BAB III
METODE PENELITIAN

3.1 Desain Penelitian

Gambar 3. 1 Desain Penelitian

Desain penelitian ini disusun melalui lima tahapan yang tersusun secara sistematis, sebagaimana digambarkan pada diagram alir penelitian berikut. Setiap tahapan dirancang untuk memastikan proses penelitian berjalan terstruktur, mulai dari kajian literatur hingga penarikan kesimpulan akhir.:

1. Studi Literatur

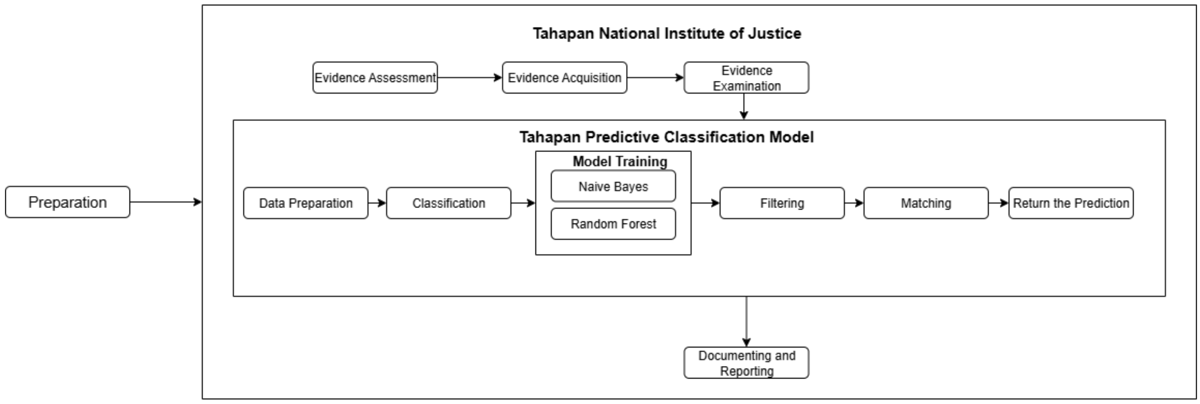
   Tahap awal dilakukan dengan mengkaji literatur terkait digital forensik, penggunaan *Predictive Classification Model (PCM)* untuk pengumpulan bukti digital, dan metode forensik berbasis standar *National Institute of Justice (NIJ).* Literatur yang dikaji meliputi teori investigasi digital, model machine learning untuk deteksi *cyberbullying*, serta implementasi PCM pada aplikasi pesan instan seperti Line dan Telegram.

2. Perancangan dan Persiapan

   Pada tahap ini dilakukan pemilihan objek penelitian, penentuan dataset, dan penyiapan alat serta perangkat lunak yang diperlukan. Peralatan yang digunakan meliputi Kali Linux, Python beserta *library* pendukung, dan tools forensik seperti ADB, Magnet AXIOM, atau Cellebrite UFED untuk membantu proses ekstraksi data. Tahap ini memastikan seluruh kebutuhan teknis siap untuk menjalankan penelitian.

3. Fase Forensik dan *Predictive Classification Model (PCM)*

   Menggunakan metode forensik berbasis NIJ yang terdiri dari beberapa tahap:



Gambar 3. 2 Tahapan NIJ dan PCM

1. *Preparation* (Pra-Forensik)

   Tahap ini melibatkan persiapan peralatan dan perangkat lunak yang diperlukan untuk penyelidikan. Peralatan dan perangkat yang dipilih harus sesuai dengan standar forensik digital dan mampu mendukung proses pengumpulan bukti digital. Contoh alat yang digunakan adalah Magnet AXIOM, Cellebrite UFED, atau Oxygen Forensic Suite dan Linux, yang berfungsi untuk memastikan data dapat diekstraksi dengan aman dari aplikasi Line dan Telegram. Sesuai panduan NIJ, tahap *Preparation* menekankan pentingnya perencanaan yang matang, pemilihan alat yang sahih secara forensik, serta pengujian awal terhadap perangkat lunak yang digunakan. Investigator diwajibkan untuk memastikan bahwa setiap perangkat dan *tool* yang dipakai telah terverifikasi, terdokumentasi, dan tidak mengubah integritas data asli. Dengan demikian, proses akuisisi bukti digital dapat berlangsung secara sistematis, aman, dan dapat dipertanggungjawabkan secara hukum.

2. *Evidence Assessment*

   Evidence Assessment merupakan tahapan awal dalam proses investigasi forensik digital berdasarkan panduan *National Institute of Justice* (NIJ). Tujuan utama dari tahap ini adalah untuk menilai, merencanakan, dan menentukan strategi terhadap data digital yang diduga mengandung bukti, sebelum proses akuisisi dilakukan. Tahap ini berfungsi sebagai fondasi awal untuk memastikan bahwa seluruh proses selanjutnya dilakukan secara efisien dan relevan, serta menjaga prinsip forensically sound. Sesuai panduan NIJ , tahap ini menekankan pada identifikasi sumber bukti digital, penentuan ruang lingkup penyelidikan, serta evaluasi potensi nilai probatif dari bukti tersebut. Investigator juga diwajibkan untuk mempertimbangkan aspek hukum, prosedural, dan teknis sejak awal, agar setiap langkah yang diambil tidak menyalahi *chain of custody* maupun merusak integritas data asli. Dengan demikian, *Evidence Assessment* menjadi pijakan penting sebelum proses akuisisi, karena memastikan hanya bukti yang relevan, sah, dan dapat dipertanggungjawabkan yang masuk ke tahap pemeriksaan berikutnya.

3. *Evidence Acquisition*

   Proses ini melibatkan pengumpulan data relevan dari perangkat atau server yang menyimpan informasi digital. Pada tahap ini, dibuat salinan digital dari objek fisik yang menyimpan bukti digital, mencakup log aktivitas, metadata, dan riwayat pesan. Pengumpulan data dilakukan dengan hati-hati untuk menjaga orisinalitasnya dan mencegah perubahan pada data asli, sehingga bukti tetap valid dan dapat dipertanggungjawabkan selama penyelidikan. Data yang terkumpul menjadi bahan input bagi PCM. Kolom-kolom seperti waktu pesan, ID pengguna, dan konten pesan akan diekstraksi untuk menganalisis pola interaksi. Sesuai dengan panduan NIJ, tahap *Evidence Acquisition* harus dilaksanakan dengan prinsip *forensically sound,* yaitu memastikan integritas bukti melalui proses pembuatan salinan bit-stream (imaging) atau metode akuisisi logis yang terverifikasi. Setiap langkah akuisisi harus didokumentasikan secara rinci, termasuk perangkat, *software*, dan teknik yang digunakan, untuk menjaga *chain of custody*. Dengan demikian, bukti digital yang diperoleh tidak hanya sah secara teknis, tetapi juga dapat dipertanggungjawabkan di ranah hukum.

4. *Evidence Examination*

   Setelah data terkumpul, tahap pemeriksaan dilakukan untuk memastikan integritas bukti digital. Pemeriksaan dapat dilakukan secara manual maupun otomatis, dengan tujuan untuk mendeteksi elemen-elemen penting dalam data, seperti waktu pesan, identitas pengguna, dan konten pesan. Tahap ini penting untuk menjaga agar bukti digital tetap orisinal dan tidak berubah dari kondisi saat ditemukan. Sesuai panduan NIJ, tahap *Examination* harus dilaksanakan secara sistematis, terdokumentasi, dan *forensically sound*. Investigator diperbolehkan menggunakan pendekatan manual maupun otomatis (misalnya melalui tools analisis atau algoritma klasifikasi), namun hasil pemeriksaan tetap perlu diverifikasi agar dapat dipertanggungjawabkan. NIJ menekankan bahwa pada tahap ini, fokus utama adalah melakukan penyaringan, validasi, dan pengorganisasian bukti digital sehingga dapat mendukung proses analisis lanjutan tanpa mengubah keaslian data. Dengan cara ini, integritas bukti tetap terjaga, sementara hasil pemeriksaan memiliki dasar ilmiah yang kuat dan dapat diuji ulang.

Setelah data terkumpul, dilakukan pemeriksaan data melalui beberapa langkah berikut:

- Data *Preparation*

  Dilakukan dengan memuat dan menyiapkan data hasil akuisisi untuk pembelajaran model. Proses ini dimulai dengan Load Data, di mana file CSV hasil ekstraksi dari aplikasi Line dan Telegram dimasukkan ke dalam lingkungan analisis. Selanjutnya dilakukan Data Cleaning untuk menghapus nilai kosong (NaN), duplikat, karakter non-ASCII, dan komentar kosong, sehingga dataset yang dihasilkan bersih dan siap dianalisis. Tahapan ini memastikan bahwa data yang digunakan dalam pemodelan merepresentasikan isi percakapan dengan kualitas terbaik.

- *Classification*

  Pada penelitian ini, proses Feature Selection juga menjadi bagian dari tahap ini, di mana fitur teks dari kolom komentar diekstraksi menggunakan Count Vectorizer agar dapat diolah oleh algoritma machine learning. Dua model dikembangkan,

Khana Yusdiana, 2025
PREDICTIVE CLASSIFICATION MODEL DALAM TAHAPAN FRAMEWORK NIJ UNTUK OTOMATISASI INVESTIGASI DIGITAL FORENSIK (STUDI KASUS: CYBERBULLYING)
Universitas Pendidikan Indonesia | repository.upi.edu | perpustakaan.upi.edu

yaitu Complement Naive Bayes dan Random Forest, untuk mengklasifikasikan komentar ke dalam kategori Bullying atau Non-bullying.

- *Filtering*
  kemudian dilakukan melalui proses Evaluation menggunakan classification report dan confusion matrix. Tahap ini berfungsi menyaring dan memvalidasi prediksi model, memastikan kolom dan fitur yang dihasilkan benar-benar relevan dengan konten percakapan yang menjadi objek analisis.
- *Matching*
  Direpresentasikan oleh analisis Feature Importance dan distribusi klasifikasi. Meskipun skripsi ini tidak melakukan pencocokan tabel Message dan Contact secara eksplisit seperti pada PCM di jurnal, tahap ini merepresentasikan pencocokan antara fitur teks dominan dan kategori komentar yang relevan untuk investigasi.
- Return the Prediction,
  Pada tahap ini, model yang sudah dilatih dan diuji disimpan menggunakan joblib, sehingga dapat digunakan untuk memprediksi komentar baru secara otomatis. Hasil prediksi ini dapat mendukung investigasi forensik digital dengan mendeteksi percakapan yang mengandung unsur cyberbullying secara efisien.

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.

4. Evaluasi dan Analisis Hasil
   Hasil klasifikasi dari kedua model dievaluasi menggunakan metrik precision, recall, f1-score, dan akurasi. Analisis ini bertujuan membandingkan kinerja

Complement Naive Bayes dan Random Forest, serta memvalidasi efektivitas PCM dalam mendeteksi cyberbullying secara otomatis.

5. Kesimpulan

   Tahap terakhir menyajikan kesimpulan penelitian berdasarkan hasil evaluasi model dan efektivitas penerapan PCM dalam mendukung investigasi forensik digital. Kesimpulan memuat temuan utama, kontribusi penelitian, dan potensi pengembangan penelitian di masa depan.

3.2 Objek Penelitian

Objek penelitian ini adalah dua aplikasi pesan instan, Line dan Telegram, yang dipilih karena tingkat popularitas dan penerapan enkripsi *end-to-end* dalam komunikasi mereka. Fokus penelitian ini adalah pada pengumpulan dan analisis bukti digital yang dihasilkan oleh kedua aplikasi tersebut, termasuk log aktivitas pengguna, metadata, riwayat pesan, serta data relevan lainnya dalam konteks investigasi forensik digital. Data ini akan dianalisis dan dibandingkan untuk menilai keakuratan masing-masing aplikasi dalam mendukung proses pengumpulan bukti digital dengan menggunakan *pendekatan Predictive Classification Model (PCM*). Selain data primer hasil akuisisi dari aplikasi LINE, penelitian ini juga memanfaatkan dataset publik Kaggle sebagai data sekunder. Dataset Kaggle dipilih karena bersifat terbuka, sudah melalui proses labeling (bullying dan non-bullying), serta dapat digunakan secara langsung untuk melatih dan menguji model klasifikasi. Dengan demikian, meskipun tidak diperlakukan sebagai objek penelitian utama, dataset ini berfungsi sebagai pondasi pendukung untuk meningkatkan reliabilitas analisis dan memperkuat validitas hasil model klasifikasi yang dibangun. Sesuai panduan NIJ, data hasil ekstraksi perangkat tetap diperlakukan sebagai bukti digital utama, sedangkan dataset publik Kaggle hanya bersifat komplementer untuk memperkaya hasil penelitian.

3.3 Pendekatan *Predictive Classification Model (PCM)*

Penelitian ini menggunakan *Predictive Classification Model (PCM)* untuk menganalisis data yang diperoleh dari aplikasi Line dan Telegram, dengan fokus utama pada deteksi dan klasifikasi kata-kata tertentu dalam percakapan yang berpotensi menjadi bukti digital. PCM akan mengidentifikasi kata kunci serta pola bahasa yang khas untuk setiap pengguna atau situasi yang relevan dengan investigasi. Pendekatan ini mengelompokkan istilah berdasarkan frekuensi kemunculannya dalam percakapan, yang memungkinkan identifikasi otomatis terhadap pesan atau pola interaksi yang

Khana Yusdiana, 2025
PREDICTIVE CLASSIFICATION MODEL DALAM TAHAPAN FRAMEWORK NIJ UNTUK OTOMATISASI INVESTIGASI DIGITAL FORENSIK (STUDI KASUS: CYBERBULLYING)
Universitas Pendidikan Indonesia | repository.upi.edu | perpustakaan.upi.edu

mencurigakan. Analisis sentimen adalah pendekatan berbasis analisis teks yang bertujuan untuk mendeteksi emosi, opini, atau sikap yang terkandung dalam suatu teks. Dalam konteks data percakapan digital, analisis sentimen memanfaatkan fitur linguistik seperti kosakata, pola emosi, dan struktur kalimat untuk mengidentifikasi pola interaksi yang relevan. Dalam konteks investigasi forensik digital, analisis sentimen membantu menyaring data percakapan yang besar, mengidentifikasi elemen-elemen linguistik yang relevan, dan mendukung proses investigasi dengan mengurangi data yang tidak relevan.

Tabel chat disusun dengan pendekatan yang mengacu pada skema data percakapan yang umumnya digunakan dalam digital forensik. Struktur tabel ini mencakup beberapa kolom utama, yaitu:

a. Konten Pesan

   Memuat teks atau kalimat dari percakapan yang telah difilter berdasarkan kata kunci tertentu.

b. Waktu Pesan

   Mencatat waktu pengiriman setiap pesan untuk memungkinkan analisis *time series* dan pola interaksi.

c. Identitas Pengguna

Menyimpan informasi pengguna yang terlibat dalam percakapan untuk keperluan identifikasi dan relasi antar-pesan.

d. Emosi dan Konteks

   Kategori tambahan ini dapat digunakan untuk menyimpan analisis emosional atau konotasi tertentu dari setiap pesan, terutama pada kalimat yang mengandung kata kunci yang bermuatan emosional.

e. Penentuan dan Pemilihan Kata Kunci

   Proses pemilihan kata kunci dalam PCM didasarkan pada literatur dan konteks percakapan yang umum ditemukan dalam investigasi forensik. Kata kunci ini memiliki kriteria sebagai berikut:

   1. Frekuensi Kemunculan

      Dipilih dari kata atau frasa yang sering digunakan dalam percakapan yang relevan dengan investigasi, misalnya kata-kata yang menunjukkan niat, tindakan, atau lokasi.

   2. Konotasi Emosional

Kata-kata yang memuat emosi tertentu seperti "takut," "sendirian," atau "temui" dipilih untuk mendeteksi pola interaksi yang mencurigakan.

3. Konteks Tindakan atau Permintaan

   Kata atau frasa yang secara kontekstual terkait dengan interaksi fisik atau permintaan.

Pemilihan kata kunci ini dimaksudkan untuk meningkatkan efisiensi PCM dalam melakukan penyaringan otomatis terhadap percakapan yang relevan. Dengan menggabungkan analisis kosakata dan pola emosional dalam percakapan, PCM berkontribusi dalam mempercepat dan meningkatkan akurasi pengumpulan bukti digital, terutama dalam konteks forensik digital. Melalui pendekatan ini, PCM diharapkan mampu memberikan tingkat keakuratan yang tinggi dalam mengumpulkan dan menganalisis bukti digital dari aplikasi Line dan Telegram, yang akan digunakan dalam konteks forensik digital untuk mendukung proses hukum.

Kategori Cyberbullying	Kata	Contoh Kalimat
Penghinaan Fisik	jelek	"Muka kamu jelek banget, mirip monster!"
	gendut	"Ih, gendut banget sih! Diet dong!"
Penghinaan Mental	bodoh	"Dasar bodoh, sekolah aja percuma!"
	gila	"Dia itu gila, jangan temenan sama dia!"
Bullying Verbal Online	anjing	"Anjing lo! Mending jangan nongol di grup ini!"
	sampah	"Hidup lo sampah, gak guna banget!"
Body Shaming	pendek	"Pendek banget sih, kayak anak SD!"
	sampah	"Hidup lo sampah, gak guna banget!"
Perundungan Sosial	hina	"Orang kayak kamu hina banget, gak pantes di sini!"
	pecundang	"Dasar pecundang, gak ada yang suka sama kamu!"

Tabel 3. 1 Istilah yang Sering Digunakan oleh Pelaku

3.4 Teknik Pengumpulan Data

Dalam penelitian ini, teknik pengumpulan data digital dilakukan dengan memanfaatkan berbagai alat forensik digital yang sesuai dengan standar *National Institute of Justice (NIJ).* Data yang dikumpulkan berasal dari aplikasi pesan instan Line dan Telegram, mencakup log aktivitas pengguna, metadata, riwayat pesan, serta data relevan lainnya. Proses pengambilan data dimulai dengan mengakses perangkat atau server yang menyimpan informasi tersebut, diikuti dengan ekstraksi data menggunakan alat forensik seperti Magnet AXIOM, Cellebrite UFED, atau Oxygen Forensic Suite. Tahapan pengumpulan data mencakup proses preservasi (untuk menjaga integritas data), akuisisi (pengambilan data secara digital), dan verifikasi untuk memastikan bahwa data yang diperoleh tidak mengalami perubahan selama proses pengambilan. Semua data yang dikumpulkan dijamin sesuai dengan prosedur standar forensik, sehingga hasilnya sah dan dapat dipertanggungjawabkan di pengadilan. Pendekatan ini membantu memastikan validitas bukti digital yang dikumpulkan dari aplikasi Line dan Telegram, yang selanjutnya akan dianalisis menggunakan *Predictive Classification Model (PCM).*

3.5 Teknik Analisi Data

Teknik analisis data dalam penelitian ini dilakukan melalui serangkaian langkah sistematis, dimulai dari pengolahan data mentah hingga klasifikasi data menggunakan *Predictive Classification Model (PCM).* Langkah pertama adalah pra-pemrosesan data, di mana data yang telah dikumpulkan dari aplikasi Line dan Telegram, seperti log aktivitas, metadata, dan riwayat pesan, dibersihkan dari elemen yang tidak relevan atau duplikat. Proses ini mencakup penghapusan karakter khusus, normalisasi data, dan pemilahan kata kunci yang berkaitan dengan investigasi. Penyusunan tabel percakapan mengikuti metodologi yang mendukung analisis data berbasis PCM, dengan setiap kolom tabel memuat atribut percakapan yang signifikan, seperti konten pesan, waktu, dan identitas pengguna.

3.6 Alat dan Bahan

Alat dan bahan yang digunakan dalam penelitian ini adalah sebagai berikut:

a. Sistem Operasi Kali Linux

   Digunakan sebagai lingkungan utama untuk melakukan proses pengumpulan (akuisisi) data digital dari perangkat Android. Kali Linux menyediakan berbagai *tools* berbasis *command line* seperti ADB (Android Debug Bridge) dan utilitas file *system* yang mendukung proses ekstraksi data dari aplikasi Line dan Telegram secara forensik.

b. Model *Predictive Classification Model (PCM)*

   Digunakan untuk melakukan analisis otomatis pada data yang diperoleh, termasuk identifikasi kata kunci, klasifikasi, dan pengelompokan data percakapan. Dalam penelitian ini, PCM diimplementasikan menggunakan dua algoritma machine learning, yaitu *Complement Naive Bayes* dan *Random Forest.*

c. Komputer

   Digunakan untuk menjalankan perangkat lunak forensik dan model PCM dalam proses pengumpulan dan analisis data.