



Filtering Konten Negatif Menggunakan Metode Algoritma C4.5 dan API Bing Penerapan di Desa Kelet

Ahmad Faidlon¹, Muhammad Miftakhul Ulum², Muhammad Nabil Mas'ud³, Fabelo Adi Putra Pradana⁴, Ibnu Maulana⁵

¹⁻⁵ Universitas Islam Nahdlatul Ulama Jepara

Alamat: Jl. Taman Siswa, Pekeng, Kauman, Tahunan, Kec. Tahunan, Kabupaten Jepara, Jawa Tengah

Korespondensi penulis: faidlon.ahm@gmail.com

Abstract. A search engine is a tool used to find information on the internet by entering the desired key words or phrases, for example, if you search for information about "cars" by using a search engine, you will get results in the form of a list of websites that contain information about "cars". Search engines are very useful because they allow you to find the information you need quickly and easily. This will be even better if you combine a word/keyword filtering system which is used to filter words or phrases to promote local products. Word filters can also be used to avoid the use of inappropriate words or phrases in a particular context. In this research, I used the API from Bing to get data. The results of this research are in the form of a web application that can filter keywords entered by the user, the keywords that have been input are filtered and if there are dirty words then these words will be deleted, after the keywords are filtered they will then be forwarded to the Bing API to search for relevant articles. can be read by the user.

Keywords: Search Engine, Word Filter, API.

Abstrak. Mesin pencari adalah alat yang digunakan untuk menemukan informasi di internet dengan memasukkan kata atau frasa kunci yang diinginkan, misalnya Anda mencari informasi tentang "mobil" dengan menggunakan mesin pencari, Anda akan mendapatkan hasil berupa daftar situs web yang mengandung informasi tentang "mobil". Mesin pencari sangat berguna karena memungkinkan Anda untuk menemukan informasi yang dibutuhkan dengan cepat dan mudah. Hal ini akan semakin bagus jika digabungkan sistem filterisasi kata / keyword kata yang digunakan untuk menyaring kata atau frasa yang untuk mempromosikan produk lokal. Penyaring kata juga dapat digunakan untuk menghindari penggunaan kata atau frasa yang tidak pantas dalam suatu konteks tertentu. Pada penelitian kali ini saya menggunakan API dari bing untuk mendapatkan data. Hasil dari penelitian ini berupa aplikasi web yang dapat memfilter keyword yang diinputkan oleh user, keyword yang sudah diinputkan tersebut difilter dan jika terdapat kata-kata jorok maka kata-kata tersebut akan dihapus, setelah keyword disaring selanjutnya akan diteruskan ke API Bing untuk mencari artikel yang dapat dibaca oleh user.

Kata kunci: Mesin Pencari, Penyaring kata, API.

LATAR BELAKANG

Saat ini internet sangat dekat dengan setiap manusia salah satu fitur yang sering dipakai adalah mesin pencari, mesin pencari adalah alat yang digunakan untuk menemukan informasi di internet dengan memasukkan kata atau frasa kunci yang diinginkan (Andayani & Rullyana, 2020. p. 131) Mesin pencari sangat berguna karena memungkinkan pengguna untuk menemukan informasi yang dibutuhkan dengan cepat dan mudah. Internet sudah menjadi gaya hidup sehari-hari bagi beberapa kalangan karena penyebaran informasi yang begitu cepat tanpa ada batasan waktu dan orang-orang di seluruh dunia dapat saling mempublikasikan dan tersebar di mana-mana (Maskur, 2015. P.124). Kecanduan pornografi internet sering melibatkan konsumsi kompulsif konten pornografi atau *role-playing* dewasa, biasanya dikaitkan dengan

isolasi sosial, reaksi stres, dan masalah hubungan (Abdullah, 2019. p. 55).

KAJIAN TEORITIS

Remaja sering menjelajahi seksualitas lewat obrolan, literatur, atau aktivitas intim, mempengaruhi perkembangan psikis mereka, padahal sebaiknya fokus pada kegiatan positif untuk pertumbuhan (Utomo & Sa'i, 2018. p. 167).

Kecanduan pornografi merupakan tindakan berulang kali mengeksplorasi materi yang membangkitkan hasrat seksual, dengan ketidakmampuan untuk mengendalikan diri dan berhenti dari kebiasaan tersebut. Berikut adalah beberapa indikator yang menunjukkan seseorang mungkin mengalami kecanduan pornografi seperti Muncul kecemasan yang nyata ketika diajak berinteraksi oleh orang lain, lebih memilih untuk menghabiskan waktu sendirian, terutama di dalam kamar, lupa atau mengabaikan kebiasaan positif yang sebelumnya dimiliki (Haidar & Apsari, 2020. p. 141).

Remaja yang sering mengakses pornografi bisa memiliki pandangan terdistorsi tentang seks, melihat wanita secara objektif, menganggap seks bebas normal, toleran terhadap kekerasan seksual, dan berisiko mengembangkan perilaku seksual menyimpang (Anggraini & Maulidya, 2020. p. 51).

Pada Penelitian ini penulis membahas tentang bagaimana cara menyaring kata kotor dalam *keyword* mesin pencari. Tema ini dipilih karena Berdasarkan data yang dikumpulkan oleh similarweb.com yang merupakan situs penyedia data mengenai statistik pangsa pasar untuk teknologi internet, situs paling teratas dan paling banyak diakses adalah situs *search engine*.

Pada Artikel ini Peneliti membuat model menggunakan model FSA (*Finite State Automata*), FSA adalah model matematis yang digunakan untuk mengikuti dan mengontrol alur kerja atau proses dari suatu sistem (Maulana, Azizah & Hanifah, 2019. p. 111). Alasan peneliti menggunakan model FSA karena mudah diimplementasikan dan memiliki kompleksitas waktu yang rendah, sehingga cocok untuk digunakan dalam sistem yang membutuhkan waktu yang cepat untuk memproses data atau mengambil keputusan.

METODE PENELITIAN

Ada beberapa alat yang digunakan dalam penelitian ini yaitu API, *javascript*, *Next JS* dan *Axios*.

Javascript merupakan bahasa pemrograman berupa skrip yang beroperasi dalam dokumen HTML, menjadi bahasa skrip pertama untuk web sejak awal sejarah internet (Sahi,

2020. p. 122), selain itu *JavaScript* memungkinkan pembuatan halaman web interaktif dengan mudah, dengan programnya yang tertanam langsung dalam *file* HTML (Lay, 2017. p. 2).

API Merupakan integrasi dua bagian dari sistem aplikasi. Bagian tersebut dijumpai oleh sebuah sistem yang disebut dengan *restAPI*, dengan menggunakan API, aplikasi dapat berkomunikasi dan berkolaborasi (Yanti & Rihyanti, 2021. p. 196) dengan sistem lain untuk menyediakan data dari kalimat *keyword* yang sudah di filter

Axios adalah *library JavaScript* yang digunakan untuk melakukan permintaan HTTP (*Hypertext Transfer Protocol*) ke *server* dan menerima respon dari *server*, berguna untuk mengambil data dari *server* atau mengirim data ke *server* (Safitri & Putro, 2021. p. 3), misalnya ketika mengirim formulir atau meminta data dari API (*Application Programming Interface*).

Penggunaan *framework NextJS* aplikasi yang akan dibuat berupa web, digunakannya web dalam pengaplikasian *filtering keyword* karena *Website* dapat menyajikan berbagai informasi, termasuk teks, gambar, animasi, dan video, yang terkoneksi melalui *hyperlink* (Siswanto & Sidik, 2019. p.1). penggunaan *website* memungkinkan penyebaran informasi secara luas tanpa batas geografis dan memudahkan akses. Informasi pun dapat diperbarui dengan cepat dan mudah (Wibisono, 2015. p. 64).

Rancangan *Prototype*

Perancangan *prototype* ini digunakan untuk memperkirakan langkah kerja dari sistem pencarian, perancangan ini dilakukan sebelum diterapkan pada model FSA.

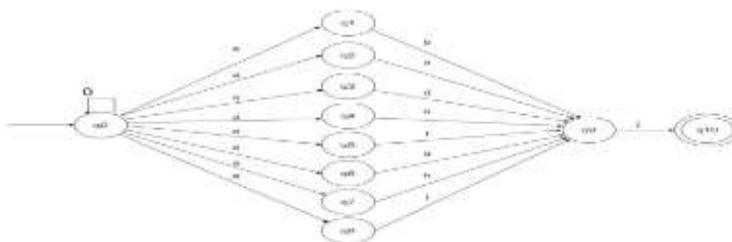


gambar 1: Rancangan

Pada rancangan tersebut terdapat input yang berupa kalimat atau kata yang selanjutnya diteruskan pada pengkondisian untuk dilakukan pemfilteran pada kalimat, pemfilteran ini dilakukan untuk menghilangkan kata jika dalam sebuah kalimat memiliki kata kotor, jika tidak maka kalimat atau kata akan diteruskan ke *API endpoint*, yang terakhir adalah output bagian ini adalah bagian yang menampilkan hasil request ke endpoint dari kalimat yang sudah difilter

Rancangan FSA

FSA merupakan model matematis yang bisa menerima *input* dan mengeluarkan output model ini digunakan peneliti untuk mensimulasikan perilaku suatu sistem dan memprediksi bagaimana sistem tersebut akan bereaksi terhadap *input* yang berbeda (Maulana, Azizah, Hanifah & Kirana, 2019. p. 111).



gambar 2: Desain FSA

Bentuk formal :

1. $Q : \{q_0, q_1, q_2, q_3, q_4, q_5, q_6, q_7, q_8, q_9, q_{10}, q_{11}, Q_{12}, q_{13}\}$
2. $\Sigma : \{a, b, c, d, e, f, g, h, i, j, k\}$
3. $S : q_0$
4. $F : q_{11}$

Tabel transisi

	a	B	C	d	E	f	g	h	i
q0									
q1	{q0}								
q2	{q0}								
q3	{q0}								
q4	{q0}								
q5	{q0}								
q6	{q0}								
q7	{q0}								
q8	{q0}								
q9	{q8}	{q1}	{q2}	{q3}	{q4}	{q5}	{q6}	{q7}	
q10									{q8}

Algoritma C4.5

Algoritma C4.5, evolusi dari ID3 yang dikembangkan oleh Quinlan, adalah teknik analisis data yang membangun pohon keputusan efisien. Menggunakan metrik seperti *information gain* dan *gain ratio*, algoritma ini memilih atribut optimal untuk klasifikasi, memungkinkan pembuatan pohon keputusan yang efektif dalam membedakan kelas data. C4.5 juga dapat menangani data tidak lengkap, menjadikannya algoritma andalan dalam analisis data dan pengambilan keputusan. (Barito, Beng & Arisandi, 2022. p. 1).

Algoritma C4.5 adalah metode klasifikasi data menggunakan pohon keputusan yang efektif untuk data numerik dan diskrit, dapat mengatasi atribut hilang, menghasilkan aturan mudah dipahami, dan merupakan salah satu yang tercepat (Pambudi, Setiawan & Indriati, 2018. p. 2637).

Proses perhitungan algoritma C4.5 melibatkan beberapa langkah penting. Tahap awalnya adalah menghitung *entropy*, untuk setiap kategori atribut. Rumus perhitungan untuk info atau *entropy*

Entropi

$$Info(D) = -\sum_{i=1}^m p_i \times \log_2(p_i)$$

$$D = \left(-\frac{7}{8} \log^2\left(\frac{7}{8}\right)\right) + \left(-\frac{1}{8} \log^2\left(\frac{1}{8}\right)\right)$$

$$D = -0.2126582926 + 0.07525749892 = -0.1374007937$$

Keterangan:

D = himpunan kasus

m = partisi data D

p_i = probabilitas dari sum(ya) dibagi dengan total kasus

Gain

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} \times Entropy(S_i)$$

Keterangan

S : himpunan kasus

A : Fitur

n : Jumlah partisi atribut A

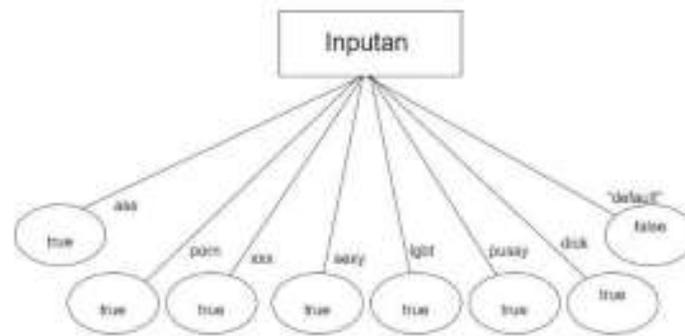
$|S_i|$: Jumlah kasus pada partisi ke-i

$|S|$: Jumlah kasus dalam S

$$-0.1374007937 - \left(\left(\frac{7}{8} \times -0.2126582926\right) + \left(\frac{1}{8} \times 0.07525749892\right)\right) = 0.03926802496$$

No	Kata	Is bad	Entropi	Total entropi	Gain
1	Ass	true	-0.2126582926	-0.1374007937	0.03926802496
2	Porn	true			
3	sexy	true			
4	Xxx	true			
5	Lgbt	true			
6	Pussy	true			
7	Dick	true			
8	"default"	false	0.07525749892		

HASIL DAN PEMBAHASAN



gambar 3: Aturan prediksi

Aturan prediksi yang diperoleh berdasarkan struktur pohon keputusan akhir, yang dikembangkan melalui analisis *Entropy* dan *Gain*, menghasilkan 8 aturan untuk memprediksi inputan yaitu sebagai berikut :

1. Menghilangkan kata yang mengandung kata “*ass*” .
2. Menghilangkan kata yang mengandung kata “*porn*”
3. Menghilangkan kata yang mengandung kata “*xxx*”
4. Menghilangkan kata yang mengandung kata “*sexy*”
5. Menghilangkan kata yang mengandung kata “*lgbt*”
6. Menghilangkan kata yang mengandung kata “*pussy*”
7. Menghilangkan kata yang mengandung kata “*dick*”
8. Jika dalam inputan tidak terdapat kata “*ass, porn, xxx, sexy, lgbt, pussy, dick*” maka script tidak melakukan penghilangan kata pada kalimat inputan

Pengujian

1. Pengujian tanpa menggunakan filter kata

Pra diterapkan pada sistem yang akan dibuat, peneliti melakukan pengujian kembali menggunakan postman pengujian ini dilakukan tanpa menggunakan filter kata. Hasil dari pengujian ini akan digunakan sebagai perbandingan untuk mengetahui seberapa efisien sistem yang akan dibuat.

a) Pengujian menggunakan *badwords*



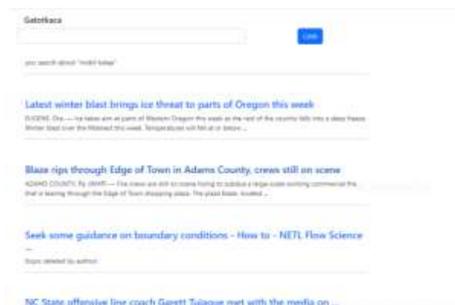
gambar 4: pengujian kedua menggunakan bad words

Pada pengujian ini peneliti menggunakan kalimat *lgbtq site*. Hasil dari API dengan menggunakan kalimat tersebut memunculkan daftar artikel dan alamat *website* yang berhubungan dengan *lgbtq*

2. Pengujian ketiga

Pasca melakukan pengujian dengan menggunakan *postman*, selanjutnya *endpoint api* diterapkan pada *script* yang sudah terdapat filter kata. peneliti menggunakan *axios* pada framework *NEXT JS* sebagai alat untuk *fetching API*, data yang didapat dari *endpoint* tersebut berupa *JSON (Javascript Object Notation)*. hasil hanya memuat *file* yang berekstensi *.json*. Ada dua elemen inti dari objek *JSON*, yaitu *key* dan *value* (Yanti & Rihyanti, 2021. p. 196).

a) Pengujian ketiga menggunakan *goodwords*



gambar 5: pengujian ketiga menggunakan good words

Pada bagian ini peneliti melakukan pengujian dengan menggunakan kalimat “mobil balap”. Kalimat tersebut bukan merupakan *badword(s)* yang artinya ketika kalimat tersebut dimasukkan tidak akan terjadi penyaringan kata, dan respon yang dihasilkan oleh API akan sesuai dengan *input*.

b) Pengujian ketiga menggunakan *badword*



gambar 6: pengujian ketiga menggunakan bad words

Pada pengujian, peneliti memasukkan "lgbt site" yang dianggap mengandung kata kasar, sehingga "lgbt" disaring dan hanya "site" yang diproses oleh API untuk menghasilkan artikel terkait "site".

Hasil dari pengujian menggunakan filter kata:

<i>Good words</i>	<i>Bad words</i>
1. Permainan Mobil Balap Bebas <i>on the App Store</i>	1. <i>No. 7 Texas vs. Louisiana gamethread - Burnt Orange Nation</i>
2. 15 Mobil Balap Terbaru 2022 & Tercepat Di Dunia Oto motifo	2. <i>Travelers and workers at MSP Airport navigate winter weather on Wednesday</i>
3. permainan mobil balap extre me - mobil balap keren - You Tube	3. <i>Eagles Reacts Survey: Week 16 - Bleeding Green Nation</i>

KESIMPULAN DAN SARAN

Pengujian pada *postman* terhadap sistem yang akan diterapkan pada desa Kelet dan sistem pencari menunjukkan bahwa tanpa filter, pencarian "lgbt site" menghasilkan konten tentang "lgbt". Berbanding terbalik dengan sistem pencari yang mempunyai filter, sistem ini menghapus kata "lgbt" dan hanya menampilkan konten yang berisi kata "site". Dari perbandingan ini dapat disimpulkan bahwa hasil pengujian pada mesin pencari dengan menggunakan *filtering* kata memberikan hasil yang positif. hal ini dikarenakan sistem yang dibuat terdapat fitur yang bisa menyaring kata yang untuk kata yang tidak relevan, dalam kasus ini Peneliti menerapkan *filtering* pada kata kotor/*badword*(s).

DAFTAR REFERENSI

- Andayani, S., Rullyana, G., Ardiansah. (2020). Pentingnya Menulis Kata Kunci Dengan Benar Pada Mesin Pencarian Untuk Melakukan Pencarian Yang Akurat. *Jurnal Iqra'*. vol 14(1). <http://dx.doi.org/10.30829/iqra.v14i1.7574>.
- Maskur. (2015). Relevansi Hasil Pencarian pada Mesin Pencari Berdasarkan Pendekatan Kata Menggunakan Ontologi. *Jurnal Gamma*, 10(1), 123–129.
- Maulana, A, S., Azizah, N, Hanifah., Kirana, C, K. (2019). Implementasi Finite State Automata (FSA) Dengan Simulasi Vending Machine Pada Aplikasi Android. *Jurnal Edukasi Elektro*. Vol 3, No2. <https://doi.org/10.21831/jee.v3i2.28332>.

- Yanti, S. N., & Rihyanti, E. (2021). Penerapan Rest API untuk Sistem Informasi Film Secara Daring. *Jurnal Informatika Universitas Pamulang*, 6(1), 195.
- Safitri, R. K., & Putro, H. P. (2021). Implementasi REST API untuk Komunikasi Antara ReactJS dan NodeJS (Studi Kasus: Modul Manajemen User Solusi247). *AUTOMATA*, 2(1).
- Barito, E. E., Beng, J. T., & Arisandi, D. (2022). Penerapan Algoritma C4. 5 untuk Klasifikasi Mahasiswa Penerima Bantuan Sosial Covid-19. *Jurnal Ilmu Komputer dan Sistem Informasi*, 10(1).
- Haidar, G., & Apsari, N. C. (2020). Pornografi pada kalangan remaja. *Prosiding Penelitian Dan Pengabdian Kepada Masyarakat*, 7(1), 136.
- Utomo, S. T., & Sa'i, A. (2018). Dampak Pornografi Terhadap Perkembangan Mental Remaja di Sekolah. *ELEMENTARY: Islamic Teacher Journal*, 6(1), 170-192.
- Siswanto, S., & Sidik, M. (2019). Perancangan Website CV. Cipta Cemerlang Abadi Semarang Sebagai Media Promosi. *Pixel: Jurnal Ilmiah Komputer Grafis*, 12(2), 1-12.
- Pambudi, R. H., Setiawan, B. D., & Indriati, I. (2018). Penerapan Algoritma C4. 5 Untuk Memprediksi Nilai Kelulusan Siswa Sekolah Menengah Berdasarkan Faktor Eksternal. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 2(7), 2637-2643.
- Wibisono, G. (2015). Perancangan Website Sebagai Media Informasi dan Promosi Batik Khas Kabupaten Kulonprogo. *EVOLUSI: Jurnal Sains dan Manajemen*, 3(2).
- Sahi, A. (2020). Aplikasi Test Potensi akademik seleksi saringan masuk LP3I berbasis web online menggunakan framework codeigniter. *Tematik*, 7(1), 120-129.
- Lay, M. E. (2017). E-COMMERCE GITAR AKUSTIK DAN SPAREPART KOTA MALANG MENGGUNAKAN METODE CUSTOMER TO CUSTOMER. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 1(2), 1-7.
- Abdullah, E. S. P. (2019). Self-control untuk mengurangi perilaku menonton film dewasa di internet. *Procedia: Studi Kasus dan Intervensi Psikologi*, 7(2).
- Anggraini, T., & Maulidya, E. N. (2020). Dampak paparan pornografi pada anak usia dini. *Al-Athfaal: Jurnal Ilmiah Pendidikan Anak Usia Dini*, 3(1), 45-55.