

INTISARI

Implementasi Algoritma *Naive Bayes* untuk Analisis Sentimen

Oleh

ANA NADYA ROSYDA

20/455492/PA/19707

Analisis sentimen merupakan sebuah bidang studi yang menganalisis opini, sikap, dan emosi seseorang terhadap suatu topik. Besarnya jumlah data yang dianalisis sering kali mengakibatkan proses analisis sentimen secara manual menjadi tidak efisien. Oleh karena itu, diperlukan sistem analisis sentimen berbasis *machine learning* yang mampu mengklasifikasikan suatu opini ke dalam sentimen positif atau negatif secara otomatis. *Naive Bayes* merupakan salah satu algoritma *machine learning* yang dapat digunakan pada masalah klasifikasi, seperti analisis sentimen. Dalam skripsi ini, penelitian hanya fokus pada metode *Naive Bayes* yang bekerja pada data diskret, yaitu *Bernoulli Naive Bayes* dan *multinomial Naive Bayes*. *Bernoulli Naive Bayes* merupakan metode yang berdasar pada asumsi data biner, yaitu muncul atau tidaknya kata, sedangkan *multinomial Naive Bayes* merupakan metode yang berdasar pada frekuensi kemunculan kata. Penelitian ini berfokus pada kajian mengenai *Bernoulli Naive Bayes* dan *multinomial Naive Bayes* secara matematis, serta implementasi kedua metode tersebut pada analisis sentimen menggunakan studi kasus mengenai *tweet* terkait program Merdeka Belajar Kampus Merdeka (MBKM) yang diunggah di media sosial X (Twitter) dalam rentang waktu November 2023 s.d. Februari 2024. Selain itu, pada penelitian ini dibahas tentang perbandingan hasil numerik dari kedua metode tersebut melalui evaluasi menggunakan perhitungan metrik berdasarkan *confusion matrix*. Hasil evaluasi menunjukkan bahwa *multinomial Naive Bayes* memiliki nilai akurasi yang lebih tinggi dibandingkan *Bernoulli Naive Bayes*. Di samping itu, dengan melihat waktu komputasi kedua metode tersebut, metode *multinomial Naive Bayes* memiliki kecepatan yang lebih tinggi dibandingkan *Bernoulli Naive Bayes*. Dengan demikian, dapat disimpulkan bahwa dalam kasus ini metode *multinomial Naive Bayes* memiliki efektivitas dan performa yang lebih baik dibandingkan metode *Bernoulli Naive Bayes*.

Kata kunci: *Bernoulli Naive Bayes*, *multinomial Naive Bayes*, *machine learning*, analisis sentimen, klasifikasi, Twitter

ABSTRACT

Implementation of the Naive Bayes Algorithm for Sentiment Analysis

By

ANA NADYA ROSYDA

20/455492/PA/19707

Sentiment analysis is a field of study that analyzes people's opinions, attitudes, and emotions toward topics. The large amount of data often leads to the inefficiency of the manual sentiment analysis process. Therefore, a machine learning-based sentiment analysis system is needed to automatically classify opinions into positive and negative sentiments. Naive Bayes is one of the machine learning algorithms that can be used in classification problems such as sentiment analysis. In this thesis, the research focuses on the methods that work on discrete data, Bernoulli Naive Bayes and multinomial Naive Bayes. Bernoulli Naive Bayes is a method based on the assumption of binary data, i.e. whether a word appears or not, while multinomial Naive Bayes is a method based on the term frequency. This research focuses on the mathematical study of Bernoulli Naive Bayes and multinomial Naive Bayes, also the implementation of both methods in sentiment analysis using a case study of tweets related to Merdeka Belajar Kampus Merdeka (MBKM) program uploaded on social media X (Twitter) in the time period November 2023 - February 2024. In addition, this research also discusses the comparison of the numerical results of both methods through quantitative evaluation using the confusion matrix. The evaluation results show that the multinomial Naive Bayes has a higher accuracy than the Bernoulli Naive Bayes. In addition, by looking at the computational time of each method, the multinomial Naive Bayes has a higher speed than the Bernoulli Naive Bayes. Thus, it can be concluded that in this case the multinomial Naive Bayes method outperforms the Bernoulli Naive Bayes method in terms of effectiveness and performance.

Keywords: Bernoulli Naive Bayes, multinomial Naive Bayes, machine learning, sentiment analysis, classification, Twitter