

Analisis Prostitusi Online Pada Aplikasi Michat Menggunakan Algoritma Naïve Bayes Dan Framework NIST

¹Readel L. B. Mangkey, ²Gladly C. Rorimpandey, ³Sondy C. Kumajas

^{1,2,3} Teknik Informatika Universitas Negeri Manado, Indonesia

¹readelmangkey@gmail.com, ²gladlycorimpandey@unima.ac.id, ³sondykumajas@unima.ac.id

Submit : 22 Okt 25 | Diterima : 28 Okt 2025 | Terbit : 30 Okt 2025

ABSTRAK

Perkembangan teknologi informasi yang pesat menghadirkan berbagai kemudahan dalam berkomunikasi. Namun, hal ini juga menciptakan potensi meningkatnya kejahatan siber, termasuk prostitusi online. Aplikasi MiChat sering dikaitkan dengan kegiatan seperti itu karena fitur *People Nearby*, yang memungkinkan pengguna dalam menemukan pengguna lain di lokasi terdekat. Tujuan dari dilakukannya penelitian ini adalah menganalisis pola percakapan yang mengarah pada praktik prostitusi online, dengan menerapkan algoritma *Naïve Bayes Classifier* dan menerapkan *National Institute of Standards and Technology* (NIST) sebagai pedoman. Autopsy dan Mobileedit Forensic Express digunakan untuk mengumpulkan dan mengekstraksi data dari skenario percakapan yang dibuat semirip mungkin dengan percakapan nyata pengguna MiChat. Preprocessing terbagi atas pembersihan data, case folding, stemming, tokenizing, dan stopwords sebelum perhitungan nilai memakai TF-IDF. Model *Naïve Bayes* dipakai dalam mengklasifikasikan percakapan ke dalam dua kelompok yaitu *positive* (tidak menunjukkan prostitusi) dan *negative* (menunjukkan prostitusi). Analisis data menghasilkan tingkat akurasi model sebesar 88%. Hasil ini membuktikan bahwa algoritma Naïve Bayes bisa dipakai secara efektif untuk menemukan percakapan yang mengindikasikan prostitusi online.

Kata Kunci: Naïve Bayes; NIST; Forensik Digital; MiChat; Prostitusi Online

PENDAHULUAN

Kemajuan teknologi sudah mengubah gaya manusia dalam berinteraksi dan bertransaksi. Internet yang awalnya berfungsi sebagai media untuk berbagi informasi, kini telah berubah menjadi sebuah tempat yang sangat aktif. Smartphone memungkinkan masyarakat untuk dapat mengakses berbagai layanan dengan mudahnya. Aplikasi pesan seperti Whatsapp, Telegram dan Michat menjadi komponen penting dari kehidupan masyarakat.

Namun dibalik kemudahan yang diberikan ada juga dampak negatif yang dapat muncul salah satunya kejahatan siber. Menurut Badan Siber dan Sandi Negara (BSSN) pada tahun 2025, terdapat lebih dari 3,64 miliar serangan siber hanya dalam periode Januari-Juli 2025 (Defara Dhanya 2025). Data ini menunjukkan meningkatnya penyalahgunaan teknologi yang termasuk didalamnya aktivitas seperti peretasan sistem, penipuan daring, penyebaran malware dan prostitusi online.

Prostitusi online adalah bentuk kejahatann yang menggunakan teknologi informasi alat utama untuk bertransaksi. Pelaku berkomunikasi dengan calon pengguna jasa dengan melalui aplikasi percakapan. Salah satu aplikasi di Indonesia yang sering dikaitkan dengan aktivitas ini adalah Michat dengan adanya fitur *People Nearby* yang memungkinkan pengguna menemukan pengguna lain berdasarkan lokasi terdekat yang dimanfaatkan untuk melakukan promosi atau mengatur layanan seksual secara terselubung.

Dengan praktik prostitusi online yang tidak lagi dilakukan dengan terbuka menyebabkan proses identifikasi terhadap kasus ini menjadi semakin susah bila dikerjakan secara manual. Oleh sebab itu diperlukan pendekatan dengan menggunakan kecerdasan buatan (*Artificial Intelligence*) menjadi suatu solusi untuk mengungkap kasus prostitusi online.

Untuk memastikan bahwa bukti digital dapat diproses secara sah dan dapat dipertanggungjawabkan framework NIST digunakan sebagai panduan dan pedoman dan juga metode Naïve bayes digunakan sebab kinerjanya yang sederhana tetapi efektif untuk mengklasifikasikan teks. Penelitian ini diharapkan dapat menghasilkan sistem analisis yang efektif untuk mendeteksi komunikasi yang mengindikasikan prostitusi online.

TINJAUAN PUSTAKA

Analisis Forensik Digital

Analisis adalah proses berpikir dimana seseorang menguraikan suatu keseluruhan menjadi bagian-bagian yang berbeda sehingga dapat mengidentifikasi hubungan dan karakteristik masing-masing bagian. (Nugroho, 2016). Digital forensik adalah disiplin ilmu yang memiliki tujuan untuk mengumpulkan bukti dan menyelidikinya sehingga dapat digunakan sebagai bukti yang sah di pengadilan. (Eunike R. Runtuwene, 2024) Oleh karena itu dapat disimpulkan bahwa analisis forensik digital adalah upaya untuk mendapatkan pemahaman yang mendalam tentang kejahatan untuk membuktikannya dengan menguraikan dan memahami data secara menyeluruh.

National Institute of Standards and Technology (NIST)

Pendekatan NIST ini umumnya digunakan pada bidang forensik digital. Metode ini digunakan untuk menjelaskan secara rinci dan tersusun bagaimana menyelesaikan masalah yang sedang diteliti. (MISHARDILA, 2020) Metode ini digunakan untuk membuat hasil yang diperoleh dapat digunakan sebagai bukti hukum. proses investigasi digital NIST terbagi menjadi empat tahapan utama:

1. *Collection*, yaitu pengumpulan data dan bukti digital dari perangkat atau media yang terlibat;a
2. *Examination*, yaitu pemeriksaan terhadap data yang diperoleh untuk memastikan integritasnya;
3. *Analysis*, yaitu proses penafsiran dan penarikan kesimpulan berdasarkan bukti digital;
4. *Reporting*, yaitu penyusunan laporan hasil analisis yang dapat dipertanggungjawabkan secara hukum (Pangghah Widiandana, 2019)

Analisis Sentimen

Analisis sentimen, juga disebut sebagai penggalian opini, adalah proses menganalisis sejumlah besar teks untuk menentukan apakah mereka mengandung sentimen positif, negatif, atau netral. Tugas utama analisis sentimen adalah mengelompokkan polaritas teks dalam dokumen, kalimat, atau opini menjadi sisi positif atau negatif. Penelitian ini bertujuan untuk menganalisis percakapan yang mengindikasikan prostitusi online dengan menggunakan metode klasifikasi yang menggunakan algoritma Naive Bayes. (Kevin, 2024). Dengan mempertimbangkan aspek dari sudut pandang yang beragam, analisis sentimen digunakan untuk mengkaji berbagai hal melalui perspektif yang berbeda guna memahami pendapat individu lain berdasarkan hasil analisis data. (Nikensia Mawikere & RHT Tangkawarow, 2024)

Naïve Bayes Classifier

Klasifikasi data adalah salah satu tugas data mining; ini berarti mengkategorikan data ke dalam kelas yang sudah diketahui sebelumnya. Naïve bayes merupakan salah satu pendekatan dalam klasifikasi, yang diciptakan oleh ilmuwan Inggris Thomas Bayes dan menggunakan perhitungan probabilitas dan statistik. Metode ini menggunakan pengalaman masa lalu untuk memprediksi kemungkinan di masa depan. (Dedi Darwis, 2021) Keunggulan Naïve Bayes kesederhanaanya, namun tetap mampu menghasilkan tingkat akurasi yang tinggi serta memerlukan jumlah data pelatihan yang relative sedikit. (ABDUL GAFUR DARUSSALAM, 2021)

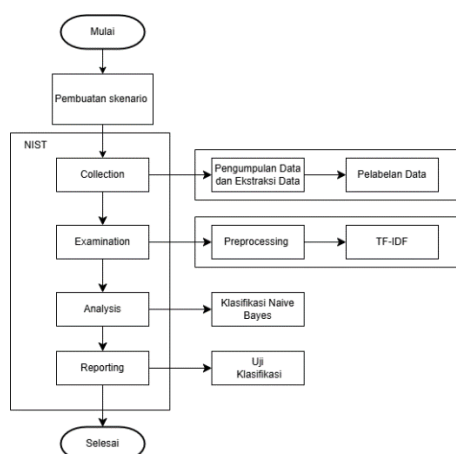
Michat

Michat merupakan salah satu aplikasi yang dikenal oleh masyarakat sebagai aplikasi media sosial, yang merupakan jenis aplikasi chat yang memungkinkan pengguna terhubung dengan orang-orang di sekitar mereka dalam radius tertentu. Dengan fitur yang ada, seperti panggilan telepon dan

massager, Anda bisa juga berbagi video dan foto dengan sesama pengguna. Namun saat ini, aplikasi ini sering digunakan untuk layanan prostitusi online, bahkan dengan anak di bawah umur. Pada awalnya, aplikasi ini tidak dimaksudkan untuk tujuan yang melanggar hukum, tetapi karena sering disalahgunakan, ia telah menjadi perhatian publik. (ANNISA, 2022)

METODE PENELITIAN

Metode penelitian menggunakan alur kerja NIST yang berfungsi menjadi tahapan-tahapan untuk melakukan proses penelitian.



Gambar 1 Alur penelitian

Pembuatan skenario

Peneliti membuat skenario percakapan di aplikasi MiChat antara dua orang: pelanggan dan penyedia jasa. Dengan menggunakan istilah yang biasa digunakan dalam prostitusi online, teks percakapan mencerminkan situasi nyata.

1. Collection

1) Ekstraksi data

Untuk menyalin database aplikasi MiChat, Mobileedit Forensic Express 7.3.0 digunakan untuk mengumpulkan data dari perangkat Android. Untuk memudahkan proses analisis, hasil pengumpulan diekstraksi menjadi format CSV melalui aplikasi Autopsy.

2) Pelabelan data

Setiap data diberi label yang didasarkan pada kamus kata yang dibuat oleh peneliti. Masing-masing sebanyak 904 data percakapan diperoleh, terdiri dari 450 positif dan 454 negatif. Jika mengandung istilah yang berkaitan dengan prostitusi yang ada pada kamus kata diberi label negative sebaliknya jika tidak diberi label positive.

2. Examination

1) preprocessing

Data teks diproses melalui tahapan *cleaning*, *case folding*, *tokenizing*, *stopword removal*, dan *stemming* dengan memanfaatkan pustaka *Sastrawi*. Proses ini bertujuan untuk menghilangkan elemen yang tidak penting untuk menyediakan data yang siap untuk dianalisis, seperti angka, tanda baca, dan kata umum.

2) Ekstraksi fitur TF-IDF

Setelah *preprocessing*, metode TF-IDF digunakan sebagai metode dalam pembobotan kata.

3. Analysis

Data yang selesai dibobot selanjutnya diklasifikasikan dengan memakai algoritma *Naïve Bayes* dengan bantuan pustaka *scikit-learn* di Google Colaboratory.

4. Reporting

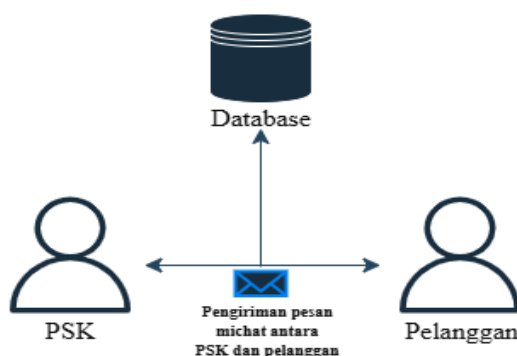
1) Evaluasi model dan Uji klasifikasi

Model diuji nilai accuracy, recall, precision dan F1-score dengan memakai meode *confusion matrix*. Uji klasifikasi dilakukan dengan menggunakan sistem yang akan menampilkan file ke dalam kategori negative atau positive

HASIL DAN PEMBAHASAN

Pembuatan skenario

Penelitian ini menganalisis percakapan terkait prostitusi online yang dilakukan melalui aplikasi Michat. Dalam pesan ini, pekerja seks komersial (PSK) menawarkan layanan prostitusi kepada calon pelanggan. Selanjutnya, PSK dan pelanggan merencanakan untuk mengatur jadwal pertemuan di lokasi yang telah ditentukan.

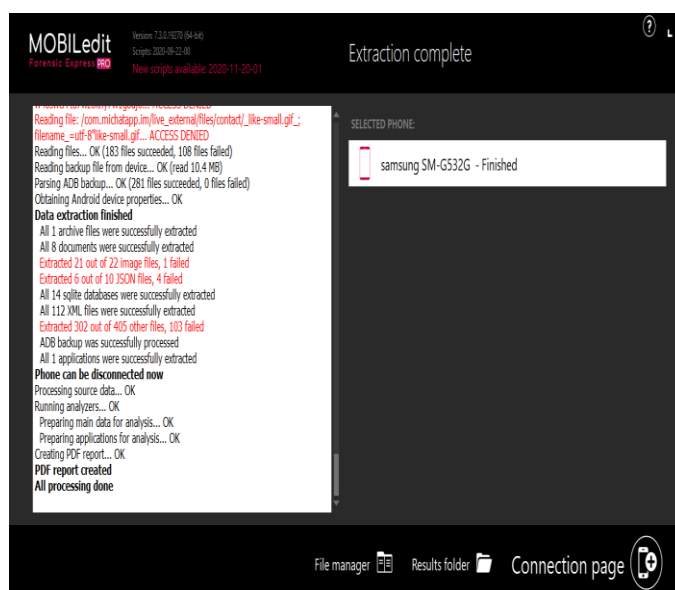


Gambar 2 Alur skenario kasus

Dalam penelitian ini dibuat delapan percakapan yang akan digunakan sebagai data pembanding untuk model naïve bayes. Kedelapan data percakapan tersebut dibagi menjadi empat percakapan yang terindikasi prostitusi dan empat percakapan yang tidak terindikasi prostitusi.

Collection dan ekstraksi data

Pengumpulan data pada Smartphone Samsung J2 Prime dengan versi Android 6.0.1 menggunakan tool MOBILEdit Forensi Express versi 7.3.0 dilakukan proses rooting, yang membuka smartphone secara keseluruhan.



Gambar 3 Proses collection dengan Mobiledit Forensic

Setelah data berhasil dikumpulkan, proses ekstraksi selanjutnya dilakukan pada table tb_message, yang berisi teks percakapan pada aplikasi Michat .

Pelabelan data

Sebelum pelabelan data dilakukan, data dilakukan proses reduction untuk mengurangi jumlah data yang tidak dibutuhkan dalam penelitian ini. atribut yang diperlukan adalah kolom message yang berisi percakapan.

1	PROSTITUSI
2	foto
3	hotel
4	main
5	pijat
6	servis
7	durasi
8	tarif
9	nego
10	harga
11	bo
12	kamar
13	kondom
14	tempat

Gambar 4 kamus kata

Labeling dilakukan dengan menggunakan kamus kata yang mengandung 13 kata yang menunjukkan indikasi prostitusi online. Data yang mengandung kata-kata ini akan dilabeli negatif jika terdapat dalam kamus kata tersebut. Se jauh ini, ada 450 label positif dan 454 label negatif yang dihasilkan dari proses labeling.

Tabel 1 Hasil Labeling data

message	label
film tadi seru banget ya, gak nyangka ending-nya gitu	Positive
iya parah sih, aku sampe merinding	Positive
servis lengkap ya? jangan setengah-setengah	Negative
udah pernah bo di sini belum? aman gak?	Negative

Examination dan preprocessing

1. Cleaning

Pembersihan data mencakup penghapusan angka, tanda baca, dan ruang kosong. Karakter ini dihapus karena tidak ada gunanya. dalam proses pemeriksaan.

Tabel 2 Cleaning data

No	Sebelum Cleaning	Sesudah Cleaning
1	Bagaimana kalau nanti sekitar jam 7 saja?	Bagaimana kalau nanti sekitar jam saja

No	Sebelum Cleaning	Sesudah Cleaning
2	Bisa main jam 10 malam ini di hotel ABC?	Bisa main jam malam ini di hotel ABC

2. Case Folding

Dalam tahap ini, agar tidak ada perbedaan antara teks digunakan metode case folding untuk mengganti semua huruf besar ke huruf kecil.

Tabel 3 Casefolding data

No	Sebelum Case Folding	Sesudah Case Folding
1	Bagaimana kalau nanti sekitar jam saja	bagaimana kalau nanti sekitar jam saja
2	Bisa main jam malam ini di hotel ABC	bisa main jam malam ini di hotel abc

3. Tokenizing

Tokenisasi adalah prosedur yang digunakan untuk memecah kata-kata dalam dokumen menjadi bagian yang tidak saling berpengaruh yang dapat diidentifikasi.

Tabel 4 Tokenizing data

No	Sebelum Tokenizing	Sesudah Tokenizing
1	Bagaimana kalau nanti sekitar jam saja	['bagaimana', 'kalau', 'nanti', 'sekitar', 'jam', 'saja']
2	bisa main jam malam ini di hotel abc	['bisa', 'main', 'jam', 'malam', 'ini', 'di', 'hotel', 'abc']

4. Stopwords

Tahapan untuk menghilangkan kata yang tidak berarti dan tidak mempengaruhi proses analisis yang dikenal sebagai stopwords removal.

Tabel 5 Stopword removal

No	Sebelum Stopwords	Sesudah Stopwords
1	['pelayanannya', 'ramah', 'dan', 'menyenangkan']	['pelayanannya', 'ramah', 'menyenangkan']
2	['servis', 'sangat', 'mengecewakan', 'tidak', 'ramah', 'sama', 'sekali']	['Servis', 'mengecewakan', 'ramah']

5. Stemming

Proses ini mengganti kata ke dalam bentuk dasarnya, dari yang sebelumnya memiliki imbuhan ke bentuk dasarnya.

Tabel 6 Stemming data

No	Sebelum Stemming	Sesudah Stemming
1	['pelayanannya', 'ramah', 'menyenangkan']	['layanannya', 'ramah', 'senang']
2	['servis', 'mengecewakan', 'ramah']	['servis', 'kecewa', 'ramah']

Pembobotan data

Sebelum pembobotan, data dibagi menjadi 20% data uji dan 80% data latihan. Metode TF-IDF menghitung seberapa penting sebuah kata dalam dokumen. Hasil dari pembobotan ini kemudian dimasukkan (dimasukkan) ke dalam proses analisis menggunakan algoritma Naive Bayes.

Tabel 7 pembobotan kata

	kata	Tf-idf
Menit pijat menit main	Menit	0.1505
	Pijat	0.07525
	Main	0.7525

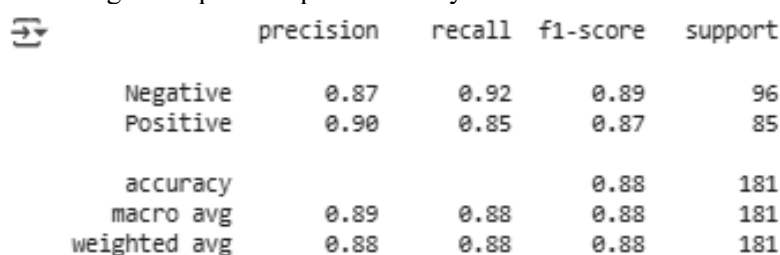
Analisis dan evaluasi model

Setelah data selesai dilakukan pembobotan, pemodelan dilakukan dengan library MultinomialNB dari Scikit-learn pada data training. Kemudian, berdasarkan data testing, teknik confusion matrix digunakan untuk menghitung akurasi klasifikasi.

Tabel 8 hasil confusion matrix

Label Aktual	Label Prediksi	
	Positive	Negative
Positive	88	8
Negative	13	72

Hasil dari matrix confusin di tabel 8 menunjukkan bahwa data positif terprediksi positif sebanyak 88, sedangkan data negatif terprediksi negatif sebanyak 72, data positif yang terprediksi negatif sebanyak 8, dan data negatif terprediksi positif sebanyak 13.



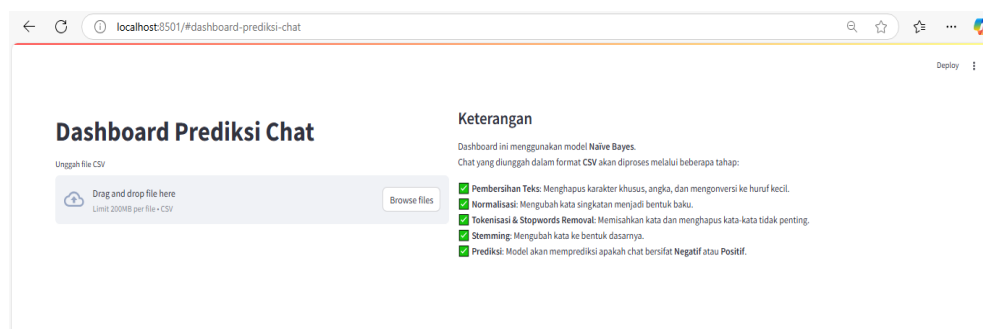
	precision	recall	f1-score	support
Negative	0.87	0.92	0.89	96
Positive	0.90	0.85	0.87	85
accuracy			0.88	181
macro avg	0.89	0.88	0.88	181
weighted avg	0.88	0.88	0.88	181

Gambar 5 hasil perhitungan akurasi

Gambar 5 menunjukkan hasil perhitungan confusion matrix. Model algoritma yang dikembangkan mendapat akurasi sebesar 88 persen, yang menunjukkan kinerja pengklasifikasian yang baik.

Pengujian klasifikasi

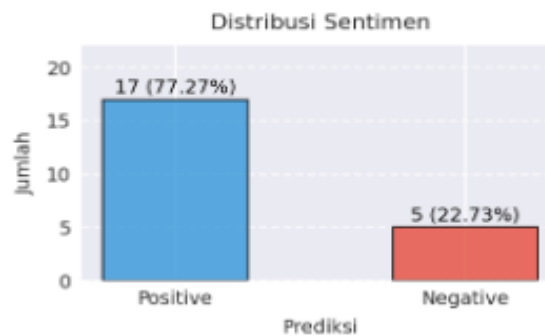
Model yang telah dibuat kemudian diuji dengan dashboard yang dibuat dengan library Python Streamlit. Library ini membuat antarmuka pengguna yang cepat dan mudah digunakan dan menampilkan hasil yang lebih mudah dipahami.



Gambar 6 Dashboard

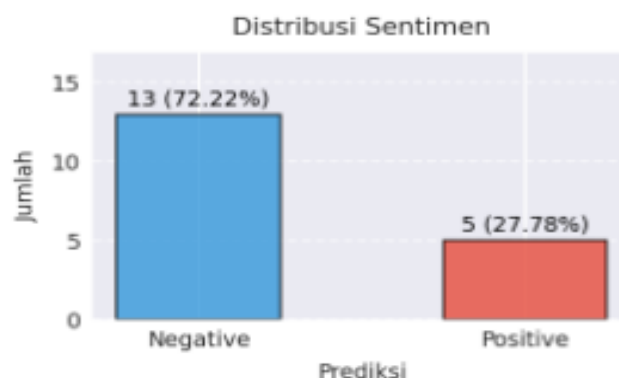
Jumlah delapan dataset digunakan untuk pengujian ini, dengan empat dataset yang menunjukkan prostitusi dan empat dataset lainnya yang tidak menunjukkan prostitusi. Hasil pengujian menunjukkan bahwa model klasifikasi naive bayes berhasil mengklasifikasikan delapan dataset dengan benar, termasuk jenis dokumen yang memiliki pola komunikasi prostitusi online dan yang bukan prostitusi online.

Hasil Klasifikasi: ⇄



Gambar 7 hasil klasifikasi data tidak prostitusi

Hasil Klasifikasi: ⇄



Gambar 8 hasil klasifikasi data prostitusi

Pembahasan

Hasil penelitian membuktikan yaitu model algoritma Naive Bayes bisa dengan tepat mengklasifikasikan percakapan yang menunjukkan prostitusi online pada aplikasi MiChat, dengan akurasi 88%. Berdasarkan pengujian terhadap delapan skenario dataset, model berhasil mengidentifikasi dengan tepat percakapan yang menunjukkan atau tidak prostitusi. Meskipun masih ada beberapa kesalahan klasifikasi, evaluasi dengan matriks kecacauan menunjukkan tingkat recall untuk percakapan prostitusi 0,92 dan percakapan normal 0,85.

KESIMPULAN

Menurut penelitian yang telah dilakukan, algoritma naive bayes dapat dipakai untuk melihat pola komunikasi yang terkait dengan prostitusi online. Pelatihan model algoritma ini, ekstraksi fitur menggunakan TF-IDF, dan proses preprocessing data memungkinkan untuk mengidentifikasi pola komunikasi yang banyak digunakan dalam komunikasi yang menunjukkan prostitusi online. Selanjutnya, model Naive Bayes yang dikembangkan membuktikan kinerja yang baik dengan tingkat accuracy 88%; dari delapan dataset yang digunakan dalam pengujian, model berhasil mengidentifikasi delapan dataset secara tepat, menunjukkan bahwa algoritma ini cukup baik untuk

mendeteksi pola komunikasi yang terkait dengan prostitusi online. Secara keseluruhan, penggabungan framework NIST dan algoritma Naïve Bayes terbukti cukup efisien untuk digunakan dalam investigasi digital, terutama untuk menemukan pola komunikasi praktik prostitusi online. Namun demikian, penelitian ini memiliki keterbatasan pada jumlah dataset yang digunakan serta analisis yang hanya menggunakan satu algoritma kalsifikasi saja, untuk penelitian selanjutnya disarankan agar melakukan eksplorasi pada algoritma lain seperti Support Vector Machine (SVM), Decision Tree, atau Random Forest untuk memperoleh perbandingan kinerja yang lebih luas.

REFERENSI

- ANNISA, D. (2022). PROSES KOMUNIKASI YANG TERJADI PADA APLIKASI MICHAAT DALAM TINDAK KEGIATAN SEKSUAL PROSTITUSI ONLINE. *Prosiding Konferensi Nasional Sosial Politik (KONASPOL)*, 300-310.
- DARUSSALAM, A. G. (2021). *ANALISIS SENTIMEN MENGGUNAKAN TEXT MINING DENGAN METODE NAÏVE BAYES DAN REGRESI LOGISTIK*. MAKASSAR: UNIVERSITAS HASANUDDIN .
- Dedi Darwis, N. S. (2021). Penerapan Algoritma Naive Bayes untuk Analisis Sentimen Review Data Twitter BMKG Nasional. *Jurnal TEKNO KOMPAK*, 131-145.
- Dhanya. (den 08 08 2025). *BSSN Catat 3,64 Miliar Serangan Siber di Indonesia Setengah Tahun Ini*. Hämtat från Tempo: <https://www.tempo.co/digital/bssn-catat-3-64-miliar-serangan-siber-di-indonesia-setengah-tahun-ini-2056396>
- Eunike R. Runtuwene, S. C. (2024). FACE IDENTIFICATION USING IMAGE PROCESING WITH THE NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY (NIST) METHOD . *Jurnal Teknik Informatika (JUTIF)* , 375-384 .
- Kevin, M. E. (2024). Analisis Sentimen Penggunaan Aplikasi Metode Naive Bayes Kinemaster Menggunakan Metode Naive Bayes. *JURNAL ILMIAH COMPUTER SCIENCE (JICS)*, 88-98.
- MISHARDILA, G. (2020). *ANALISA DAN PENCARIAN BUKTI FORENSIKA DIGITAL PADA APLIKASI MEDIA SOSIAL FACEBOOK DAN TWITTER MENGGUNAKAN METODE STATIK FORENSIK*. PEKANBARU: UNIVERSITAS ISLAM RIAU.
- Nugroho, K. (2016). MODEL ANALISIS PREDIKSI MENGGUNAKAN METODE FUZZY TIME SERIES. *INFOKAM*, 46-50.
- Panggah Widiandana, I. R. (2019). Implementasi Metode Jaccard pada Analisis Investigasi Cyberbullying WhatsApp Messenger Menggunakan Kerangka Kerja National Institute of Standards and Technology. *RESTI (Rekayasa Sistem dan Teknologi Informasi)*, 1046 – 1051.
- Sarah Nikensia Mawikere, I. R. (2024). Analisis Sentimen Pada Berita Online Terhadap Coklit Pemilu 2024 Dengan Metode Naïve Bayes. *JOURNAL OF INFORMATICS, BUSSINES, EDUCATION, AND INNOVATION TECHNOLOGY*, 59-67.