

# Deteksi Perilaku Depresi dengan Sentiment Analysis pada Media Sosial Reddit Menggunakan Feature Selection Information Gain dan Categorical Proportional Difference dengan Algoritma Multinomial Naïve Bayes

M Reza Prawira S  
1301161771

Prof. Dr. Adiwijaya, S.Si, M.Si  
00740046

Said Al Faraby, S.T., M.Sc  
15890019

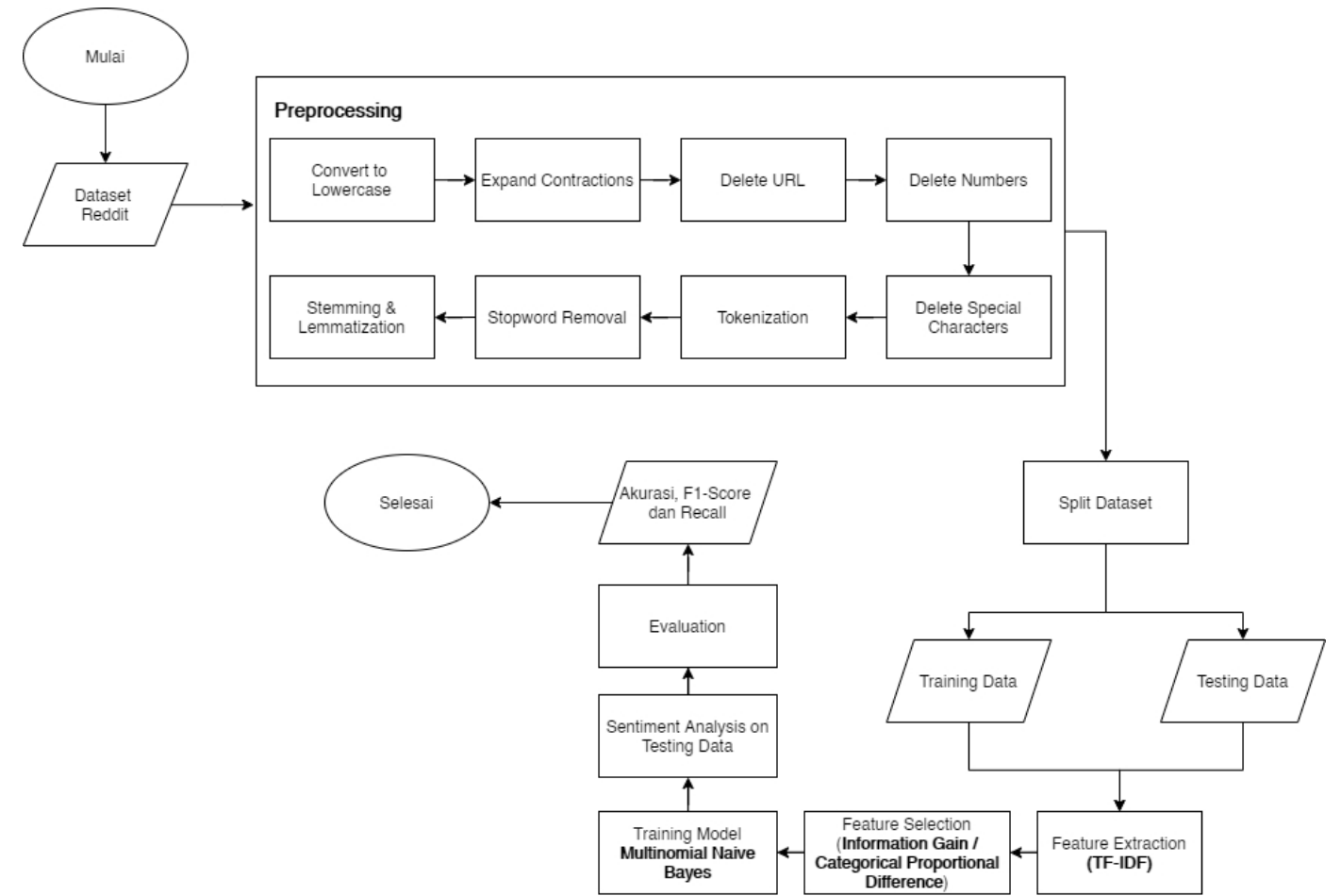
## Abstrak

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 terhadap pola penggunaan Bahasa pada data depresi dan eksperimen pengaruh performa model dengan parameter jenis preprocessing, feature selection, dan laplace smoothing dilakukan.

Hasil penelitian menunjukkan bahwa terdapat peningkatan performa model akibat dari stopwords 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%.

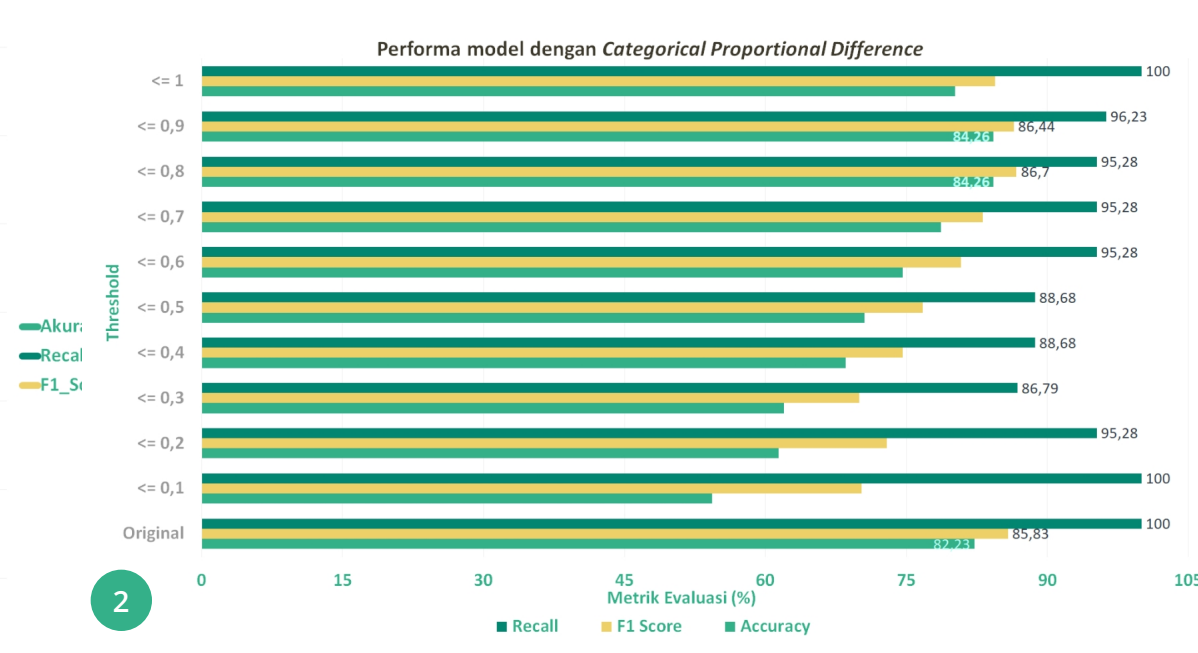
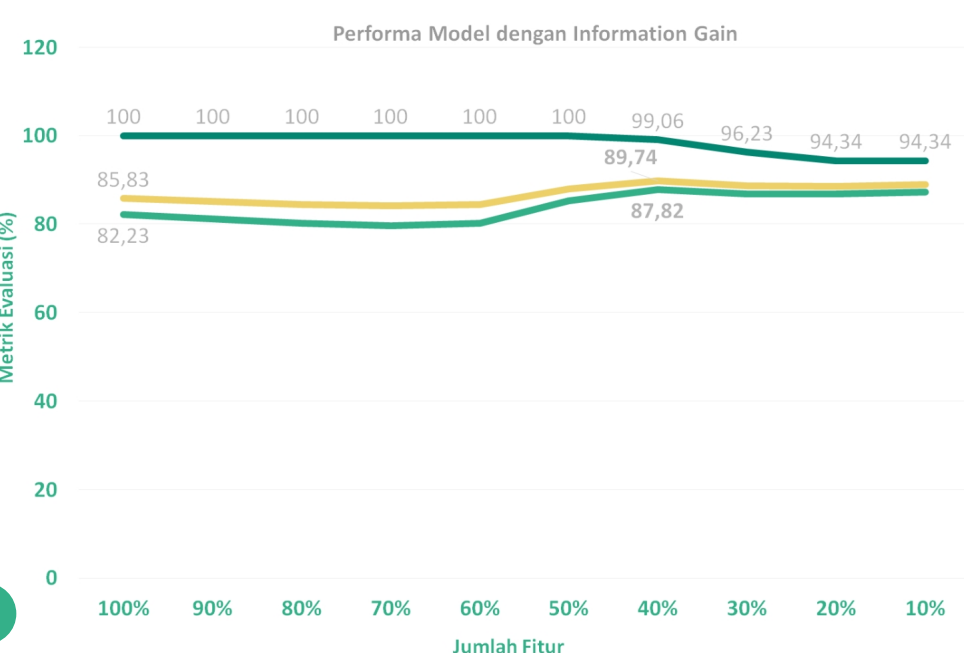
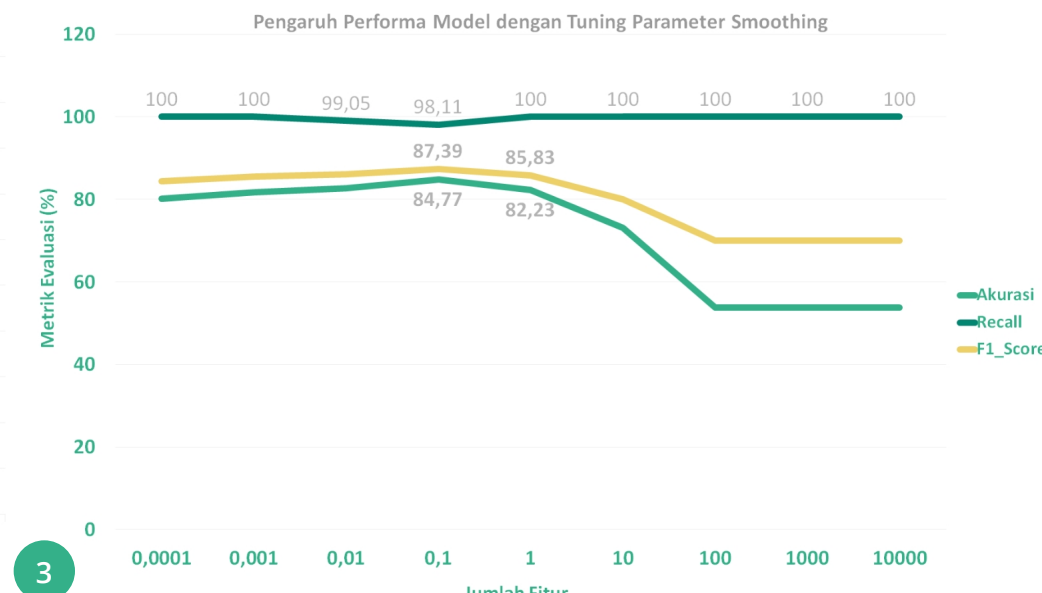
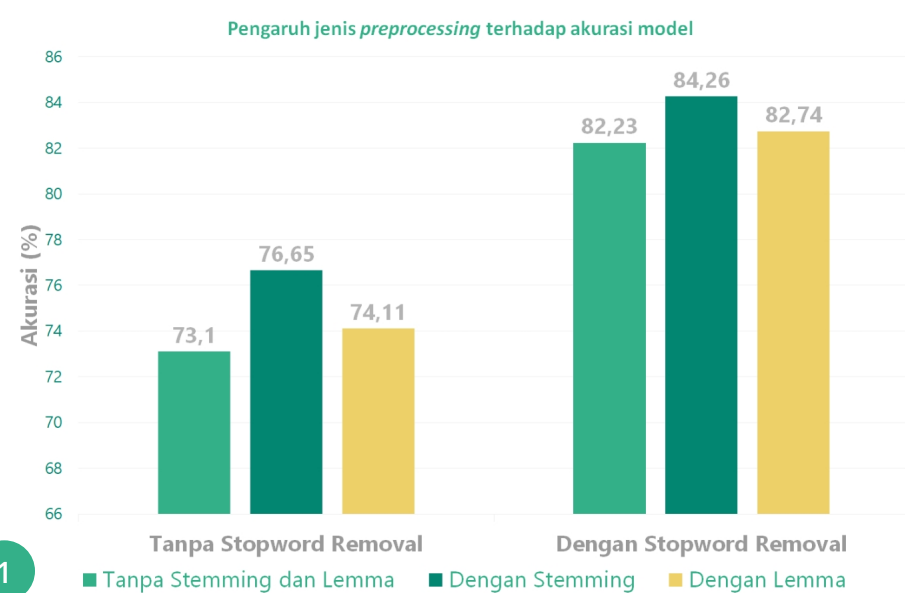
## Skema Sistem yang dibangun



## Pola Penggunaan Bahasa pada Data Depresi



## Hasil Pengujian



Pada **Grafik 1** terdapat pengaruh stopwords removal terhadap performa model sebanyak 7-9% pada akurasi. Stemming juga memberikan peningkatan akurasi sebanyak 2-3.5%.

Pada **Grafik 3** terdapat peningkatan performa model menjadi 84.77% pada akurasi dan 87.39% pada f1-score.

Pada **grafik 2**, performa model dengan subset hasil seleksi Information Gain memberikan peningkatan akurasi terbaik hingga 87.82% dan f1-score hingga 89.74%.