

## **BAB I**

### **PENDAHULUAN**

#### **1.1 Latar Belakang Masalah**

Maraknya kasus kekerasan seksual berbasis gender terutama terhadap perempuan masih menjadi permasalahan serius yang berdampak pada kesehatan, keamanan, dan kesejahteraan sosial di berbagai negara. Studi global menunjukkan bahwa kekerasan berbasis gender masih terjadi secara luas dan menghambat pencapaian kesetaraan gender di berbagai wilayah dunia [1]. Di Indonesia, gambaran kondisi tersebut dapat dilihat melalui data resmi Sistem Informasi Online Perlindungan Perempuan dan Anak (SIMFONI-PPA, Kementerian PPPA, 2025) yang dikelola oleh Kementerian Pemberdayaan Perempuan dan Perlindungan Anak. Berdasarkan halaman ringkasan SIMFONI-PPA yang diakses pada tanggal 27 Januari 2026, tercatat sebanyak 35.131 kasus laporan kekerasan. Sistem tersebut mencatat bahwa data bersifat real-time dan mencakup data terverifikasi serta data yang masih dalam proses verifikasi pada periode berjalan. Jumlah laporan ini tidak secara langsung merepresentasikan seluruh kejadian, karena dapat dipengaruhi oleh faktor pelaporan dan mekanisme pencatatan. Oleh sebab itu, analisis tidak hanya berfokus pada jumlah kasus, tetapi juga pada persepsi dan pembenaran masyarakat terhadap kekerasan seksual berbasis gender.

Salah satu aspek dasar dalam memahami akar kekerasan seksual adalah menganalisis faktor demografi yang memengaruhi perilaku atau pembenaran terhadap tindakan tersebut. Penelitian terkait, menunjukkan bahwa variabel seperti tingkat pendidikan, status ekonomi, dan usia memiliki hubungan signifikan terhadap risiko dan persepsi kekerasan seksual [2]. Dataset *'Violence Against Women & Girls'* menyediakan data survei lintas negara yang memuat respons masyarakat terhadap berbagai situasi pembenaran kekerasan domestik beserta karakteristik demografinya. Pemanfaatan data tersebut memungkinkan analisis pola pembenaran kekerasan secara lebih luas sehingga dapat membantu mengidentifikasi kelompok masyarakat yang memiliki kecenderungan penerimaan lebih tinggi terhadap tindakan kekerasan berbasis gender.

Seiring dengan kompleksitas data sosial yang tersedia, metode statistik tradisional mulai beralih ke pendekatan *Machine Learning* untuk kinerja prediksi lebih baik. Studi terbaru menyoroti bahwa pemanfaatan kecerdasan buatan dan *machine learning* dapat mempercepat pencapaian *Sustainable Development Goal 5* (SDG 5) dengan mendeteksi pola kekerasan dan pemberdayaan perempuan secara lebih efektif [3]. Selain itu, perkembangan teknologi *Natural Language Processing* (NLP) juga memungkinkan analisis konteks teks yang berkaitan dengan opini dan sikap masyarakat terhadap isu kekerasan secara lebih efektif, baik melalui data survei maupun sumber teks lainnya [4]. Namun, tantangan utama tetap terletak pada integrasi data teks dengan data demografi agar model mampu menghasilkan prediksi yang stabil sekaligus dapat diinterpretasikan.

Untuk menjawab tantangan tersebut, penelitian ini mengusulkan pendekatan *hybrid BERT-LightGBM* (*Bidirectional Encoder Representations from Transformers – Light Gradient Boosting Machine*) yang diintegrasikan dengan *dashboard* interaktif berbasis Streamlit. Pendekatan hibrida ini mengombinasikan representasi vektor semantik dari model BERT sebagai *feature extractor* teks dengan algoritma LightGBM berbasis *gradient boosting* untuk tugas regresi nilai kontinu. Kombinasi tersebut memungkinkan model memanfaatkan informasi kontekstual teks dan fitur demografi secara bersamaan sehingga menghasilkan kinerja prediksi yang lebih baik dibandingkan model tunggal [5]. Model yang dihasilkan kemudian diintegrasikan ke dalam *dashboard* interaktif berbasis Streamlit untuk memvisualisasikan hasil analisis dan prediksi secara informatif. Melalui visualisasi ini, pengguna diharapkan dapat memperoleh pemahaman yang lebih komprehensif mengenai pola pembenaran kekerasan berbasis gender serta mendukung penyusunan strategi pencegahan berbasis data.

## 1.2 Rumusan Masalah

1. Bagaimana hubungan variabel demografi terhadap tingkat pembenaran tindak kekerasan seksual berbasis gender berdasarkan analisis data?

2. Bagaimana kinerja model *Hybrid* BERT-LightGBM dalam memprediksi tingkat pembenaran kekerasan seksual berbasis gender dibandingkan dengan model tunggal?
3. Bagaimana implementasi *dashboard* interaktif berbasis Streamlit untuk memvisualisasikan hasil analisis faktor demografi dan prediksi model secara efektif?

### 1.3 Batasan Masalah

1. Data : Penelitian ini menggunakan dataset sekunder yang diperoleh dari platform Kaggle, yaitu dataset "*Violence Against Women & Girls*". Dataset tersebut memuat variabel demografi responden (seperti usia, tingkat pendidikan, dan status ekonomi) serta data pernyataan survei yang merepresentasikan pembenaran kekerasan seksual berbasis gender dalam konteks hubungan suami dan istri.
2. Metode : Metode utama yang digunakan dalam penelitian ini adalah pendekatan *hybrid* BERT-LightGBM. BERT (*Bidirectional Encoder Representations from Transformers*) dimanfaatkan sebagai *feature extractor* untuk mengekstraksi representasi semantik dari teks pernyataan survei. Selanjutnya, LightGBM digunakan sebagai model regresi untuk memprediksi tingkat pembenaran kekerasan berbasis gender dalam bentuk nilai persentase. Pendekatan ini diformulasikan sebagai tugas regresi nilai kontinu, bukan klasifikasi kategori.
3. Visualisasi : Implementasi antarmuka pengguna (*User Interface*) dibatasi pada pembuatan *dashboard* berbasis Python menggunakan *framework* Streamlit sebagai media demonstrasi hasil prediksi dan analisis data.
4. Lingkup Analisis : Penelitian ini berfokus pada analisis pola pembenaran kekerasan seksual berbasis gender berdasarkan variabel demografi. Penelitian ini tidak mencakup sistem pelaporan kasus hukum, analisis pada tingkat individu, maupun pemantauan kejadian secara *real-time*.

#### 1.4 Tujuan Penelitian

1. Menganalisis hubungan antara faktor demografi dengan tingkat pembenaran kekerasan seksual berbasis gender untuk mengidentifikasi variabel yang paling dominan.
2. Mengimplementasikan model *Hybrid* BERT-LightGBM dan mengevaluasi kinerja dalam memprediksi tingkat pembenaran kekerasan seksual berbasis gender menggunakan metrik regresi.
3. Mengimplementasikan *dashboard* visualisasi data berbasis Streamlit untuk menampilkan hasil analisis faktor demografi dan prediksi model secara interaktif dan komunikatif.

#### 1.5 Manfaat Penelitian

1. Manfaat Teoretis
  - a. Memberikan kontribusi pada pengembangan ilmu komputer, khususnya dalam penerapan pendekatan *hybrid* antara model *deep learning* (BERT) dan *machine learning* berbasis *boosting* (LightGBM) yang dikombinasikan dengan teknik visualisasi data (*dashboard*) untuk penyelesaian permasalahan sosial (*social computing*).
  - b. Menjadi referensi akademis terkait efektivitas integrasi model bahasa berbasis *Hybrid* BERT-LightGBM dalam analisis data campuran berupa data teks dan data tabular.
2. Manfaat Praktis
  - a. *Dashboard* yang dihasilkan dapat digunakan sebagai alat bantu pendukung pengambilan keputusan (*Decision Support System*) untuk memantau tren persepsi masyarakat dan mendukung perancangan kebijakan sosialisasi yang lebih tepat sasaran.

- b. Memberikan edukasi visual yang mudah dipahami mengenai bahaya normalisasi kekerasan seksual serta faktor-faktor demografi yang berkontribusi terhadap fenomena tersebut.
- c. Menyediakan *framework* atau *prototype* sistem analisis data sosial yang dapat dikembangkan lebih lanjut dengan penambahan fitur dan model yang lebih kompleks.

## 1.6 Sistematika Penulisan

BAB I PENDAHULUAN, berisi Latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, dan manfaat penelitian, serta sistematika penulisan skripsi.

BAB II TINJAUAN PUSTAKA, berisi tinjauan pustaka, dasar-dasar teori yang relevan, penelitian-penelitian sebelumnya yang berkaitan, dan landasan teori untuk metode yang digunakan, termasuk konsep-konsep terkait kekerasan berbasis gender, faktor demografi, Natural Language Processing (NLP), dan algoritma BERT serta LightGBM *Regressor*.

BAB III METODE PENELITIAN, didalamnya terdapat tinjauan umum tentang objek penelitian, analisis masalah, solusi yang ditawarkan, rancangan, seperti pengumpulan dan pra-pemrosesan data, arsitektur model menggunakan Hybrid BERT-LightGBM, tahapan pelatihan dan evaluasi model, serta metode visualisasi data.

BAB IV HASIL DAN PEMBAHASAN, bab ini merupakan tahapan yang penulis lakukan dalam mengembangkan model, hingga hasil visualisasi dari dashboard interaktif.

BAB V PENUTUP, berisi kesimpulan dan saran yang dapat peneliti rangkum selama proses penelitian.