UNIVERSITAS
GADJAH MADA
Klasifikasi Gejala Depresi pada Teks Postingan Reddit dengan Pendekatan XGBoost dan Random Forest
Shevrilla Vilnafa Bilbina S., Dr. Sigit Priyanta, S.Si., M.Kom.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

ABSTRAK

KLASIFIKASI GEJALA DEPRESI PADA TEKS POSTINGAN REDDIT DENGAN XGBOOST DAN RANDOM FOREST

Oleh

SHEVRILLA VILNAFA BILBINA S.
22/492511/PA/21118

Perkembangan zaman yang semakin maju dan meningkatnya beban tanggung jawab hidup membuat banyak individu mengalami tekanan psikologis yang berpotensi memicu munculnya gangguan mental, termasuk depresi. Depresi, gangguan mental umum dan serius, mempengaruhi lebih dari 264 juta orang di seluruh dunia, mengakibatkan berbagai masalah emosional dan fisik, serta meningkatkan risiko bunuh diri. Sering kali, depresi tidak disadari oleh penderita maupun orang-orang di sekitarnya, terutama karena gejala awalnya sering dianggap sepele. Di era digital saat ini, banyak individu mencurahkan perasaan mereka di media sosial, yang bisa menjadi tanda awal depresi. Mengingat pentingnya deteksi dini, pendekatan machine learning dinilai mampu meningkatkan akurasi dalam mengidentifikasi dan memprediksi gangguan mental melalui analisis pola data secara otomatis (Islam et al.). Fitur yang dimanfaatkan adalah *n-gram*, *TF-IDF*, dan *Word2Vec*, serta model *machine learning Random Forest*dan *XGBoost*. Untuk *dataset* latih, pengambilan data dilakukan dari *website* Kaggle. Data tersebut terdiri dari postingan berbahasa Inggris yang diambil dari platform diskusi Reddit. Dari proses ekstraksi fitur, dapat ditunjukkan kata "*depression*" sebagai yang paling penting, diikuti oleh "*feel*", "*life*", dan "*anxiety*". Sedangkan, dari hasil model eksperimen yang telah dilakukan, model *Random Forest* dan *XGBoost* menunjukkan nilai *recall* tertinggi mencapai 96% menggunakan fitur *unigram* dan *TF-IDF*.

Kata Kunci: *Natural Language Processing* (NLP), Depresi, *TF-IDF*, *Word2Vec*, *n-gram*, Reddit.

UNIVERSITAS
GADJAH MADA
Klasifikasi Gejala Depresi pada Teks Postingan Reddit dengan Pendekatan XGBoost dan Random Forest
Shevrilla Vilnafa Bilbina S., Dr. Sigit Priyanta, S.Si., M.Kom.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

ABSTRACT

CLASSIFICATION OF DEPRESSIVE SYMPTOMS IN REDDIT POSTS USING XGBOOST AND RANDOM FOREST

By
SHEVRILLA VILNAFA BILBINA S.
22/492511/PA/21118

The rapid advancement of modern society and the increasing burden of life responsibilities have caused many individuals to experience psychological distress, which may trigger mental health disorders, including depression. Depression is a common and serious mental disorder affecting more than 264 million people worldwide, leading to various emotional and physical problems while significantly increasing the risk of suicide. In many cases, depression goes unrecognized by both individuals and those around them because its early symptoms are often considered insignificant. In the digital era, many people express their emotions through social media, providing valuable indicators for the early detection of depression. Given the importance of early identification, machine learning has emerged as a promising approach to improve the accuracy of detecting and predicting mental health disorders through the automatic analysis of textual patterns (Islam et al.). This study utilizes n-gram, TF-IDF, and Word2Vec as feature extraction techniques, while Random Forest and XGBoost are employed as the machine learning classification models. The training dataset was obtained from Kaggle and consists of English-language posts collected from the Reddit discussion platform. The feature extraction results indicate that the word *"depression"* is the most influential feature, followed by *"feel," "life,"* and *"anxiety."* Experimental results demonstrate that both the Random Forest and XGBoost models achieved the highest recall score of 96% when using unigram features with TF-IDF weighting.

Keywords: Natural Language Processing (NLP), Depression, TF-IDF, Word2Vec, n-gram, Reddit.