

Implementasi *Machine Learning* Pada Sistem Pendeteksi URL Bermuatan Konten Negatif Menggunakan Metode Algoritma *Naive Bayes* Dan *Support Vector Machine*

Rada Rasi Saputri^{1*}, Agus Heri Yunial¹

^{1,2}Fakultas Teknik, Teknik Informatika, Universitas Pamulang, Jl. Raya Puspiptek No. 46, Kel. Buaran, Kec. Serpong, Kota Tangerang Selatan. Banten 15310, Indonesia

Email: ^{1*}radarrs18@gmail.com, ²Dosen02525@unpam.ac.id

(* : coressponding author)

Abstrak– Sistem untuk melakukan penyaringan situs yang mengandung konten negatif telah banyak dilakukan oleh para peneliti-peneliti sebelumnya. Namun sistem-sistem tersebut lebih banyak berfokus hanya pada 1 jenis konten negatif saja dan sebagian besar dibangun untuk situs yang berbahasa Inggris. Adapun sistem yang dapat melakukan penyaringan URL menggunakan Bahasa Indonesia hanya berfokus pada konten negatif. Penelitian ini bertujuan untuk menciptakan sebuah sistem pendeteksi URL yang mengandung konten negatif dengan menggunakan model *Machine Learning*. Sistem pada penelitian ini melakukan penyaringan konten pada URL yang menggunakan bahasa Indonesia. Penelitian ini menggunakan 2 model utama yaitu *Naive Bayes*, *Support Vector Machine*. Dari keseluruhan model yang digunakan, model *SVM* menghasilkan akurasi tertinggi yaitu sebesar 96.161%.

Kata Kunci: *Machine Learning*, Klasifikasi, Konten Negatif

Abstract– *Systems for filtering sites that contain negative content have been carried out by many previous researchers. However, these systems focus more on only 1 type of negative content and are mostly built for sites that are in English. The system that can filter URLs using Indonesian only focuses on negative content. This study aims to create a URL detection system that contains negative content using a Machine Learning model. The system in this study filters content on URLs that use Indonesian. This study uses 2 main models, namely Naive Bayes, Support Vector Machine. Of all the models used, the SVM model produces the highest accuracy of 96.161%.*

Keywords: *Machine Learning, Classification, Negative Content*

1. PENDAHULUAN

Perkembangan teknologi informasi yang semakin pesat seperti saat ini menyebabkan berbagai dampak positif dan dampak negatif. Segala informasi (konten positif dan negatif) tersedia di internet dan dengan mudah diakses oleh berbagai kalangan masyarakat diantaranya adalah kalangan pelajar. Konten negatif (pornografi) yang terkandung di dalam internet dapat berdampak buruk dan mempengaruhi keadaan psikis maupun mental khususnya para kalangan pelajar yang mengaksesnya. Sebagai generasi penerus bangsa, generasi pelajar haruslah dibekali dengan kesiapan mental dan keteguhan iman sehingga generasi ini akan lebih produktif, kreatif dan memajukan Negara Indonesia. Pengguna internet di Indonesia termasuk dalam kategori medium yang artinya penggunaan internet merata di berbagai wilayah di Indonesia. Sebagaimana data yang telah disajikan oleh Kementerian Komunikasi dan Informatika RI bahwa Indonesia menduduki peringkat enam dunia sebagai pengguna internet (Hardiyanto and Sartika 2017).

Pendeteksian konten negatif dari suatu halaman *website* dapat dilakukan dengan cara mengidentifikasi keseluruhan isi *website* apakah mengandung kalimat atau kata berunsur negatif. Konten negatif yang diatur di UU ITE adalah sebagai berikut yaitu, konten ilegal (pornografi, perjudian, penghinaan / pencemaran nama baik, pengancaman dan pemerasan), akses ilegal, intersepsi ilegal, *data interference*, *system interference*, dan *misuse of device*. Pelaku MS dijerat Pasal 27 dan Pasal 45 Undang-Undang Nomor 19 Tahun 2016 tentang Perubahan atas Undang-Undang Nomor 11 tahun Pornografi. Identifikasi referensi atau dataset dilakukan dengan data dari 2008 tentang Informasi dan Transaksi Elektronik, juga Pasal 4 ayat (1) Undang-Undang Nomor 44 Tahun 2008 tentang hasil pengumpulan secara manual dengan tehnik *dork* pada google.

Machine Learning (ML) atau pembelajaran mesin merupakan pendekatan dalam AI yang banyak digunakan untuk menggantikan atau menirukan perilaku manusia untuk menyelesaikan

masalah atau melakukan otomatisasi. Sesuai namanya, ML mencoba menirukan bagaimana proses manusia atau makhluk cerdas belajar dan mengeneralisasi. Setidaknya ada dua aplikasi utama dalam ML yaitu, klasifikasi dan prediksi. Ciri khas dari ML adalah adanya proses pelatihan, pembelajaran, atau training. Oleh karena itu, ML membutuhkan data untuk dipelajari yang disebut sebagai data *training*. Klasifikasi adalah metode dalam ML yang digunakan oleh mesin untuk memilah atau mengklasifikasikan obyek berdasarkan ciri tertentu sebagaimana manusia mencoba membedakan benda satu dengan yang lain (Ahmad 2017).

Kelemahan-kelemahan yang dimiliki oleh metode manual, mendorong para peneliti untuk mengembangkan suatu metode yang tidak bergantung 100% pada manusia. Metode yang dikembangkan memanfaatkan komputer sebagai alat bantu untuk menganalisa konten media sosial. Hal ini merupakan sebuah tantangan karena komputer hanya mengenali bahasa mesin, sedangkan konten yang beredar di dunia media sosial menggunakan bahasa yang dimengerti manusia. Permasalahan tersebut dapat diatasi dengan mengkuantifikasi data. Data yang telah terkuantifikasi selanjutnya dipisahkan, dikelompokkan, dengan pendekatan *Machine learning* (Laksana Utama 2018).

Di negara kita pornografi merupakan sesuatu yang ilegal karena sifatnya yang merusak bagi setiap orang yang mengaksesnya, terutama kalangan remaja dan anak-anak. Karena itu di Indonesia sudah dibentuk Undang-Undang yang mengatur tentang pornografi yaitu UU No. 44 Tahun 2008 Tentang Pornografi (UU Pornografi). Sementara itu juga melalui Kementerian Komunikasi dan Informatika pemerintah sudah berusaha mencegah penyebaran konten negatif khususnya konten pornografi di Indonesia melalui internet, salah satunya dengan membuat Layanan DNS (Tullah, Maisaroh, and Sumpena 2020).

Algoritma ini sudah sangat terkenal dan disukai karena memiliki banyak kelebihan. Kelebihan ini misalnya dapat mengolah data numerik dan *diskret*, dapat menangani nilai atribut yang hilang satu yang tercepat dibandingkan dengan algoritma lain. Ide dasar dari algoritma ini adalah pembuatan pohon keputusan berdasarkan pemilihan atribut yang memiliki prioritas tertinggi atau dapat memiliki gain tertinggi berdasarkan nilai *entropy* atribut tersebut sebagai poros atribut. Kemudian secara rekursif cabang-cabang pohon diperluas sehingga seluruh pohon terbentuk. *entropy* adalah jumlah data yang tidak relevan terhadap informasi dari suatu kumpulan data (Tullah, Maisaroh, and Sumpena 2020).

2. METODOLOGI PENELITIAN

2.1 Google Dork

Berikut cara memakai Google Dork :

a. Site

Parameter pencarian site digunakan buat memunculkan isu dari sebuah situs eksklusif dengan ekstensi yg Anda cari.

Rumus = Keyword + Site : (Nama situs)

contohnya: ekamatra kuantum site : wikipedia.org

b. Intitle

Parameter pencarian intitle dipergunakan buat menampilkan informasi berasal seluruh website yang mengandung keyword yang Anda masukkan.

Rumus = intitle: (Keyword)

contoh = intitle : ekamatra kuantum

c. Filetype

Parameter filetype dipergunakan buat menampilkan isu dalam berbagai situs menggunakan format khusus.

Rumus = filetype:(pdf/doc) + keyword

model: filetype:pdf fisika kuantum

2.2 Python Library Scikit Learn

Berikut cara menggunakan perpustakaan:

a. *NumPy*: Untuk pekerjaan apa pun dengan matriks, terutama operasi matematika

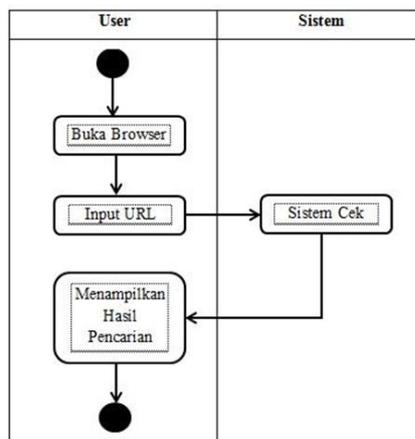
b. *SciPy*: Komputasi ilmiah dan teknis

- c. *Matplotlib*: Visualisasi data
- d. *IPython*: Konsol interaktif untuk Python
- e. *Sympy*: Matematika simbolik
- f. *Pandas*: Penanganan, manipulasi, dan analisis data

3. ANALISA DAN PEMBAHASAN

3.1 Analisa Sistem Berjalan

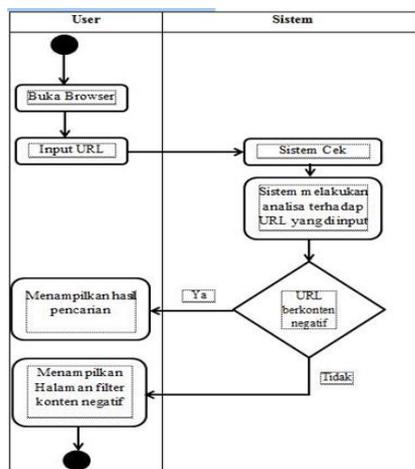
Analisa sistem yang berjalan sangat diperlukan untuk analisa masalah yang akan terjadi dan menemukan jawaban dari masalah yang timbul. Pada sistem berjalan dilakukan oleh beberapa pelaku sistem, yaitu user melakukan input URL yang akan diakses. Setelah input URL dimasukan, URL tersebut akan otomatis terdeteksi apakah URL tersebut Negatif atau Positif. Adapun analisa sistem saat ini yang digambarkan dalam activity diagram adalah sebagai berikut:



Gambar 1. Activity Diagram Sistem Berjalan

3.2 Analisa Sistem Usulan

Dari hasil analisa sistem saat ini, penulis memberikan sistem usulan yaitu, user atau admin dapat membuka *browser* dengan memasukan perintah cmd lalu user / admin memasukan perintah agar bisa masuk ke halaman *browser*. Setelah masuk ke halaman *browser* melakukan *input URL*. Setelah terinput *URL* tersebut akan otomatis terblokir jika *URL* memuat konten negatif dan apabila terdeteksi positif *URL* tersebut bisa langsung di akses. Adapun analisa sistem saat ini yang digambarkan dalam *activity* diagram adalah sebagai berikut:



Gambar 2. Activity Diagram Sistem Usulan

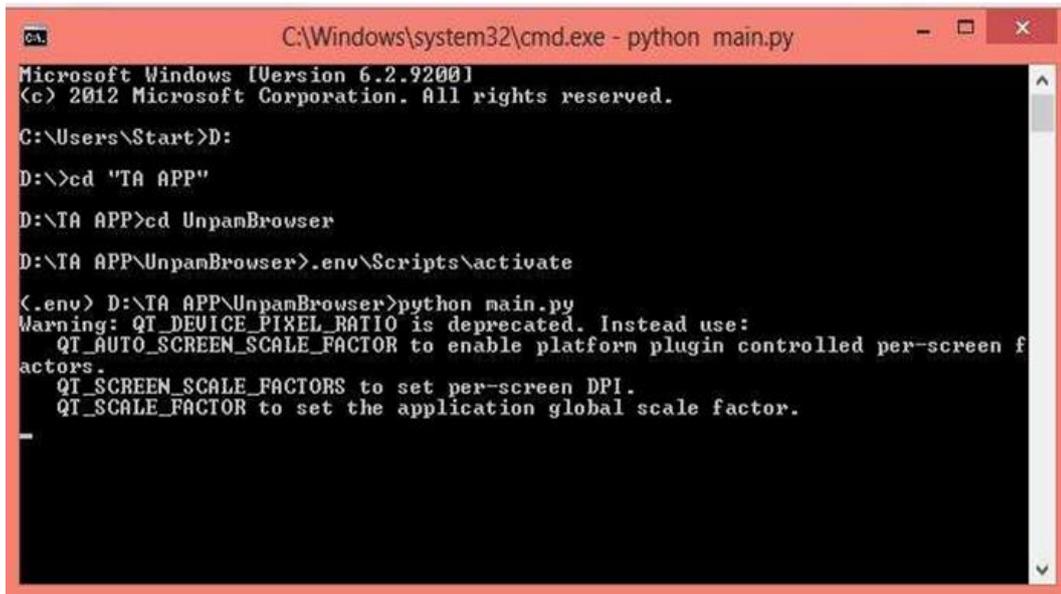
4. IMPLEMENTASI

4.1 Implementasi Program

Berikut merupakan hal yang dibutuhkan dalam membuat sistem cerdas untuk diagnosa penyakit kanker mulut rahim.

4.1.1 *Command Prompt*

Halaman *Command Propmt* sebagai suatu perintah *DOS* yang disampaikan dalam bentuk text.



```
C:\Windows\system32\cmd.exe - python main.py
Microsoft Windows [Version 6.2.9200]
(c) 2012 Microsoft Corporation. All rights reserved.
C:\Users\Start>D:
D:\>cd "TA APP"
D:\TA APP>cd UnpamBrowser
D:\TA APP\UnpamBrowser>.env\Scripts\activate
(.env) D:\TA APP\UnpamBrowser>python main.py
Warning: QT_DEVICE_PIXEL_RATIO is deprecated. Instead use:
  QT_AUTO_SCREEN_SCALE_FACTOR to enable platform plugin controlled per-screen factors.
  QT_SCREEN_SCALE_FACTORS to set per-screen DPI.
  QT_SCALE_FACTOR to set the application global scale factor.
```

Gambar 3. *Command Prompt*

4.1.2 *Halaman Menu Home*

Menu Home adalah tampilan menu utama pada aplikasi ini, tampilann ini muncul setelah pengguna memasukan perintah. Menampilkan beberapa menu yang akan digunakan.



Gambar 4. *Halaman Menu Home*

4.1.3 Tampilan URL Positif

Pada tampilan *URL* Positif, berisi informasi data *URL* yang di akses di *Unpam Browser* bersifat aman.



Gambar 5. Tampilan URL Positif

4.1.3 Tampilan URL Negatif

Pada tampilan *URL* Positif, berisi informasi data *URL* yang di akses di *Unpam Browser* bersifat tidak aman atau di blokir.



Gambar 6. Tampilan URL Negatif

4.2 Pengujian Sistem

4.2.1 Pengujian *Black-Box*

Pada pengujian *Black Box*, dilakukan pengujian terhadap fungsi tombol yang ada pada aplikasi. Setiap simbol akan diuji kesesuaian hasil *output* dan tidak adanya *error* yang didapati. Pengujian ini dirangkum dalam tabel skenario pengujian dibawah ini.

Tabel 6. Pengujian *Black-Box*

No	Menu	Fungsi	Kasus	Hasil	Keterangan
1.	<i>Command Prompt</i>	Untuk masuk ke dalam menu utama	Pengguna mengetik atau memasukan URL yang akan di akses	Masuk ke dalam Menu Utama	Sukses
2.	Menu <i>Hpme</i>	Menampilkan menu utama pada aplikasi	-	Menampilkan Menu Utama	Sukses
3.	Tampilan URL Positif	Menampilkan data URL yang di akses bersifat aman	Pengguna dapat melihat atau beralih ke URL lain	URL dapat diakses dengan aman	Sukses
4.	Tampilan URL Positif	Menampilkan data URL yang di akses bersifat tidak aman atau diblokir	Pengguna dapat melihat atau beralih ke URL lain	URL di Blokir	Sukses

4.2.2 Analisa Hasil Pengujian

Setelah melakukan pengujian menggunakan metode *Black Box* dan menjalankan aplikasi pengolahan data, dapat dianalisa bahwa setiap proses pada menu berjalan dengan baik dan berfungsi sesuai dengan raancangan yang dibuat.

5. KESIMPULAN

Berdasarkan analisis, perancangan, dan pengujian dari sistem pendeteksi URL yang mengandung konten negatif dengan menggunakan *Machine Learning*, maka diperoleh beberapa kesimpulan:

1. Penggunaan model *Support Vector Machine* saat proses *training* dan *testing* menghasilkan akurasi rata-rata tertinggi dengan nilai akurasi rata-rata mencapai lebih dari 96%
2. Pengujian menggunakan sistem yang telah dibangun menjadi bentuk *website* dengan menerima *URL* inputan dari pengguna memberikan hasil dengan persentase keberhasilan 96%
3. Sistem pendeteksi URL yang bermuatan konten negatif ini dapat digunakan sebagai alat bantu untuk mendeteksi apakah sebuah situs mengandung konten negatif atau tidak.

REFERENCES

- Kusuma, S. D. Y., Al Islami, H., Wicaksono, B. S., Nurlaila, F., & Apriyanto, B. (2020). Peningkatan Kemampuan dan Keterampilan Siswa di Bidang TIK Berbasis SKKNI Pada SMK YPUI Parung. *KOMMAS: Jurnal Pengabdian Kepada Masyarakat*, 1(3), 123-130.
- Kusuma, S. D. Y., Al Islami, H., Wicaksono, B. S., Nurlaila, F., & Apriyanto, B. (2020). Pelatihan Web Desain dan Help Desk Bagi Siswa Prodi Teknik Komputer Jaringan (TKJ) Pada SMK YPUI Parung. *JAMAICA: Jurnal Abdi Masyarakat*, 1(1), 101-114.
- Husain, T., & Ardiansyah, M. (2020). Pair-Samples T Test: Simulation Model of Financial Ratio's Measurement with Decision Support Systems (DSS) Approach. *International Journal of Advanced Trends in Engineering, Science and Technology (IJATEST)*, 5(4), 13-17.
- Zailani, A. U., Perdananto, A., Nurjaya, N., & Sholihin, S. (2020). Pengenalan Sejak Dini Siswa SMP tentang Machine Learning untuk Klasifikasi Gambar dalam Menghadapi Revolusi 4.0. *KOMMAS: Jurnal Pengabdian Kepada Masyarakat*, 1(1).