

# KLASIFIKASI TEKS PORNOGRAFI BERBASIS MACHINE LEARNING MENGGUNAKAN METODE NAÏVE BAYES

Samsudin Tambung<sup>1</sup>, Kartika C. Pelangi<sup>2</sup>, Muis Nanja<sup>3</sup>, Yusrianto Malago<sup>4</sup>

<sup>1,2,3</sup>Fakultas Ilmu Komputer, Program Studi Teknik Informatika, Ichsan Gorontalo, Kota Gorontalo, Indonesia  
Email: <sup>1</sup>samsudintambung71@gmail.com, <sup>2</sup>apelangie@gmail.com, <sup>3</sup>muisnaja@gmail.com,  
<sup>4</sup>yriant.iiixvix.fikom@gmail.com

**Abstrak-** Penyebaran informasi melalui internet yang cukup pesat, Tidak hanya dampak positif, internet juga memberikan dampak negatif, salah satunya adalah mudahnya dalam menemukan informasi yang mengandung hal negatif seperti pornografi. Hal tersebut dapat ditanggulangi dengan penelitian mengenai klasifikasi teks pornografi berbasis machine learning menggunakan metode Naive Bayes. Naive Bayes merupakan pengklasifikasi probalistik yang dimana menghitung sekumpulan probabilitas berdasarkan dengan jumlah frekuensi dan kombinasi setiap nilai dari dataset yang diberikan. Dengan adanya Naive Bayes, diharapkan dapat melakukan klasifikasi teks judul cerita yang mengandung pornografi yang sudah didownload melalui laman web. Jenis data yang digunakan berupa judul cerita pornografi yang sudah didownload melalui situs porno. Ada 100 dataset judul cerita yang di gunakan. Tahapan-tahapan dalam melakukan klasifikasi menggunakan Algoritma Naive Bayes Classifier antara lain, Preprocessing, pembobotan TF-IDF, klasifikasi Naive Bayes, dan pengujian menggunakan Classification Accuration, Precision, dan Recall. Hasil dari pengujian ini dengan menggunakan metode Data Splitting yang diujikan menjadi 5 perbandingan dengan nilai TF-IDF dengan setiap Level Akurasi, Presisi, dan Recall hingga 76,44%, 82,43%, 73,43%. Terlebih lagi, karakterisasi tanpa menggunakan esteem TF-IDF masing-masing mendapatkan esteem Accuration, Precision, Recall hingga 97%, 96%, dan 93%. Akhirnya adalah bahwa teknik Naive Bayes tanpa TFIDF menciptakan eksekusi yang lebih disukai daripada TF-IDF.

Kata kunci : Teks Pornografi, Klasifikasi, Naive Bayes

**Abstract-** The information via the internet spreads rapidly. Not only does it have a positive impact, but also a negative impact, one of which is the ease of finding information that contains negative things such as pornography. That problem can be overcome by research on machine learning-based classification of pornographic texts using the Naive Bayes method. Naive Bayes is a probabilistic classifier that calculates a set of probabilities based on the number of frequencies and combinations of each value from a given dataset. Through Naive Bayes, it is expected that there is a possibility to classify the text of story titles containing pornography downloaded via a web page. The type of data used is pornographic story titles downloaded via porn sites. There are 100 story title datasets employed. The classification stages using the Naive Bayes Classifier Algorithm include Preprocessing, TF-IDF weighting, Naive Bayes classification, and testing using Classification Accuracy, Precision, and Recall. The test results using the Data Splitting method are examined into five comparisons with TF-IDF values with each level of Accuracy, Precision, and Recall up to 76.44%, 82.43%, and 73.43%. The characterization without using TF-IDF, respectively, obtained Accuracy, Precision, and Recall values up to 97%, 96%, and 93%. The result is that the Naive Bayes technique without TF-IDF creates a more favorable execution than TF-IDF.

Keywords: Pornographic Text, classification, Naive Bayes

## 1. PENDAHULUAN

Pornografi telah menjadi hal yang umum karena sangat mudah diakses oleh setiap kalangan usia. Menurut Laporan Terakhir Tentang Pornografi, Konsumen utama pornografi (baik dari majalah, internet, tabloid, dan lainnya) adalah anak-anak berusia 12 hingga 17 tahun [1]. Di era mekanis saat ini, hiburan erotis secara efektif terbuka melalui media, khususnya media virtual, Biaya sewa web

yang terjangkau untuk remaja dan anak-anak, serta ponsel yang memiliki aplikasi web membuat hiburan erotis semakin mudah dibuka melalui media virtual. Tidak hanya itu, tidak sedikit buku, majalah, film dan komik yang sengaja atau tidak mengandung unsur hiburan seksual untuk membangun nilai jualnya [2]

Ketergantungan pada pornografi adalah cara hidup negatif lainnya yang merupakan kondisi medis umum di Indonesia. Hiburan seksual adalah ketergantungan lain (perbudakan) yang tidak terlihat oleh mata, tidak terdengar oleh telinga, kebiasaan ini sering diabaikan meskipun efeknya merusak pikiran, Hiburan seksual merusak lima bagian otak manusia, hiburan erotis dapat mencukupi kebutuhan baru mereka dengan lebih efektif, kapan pun, di mana pun, bahkan melalui ponsel. Oleh karena itu, diperlukan bimbingan dan pengawasan dari berbagai latar belakang, terutama untuk anak-anak, remaja, dan dewasa muda, sehingga mereka dapat menghindari resiko kecanduan baru, terutama dengan menyebarkan informasi tentang dampak pornografi yang sangat tahan lama pada otak pecandu (Layanan Kesejahteraan RI, 2012). [3]

Di Indonesia ada beberapa upaya yang dilakukan untuk mengalahkan isu pornografi di web, salah satunya melalui undang-undang tidak resmi. Dalam pasal 1 pedoman imam korespondensi dan Data Negara Republik Indonesia (PERMENKOMINFO RI) Nomor 19 Tahun 2014 tentang Perlakuan Tujuan Web Substansi Negatif, dimaklumi bahwa halaman web tersebut adalah situs yang berisi hiburan erotis dan operasi kriminal lainnya sesuai dengan peraturan dan pedoman. Kominfo memblokir dengan memboikot tujuan cabul di DNS Nawala, DNS Nawala adalah administrasi penyaringan DNS yang gratis dan dapat di gunakan oleh semua klien web. [4]

## 2. TINJAUAN PUSTAKA

### 2.1 Pengertian Pornografi

Pornografi adalah gambar, sketsa, ilustrasi, foto, tulisan, suara, bunyi, gambar bergerak, animasi, kartun, percakapan, gerak tubuh, atau bentuk pesan lainnya melalui berbagai bentuk media komunikasi dan/atau pertunjukan di muka umum, yang memuat kecabulan atau eksploitasi seksual yang melanggar norma kesusilaan dalam masyarakat permenkes RI No. 129 tahun 2018. [6].

### 2.2. Machine Learning

Machine Learning (ML) adalah studi ilmiah tentang algoritma dan model statistik yang menggunakan sistem komputer untuk melakukan tugas tertentu secara efektif tanpa menggunakan instruksi yang eksplisit, dengan mengandalkan pola dan inferensi sebagai gantinya. (Koza, dkk., 1996). Hal itu dapat dikatakan sebagai Kecerdasan Buatan (Artificial Intelligence). Dalam ML, tugas umum yang dilakukannya adalah mempelajari dan membangun algoritma sehingga menghasilkan prediksi dari data yang telah dipelajari. (Snijders, C., Matzat, U., & Reips., 2012). [11]

### 2.3 Text Mining

*Text mining* dapat diartikan sebagai suatu alur proses dimana pengguna berinteraksi dengan kumpulan dokumen menggunakan aplikasi analisis. *Text mining* merupakan suatu proses untuk mendapatkan informasi yang berguna dari sumber data dengan mengidentifikasi dan mengeksplorasi pola tertentu. Menurut Harlian (2006) *Text mining* adalah menambang data yang berupa teks dimana sumber data biasanya didapatkan dari dokumen dan tujuannya adalah mencari kata-kata yang dapat mewakili isi dari dokumen sehingga dapat dilakukan analisa keterhubungan antar dokumen. (Harlian, 2006). [12]

### 2.4 Klasifikasi

Klasifikasi merupakan proses pencarian dalam sekumpulan model/fungsi untuk menggambarkan dan membedakan kelas data dengan tujuan sehingga model tersebut digunakan untuk suatu objek yang belum diketahui kelasnya. klasifikasi juga mempunyai dua proses yaitu membangun model klasifikasi dari sekumpulan kelas data yang sudah didefinisikan sebelumnya (*training data set*) dan menggunakan model untuk klasifikasi data set untuk mengukur akurasi dari model [13].

## 2.4 Algoritma Naïve Bayes

Naïve bayes Classifier dengan istilah Bayesian Classification merupakan metode pengklasifikasian statistik yang didasarkan pada teorema bayes digunakan sebagai memprediksi probabilitas keanggotaan suatu kelas. Sudah terbukti bahwa Bayesian classification memiliki akurasi cepat dalam pengaplikasian ke dalam databases yang besar [14].

Bentuk persamaan teorema bayes adalah sebagai berikut:

$$P(H|X) = \frac{P(X|H) \cdot P(H)}{P(X)} \dots \dots \dots (1)$$

Dimana :

X = Data dengan kelas yang belum diketahui

H = Hipotesis data X merupakan suatu kelas spasifik

P(H|X) = Probabilitas hipotesis H berdasarkan kondisi X (posterior probability)

P(H) = Probabilitas hipotesis H (prior probability)

P(X|H) = Probabilitas X berdasarkan kondisi pada hipotesis H

P(X) = Probabilitas X

Langkah - langkah perhitungan Metode Naïve Bayes : [15].

- 1) Menghitung prior masing-masing kelas yaitu dengan cara menghitung total masing-masing label kelas pada data latih dan membaginya dengan total data latih.
- 2) Menghitung likelihood, yaitu menghitung probabilitas masing-masing atribut
- 3) Menghitung posterior
- 4) Menentukan label kelas dengan melakukan perbandingan antar nilai posterior. Label kelas dengan nilai posterior terbesar akan menjadi label kelas data yang diuji

## 3. METODE PENELITIAN

Dipandang dari tingkat penerapan maka, penelitian ini merupakan penelitian terapan Dipandang dari jenis informasi yang di olah maka, penelitian ini merupakan penelitian kuantitatif. berdasarkan dari perilaku terhadap data maka penelitian ini merupakan penelitian konfirmatori.

### 3.1 Pengumpulan Data

Untuk mengumpulkan data digunakan 2 (dua) jenis data yaitu data primer dan data sekunder. Data primer yaitu data yang berasal dari penelitian lapangan dan data sekunder berasal dari penelitian kepustakaan.

1. Penelitian Data Primer (Lapangan)

Untuk memperoleh data primer yang merupakan data langsung dari objek penelitian yaitu beberapa situs internet.

2. Penelitian Data Sekunder (Kepustakaan)

Metode kepustakaan diperlukan untuk mendapatkan data sekunder dengan tujuan melengkapi data primer. Data sekunder didapatkan dari pengkajian kepustakaan yang berisi dasar-dasar teori. Metode kepustakaan digunakan oleh analis sistem dengan cara mengambil contoh dokumen-dokumen yang berhubungan dengan materi penelitian. Selain itu, analis sistem mencari data mengenai hal-hal atau variabel yang berupa catatan, buku, majalah, dan sebagainya yang berhubungan dengan penelitian

#### 3.3.1 Pengembangan Model

Prosedur atau langkah-langkah pokok dalam prediksi menggunakan metode Naïve Bayes. untuk klasifikasi teks pornografi dengan menggunakan alat bantu tools PHP, Database MySQL serta *White Box Testing* dan *Black Box Testing* untuk menguji kinerja sistemnya.

## 4. HASIL DAN PEMBAHASAN

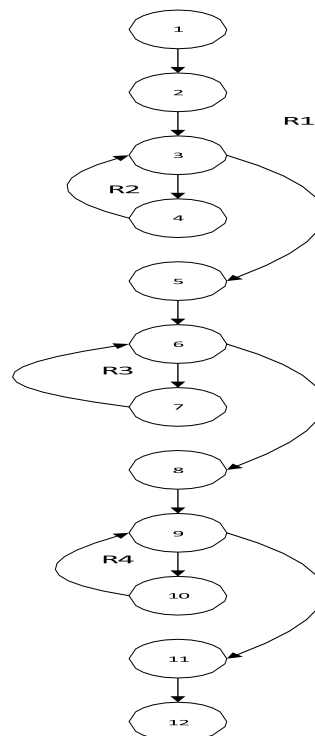
### 4.1 Hasil Pengumpulan Data

Pengembangan sistem dari aplikasi ini dimulai dari pengumpulan data adapun data yang berhasil di kumpulkan adalah sebagai berikut.

**Tabel 4.1** Data Training

No	Text	Label
1	Cerita Dewasa Menikmati DiPerkosa Anak Kost Ku Sendiri	Pornografi
2	Bareskrim Periksa Zul Zivilia Terkait Peredaran Narkoba Fredy Pratama	Non Pornografi
3	Cerita Bokep Anak Gadis SMA Ngentot Di Semak Semak	Pornografi
4	Pengasuh Rayyanza Ketar-ketir Diberi Kejutan Ultah oleh Raffi dan Nagita	Non Pornografi
.....	.....	.....
93	Ammar Zoni Merugi Ratusan Juta Rupiah Akibat Tersandung Kasus Narkoba	Non Pornograffi
94	Cerita Sex Nafsu Birahi FannyYang Sangat Binal Dan Menggoda	Pornografi
95	14 Fakta Pilu Dini Sera Afrianti Tewas Dianiaya Pacar	Non Pornografi
96	Cerita Panas Aku Di Ajarin Bersetubuh Dengan Mbak Linda	Pornografi
97	Bupati Gorontalo Mintah Anak Sekolah Jauhi Narkoba	Non Pornografi
98	cerita panas aku sungguh enak menikmati memek segar muridku	Pornografi
99	Apa Salah Dini Dianiaya hingga Tewas oleh Ronald Anak DPR RI?	Non Pornografi
100	cerita tante yani guru sex yang enak ku entot berbagai gaya	Pornografi

### 4.1.2 Flowgraph Untuk Pengujian *White Box*



**Gambar 1.** Flowgraph

### 4.1.3 Perhitungan CC pada Pengujian *White Box*

Diketahui :

Region (R) = 4  
 Node (N) = 12  
 Edge (E) = 14  
 Predicate Node (P) = 3

Rumus :

$$V(G) = (E - N) + 2 \text{ atau } VG = P + 1$$

Penyelesaian :

$$V(G) = (14 - 12) + 2 = 4$$

$$V(G) = 3 + 1 = 4$$

(R1, R2, R3, R4)

### 4.1.4 *Path* pada Pengujian *White Box*

**Tabel 4.7** *Path* Pengujian *White Box*

No	<i>Path</i>	Ket
1	1-2-3-5-6-8-9-11-12	Ok
2	1-2-3-4-3....	Ok
3	1-2-3-5-6-7-6.....	Ok
4	1-2-3-5-6-8-9-10-9.....	Ok

## KESIMPULAN

### 5.1. Kesimpulan

Dari hasil penelitian dengan Program Klasifikasi teks pornografi berbasis machine learning menggunakan metode Naive Bayes, maka pada akhir laporan penelitian ini penulis menyimpulkan bahwa :

1. dari hasil Penelitian ini peneliti dapat mengetahui bagaimana cara mengklasifikasi teks pornografi menggunakan metode Naive Bayes.
2. Metode Naive Bayes digunakan untuk mengklasifikasi Teks Pornografi. Hal ini dapat dilihat dari pengujian yang dilakukan pada prosedur dengan mendapat nilai  $CC=VG=R$  yaitu 3. Dan juga telah dilakukan pengujian Akurasi hasil prediksi dengan nilai mape sebesar 100.%.  
3.

## DAFTAR PUSTAKA

- [1] Amalina, Nur. Sistem Rekomendasi Dosen Pembimbing Tugas Akhir Berbasis Text Mining Menggunakan Vector Space Model. Jawa Tengah: Universitas Diponegoro, 2017.
- [2] Arifin. 2018a. "adrian satria putra." Klasifikasi status gizi balita menggunakan naive bayes classification (studi kasus posyandu ngudi luhur) 1(1):1-8.
- [3] Arifin. 2018b. "zia ayu nuansa gumilang." Implementasi naive bayes classifier dan asosiasi untuk analisis sentimen data ulasan aplikasi e-commerce shopee pada situs google play 1(1):1-8.

- [4] Brownlee, J. (2013). How To Use Word Embedding Layers For Deep Learning With Keras. *International Journal On Computational Sciences & Applications (Ijcsa)*, 3
- [5] Fitri, Meisya. "Perancangan Sistem Temu Balik Informasi Dengan Metode Pembobotan Kombinasi Tf-Idf Untuk Pencarian Dokumen Berbahasa Indonesia." *Jurnal Sistem dan Teknologi Informasi (JustIN)*, 2013: 1-6.
- [6] Gumilang, z. A. (2018). Implementasi naïve bayes classifier dan asosiasi untuk analisis sentimen data ulasan aplikasi e-commerce shopee pada situs google play. Yogyakarta: universitas islam indonesia.
- [7] Harlian, m. (2006). *Machine learning text*. University of texas, 14-22
- [8] Mahardika, rendra. 2018. "rendra mahardika." Identifikasi konten pornografi berbahasa indonesia menggunakan algoritma support vector machine. Permenkes ri no 129 tahun. 2008. "undang-undang republik indonesia nomor 44 tahun 2008 tentang pornografi." Undang-undang republik indonesia nomor 44 tahun 2008 tentang pornografi (may).
- [9] Mooney, Raymond J. *Machine Learning Text Categorization*. Texas: University of Texas, 2006.
- [10] Pratama, Rizki Rino. "Analisis Model Machine Learning Terhadap Pengenalan Aktifitas Manusia ." *Matrik : Jurnal Manajemen,TEKNIK*, 2020: 302-311.
- [11] Pratama, rizki rino. 2020. "analisis model machine learning terhadap pengenalan aktifitas manusia." *Matrik : jurnal manajemen, teknik informatika dan rekayasa komputer* 19(2):302–11. Doi: 10.30812/matrik.v19i2.688.
- [12] Pratama, rizki rino. 2020. "analisis model machine learning terhadap pengenalan aktifitas manusia." *Matrik : jurnal manajemen, teknik informatika dan rekayasa komputer* 19(2):302–11. Doi: 10.30812/matrik.v19i2.688.