

ABSTRAK

PERBANDINGAN ALGORITMA *STATISTICAL BASED* DAN *RULE BASED* PADA ANALISIS TEKS KALIMAT DEPRESIF DI TWITTER

Dalam penelitian ini penulis melakukan penelitian pada klasifikasi kalimat depresif di twitter. Penelitian ini menggunakan 1500 *tweets* yang diambil dari 30 akun *twitter* yang telah menuliskan klaim bahwa pengguna tersebut divonis mengalami gangguan depresi oleh psikolog/psikiater. Dokumen tersebut dilabeli secara manual berdasarkan 6 gejala (gangguan tidur, nafsu makan, agitasi, kehilangan energi, perasaan tidak berharga, keinginan bunuh diri) yang telah diverifikasi oleh ahli. Penelitian ini membandingkan kinerja algoritma *rule based* dan *statistical based* berdasarkan nilai akurasi algoritma. Adapun algoritma *rule based* yang digunakan adalah algoritma *Synesketch* dengan beberapa rule yang digunakan untuk klasifikasi. Sedangkan algoritma *statistical based* yang digunakan ialah *Support Vector Machine* (SVM) dan *Naïve Bayes*. Pada *pre-processing* di dalam penelitian ini menggunakan beberapa tahapan yaitu normalisasi hingga konversi kata berdasarkan kamus kata. Hasil penelitian ini menunjukkan algoritma *statistical based* memiliki tingkat akurasi lebih tinggi yaitu 98,9% untuk SVM dan 91% untuk *Naïve Bayes*. Sedangkan *Synesketch* mendapat nilai akurasi 86,91%.

Kata Kunci: analisis teks, *data mining*, *support vector machine*, *naïve bayes*,
synesketch, *statistical based*, *rule based*

ABSTRACT

COMPARISON OF STATISTICAL BASED AND RULE BASED ALGORITHM ON TEXT ANALYSIS OF DEPRESSIVE SENTENCES ON TWITTER

In this study the authors conducted research on the classification of depressive sentences on twitter. This study uses 1500 tweets taken from 30 twitter accounts that have written claims that the user has been diagnosed with depressive disorders by a psychologist / psychiatrist. The documents are manually labeled based on 6 symptoms (sleep, appetite, agitation, loss of energy, feelings of worthlessness, suicidal thoughts) which have been verified by experts. This study compares the performance of rule-based and statistical based algorithms based on the algorithm's accuracy value. The rule-based algorithm that this study use is Synesketch algorithm with several rules for classification. For statistical based algorithms that this study use are Support Vector Machine (SVM) and Naïve Bayes. The pre-processing in this study uses several stages, normalization and word conversion based on a word dictionary. The results of this study indicate that the statistical based algorithm has a higher accuracy, with 98.9% accuracy score for SVM and 91% for Naïve Bayes. Meanwhile, Synesketch only 86.91%.

Keywords: text analysis, data mining, support vector machine, naïve bayes, synesketch, statistical based, rule based