri dapat dicegah dengan melakukan identifikasi dan deteksi dini, dan fenomena ini baru-baru ini digunakan sebagai domain penelitian terkait deteksi pemikiran ideasi bunuh diri di media sosial. Namun, menemukan dan memahami pola ideasi bunuh diri merupakan tugas yang menantang dan pada saat yang sama perlu mengatasi masalah bahasa alami yang ada media sosial. Oleh karena itu, penelitian ini bertujuan untuk meningkatkan deteksi ideasi bunuh diri menggunakan Machine Learning di Reddit dengan menggunakan C-SSRS Reddit Suicide Dataset dan Subreddit non-mental health lainnya dengan memanfaatkan berbagai metode feature enrichment seperti fitur Berbasis Konten, Linguistic Inquiry and Word Count (LIWC) untuk menangkap pola, dan Feature Expansion untuk mengatasi masalah bahasa alami di media sosial kemudian menerapkan dasar hasil penelitian sebelumnya[10]. Hasil terbaik dari penelitian ini mencapai Akurasi 85% dan Recall 82% dengan menggabungkan seluruh metode *feature enrichment* karena masing-masing metode berhasil memberikan model informasi tambahan seperti karakteristik dari Fitur Berbasis Konten, dimensi psikologis isi pesan dari LIWC serta menangani ketidakcocokan kosakata yang terjadi di media sosial dengan menggunakan Ekspansi Fitur.

Kata kunci: deteksi ideasi bunuh diri, Reddit, Machine Learning, Feature Enriching, fitur Berbasis Konten, Linguistic Inquiry and Word Count (LIWC), Feature Expansion