



SNESTIK

Seminar Nasional Teknik Elektro, Sistem Informasi,
dan Teknik Informatika

<https://ejurnal.itats.ac.id/snestik> dan <https://snestik.itats.ac.id>



Informasi Pelaksanaan :

SNESTIK IV - Surabaya, 27 April 2024

Ruang Seminar Gedung A, Kampus Institut Teknologi Adhi Tama Surabaya

Informasi Artikel:

DOI : 10.31284/p.snestik.2024.5882

Prosiding ISSN 2775-5126

Fakultas Teknik Elektro dan Teknologi Informasi-Institut Teknologi Adhi Tama Surabaya
Gedung A-ITATS, Jl. Arief Rachman Hakim 100 Surabaya 60117 Telp. (031) 5945043
Email : snestik@itats.ac.id

Deteksi Pelecehan Seksual dan Predator Obrolan Media Sosial Menggunakan Naive Bayes

Muhammad Akmal Fijar Riyadi, Erie Kresna Andana, Muhamad Amirul Haq

Universitas Muhammad Surabaya

e-mail: fijarriyadi@gmail.com

ABSTRACT

In recent years, advances in media technology have facilitated interaction between millions of people through social networking platforms and online gaming. However, behind the convenience offered by the internet, there are big challenges regarding the safety of children who use the internet. This research aims to develop an intelligent system that can identify predatory behavior and online sexual harassment. Leveraging text analysis, artificial intelligence, and Naive Bayes, this research introduces an original and potentially effective approach to detecting suspicious behavior. However, this study has certain limitations, including the limitations of current resources and technology in identifying suspicious online behavior. It is hoped that this research can make a significant contribution in improving children's safety when using the internet.

Keywords: Online sexual harassment; Naive Bayes; Identification of predatory behavior; Protection of children;

ABSTRAK

Dalam beberapa tahun terakhir, kemajuan teknologi media telah memfasilitasi interaksi antara jutaan orang melalui platform-platform jejaring sosial dan game online. Namun, di balik kemudahan yang ditawarkan oleh internet, terdapat tantangan besar terkait keselamatan anak-anak yang menggunakan internet. Penelitian ini bertujuan untuk mengembangkan sistem cerdas yang dapat mengidentifikasi perilaku predator dan pelecehan seksual online. Dengan memanfaatkan analisis teks, kecerdasan buatan, dan Naive Bayes, penelitian ini memperkenalkan pendekatan orisinal dan berpotensi efektif untuk mendeteksi perilaku yang mencurigakan. Meskipun demikian, penelitian ini memiliki batasan tertentu, termasuk keterbatasan sumber daya dan teknologi saat ini dalam mengidentifikasi perilaku online yang mencurigakan. Diharapkan

penelitian ini dapat memberikan kontribusi yang signifikan dalam meningkatkan keselamatan anak-anak dalam menggunakan internet.

Keywords: Pelecehan seksual online; Naive Bayes; Identifikasi perilaku predator; Perlindungan anak-anak;

PENDAHULUAN

Dalam beberapa tahun terakhir, kemajuan teknologi media telah memfasilitasi interaksi antara jutaan orang, baik secara pribadi maupun profesional, melalui platform-platform jejaring sosial dan game online. Fenomena ini telah menciptakan sebuah dunia di mana berbagi pendapat, pengalaman, foto, dan video telah menjadi bagian tak terpisahkan dari kehidupan sehari-hari. Namun, di balik kemudahan dan keuntungan yang ditawarkan oleh internet, ada tantangan besar yang harus diatasi, terutama dalam hal keselamatan anak-anak yang menggunakan internet. Pada Oktober 2020, jumlah pengguna internet aktif di seluruh dunia mencapai sekitar 4,66 miliar orang, menunjukkan betapa luasnya dampak dan jangkauan internet dalam kehidupan modern [1]. Namun, di sisi lain, studi dari EU Kids Online pada tahun 2020 menunjukkan bahwa anak-anak berisiko ketika menggunakan internet, dengan rata-rata anak usia 9 hingga 16 tahun menghabiskan dua jam online setiap hari, dan sebagian besar tanpa pengawasan orang dewasa [2].

Studi oleh Parlemen Eropa telah mengidentifikasi risiko utama yang terkait dengan penggunaan internet pada anak-anak, termasuk paparan terhadap konten berbahaya dan ilegal serta potensi eksploitasi seksual melalui media sosial [3]. Untuk mengatasi tantangan ini, diperlukan pendekatan yang melibatkan berbagai bidang ilmu seperti forensik standar, forensik digital, kecerdasan buatan, pemodelan komputer, dan analisis jaringan sosial [4]. Dalam konteks ini, penelitian ini bertujuan untuk mengembangkan sistem cerdas yang dapat membantu dalam mengidentifikasi perilaku predator dan pelecehan seksual online. Diharapkan bahwa sistem ini akan memberikan manfaat yang signifikan bagi orang tua dan otoritas setempat seperti petugas polisi, dengan menyediakan alat untuk mendeteksi dan menangani kasus-kasus tersebut secara lebih efektif.

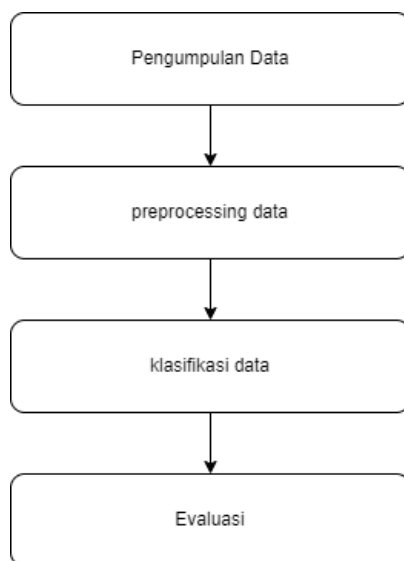
Penelitian ini juga memiliki kontribusi penting dalam mengembangkan metode cerdas untuk mendukung penyelidikan polisi dalam perlindungan anak-anak dan mempercepat proses menemukan bukti digital terkait kasus pelecehan seksual online. Dengan memanfaatkan analisis teks, kecerdasan buatan, dan wawasan ilmiah sosial seperti teori linguistik, kriminologi, dan psikologi, penelitian ini memperkenalkan pendekatan orisinal dan berpotensi efektif untuk mendeteksi perilaku yang mencurigakan [5]. Meskipun demikian, penelitian ini memiliki batasan tertentu, termasuk keterbatasan sumber daya, akses terhadap data yang tersedia, dan keterbatasan teknologi saat ini dalam mengidentifikasi dengan akurat perilaku online yang mencurigakan. Dengan memahami dan mengatasi batasan-batasan ini, diharapkan penelitian ini dapat memberikan kontribusi yang berarti dalam meningkatkan keselamatan anak-anak dalam menggunakan internet.

Penelitian dahulu Para penulis di Ref. [6] mengusulkan sistem yang melibatkan metode kategorisasi dua rantai untuk mendeteksi apakah percakapan menyertakan perilaku atau kosakata yang mencurigakan. Hal ini diperlukan untuk menyaring sebanyak mungkin peserta non-predator. Pada saat yang sama, pengguna predator sebenarnya dipisahkan dari pengguna non-predator dalam percakapan yang mencurigakan.

Penelitian Studi di Ref. [7] menarik perbandingan antara temuan algoritma yang diawasi ML untuk mengkategorikan pelecehan seksual online di Twitter. SVM Gaussian dan Polinomial, serta MultiLayer Perceptron, digunakan dengan vektor TF-IDF dan disematkan dengan Word2Vec.

METODE

Penelitian ini menggunakan dua jenis dataset terkait pelecehan seksual. Pertama, komentar yang termasuk dalam Kategori Pelecehan Seksual (KPS). Dataset kedua berupa percakapan dari Chat Sexual Predators (CSP). Kedua dataset ini terdiri dari teks pendek yang meliputi singkatan, emotikon, angka, simbol, pengulangan karakter, URL, dan teks dengan panjang yang bervariasi, yang semuanya dapat mempengaruhi kinerja klasifikasi. Ini sering disebut sebagai kebisingan[8]. Pra-pemrosesan data, membantu mengatasi kebisingan tersebut dan meningkatkan kebersihan data, sehingga meningkatkan kinerja model klasifikasi. Sebelum mewakili kata-kata dalam bentuk numerik, langkah-langkah pra-pemrosesan tertentu harus diterapkan. Pada tahap awal pra-pemrosesan, setiap karakter dalam dataset diubah menjadi huruf kecil. Informasi yang berkaitan dengan URL dihapus karena konten teks seksual mungkin tidak tersedia di dalamnya. Tanda baca, angka, karakter teks yang tidak dikenali, dan kata-kata dalam bahasa asing (seperti Arab dan Prancis) semuanya dihapus.



Gambar 1. Blok Diagram

Terkait dengan ukuran kalimat, kalimat dengan jumlah kata di bawah lima atau lebih dari 400 dikecualikan. Hal ini karena kalimat yang terlalu pendek cenderung tidak mengandung makna yang signifikan (seperti "Hai, apa kabar?"), sementara kalimat yang terlalu panjang dapat menjadi seperti artikel yang ditemukan yang tidak memberikan indikasi yang jelas terkait pelecehan seksual. Stemming digunakan untuk mengembalikan kata-kata ke bentuk dasarnya [9]. Setelah operasi pra-pemrosesan, jumlah kalimat yang terkait dengan pelecehan seksual dan yang tidak terkait dalam kedua dataset (KPS, CSP) dirangkum.

HASIL DAN PEMBAHASAN

1. Pengaturan Dataset.

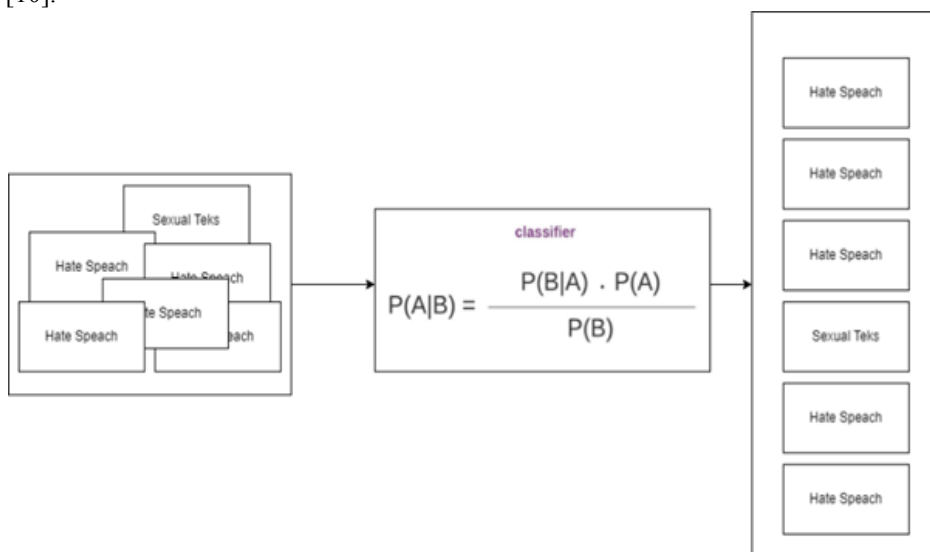
Penelitian ini memanfaatkan dataset dari media sosial dengan pengaturan yang dikategorikan berdasarkan Nama, Jenis, Kategori, dan Nilai.

2. Data Preprocessing.

Pemrosesan awal data adalah tahap penting dalam penelitian apa pun. Dengan teknik preprocessing yang tepat maka dapat diperoleh data yang bersih dan berkualitas untuk diproses lebih lanjut pada tahap selanjutnya.

3. Klasifikasi.

Tahap klasifikasi untuk mendeteksi kata-kata yang mengandung bahasa seksual dilakukan dengan menggunakan metode Naïve Bayes dipilih karena beberapa keunggulan, antara lain pemrosesan paralel, toleransi kesalahan yang tinggi, dan kemampuan nonlinier yang sangat baik [10].



Gambar 2. Desain Sistem

Tabel 1. Kategori dan Jenis

kategori	Nama	Jenis	Nilai
kategori	HS Ind	Feature	0.1
	HS Str		0.1
	HS Grp		0.1
	HS Rac		0.1
	HS Phy		0.1
	HS Gen		0.1
	HS Wek		0.1
	SEXUAL	Target	0.1
	SEXUAL		0.1

4. Evaluasi

Evaluasi dilakukan dengan mengukur kinerja menggunakan metrik Akurasi dan Presisi.

KESIMPULAN

Penelitian tentang Kategori Pelecehan Seksual (KPS) dan Chat Sexual Predators (CSP) di media sosial menjadi topik penelitian yang menarik karena jumlah pengguna media sosial

yang terus meningkat dari tahun ke tahun. Penggunaan metode yang tepat dalam mendeteksi (KPS) dan (CSP) dapat sangat membantu interaksi di media sosial dalam masyarakat. Dalam penelitian ini, pengolahan data berbasis teks pada Media Sosial menggunakan metode preprocessing yang tepat serta menerapkan Naïve Bayes sebagai salah satu metode unggulan dalam Machine Learning. Untuk penelitian ke depan akan mengambil data teks dari media sosial lainnya.

UCAPAN TERIMA KASIH

Penelitian ini didukung oleh Prodi Informatika Universitas Muhammadiyah Surabaya.

DAFTAR PUSTAKA

- [1] I. A. Putri, M. A. Komalasari, and . S., “Pengaruh Intensitas Penggunaan Gadget Terhadap Kondisi Sosial Ekonomi Masyarakat Sukamulia Timur Saat Pandemi Covid-19,” *Jurnal Dinamika*, vol. 3, no. 1, 2022, doi: 10.18326/dinamika.v3i1.1-22.
- [2] D. Smahel *et al.*, “EU Kids Online 2020: Survey results from 19 countries.,” *EU Kids Online*, no. February, 2020.
- [3] E. Quayle and N. Koukopoulos, “Deterrence of Online Child Sexual Abuse and Exploitation,” *Policing (Oxford)*, vol. 13, no. 3. 2019. doi: 10.1093/police/pay028.
- [4] A. Irons and H. Lallie, “Digital Forensics to Intelligent Forensics,” *Future Internet*, vol. 6, no. 3, 2014, doi: 10.3390/fi6030584.
- [5] M. Koppel, J. Schler, and S. Argamon, “Authorship attribution in the wild,” *Lang Resour Eval*, vol. 45, no. 1, 2011, doi: 10.1007/s10579-009-9111-2.
- [6] C. Cardei and T. Rebedea, “Detecting sexual predators in chats using behavioral features and imbalanced learning,” *Nat Lang Eng*, vol. 23, no. 4, 2017, doi: 10.1017/S1351324916000395.
- [7] M. Saeidi, S. B. Samuel, E. Milios, N. Zeh, and L. Berton, “Categorizing Online Harassment on Twitter,” in *Communications in Computer and Information Science*, 2020. doi: 10.1007/978-3-030-43887-6_22.
- [8] A. Sarker, “A customizable pipeline for social media text normalization,” *Soc Netw Anal Min*, vol. 7, no. 1, 2017, doi: 10.1007/s13278-017-0464-z.
- [9] A. Jabbar, S. Iqbal, M. I. Tamimy, S. Hussain, and A. Akhunzada, “Empirical evaluation and study of text stemming algorithms,” *Artif Intell Rev*, vol. 53, no. 8, 2020, doi: 10.1007/s10462-020-09828-3.
- [10] B. Santoso, A. Azis, and Zohrahayaty, *Machine Learning & Reasoning Fuzzy logic Algoritma, Manual, Matlab & Rapid Miner*, vol. 1, no. 1. 2020.