

INTISARI

**KLASIFIKASI KONTEN PORNOGRAFI BERBASIS TEXT MINING
MENGUNAKAN ALGORITMA NAÏVE BAYES**

Oleh :
ALDWI MANDAK
1911010070

Bandar Lampung, Telp 098676086001

Teknik Informatika, Institut Informatika dan Bisnis Darmajaya, Bandar Lampung

E-Mail: aldwimandak.1911010070@mail.darmajaya.ac.id

Konten *website* atau *blog* yang mengandung pornografi adalah jenis konten yang memiliki dampak buruk terhadap moralitas dan integritas individu, terutama jika di konsumsi oleh anak-anak dan remaja. Oleh karena itu, dalam penelitian ini, dilakukan pengklasifikasian konten website atau blog ke dalam kategori pornografi dan non-pornografi menggunakan algoritma *Naïve Bayes* menggunakan model *Multinomial Naïve Bayes* dengan metode *Machine Learning*. Penelitian ini menggunakan 49.702 dokumen teks dari *website* atau *blog* dan menganalisis pola dari setiap dokumen teks untuk mengembangkan model klasifikasi *Multinomial Naïve Bayes*. Metode *Machine Learning Life Cycle* digunakan sebagai tahapan pembuatan model klasifikasi *Multinomial Naïve Bayes*, dimulai dari pengumpulan *dataset*, pembersihan data, pelabelan data, penambahan fitur pada model untuk meningkatkan performa model klasifikasi, pelatihan model, evaluasi model, hingga pembuatan sistem klasifikasi. Hasil dari penelitian ini menunjukkan bahwa dengan menggunakan data training berjumlah 34.791 data yang kemudian diterapkan teknik *SMOTE* sehingga data *training* yang digunakan berjumlah 45.366 data model *Multinomial Naïve Bayes* berhasil mengklasifikasikan dokumen teks dari *website* atau *blog* ke dalam kategori pornografi dan non-pornografi dengan akurasi sebesar 94.61%, *Precision* sebesar 92.14%, *Recall* sebesar 97.53%, dan *F1-Score* sebesar 94.76%.

Kata Kunci: *Naïve Bayes, Multinomial Naïve Bayes, Pornografi, Machine Learning, Klasifikasi Konten, Website, Blog*

ABSTRACT

CLASSIFICATION OF PORNOGRAPHY CONTENT BASED ON TEXT MINING USING THE NAÏVE BAYES ALGORITHM

By:

ALDWI MANDAK

1911010070

Bandar Lampung, Telp 098676086001

Teknik Informatika, Institut Informatika dan Bisnis Darmajaya, Bandar Lampung

E-Mail: aldwimandak.1911010070@mail.darmajaya.ac.id

The content of websites or blogs containing pornography is a type of content that has a negative impact on the morality and integrity of individuals, especially when consumed by children and adolescents. Therefore, in this study, website or blog content was classified into pornographic and non-pornographic categories using Naïve Bayes algorithm with Multinomial Naïve Bayes model and Machine Learning method. The study used 49.702 text documents from websites or blogs and analyzed the patterns of each text document to develop a Multinomial Naïve Bayes classification model. The Machine Learning Life Cycle method was used as a stage in the development of the Multinomial Naïve Bayes classification model, starting from the collection of the dataset, data cleaning, labeling data, adding features to the model to improve its performance, model training, model evaluation, and finally creating a classification system. The results of this study show that the Multinomial Naïve Bayes model successfully classified text documents from websites or blogs into pornographic and non-pornographic categories with an accuracy of 94.61%, precision of 92.14%, recall of 97.53%, and F1-Score of 94.76%.

Key Words: Naïve Bayes, Multinomial Naïve Bayes, Pornographic, Machine Learning, Content Classification, *Website, Blog*