

**ANALISIS FAKTOR DEMOGRAFI TERHADAP TINGKAT
PEMBENARAN KEKERASAN SEKSUAL BERBASIS
GENDER MENGGUNAKAN PENDEKATAN *HYBRID* BERT-
LIGHTGBM**

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

HAYYA RASYKAH NINGRUM

22.11.5303

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2026

**ANALISIS FAKTOR DEMOGRAFI TERHADAP TINGKAT
PEMBENARAN KEKERASAN SEKSUAL BERBASIS
GENDER MENGGUNAKAN PENDEKATAN *HYBRID* BERT-
LIGHTGBM**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

HAYYA RASYKAH NINGRUM

22.11.5303

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2026

HALAMAN PERSETUJUAN

SKRIPSI

**ANALISIS FAKTOR DEMOGRAFI TERHADAP TINGKAT
PEMBENARAN KEKERASAN SEKSUAL BERBASIS GENDER
MENGUNAKAN PENDEKATAN *HYBRID BERT-LIGHTGBM***

yang disusun dan diajukan oleh

Hayya Rasykah Ningrum

22.11.5303

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 18, Februari 2026

Dosen Pembimbing


Prof. Dr. Ema Utami S.Si, M.Kom.
NIK. 190302037

HALAMAN PENGESAHAN

SKRIPSI

**ANALISIS FAKTOR DEMOGRAFI TERHADAP TINGKAT
PEMBENARAN KEKERASAN SEKSUAL BERBASIS GENDER
MENGUNAKAN PENDEKATAN *HYBRID* BERT-LIGHTGBM**

yang disusun dan diajukan oleh

Hayya Rasykah Ningrum

22.11.5303

Telah dipertahankan di depan Dewan Penguji
pada tanggal 18 Februari 2026

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Theophilus Bayu Sasongko, S.Kom., M.Eng.
NIK. 190302375

Bayu Setiaji, M.Kom.
NIK. 190302216

Prof. Dr. Ema Utami, S.Si., M.Kom.
NIK. 190302037

Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 18 Februari 2026

DEKAN FAKULTAS ILMU KOMPUTER



Prof. Dr. Kusriani, M.Kom.
NIK. 190302106

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Hayya Rasykah Ningrum
NIM : 22.11.5303

Menyatakan bahwa Skripsi dengan judul berikut:

Analisis Faktor Demografi Terhadap Tingkat Pembenaaran Kekerasan Seksual Berbasis Gender Menggunakan Pendekatan Hybrid Bert-LightGBM

Dosen Pembimbing : Prof. Dr. Ema Utami, S.Si., M.Kom.

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 18 Februari 2026

Yang Menyatakan,



Hayya Rasykah Ningrum

HALAMAN PERSEMBAHAN

Dengan penuh rasa syukur dan cinta, skripsi ini penulis persembahkan kepada Mama dan Papa tercinta. Terima kasih atas doa yang tak pernah putus mengiringi setiap langkah, dukungan tanpa batas, dan pengorbanan yang tak ternilai harganya. Nilai-nilai kehidupan yang Mama dan Papa tanamkan akan selalu menjadi fondasi terkuat dalam setiap proses pembelajaran dan perjalanan akademik ini. Keluarga yang senantiasa memberikan motivasi, kepercayaan, dan dukungan moral dalam setiap tahap penyelesaian studi.

Persembahan ini juga penulis tuju kepada keluarga yang senantiasa menjadi tempat pulang, serta memberikan motivasi dan dukungan moral dalam setiap tahap penyelesaian studi.

Rasa terima kasih yang mendalam juga penulis haturkan kepada Dosen Pembimbing dan seluruh civitas akademika Universitas AMIKOM Yogyakarta yang telah dengan sabar memberikan arahan, ilmu yang bermanfaat, serta lingkungan akademik yang sangat mendukung perkembangan intelektual penulis dari awal hingga akhir.

Dalam perjalanan yang tidak selalu mudah ini, penulis merasa sangat bersyukur memiliki teman-teman terdekat, Niken dan Nadiya. Terima kasih telah menjadi tempat bersandar, kawan bertukar pikiran, dan menemani setiap tawa maupun keluh kesah sehingga beban perjalanan ini terasa jauh lebih ringan.

Rasa terima kasih juga penulis sampaikan kepada Kim Sunoo (EN-HYPEN) yang selalu menjadi sumber penyemangat dan energi positif di kala lelah. Tidak lupa, lagu "Matilda" karya Harry Styles yang telah menjadi teman setia, memberikan ketenangan, dan mengingatkan penulis bahwa selalu ada ruang yang aman untuk merayakan proses bertumbuh.

Akhir kata, persembahan ini secara khusus penulis tuju kepada diri sendiri. Terima kasih sudah bertahan, berjuang, dan menolak untuk menyerah pada keadaan. Halaman ini adalah wujud apresiasi tertinggi atas tanggung jawab, ketekunan, dan komitmen yang telah dijaga dengan sebaik-baiknya untuk menyelesaikan amanah pendidikan ini hingga garis akhir.

KATA PENGANTAR

Bismillahirrahmanirrahim.

Puji dan syukur penulis panjatkan ke hadirat Allah SWT atas segala rahmat dan karunia-Nya sehingga skripsi yang berjudul “Analisis Faktor Demografi Terhadap Tingkat Pembenaan Kekerasan Seksual Berbasis Gender Menggunakan Pendekatan Hybrid BERT-LightGBM” ini dapat diselesaikan dengan baik sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer pada Program Studi Informatika, Fakultas Ilmu Komputer, Universitas AMIKOM Yogyakarta.

Penulisan skripsi ini tidak terlepas dari dukungan dan bantuan berbagai pihak. Oleh karena itu, penulis menyampaikan ucapan terima kasih yang sebesar-besarnya kepada:

1. Prof. Dr. Ema Utami, S.Si., M.Kom. selaku Dosen Pembimbing yang telah memberikan arahan, bimbingan, dan masukan selama proses penyusunan skripsi ini.
2. Bapak/Ibu Dosen Penguji yang telah memberikan saran dan evaluasi demi penyempurnaan penelitian ini.
3. Prof. Dr. Kusriani, M.Kom. selaku Dekan Fakultas Ilmu Komputer Universitas AMIKOM Yogyakarta.
4. Seluruh dosen dan staf Program Studi Informatika yang telah memberikan ilmu dan pengalaman selama masa perkuliahan.
5. Kedua orang tua dan keluarga tercinta atas doa, dukungan moral, dan semangat yang tidak pernah putus.

Penulis menyadari bahwa skripsi ini masih memiliki keterbatasan dan kekurangan. Oleh karena itu, kritik dan saran yang membangun sangat diharapkan untuk pengembangan penelitian selanjutnya. Semoga skripsi ini dapat memberikan manfaat bagi pengembangan ilmu pengetahuan, khususnya dalam bidang machine learning dan analisis data sosial.

Yogyakarta, 18 Februari 2026

Penulis

DAFTAR ISI

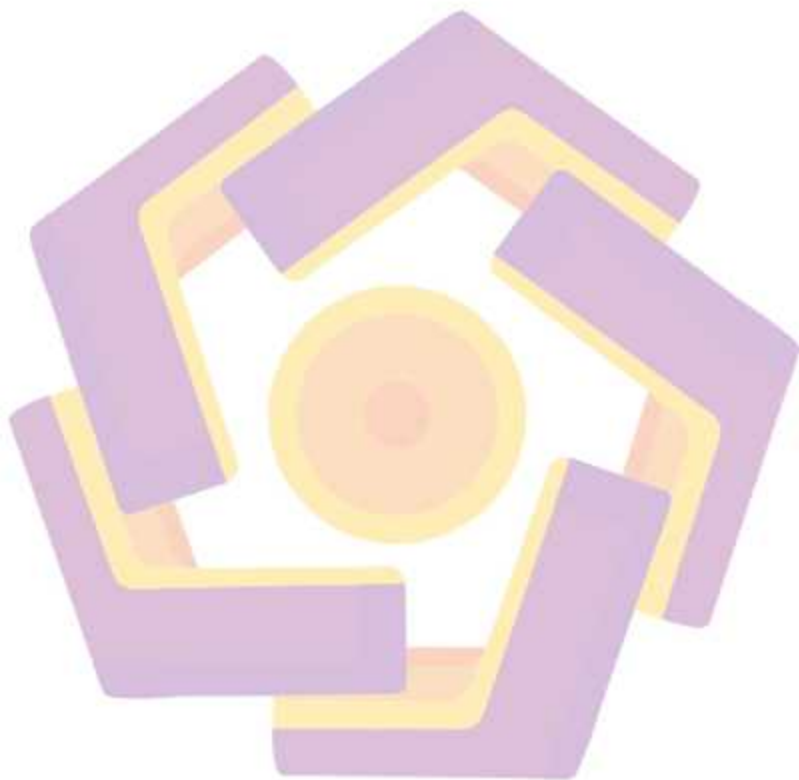
HALAMAN JUDUL	i
HALAMAN PERSETUJUAN	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	iv
HALAMAN PERSEMBAHAN	v
KATA PENGANTAR	vi
DAFTAR ISI	vii
DAFTAR TABEL	x
DAFTAR GAMBAR	xi
DAFTAR LAMPIRAN	xii
DAFTAR LAMBANG DAN SINGKATAN	xiii
DAFTAR ISTILAH	xiv
INTISARI	xv
<i>ABSTRACT</i>	xvi
BAB I PENDAHULUAN	1
1.1 Latar Belakang Masalah	1
1.2 Rumusan Masalah	2
1.3 Batasan Masalah	3
1.4 Tujuan Penelitian	4
1.5 Manfaat Penelitian	4
1.6 Sistematika Penulisan	5
BAB II TINJAUAN PUSTAKA	6
2.1 Studi Literatur	6

2.2	Dasar Teori	13
2.2.1	Kekerasan Seksual Berbasis Gender	13
2.2.2	<i>Natural Language Processing (NLP)</i>	15
2.2.2.1	Tahapan Pra-pemrosesan Teks (<i>Text Preprocessing</i>)	15
2.2.2.2	Representasi dan Pemodelan Teks	16
2.2.2.3	Evolusi Menuju Arsitektur <i>Transformer</i> (BERT)	16
2.2.3	Arsitektur <i>Transformer</i> & BERT	17
2.2.3.1	<i>Transformer</i> dan Mekanisme Atensi (<i>Attention Mechanism</i>)	17
2.2.3.2	Arsitektur BERT (<i>Bidirectional Encoder Representations from Transformers</i>)	18
2.2.3.3	Mekanisme <i>Pre-training</i> dan <i>Fine-Tuning</i>	18
2.2.3.4	Output BERT: <i>Contextual Word Embedding</i>	19
2.2.4	<i>Light Gradient Boosting Machine (LightGBM)</i>	20
2.2.4.1	Konsep Dasar dan Efisiensi	20
2.2.4.2	Strategi Pertumbuhan Pohon (<i>Leaf-wise Growth</i>)	20
2.2.4.3	Formulasi Matematis Pembagi Node (<i>Split Finding</i>)	21
2.2.4.4	Fungsi Objektif Regresi	22
2.2.5	Metode Pendukung	23
2.2.5.1	<i>Hybrid Modeling</i> (BERT-LightGBM)	23
2.2.5.2	Evaluasi Model	24
2.2.6	Streamlit	25
BAB III METODE PENELITIAN		27
3.1	Objek Penelitian	27
3.2	Alur Penelitian	27
3.3	Alat dan Bahan	31

3.3.1	Data Penelitian	31
3.3.2	Perangkat Lunak dan <i>Library</i>	31
3.3.3	Perangkat Keras	32
BAB IV HASIL DAN PEMBAHASAN		33
4.2	Seleksi Fitur	34
4.3	<i>Cleaning</i> Data	35
4.4	<i>Exploratory Data Analysis</i> (EDA)	36
4.5	Visualisasi Data Analisis	39
4.6	<i>Encoding</i> Fitur Kategorikal	43
4.7	Ekstraksi Fitur Teks dengan BERT	43
4.8	Train Test Split	45
4.9	Standarisasi Fitur Demografi	46
4.10	Reduksi Dimensi Embedding (PCA)	46
4.11	Pembentukan Fitur Hybrid	47
4.12	Pembangunan Model	48
4.13	Evaluasi dan Validasi Model	50
4.14	Validasi Stabilitas Model Menggunakan <i>K-Fold Cross-Validation</i>	54
4.15	Bias & Interpretabilitas	56
4.16	Implementasi <i>Streamlit</i>	57
BAB V PENUTUP		59
5.1	Kesimpulan	59
5.2	Saran	59
REFERENSI		61
LAMPIRAN		65

DAFTAR TABEL

Tabel 2.1. Keaslian Penelitian	9
Tabel 4.1. Perbandingan Evaluasi Model	53



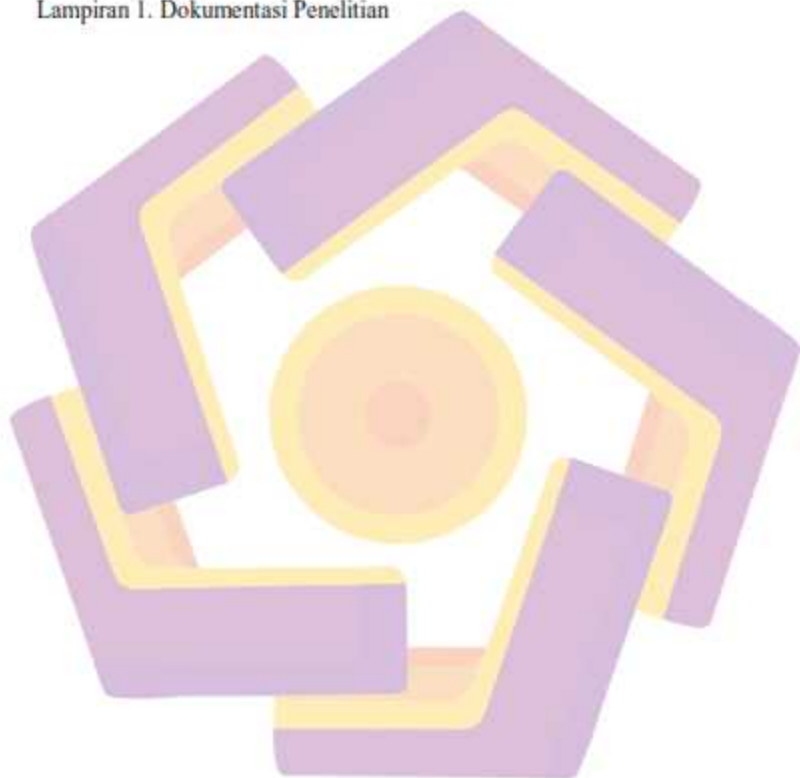
DAFTAR GAMBAR

Gambar 3.1. Alur Penelitian	27
Gambar 4.1. Memasukkan Data	32
Gambar 4.2. Verifikasi Data	33
Gambar 4.3. Lima Baris Data Teratas	33
Gambar 4.4. Fitur yang Digunakan	34
Gambar 4.5. Fitur Target	34
Gambar 4.6. Normalisasi Label Gender	35
Gambar 4.7. Penanganan <i>Missing Value</i>	35
Gambar 4.8. Output	36
Gambar 4.9. Deskripsi Fitur Target (<i>Value</i>)	37
Gambar 4.10. Fitur <i>Question</i>	37
Gambar 4.11. Fitur Gender	37
Gambar 4.12. Histogram & <i>Boxplot</i>	38
Gambar 4.13. <i>Violin Plot</i>	39
Gambar 4.14. <i>Boxplot Demographic Question</i>	39
Gambar 4.15. Top 15 Negara dengan Rata-rata Pembelian Tertinggi	40
Gambar 4.16. Tren Tahun	41
Gambar 4.17. Korelasi Fitur Demografi dengan Target	41
Gambar 4.18. Label <i>Encoding</i>	42
Gambar 4.19. Load Model	43
Gambar 4.20. Load Tokenizer	43
Gambar 4.21. Ekstraksi Fitur Kolom <i>Question</i>	44
Gambar 4.22. <i>Train-Test Split</i>	44
Gambar 4.23. <i>Scaler-Train Only</i>	45
Gambar 4.24. Reduksi Dimensi <i>Embedding (PCA)</i>	45
Gambar 4.25. <i>Combine Features</i>	46
Gambar 4.26. Model <i>Training Demo</i>	47
Gambar 4.27. Model <i>Training Text</i>	48
Gambar 4.28. Model <i>Training Hybrid</i>	49
Gambar 4.29. Visualisasi Perbandingan Evaluasi Model	50
Gambar 4.30. <i>Actual vs Predict Plot</i>	50
Gambar 4.31. Residual <i>Hybrid</i>	52
Gambar 4.32. <i>K-Fold Cross-Validation</i>	53
Gambar 4.33. SHAP-Fitur Demografi	55
Gambar 4.34. <i>Gender Bias Analysis</i>	56
Gambar 4.35. Visualisasi Bias Gender	56
Gambar 4.36. Implementasi Streamlit	57

DAFTAR LAMPIRAN

Lampiran 1. Dokumentasi Penelitian

65



DAFTAR LAMBANG DAN SINGKATAN

g_i	Turunan pertama (<i>gradient</i>) dari fungsi kerugian terhadap prediksi pada sampel ke- i .
h_i	Turunan kedua (<i>hessian</i>) dari fungsi kerugian terhadap prediksi pada sampel ke- i .
$L(y_i, \hat{y}_i)$	Fungsi kerugian (<i>loss function</i>) antara nilai aktual dan nilai prediksi.
y_i	Nilai aktual pada sampel ke- i .
$\hat{y}_i$	Nilai prediksi model pada sampel ke- i .
L	<i>Node</i> kiri (<i>Left node</i>) pada proses pemisahan pohon keputusan.
R	<i>Node</i> kanan (<i>Right node</i>) pada proses pemisahan pohon keputusan. memperkirakan nilai numerik kontinu.
$\sum_{i \in L} g_i$	Jumlah <i>gradient</i> pada <i>node</i> kiri.
$\sum_{i \in R} g_i$	Jumlah <i>gradient</i> pada <i>node</i> kanan.
$\sum_{i \in L} h_i$	Jumlah <i>hessian</i> pada <i>node</i> kiri.
$\sum_{i \in R} h_i$	Jumlah <i>hessian</i> pada <i>node</i> kanan.
λ	Parameter regularisasi L2 untuk mengontrol kompleksitas model.
γ	Ambang batas minimum <i>gain</i> (<i>min split gain</i>) untuk menerima pemisahan <i>node</i> .
<i>Gain</i>	Nilai peningkatan kualitas model akibat pemisahan <i>node</i> .
L2	Regulasi berbasis penalti kuadrat bobot parameter.
GBDT	<i>Gradient Boosting Decision Tree</i> .
<i>Split</i>	Proses Pembagian <i>node</i> dalam pohon keputusan.
<i>Leaf</i>	<i>Node</i> akhir pada pohon keputusan.

DAFTAR ISTILAH

<i>Embedding</i>	Representasi numerik dari teks dalam bentuk vektor yang menangkap makna semantik.
<i>Feature Extractor</i>	Komponen model yang digunakan untuk mengekstraksi representasi fitur dari data mentah.
<i>Hybrid Model</i>	Model yang menggabungkan dua atau lebih pendekatan berbeda dalam satu kerangka pemodelan.
Regresi	Metode prediksi untuk memperkirakan nilai numerik kontinu.
<i>Contextual Representation</i>	Representasi kata atau kalimat yang mempertimbangkan konteks sekitarnya.
<i>Gradient Boosting</i>	Metode <i>ensemble learning</i> yang membangun model secara bertahap untuk memperbaiki kesalahan sebelumnya.
<i>Overfitting</i>	Kondisi ketika model terlalu menyesuaikan diri terhadap data latih sehingga performa menurun pada data baru.
Stratifikasi	Teknik pembagian data yang menjaga proporsi kategori tertentu pada data latih dan uji.
<i>Data Leakage</i>	Kondisi ketika informasi dari data uji ikut digunakan dalam proses pelatihan model.
Interpretabilitas Model	Kemampuan untuk menjelaskan bagaimana model menghasilkan suatu prediksi.

INTISARI

Penelitian ini membahas analisis faktor demografi terhadap tingkat pembenaran kekerasan berbasis gender menggunakan pendekatan *machine learning* berbasis data survei agregat. Permasalahan yang diangkat adalah masih tingginya variasi tingkat pembenaran kekerasan pada berbagai kelompok demografi, yang berpotensi mempengaruhi efektivitas strategi pencegahan berbasis data. Penelitian ini menggunakan dataset 'Violence Against Women & Girls' yang memuat variabel demografi dan teks pertanyaan survei, dengan target berupa nilai persentase pembenaran pada tingkat kelompok responden. Metode yang diusulkan adalah pendekatan *hybrid* BERT-LightGBM, di mana BERT digunakan sebagai *feature extractor* untuk menghasilkan *embedding* semantik teks, dan LightGBM digunakan sebagai model regresi untuk memprediksi nilai kontinu persentase pembenaran. Tahapan penelitian meliputi pembersihan data, eksplorasi data, *encoding* fitur, ekstraksi *embedding* teks, reduksi dimensi dengan PCA, pembentukan fitur *hybrid*, pelatihan model, evaluasi menggunakan metrik MAE, RMSE, dan R^2 , serta analisis interpretabilitas menggunakan SHAP. Hasil penelitian menunjukkan bahwa model *hybrid* memberikan kinerja prediksi yang lebih baik dan stabil dibandingkan model tunggal berbasis demografi atau teks saja. Model kemudian diimplementasikan dalam *dashboard* interaktif berbasis Streamlit untuk mendukung eksplorasi hasil analisis secara visual. Hasil penelitian ini dapat dimanfaatkan sebagai kerangka analisis data sosial berbasis *machine learning* untuk mendukung pemahaman pola pembenaran kekerasan berbasis gender pada level agregat.

Kata kunci: BERT, LightGBM, regresi, kekerasan berbasis gender, *dashboard*

ABSTRACT

This study analyzes demographic factors related to the level of justification of gender-based violence using a machine learning approach on aggregated survey data. The problem addressed is the variation in justification levels across demographic groups, which may affect the effectiveness of data-driven prevention strategies. The research uses the 'Violence Against Women & Girls' dataset containing demographic variables and survey question texts, with the target defined as an aggregated percentage value of justification at the group level. A hybrid BERT–LightGBM approach is proposed, where BERT is utilized as a text feature extractor to generate semantic embeddings, and LightGBM is applied as a regression model to predict continuous percentage values. The research stages include data cleaning, exploratory data analysis, categorical encoding, BERT embedding extraction, dimensionality reduction using PCA, hybrid feature construction, model training, and evaluation using MAE, RMSE, and R^2 metrics, along with interpretability analysis using SHAP. The results show that the hybrid model achieves more stable and better predictive performance compared to single-feature models based only on demographics or text. The final model is implemented in an interactive Streamlit dashboard to support visual exploration of analytical results. This study contributes a practical framework for social data modeling using hybrid NLP and boosting-based regression to support aggregated-level understanding of gender-based violence justification patterns.

Keyword: BERT, LightGBM, regression, gender-based violence, dashboard