JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 4 No 2 September 2025
Hal: 836-844
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v4i2.521

Predictive Classification Model dalam Tahapan Framework NIJ untuk Otomatisasi Investigasi Digital Forensik (Studi Kasus: Cyberbullying)

Khana Yusdiana[1], Rizky Rahman J.P [2], Eddy Prasetyo Nugroho[3]

[1,2] Fakultas Pendidikan Matematika dan Ilmu Pengetahuan Alam, Ilmu Komputer, Universitas Pendidikan Indonesia, Bandung, Indonesia
Email: [1]khanayusdiana@gmail.com, [2]rizkyrjp@upi.edu, [3]eddypn@upi.edu
(* Email Corresponding Author: khanayusdiana@gmail.com)
Received: July 30, 2025 | Revision: August 4, 2025 | Accepted: August 6, 2025

Abstrak

Penelitian ini menerapkan kerangka kerja *National Institute of Justice* dalam proses penyelidikan kasus *cyberbullying* forensik digital pada bukti digital berupa percakapan pada aplikasi LINE dan Telegram, serta mengeksplorasi pemanfaatan *Predictive Classification Model* untuk klasifikasi otomatis komentar berbasis teks dalam kasus *cyberbullying*. *Cyberbullying* merupakan salah satu bentuk kejahatan digital yang semakin meningkat, khususnya pada platform pesan instan yang bersifat privat dan sulit dipantau. Penelitian ini menggunakan dua algoritma machine learning dalam Predictive Classification Model, yaitu *Complement* Naive Bayes dan Random Forest, untuk mendeteksi komentar dengan potensi perundungan. Proses forensik mencakup tahapan *Preperation, Evidence Assessment, Evidence Acquisition, Evidence Examination, Documenting and Reporting*, dengan pendekatan ekstraksi data yang aman dan *forensically sound* dari kedua aplikasi, Pemanfaatan *Predictive Classification Model* dengan model *Complement Naive Bayes* dan *Random Forest* diletakkan pada tahapan *evidence examination*. Hasil evaluasi menunjukkan bahwa model *Complement Naive Bayes* memiliki performa lebih unggul dengan akurasi sebesar 86% dan F1-score yang seimbang, sementara *Random Forest* mencatatkan akurasi sebesar 75%. Temuan ini mendukung penggunaan PCM sebagai solusi pendukung dalam identifikasi otomatis konten berisiko tinggi di media sosial. Integrasi digital forensik dengan kecerdasan buatan berpotensi meningkatkan efektivitas investigasi kasus cyberbullying secara signifikan.

Kata kunci: Cyberbullying, Predictive Classification Model, Complement Naive Bayes, Random Forest , Line, Telegram, Digital Forensik, National Institute of Justice

Abstract

This study aims to apply the National Institute of Justice framework in the digital forensic process for conversations retrieved from LINE and Telegram applications, as well as to explore the utilization of a Predictive Classification Model for automated text-based comment classification in cyberbullying cases. Cyberbullying is a growing form of digital crime, particularly on private and encrypted instant messaging platforms that are difficult to monitor. The research employs two machine learning algorithms within the PCM framework Complement Naive Bayes and Random Forest to detect potentially abusive comments. The forensic process follows several stages: Preparation, Evidence Assessment, Evidence Acquisition, Evidence Examination, and Documenting and Reporting, with a secure and forensically sound data extraction approach from both applications. Due to data limitations from LINE and Telegram, the classification analysis is conducted using an Instagram comment dataset that reflects the cyberbullying context. Evaluation results show that the Complement Naive Bayes model outperforms Random Forest, achieving an accuracy of 86% with balanced F1-scores, while Random Forest achieves 75% accuracy. These findings support the use of PCM as an effective aid for automatically identifying high-risk content on social media. The integration of digital forensics and artificial intelligence has significant potential to enhance the effectiveness of cyberbullying investigations.

Keywords: *Cyberbullying, Predictive Classification Model, Complement Naive Bayes, Random Forest, LINE, Telegram, Digital Forensics, National Institute of Justice*

1. PENDAHULUAN

Pesatnya perkembangan teknologi telah mendorong perubahan besar dalam cara berkomunikasi masyarakat, terutama melalui aplikasi pesan instan seperti Telegram dan Line. Aplikasi ini tidak hanya memfasilitasi komunikasi pribadi dan kelompok, tetapi juga menjadi media potensial bagi penyalahgunaan teknologi dalam aktivitas kejahatan siber, seperti penipuan, pemerasan digital, dan penyebaran konten ilegal[1]. Karakteristik *end-to-end encryption* yang pada kedua platform ini untuk menjaga privasi, justru menjadi tantangan serius dalam proses penegakan hukum karena menyulitkan pelacakan data oleh investigator forensik. Seiring meningkatnya serangan siber dan kejahatan digital, kebutuhan akan forensik yang efisien, akurat, dan dapat bekerja secara otomatis semakin mendesak. sekitar 15% pengguna media sosial pernah mengalami kejahatan digital, termasuk penyalahgunaan data, cyberbullying, hingga pemalsuan identitas. Oleh karena itu, investigasi digital terhadap aplikasi pesan instan tidak hanya memerlukan kemampuan ekstraksi data, tetapi juga kemampuan klasifikasi bukti digital yang relevan secara otomatis dan dapat dipertanggungjawabkan secara hukum[2]. Pesatnya pertumbuhan penggunaan aplikasi pesan instan seperti Line, Messenger, WhatsApp, dan Telegram telah menimbulkan tantangan baru dalam dunia digital forensik. Dengan lebih dari 6 miliar pengguna ponsel pintar di seluruh dunia dan prediksi pengguna aplikasi pesan instan yang mencapai 3,51 miliar pada tahun 2025, volume data yang dihasilkan semakin meningkat. informasi percakapan dari aplikasi ini berpotensi menjadi bukti penting dalam penyelidikan kriminal, kasus hukum, hingga litigasi digital [3]. Pendekatan manual dalam pencarian bukti digital terbukti sangat tidak efisien untuk volume data yang besar.

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 4 No 2 September 2025
Hal: 836-844
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v4i2.521

Untuk itu, integrasi antara teknik *machine learning* dan prosedur forensik menjadi solusi yang menjanjikan. Digital forensik saat ini mulai memanfaatkan *predictive modeling* untuk membantu investigator dalam mendeteksi pola dan anomali dari data digital[4].

Seiring berkembangnya teknologi, bukti digital menjadi elemen penting dalam investigasi forensik. Aplikasi pesan instan seperti Line dan Telegram menggunakan enkripsi end-to-end untuk melindungi privasi pengguna, yang pada gilirannya menyulitkan proses pengumpulan bukti. Enkripsi Telegram hanya berlaku untuk *SecretChat*, sementara Line menerapkannya secara menyeluruh[5]. Tantangan ini diperparah oleh sifat bukti digital yang mudah berubah, sehingga perlu ditangani dengan metode forensik yang aman dan sah[6]. Meski demikian, Line dan Telegram memiliki arsitektur penyimpanan data yang mendukung proses forensik. Telegram menyimpan data di cloud (kecuali *SecretChat*), sedangkan Line menyimpan histori pesan dalam format .txt di direktori lokal. Kedua aplikasi ini juga relatif terbuka terhadap ekspor data Telegram menyediakan fitur ChatExport, dan Line dapat diakses via ADB tanpa perlu *rooting*, sehingga mempermudah ekstraksi data secara *forensically sound* [7]. Oleh karena itu dalam proses mengumpulkan, penyimpan dan pengujian bukti digital harus menggunakan standar *National Institute of Justice (NIJ)* agar sah di pengadilan. *National Institute of Justice (NIJ)* adalah sebuah Lembaga yang menetapkan standar untuk forensik digital [8]. Untuk mendukung klasifikasi otomatis dalam *Predictive Classification Model (PCM)*. Penelitian ini menggunakan dua algoritma machine learning, yaitu *Complement* Naive Bayes dan Random Forest, yang masing-masing dipilih berdasarkan keunggulannya dalam konteks klasifikasi teks untuk investigasi forensik digital. *Complement* Naive Bayes dipilih karena kemampuannya dalam menangani data teks yang tidak seimbang serta efisiensinya dalam mengidentifikasi kolom dan konten percakapan pada aplikasi pesan instan. Model ini dapat digunakan untuk memfilter data yang tidak relevan secara otomatis dan efektif, sehingga sesuai untuk diterapkan dalam kerangka investigasi digital yang membutuhkan kecepatan dan presisi pada data berbasis percakapan [3]. Sementara itu, Random Forest digunakan karena kekuatannya dalam mengolah relasi kompleks antar fitur dan kestabilannya dalam melakukan klasifikasi menggunakan pendekatan berbasis voting dari sejumlah pohon keputusan. Algoritma ini digunakan untuk mendeteksi kejahatan siber pada media sosial, dengan pendekatan yang memanfaatkan parameter-parameter seperti informasi pengguna dan ciri linguistik dalam komentar. Random Forest terbukti mampu mengenali pola-pola yang menunjukkan ancaman atau aktivitas mencurigakan dalam data teks publik, serta bersifat robust terhadap data berisik dan bervariasi, sehingga sangat relevan untuk kebutuhan analisis forensik digital [9]. Kedua model diterapkan secara terpisah untuk keperluan perbandingan performa dalam mendeteksi komentar cyberbullying. Pendekatan ini dipilih agar dapat memberikan gambaran menyeluruh mengenai efektivitas masing-masing algoritma dalam menangani variasi karakteristik data percakapan, serta mendukung pengambilan keputusan dalam pemilihan model klasifikasi terbaik yang sesuai dengan kerangka kerja investigasi digital berbasis standar National Institute of Justice (NIJ).

Teknologi informasi saat ini menjadi pedang bermata dua. Di satu sisi, teknologi mendorong kesejahteraan, kemajuan, dan perkembangan peradaban manusia. Namun di sisi lain, kemajuan ini juga membuka peluang bagi berbagai bentuk kejahatan digital [10]. Salah satu tantangan dalam perkembangan komunikasi digital adalah meningkatnya kasus cyberbullying, yaitu kekerasan verbal atau psikologis melalui pesan teks, komentar, atau percakapan pribadi. Perundungan ini sering terjadi di aplikasi pesan instan karena sifat komunikasinya yang cepat, privat, dan sulit dipantau [11]. Banyak pelaku memanfaatkan ruang chat pribadi untuk memberikan tekanan psikologis kepada korban [12]. Telegram menjadi salah satu platform yang rawan disalahgunakan karena menyediakan grup publik dan channel terbuka, yang sering digunakan untuk menyebarkan ujaran kebencian, body shaming, hingga pelecehan seksual [13]. Sementara itu, kasus perundungan melalui aplikasi Line juga meningkat, terutama di kalangan pelajar, dalam bentuk hinaan, pengucilan, dan penyebaran konten memalukan melalui grup tertutup. Karena sifat komunikasi yang terenkripsi dan sulit dipantau, deteksi dini terhadap konten bermuatan perundungan menjadi penting dalam forensik digital [14]. Dengan tingginya volume percakapan dan beragamnya bahasa informal, pendekatan manual menjadi tidak efisien

Berbagai studi telah berupaya menjawab tantangan dalam klasifikasi data percakapan digital. Salah satu penelitian mengembangkan model *Predictive Classification Model (PCM)* berbasis *Complement Naive Bayes* untuk mengidentifikasi kolom percakapan dalam database aplikasi pesan instan seperti LINE, WhatsApp, dan Messenger. Model ini mampu menyaring lebih dari 88% data yang tidak relevan serta secara otomatis mengenali kolom penting seperti isi pesan, timestamp, dan ID pengguna. Namun, fokusnya masih terbatas pada identifikasi struktur data, tanpa memperhatikan konteks emosional dalam isi percakapan seperti ujaran kebencian atau perundungan digital [3]. Penelitian lain mengusulkan sistem deteksi predator daring menggunakan kombinasi fitur emosi dan kosakata. Meskipun pendekatan ini efektif dalam mengenali konten berisiko, sistem tersebut tidak menggunakan kerangka kerja forensik yang sah secara hukum seperti standar *National Institute of Justice (NIJ)* [15]. Penelitian terhadap artefak forensik dari aplikasi Telegram juga telah dilakukan dengan menggunakan lingkungan *virtual Android*, namun pendekatannya belum mencakup klasifikasi otomatis berbasis *machine learning* [2]. Penggunaan *framework National Institute of Justice (NIJ)* secara parsial juga pernah diterapkan untuk analisis Telegram, tetapi hanya sebatas pada tahap pengumpulan dan konversi data, tanpa adanya eksplorasi terhadap klasifikasi konten secara mendalam [7]. Selain itu, studi mengenai forensik media sosial menyoroti pentingnya alat bantu analisis dalam mengungkap kasus seperti perundungan, pemalsuan identitas, dan ujaran kebencian. Namun demikian, pendekatan tersebut belum mengintegrasikan metode klasifikasi otomatis berbasis *machine learning* secara eksplisit [16].

Kombinasi antara sistem predictive modeling dan standar prosedur forensik seperti yang diatur oleh *National Institute of Justice (NIJ)* memberikan pendekatan yang sistematis dan berbasis hukum agar tetap sah di mata hukum. Namun, penelitian-penelitian tersebut masih terbatas pada pengumpulan dan analisis artefak tanpa integrasi eksplisit model klasifikasi

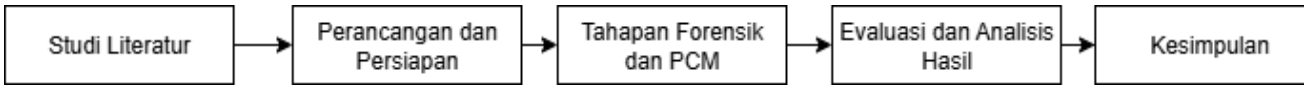
JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 4 No 2 September 2025
Hal: 836-844
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v4i2.521

berbasis *machine learning* yang dapat digunakan secara *real-time* untuk mendeteksi bukti digital pada pesan instan. Selain itu, belum banyak penelitian yang mengadopsi *framework National Institute of Justice (NIJ)* secara penuh dalam alur kerja forensik digital modern. Maka dari itu, penelitian ini berfokus pada *pengembangan Predictive Classification Model (PCM)* berbasis *Complement Naive Bayes* dan *Random Forest* yang diintegrasikan dengan pendekatan *National Institute of Justice (NIJ)*, untuk membantu investigator dalam mengidentifikasi konten bukti digital secara otomatis dan sah secara prosedural. Oleh karena itu, penelitian ini tidak hanya berfokus pada proses ekstraksi dan pengolahan bukti digital dari aplikasi Line dan Telegram, tetapi juga membandingkan performa dua algoritma *machine learning*, yaitu *Complement Naive Bayes* dan *Random Forest*, yang masing-masing diimplementasikan secara terpisah dalam metode *Predictive Classification Model (PCM)*. Perbandingan ini bertujuan untuk menentukan model yang paling efektif dalam mengklasifikasikan komentar berbasis teks dalam konteks investigasi forensik digital terhadap kasus *cyberbullying*.

Tujuan penelitian ini adalah untuk mengidentifikasi dan menganalisis proses pengumpulan metadata bukti digital dari aplikasi Line dan Telegram dengan mengacu pada standar *National Institute of Justice (NIJ)*, sehingga dapat menjamin validitas dan legalitas bukti digital dalam proses investigasi forensik. Selain itu, penelitian ini juga bertujuan menerapkan metode *Predictive Classification Model (PCM)* dalam proses pemeriksaan dan klasifikasi data digital hasil akuisisi dengan menggunakan dua algoritma machine learning, yaitu *Complement Naive Bayes* dan *Random Forest*, guna mendukung efektivitas investigasi forensik digital terhadap kasus *cyberbullying*. Selanjutnya, penelitian ini bermaksud mengevaluasi serta membandingkan performa kedua algoritma tersebut dalam penerapan PCM untuk klasifikasi otomatis terhadap bukti digital berbasis teks.

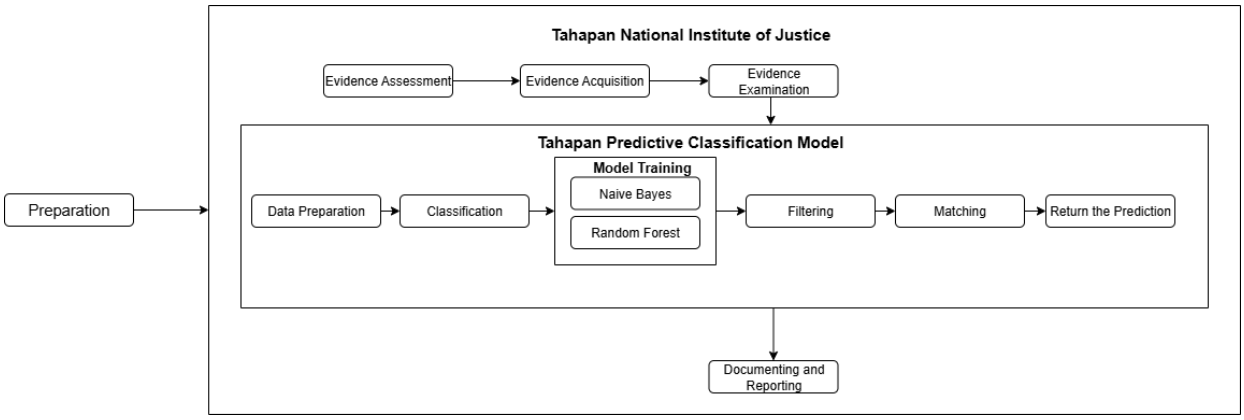
2. METODOLOGI PENELITIAN

2.1 Tahap Penelitian



Gambar 1. Metode Penelitian

Desain penelitian ini disusun melalui lima tahapan yang tersusun secara sistematis, sebagaimana digambarkan pada diagram alir penelitian berikut. Penelitian ini menggunakan kerangka kerja *National Institute of Justice* yang terdiri dari lima tahapan utama, yaitu [8]:



Gambar 2. Tahapan NIJ dan PCM

1. *Preparation (Pra-Forensik)*
   Tahap ini diawali dengan menyiapkan lingkungan sistem forensik berbasis *Kali Linux*. Dilakukan instalasi tools *ADB* dan *Fastboot* untuk interaksi dengan perangkat *Android*, serta pemasangan *library* pendukung seperti *BeautifulSoup* untuk pengolahan file *HTML* hasil ekspor Telegram. Selain itu, dibuat skrip *Python* khusus untuk mengonversi file *messages.html* Telegram ke format CSV, guna mempermudah analisis lanjutan.
2. *Evidence Assessment*
   Pada tahap ini dilakukan identifikasi dan penilaian awal terhadap dua aplikasi target, yaitu LINE dan Telegram, termasuk jenis data yang tersedia dan metode ekstraksi yang sesuai. Untuk Telegram, digunakan fitur *ChatExport* yang menghasilkan file *messages.html*, sementara LINE menyimpan histori pesan dalam file *Line_Chat.txt* di direktori internal. Penilaian mencakup aspek format file, struktur metadata, serta kemungkinan akses tanpa perlu *rooting*, untuk memastikan metode ekstraksi *forensically sound*.
3. *Evidence Acquisition*

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 4 No 2 September 2025
Hal: 836-844
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v4i2.521

Tahap ini merupakan proses pengambilan bukti digital dari perangkat *Android*. Perangkat dihubungkan ke sistem forensik, kemudian dilakukan proses *backup*, konversi file .ab ke .tar, dan ekstraksi ke direktori kerja. File *Line_Chat.txt* disalin menggunakan perintah *ADB* dan diolah dengan *awk* untuk dikonversi menjadi CSV. *File messages.html* dari Telegram diolah dengan skrip *Python* untuk diubah ke dalam bentuk CSV. Semua proses dilakukan dengan menjaga integritas data asli.

4. *Evidence Examination*
   Setelah proses akuisisi selesai, data hasil ekstraksi kemudian diperiksa dan dipersiapkan untuk analisis lebih lanjut. Tahap pertama dilakukan dengan memuat file CSV hasil konversi ke dalam lingkungan pemrograman *Python*. Selanjutnya, dilakukan proses data cleaning yang mencakup penghapusan nilai kosong (*missing values*), baris duplikat, karakter *non-ASCII*, serta *string* kosong yang berpotensi mengganggu hasil analisis. Setelah data dibersihkan, dilakukan tahap feature selection melalui teknik *preprocessing* teks, seperti normalisasi dan pemilahan kata kunci yang relevan terhadap konteks klasifikasi. Seluruh proses ini dilaksanakan dengan bantuan pustaka *Python* seperti *pandas, re, dan scikit-learn*, serta disusun untuk mendukung tahap pelatihan model klasifikasi secara optimal.
5. *Documenting and Reporting*
   Tahap akhir dari proses forensik adalah penyusunan laporan yang mendokumentasikan hasil analisis dan proses yang telah dilakukan. Laporan ini mencakup rangkuman kegiatan, alat dan metode yang digunakan, hasil analisis bukti digital, serta rekomendasi dan evaluasi yang berguna untuk penyelidikan forensik digital selanjutnya.

2.2 Proses Implementasi *Predictive Classification Model*

Predictive Classification Model (PCM) dibangun dengan dua pendekatan terpisah, yaitu menggunakan algoritma *Complement* Naive Bayes dan Random Forest. Kedua model tersebut diterapkan secara independen untuk mengevaluasiefektivitas masing-masing dalam mengklasifikasikan komentar ke dalam kategori bullying dan non-bullying berdasarkan fitur linguistik. Tujuan utama dari penggunaan dua algoritma ini adalah untuk mengetahui model mana yang paling efektif dalam melakukan klasifikasi terhadap komentar berbasis teks yang telah diekstrak dari hasil akuisisi bukti digital. Hasil evaluasi dari kedua model ini kemudian menjadi dasar untuk menentukan pendekatan klasifikasi yang paling sesuai dalam konteks investigasi forensik digital terhadap kasus cyberbullying.

Tabel 1. Contoh Kata dan Kalimat untuk Kategori Bullying dan Non-Bullying

Kategori	Kata Kunci	Contoh Kalimat
Bullying	jelek, bodoh, hina, gendut, idiot, banci, kafir	"Dasar lo jelek banget, pantas gak punya temen!"
Bullying	mati aja, goblok, gak laku, gak guna	"Orang kayak kamu mendingan mati aja, gak guna hidup!"
Non-Bullying	semangat, sukses, cakep, keren, pintar, baik	"Kamu keren banget, terus semangat ya belajarnya!"
Non-Bullying	cantik, sopan, lucu, rajin, hebat	"Postingan kamu lucu banget, sukses terus ya!"
Non-Bullying	aamiin, sehat selalu, tetap jaga diri	"Semoga kamu sehat selalu dan diberi kelancaran rezeki."

2.2.1 Proses Pelatihan dan Pengujian Model

Dataset yang digunakan dalam penelitian ini merupakan kumpulan 650 komentar publik dari platform Instagram yang telah diberi label ke dalam dua kategori, yaitu bullying dan non-bullying. Label diberikan secara manual berdasarkan analisis konten linguistik yang mengandung unsur penghinaan, ujaran kebencian, atau dukungan positif. Dataset tersebut terdiri atas dua atribut utama, yaitu Komentar sebagai data teks yang menjadi fitur input, dan Kategori sebagai label target klasifikasi. Untuk keperluan pelatihan dan pengujian model, dataset dibagi menjadi dua subset dengan rasio 80% untuk data latih (training) dan 20% untuk data uji (testing) menggunakan fungsi train_test_split dari pustaka scikit-learn. Pembagian dilakukan secara acak untuk memastikan variasi data yang memadai pada kedua subset dan menghindari overfitting. Tahapan preprocessing dilakukan terhadap data teks sebelum pelatihan, yang meliputi penghapusan karakter non-alfabet, normalisasi huruf menjadi lowercase, penghapusan duplikasi, serta konversi teks ke dalam representasi numerik menggunakan metode bag-of-words melalui CountVectorizer.

Selanjutnya, proses pelatihan dilakukan terhadap dua algoritma klasifikasi yang diimplementasikan secara terpisah, yaitu *Complement* Naive Bayes dan Random Forest Classifier. Keduanya dilatih menggunakan data latih untuk mengenali pola linguistik yang membedakan komentar bullying dan non-bullying. Setelah pelatihan, model diuji menggunakan data uji untuk mengevaluasi kemampuan generalisasi terhadap data baru. Evaluasi dilakukan dengan menghitung metrik performa

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 4 No 2 September 2025
Hal: 836-844
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v4i2.521

seperti precision, recall, f1-score, dan akurasi, yang bertujuan untuk menilai keandalan masing-masing model dalam mendeteksi komentar bermuatan bullying secara otomatis.

2.3 Evaluasi Model

Evaluasi dilakukan terhadap dua model klasifikasi teks, yaitu *Complement* Naive Bayes dan Random Forest, dengan tujuan mengukur kinerja dalam mendeteksi komentar bullying dan non-bullying. Evaluasi dilakukan menggunakan metrik precision, recall, f1-score, dan support.

Rincian hasil evaluasi model *Complement Naive Bayes* sebagai berikut:

Tabel 2. Model Complement Naive Bayes

Kelas	Precision	Recall	F1-Score	Support
Bullying	0.85	0.86	0.86	63
Non-bullying	0.88	0.85	0.86	67
accuracy	-	-	0.86	130
macro avg	0.86	0.86	0.86	130
weighted avg	0.86	0.86	0.86	130

Model *Complement* Naive Bayes menunjukkan performa yang seimbang antara dua kelas dengan akurasi keseluruhan sebesar 86%. Precision dan recall untuk kedua kelas hampir seragam, menandakan klasifikasi yang stabil dan tidak bias terhadap salah satu kategori.

Rincian hasil evaluasi model *Random Forest* sebagai berikut:

Tabel 3. Model Random Forest

Kelas	Precision	Recall	F1-Score	Support
Bullying	0.70	0.87	0.77	63
Non-bullying	0.84	0.64	0.73	67
accuracy	-	-	0.75	130
macro avg	0.77	0.76	0.75	130
weighted avg	0.77	0.75	0.75	130

Meskipun model Random Forest mampu mendeteksi komentar bullying dengan recall tinggi (87%), model ini cenderung bias terhadap kelas bullying, karena recall pada kelas non-bullying hanya sebesar 64%. Hal ini menyebabkan false positive lebih tinggi untuk kategori bullying, yang dapat menurunkan akurasi keseluruhan ke 75%.

3. HASIL DAN PEMBAHASAN

Bagian ini menyajikan hasil evaluasi kinerja dua model klasifikasi, yaitu *Complement* Naive Bayes dan Random Forest, dalam mengidentifikasi komentar bermuatan bullying. Hasil yang diperoleh dibahas berdasarkan metrik evaluasi seperti precision, recall, dan f1-score untuk masing-masing kelas. Selain evaluasi terhadap data uji, model yang telah dibangun juga diimplementasikan ke dalam sistem klasifikasi real-time guna menguji kemampuannya dalam mendeteksi komentar secara otomatis. Pembahasan difokuskan pada interpretasi performa model terhadap data uji yang telah diproses serta hasil uji coba implementasinya.

3.1 Hasil Evaluasi Model

a. *Predictive Classification Model* diterapkan untuk mendeteksi pola linguistik dalam komentar digital, khususnya ujaran yang mengandung unsur perundungan (cyberbullying).
b. Dua model dibandingkan, yaitu:
    1. *Complement* Naive Bayes: cocok untuk teks pendek dan data tidak seimbang.
    2. Random Forest: model berbasis pohon keputusan yang kuat terhadap overfitting.

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 4 No 2 September 2025
Hal: 836-844
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v4i2.521

3.1.1 Pembahasan Hasil Evaluasi

Hasil evaluasi menunjukkan bahwa model *Complement* Naive Bayes memiliki performa yang lebih stabil dan seimbang dalam mengklasifikasikan komentar ke dalam kategori bullying dan non-bullying. Hal ini terlihat dari nilai precision dan recall yang cukup tinggi dan merata pada kedua kelas. Untuk kelas bullying, precision mencapai 0.85, sedangkan recall-nya adalah 0.86. Artinya, dari semua komentar yang diprediksi sebagai bullying oleh model, 85% memang benar merupakan komentar bullying, dan dari seluruh komentar bullying dalam dataset, 86% berhasil dikenali oleh model. Nilai F1-Score yang merupakan rata-rata harmonis dari precision dan recall berada pada angka 0.86, menunjukkan keseimbangan antara kemampuan model untuk mengidentifikasi komentar yang benar dan menghindari kesalahan klasifikasi. Untuk kelas non-bullying, precision bahkan lebih tinggi yaitu 0.88, meskipun recall sedikit lebih rendah, yakni 0.85. Namun, secara umum, performa ini sangat baik dan konsisten, dengan akurasi keseluruhan mencapai 86%.

Model Random Forest, di sisi lain, menunjukkan performa yang sedikit kurang stabil. Precision pada kelas bullying lebih rendah yaitu 0.70, meskipun recall-nya cukup tinggi, yakni 0.87. Ini mengindikasikan bahwa model mampu mengenali sebagian besar komentar bullying, tetapi cukup sering mengklasifikasikan komentar non-bullying sebagai bullying. Sebaliknya, precision pada kelas non-bullying lebih tinggi yaitu 0.84, tetapi recall-nya justru rendah di angka 0.64. Artinya, hanya 64% dari komentar non-bullying yang berhasil dikenali dengan benar, sementara sisanya diklasifikasikan keliru sebagai bullying. Akurasi keseluruhan dari Random Forest hanya 75%, yang secara signifikan lebih rendah dibandingkan dengan *Complement* Naive Bayes. Hal ini menunjukkan adanya kecenderungan bias dari model Random Forest terhadap kelas bullying, yang dapat menyebabkan meningkatnya false positive rate.

Model Random Forest, di sisi lain, menunjukkan performa yang sedikit kurang stabil. Precision pada kelas bullying lebih rendah yaitu 0.70, meskipun recall-nya cukup tinggi, yakni 0.87. Ini mengindikasikan bahwa model mampu mengenali sebagian besar komentar bullying, tetapi cukup sering mengklasifikasikan komentar non-bullying sebagai bullying. Sebaliknya, precision pada kelas non-bullying lebih tinggi yaitu 0.84, tetapi recall-nya justru rendah di angka 0.64. Artinya, hanya 64% dari komentar non-bullying yang berhasil dikenali dengan benar, sementara sisanya diklasifikasikan keliru sebagai bullying. Akurasi keseluruhan dari Random Forest hanya 75%, yang secara signifikan lebih rendah dibandingkan dengan *Complement* Naive Bayes. Hal ini menunjukkan adanya kecenderungan bias dari model Random Forest terhadap kelas bullying, yang dapat menyebabkan meningkatnya false positive rate.

Secara kontekstual, performa *Complement* Naive Bayes yang seimbang sangat penting untuk kasus seperti cyberbullying, di mana salah satu risiko utama adalah terjadinya over detection (terlalu banyak komentar non-bullying yang dianggap sebagai bullying). Hal ini sangat mungkin terjadi pada sistem klasifikasi yang tidak akurat, dan dapat berdampak pada kebebasan berekspresi atau kesalahan deteksi dalam sistem moderasi otomatis. Dengan kemampuan model *Complement* Naive Bayes yang mampu membedakan komentar berdasarkan karakteristik linguistik secara efisien, model ini menjadi pilihan yang lebih tepat dalam implementasi nyata.

Dari hasil klasifikasi real-time menggunakan model *Complement* Naive Bayes, dapat dilihat bahwa komentar seperti “Kamu jelek banget, pantas gak punya temen!” terdeteksi sebagai bullying dengan tingkat keyakinan tinggi, sedangkan komentar seperti “Semangat terus ya, kamu pasti bisa lulus.” secara konsisten diklasifikasikan sebagai non-bullying. Hal ini menunjukkan bahwa model tidak hanya mengandalkan kata kasar secara eksplisit, tetapi juga mampu mengenali konteks dan intensi yang terkandung dalam komentar.

Sebaliknya, model Random Forest dalam beberapa kasus terlihat melakukan kesalahan klasifikasi terhadap komentar yang ambigu. Misalnya, komentar yang secara sarkastik menggunakan kata positif namun dalam konteks merendahkan, sering kali gagal ditangani dengan baik oleh Random Forest. Hal ini mengindikasikan bahwa model tersebut belum mampu menangkap nuansa semantik secara menyeluruh.

Dalam konteks penelitian serupa, hasil ini sejalan dengan temuan dari yang menyebutkan bahwa pendekatan berbasis *Complement* Naive Bayes cukup efektif dalam klasifikasi teks pendek karena efisiensinya dalam menangani frekuensi kata[15]. Selain itu, pendekatan ini tidak terlalu terpengaruh oleh fitur yang tidak relevan atau noise dalam data. Studi juga menunjukkan bahwa PCM yang digabungkan dengan metode probabilistik memberikan hasil klasifikasi yang akurat dalam konteks analisis percakapan[16].

Berdasarkan hasil pengujian tersebut, model *Complement* Naive Bayes dapat diintegrasikan ke dalam sistem moderasi konten digital, seperti aplikasi pelaporan kasus cyberbullying atau sistem filter komentar otomatis. Dengan akurasi tinggi dan performa yang seimbang, model ini tidak hanya mampu mendeteksi komentar negatif, tetapi juga menghindari kesalahan dalam mendeteksi komentar yang bersifat netral atau mendukung. Kemampuannya dalam mempertahankan nilai precision dan recall yang stabil menjadikannya cocok diterapkan pada lingkungan yang membutuhkan sensitivitas tinggi, seperti pendidikan, platform sosial, atau layanan konseling daring.

Secara teknis, penerapan model ini memanfaatkan pipeline *CountVectorizer* dan *ComplementNB* dalam *scikit-learn*, dengan waktu pelatihan yang relatif singkat dan kebutuhan komputasi yang rendah. Hal ini menjadi keunggulan tersendiri dalam implementasi sistem berbasis real-time. Di sisi lain, meskipun Random Forest memiliki potensi untuk menangkap hubungan non-linier, performanya pada data yang sangat bergantung pada kata dan konteks semantik seperti komentar sosial masih perlu ditingkatkan, misalnya dengan penggabungan metode *word embedding* seperti *Word2Vec* atau *BERT* agar lebih memahami konteks kalimat.

Dengan mempertimbangkan keseluruhan hasil dan konteks implementasi, dapat disimpulkan bahwa model *Complement* Naive Bayes lebih layak diterapkan sebagai komponen utama dalam sistem deteksi cyberbullying otomatis,

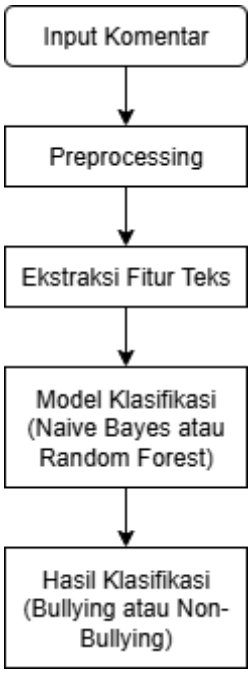
JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 4 No 2 September 2025
Hal: 836-844
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v4i2.521

khususnya pada platform dengan interaksi komentar teks pendek. Sementara Random Forest tetap berguna sebagai perbandingan atau sebagai lapisan tambahan untuk validasi, namun tidak direkomendasikan sebagai model utama tanpa dilakukan penyempurnaan.

3.2 Implementasi Model

Setelah model *Predictive Classification Model* dibangun dan dievaluasi, tahap selanjutnya adalah implementasi model ke dalam sistem klasifikasi komentar secara real-time. Tujuan dari implementasi ini adalah untuk memungkinkan sistem mengenali komentar yang berpotensi mengandung unsur cyberbullying secara otomatis, tanpa perlu campur tangan manusia dalam setiap proses deteksinya.

a. Arsitektur Sistem Implementa
Sistem implementasi dibangun menggunakan Python dengan pustaka seperti scikit-learn, pandas, joblib, dan matplotlib. Model *Complement* Naive Bayes dan Random Forest yang telah dilatih disimpan menggunakan joblib untuk dapat digunakan kembali tanpa pelatihan ulang. Proses dimulai dari input komentar mentah, dilanjutkan dengan preprocessing (normalisasi teks), lalu dikonversi menjadi format numerik menggunakan CountVectorizer. Data tersebut kemudian diprediksi oleh model untuk menentukan apakah termasuk bullying atau tidak. Hasil klasifikasi ditampilkan sebagai output dan dapat digunakan untuk moderasi otomatis secara real-time dengan akurasi tinggi.



Gambar 3. Arsitektur Sistem Implementasi PCM

b. Contoh Penggunaan Real-Time
Pada tahap implementasi real-time, model yang telah dilatih digunakan untuk mengklasifikasikan komentar sebagai bullying atau non-bullying. *Complement* Naive Bayes menunjukkan kinerja yang stabil, mampu mengenali komentar negatif secara akurat, dan membedakan komentar positif dengan baik. Sebaliknya, Random Forest cukup akurat namun cenderung salah mengklasifikasikan komentar sarkastik atau yang mengandung kata kasar dalam konteks bercanda. Misalnya, komentar seperti "Postingan ini norak banget" dianggap non-bullying meskipun bernada menyerang. Hal ini menunjukkan bahwa *Complement* Naive Bayes lebih konsisten dalam menangkap nuansa negatif halus dibanding Random Forest.

c. Kelebihan dan Keterbatasan Implementasi
Implementasi Predictive Classification Model (PCM) memiliki keunggulan dalam klasifikasi komentar berbasis teks karena ringan, responsif, dan dapat berjalan secara real-time tanpa beban komputasi besar. Model ini juga mudah diintegrasikan ke berbagai platform seperti media sosial atau sistem pelaporan. Selain klasifikasi, hasilnya dapat diperkaya dengan analisis toksisitas dan sentimen. Namun, PCM masih terbatas pada teks dan belum mampu memahami konteks visual atau bahasa sarkastik, ambigu, dan slang lokal. Random Forest khususnya menunjukkan kecenderungan

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 4 No 2 September 2025
Hal: 836-844
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v4i2.521

salah klasifikasi pada komentar netral. Oleh karena itu, meskipun menjanjikan, model tetap perlu penyempurnaan agar optimal dalam konteks yang kompleks.

d. Rencana Pengembangan Lanjutan
Untuk meningkatkan akurasi dan cakupan deteksi, pengembangan lanjutan mencakup integrasi teknik word embedding seperti Word2Vec, GloVe, atau BERT agar model lebih memahami konteks makna kata. Sistem juga dapat dikembangkan menjadi multi-label untuk mengklasifikasikan jenis cyberbullying secara spesifik, serta memberi peringatan otomatis dan saran kalimat etis kepada pengguna. Penggunaan dashboard analitik akan membantu moderator memantau tren dan aktivitas berisiko. Selain itu, penguatan dataset dari komentar asli berbahasa Indonesia diperlukan agar model lebih adaptif terhadap bahasa lokal, slang, dan emoji. Strategi ini diharapkan membuat sistem lebih akurat, kontekstual, dan berkelanjutan.

Dengan implementasi yang tepat dan evaluasi yang mendalam, sistem klasifikasi komentar berbasis machine learning ini dapat menjadi alat bantu yang signifikan dalam mencegah penyebaran cyberbullying, menjaga kenyamanan pengguna digital, dan mendukung proses investigasi forensik digital. Model yang dibangun memiliki potensi luas untuk diintegrasikan ke dalam berbagai sistem keamanan konten berbasis AI, baik di sektor pendidikan, platform sosial, maupun layanan pemerintahan.

4. KESIMPULAN

Penelitian ini berhasil mengimplementasikan Predictive Classification Model (PCM) sebagai metode klasifikasi komentar berbasis teks dalam mendukung proses investigasi forensik digital terhadap kasus cyberbullying. Dengan mengadopsi kerangka kerja National Institute of Justice (NIJ), proses forensik dilakukan secara sistematis mulai dari ekstraksi hingga analisis data digital. Dua algoritma machine learning, yaitu Complement *Complement* Naive Bayes dan Random Forest, diuji secara terpisah untuk menilai keefektifan PCM dalam mendeteksi komentar bermuatan cyberbullying. Hasil evaluasi menunjukkan bahwa model *Complement* Naive Bayes memiliki performa yang lebih stabil dan seimbang, dengan nilai precision dan recall yang konsisten pada kedua kelas, serta akurasi keseluruhan sebesar 86%. Sementara itu, model Random Forest mencatatkan akurasi 75%, namun cenderung bias terhadap kelas cyberbullying dengan tingkat kesalahan klasifikasi lebih tinggi pada komentar netral. Implementasi model secara real-time juga memperkuat efektivitas *Complement* Naive Bayes dalam mengenali komentar negatif dan membedakan komentar non-bullying secara akurat dan efisien.Kendala utama dalam penelitian ini terletak pada keterbatasan akses data asli dari aplikasi Line dan Telegram karena faktor enkripsi dan kebijakan privasi. Sebagai solusi, digunakan dataset komentar Instagram yang memiliki struktur dan konteks sosial yang relevan untuk simulasi deteksi cyberbullying. Meski demikian, hasil penelitian ini tetap memberikan kontribusi yang signifikan dalam menunjukkan potensi integrasi PCM dan machine learning dalam digital forensik. Dengan demikian, dapat disimpulkan bahwa pendekatan PCM dengan algoritma *Complement* Naive Bayes layak diterapkan sebagai komponen utama dalam sistem klasifikasi komentar berbasis teks, khususnya untuk mendeteksi cyberbullying secara otomatis dan sah secara prosedural. Sistem ini berpotensi mendukung berbagai platform digital dalam upaya menjaga ruang komunikasi yang aman dan sehat.

REFERENCES

[1] A. Raza and M. Bilal Hassan, “Digital Forensic Analysis of Telegram Messenger App in Android Virtual Environment,” *Mobile and Forensics*, vol. 4, no. 1, pp. 31–43, Mar. 2022, doi: 10.12928/mf.v4i1.5537.

[2] Dr. Vivekananth.P, “The Role of Social Media Forensics in Digital Forensics,” *International Journal of Engineering and Management Research*, vol. 12, no. 4, pp. 1–3, Aug. 2022, doi: 10.31033/ijemr.12.4.1.

[3] W. C. Lee, H. Y. Chen, and T. N. Lin, “A Predictive Classification Model for Identifying Conversation-related Mobile Forensic Data in Instant Messaging Applications,” *ISDFS 2023 - 11th International Symposium on Digital Forensics and Security*, pp. 0–5, 2023, doi: 10.1109/ISDFS58141.2023.10131853.

[4] O. Abudu, O. Scholastica Onyenaucheya, S. Erinfolami, A. Adams, and O. Esther Abudu, “Digital Forensics in Cybersecurity,” 2024. [Online]. Available: https://www.researchgate.net/publication/387467023

[5] C.-E. Bogos, R. Mocanu, and E. Simion, “A security analysis comparison between Signal, WhatsApp and Telegram,” *Cryptology ePrint Archive*, no. January, pp. 1–15, 2023.

[6] M. Riskiyadi, “Investigasi Forensik Terhadap Bukti Digital Dalam Mengungkap Cybercrime,” *Cyber Security dan Forensik Digital*, vol. 3, no. 2, pp. 12–21, 2020, doi: 10.14421/csecurity.2020.3.2.2144.

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 4 No 2 September 2025
Hal: 836-844
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v4i2.521

[7] M. E. Apriyani, R. A. Maskuri, M. H. Ratsanjani, A. Pramudhita, and R. Rawansyah, "Forensic Digital Analysis of Telegram Applications Using the National Institute Of Justice and Naïve Bayes Methods," *Mobile and Forensics*, vol. 5, no. 2, pp. 21–30, 2023, doi: 10.12928/mf.v5i2.7893.

[8] N. Institute of Justice, "Forensic Examination of Digital Evidence: A Guide for Law Enforcement." [Online]. Available: http://www.ojp.usdoj.gov/nij

[9] T. Arora, M. Sharma, and S. Khatri, *Detection of Cyber Crime on Social Media using Random Forest Algorithm*. IEEE, 2019.

[10] Y. Daeng, J. Levin, M. Razzaq Prayudha, N. Putri Ramadhani, S. Imanuel, and A. Penerapan Sistem Keamanan Siber Terhadap Kejahatan Siber Di Indonesia Yusuf Daeng, "Analisis Penerapan Sistem Keamanan Siber Terhadap Kejahatan Siber Di Indonesia," *Journal Of Social Science Research*, vol. 3, no. 6, pp. 1135–1145, 2023.

[11] E. Aboujaoude, M. W. Savage, V. Starcevic, and W. O. Salame, "Cyberbullying: Review of an old problem gone viral," Jul. 01, 2015, *Elsevier USA*. doi: 10.1016/j.jadohealth.2015.04.011.

[12] T. Ruslan, I. Riadi, and S. Sunardi, "Analisis Forensik Digital Pada Whatsapp Dan Facebook Menggunakan Metode NIST," *Jurnal Fasilkom*, vol. 13, no. 02, pp. 286–292, 2023, doi: 10.37859/jf.v13i02.5540.

[13] C. Negi, ""An overview of worldwide cyberbullying and cyberviolence against Women." [Online]. Available: https://ssrn.com/abstract=4529613

[14] M. R. D. Qibriya, A. Ambarwati, and K. E. Susilo, "Analisis Forensik Digital Pada Aplikasi Instant Messaging Di Smartphone Berbasis Android Untuk Bukti Digital," *Jurnal Teknologi Informasi*, vol. 5, no. 2, pp. 114–121, 2021, doi: 10.36294/jurti.v5i2.2200.

[15] M. A. Wani, N. Agarwal, and P. Bours, "Sexual-predator Detection System based on Social Behavior Biometric (SSB) Features," *Procedia CIRP*, vol. 189, pp. 116–127, 2021, doi: 10.1016/j.procs.2021.05.075.

[16] M. A. Aziz, I. Riadi, and R. Umar, "Analisis Forensik Line Messenger Berbasis Web Menggunakan Framework National Institute of Justice (Nij)," *Seminar Nasional Informatika*, vol. 2018, no. November, pp. 51–57, 2018.