

INTISARI

Analisis Sentimen Mahasiswa Ugm Terhadap Tweet Simaster Menggunakan Algoritma Naive Bayes Classifier

Oleh

Hanyel Daryus Bancin
20/455445/PA/19660

Universitas Gadjah Mada (UGM), sebagai salah satu perguruan tinggi terkemuka di Indonesia, memiliki Sistem Informasi Terintegrasi (SIMASTER) yang berguna dalam menunjang aktivitas akademik perkuliahaan. Dalam penerapannya terdapat sentimen-sentimen baik positif maupun negatif yang diunggah ke dalam Twitter. Penelitian ini bertujuan untuk membuat pemodelan analisis sentimen mahasiswa UGM mengenai SIMASTER menggunakan algoritma *Naive Bayes Classifier*. *Naive Bayes* dipilih bukan hanya karena popularitasnya tetapi karena keunggulannya yang relatif sederhana dan efektif dalam klasifikasi teks pendek seperti tweet.

Implementasi analisis sentimen melalui beberapa tahapan. Pertama, dilakukan pengumpulan data dengan metode *scraping* pada tweet yang terkait dengan SIMASTER dalam periode waktu September 2023 - Februari 2024. Kedua, dataset tweet SIMASTER disiapkan melalui tahapan *pre-processing*, termasuk *filtering*, *case folding*, *stopword*, *tokenizer*, *stemming*, *normalization* dan *stemming*. Ketiga, dilakukan pelabelan data secara manual dan *TextBlob*. Keempat, dilakukan juga pembobotan kata menggunakan metode TF-IDF lalu melakukan klasifikasi teks menggunakan *Naive Bayes Classifier*.

Evaluasi menunjukkan bahwa metode pelabelan manual memiliki akurasi lebih tinggi dibandingkan dengan *TextBlob*. Pada data manual yang seimbang, diperoleh akurasi 85,29%, presisi 92,85%, recall 76,47%, dan f1-score 86,49%. *Unigram* terbukti lebih efektif untuk teks pendek seperti tweet dibanding *bigram* dan *trigram*. Pengaturan parameter *alpha* sebesar 5.0 pada unigram memberikan akurasi tertinggi. Uji coba sistem dengan 2085 data menunjukkan mayoritas mahasiswa UGM menyampaikan sentimen negatif terhadap SIMASTER, dengan hanya 6,6% sentimen positif dan 93,4% sentimen negatif. Visualisasi hasil analisis menggunakan wordcloud menunjukkan kata-kata seperti "SIMASTER", "buka", "nilai", dan "notif" dalam sentimen positif, sedangkan sentimen negatif didominasi oleh kata-kata seperti "SIMASTER", "error", dan "lemot".

Kata kunci: *Naive Bayes*, *scraping*, *case folding*, *stopword*, *tokenizer*, *stemming*, *filtering*, *normalization*

ABSTRACT

Sentiment Analysis Of Ugm Students On Simaster Tweets Using Naive Bayes Algorithm

By

Hanyel Daryus Bancin

20/455445/PA/19660

Gadjah Mada University (UGM), as one of the leading universities in Indonesia, has an Integrated Information System (SIMASTER) that is useful in supporting academic activities. In its implementation, of course there are sentiments both positive and negative uploaded to Twitter. This research aims to model the sentiment analysis of UGM students regarding SIMASTER using the Naive Bayes Classifier algorithm. Naive Bayes was chosen not only because of its popularity but because of its relatively simple and effective advantages in the classification of short texts such as tweets.

The implementation of sentiment analysis goes through several stages. First, data collection was carried out using the scraping method on tweets related to SIMASTER in the time period September 2023 - February 2024. Second, the SIMASTER tweet dataset is prepared through pre-processing stages, including filtering, case folding, stopword, tokenizer, stemming, normalization and stemming. Third, data labeling is done manually and TextBlob. Fourth, word weighting using TF-IDF method is also done and then text classification using Naive Bayes Classifier.

The evaluation shows that the manual labeling method has higher accuracy than TextBlob. On balanced manual data, 85.29% accuracy, 92.85% precision, 76.47% recall, and 86.49% f1-score were obtained. Unigrams proved to be more effective for short texts such as tweets than bigrams and trigrams. Setting the alpha parameter of 5.0 on unigram gives the highest accuracy. The system test with 2085 data showed that the majority of UGM students expressed negative sentiments towards SIMASTER, with only 6.6% positive sentiments and 93.4% negative sentiments. Visualization of analysis results using wordcloud shows words such as "SIMASTER", "open", "value", and "notif" in positive sentiments, while negative sentiments are dominated by words such as "SIMASTER", "error" and "slow".

Keywords: Naive Bayes, scraping, case folding, stopword, tokenizer, stemming, filtering, normalization