



KAJIAN KINERJA *GLOBAL SURROGATE MODEL* UNTUK MENJELASKAN MEKANISME INDOBERT DALAM MENGIDENTIFIKASI SENTIMEN AJAKAN BOIKOT PRODUK PRO ISRAEL

GITA CAHYO NOMI



**PROGRAM STUDI STATISTIKA DAN SAINS DATA
SEKOLAH SAINS DATA, MATEMATIKA, DAN INFORMATIKA
INSTITUT PERTANIAN BOGOR
BOGOR
2025**



@Hak cipta milik IPB University

IPB University



IPB University
— Bogor Indonesia —

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

Perpustakaan IPB University

PERNYATAAN MENGENAI SKRIPSI DAN SUMBER INFORMASI SERTA PELIMPAHAN HAK CIPTA

Dengan ini saya menyatakan bahwa skripsi dengan judul “Kajian Kinerja *Global Surrogate Model* Untuk Menjelaskan Mekanisme IndoBERT Dalam Mengidentifikasi Sentimen Ajakan Boikot Produk Pro Israel” adalah karya saya dengan arahan dari dosen pembimbing dan belum diajukan dalam bentuk apa pun kepada perguruan tinggi mana pun. Sumber informasi yang berasal atau dikutip dari karya yang diterbitkan maupun tidak diterbitkan dari penulis lain telah disebutkan dalam teks dan dicantumkan dalam Daftar Pustaka di bagian akhir skripsi ini.

Dengan ini saya melimpahkan hak cipta dari karya tulis saya kepada Institut Pertanian Bogor.

Bogor, Januari 2025

Gita Cahyo Nomi
G14190011



ABSTRAK

GITA CAHYO NOMI. Kajian Kinerja *Global Surrogate Model* Untuk Menjelaskan Mekanisme IndoBERT Dalam Mengidentifikasi Sentimen Ajakan Boikot Produk Pro Israel. Dibimbing oleh HARI WIJAYANTO dan GERRY ALFA DITO.

Seiring berkembangnya teknologi, *artificial intelligence* (AI) menjadi sebuah model yang kompleks dan sulit dimengerti oleh manusia. *Explainable AI* (XAI) dapat digunakan untuk mengatasi masalah ini. *Global surrogate model* adalah metode XAI yang digunakan dalam penelitian ini untuk mengidentifikasi kemampuan *white box model* dalam mengukur kedekatan dan kemiripan prediksi dengan *black box model*. Perpaduan *black box* dan *white box model* yang digunakan adalah IndoBERT dan regresi logistik multinomial dengan penalti *least absolute shrinkage and selection operator* (LASSO). Metode tersebut cocok diterapkan pada data sentimen terhadap ajakan boikot produk pro Israel. Kejahatan terhadap Palestina kembali menggerakkan hati masyarakat global untuk membantu mengakhiri kejahatan tersebut dengan memboikot produk pro Israel. Realitanya, gerakan boikot ini tidak mudah karena masifnya informasi menimbulkan respon beragam dari masyarakat. Hasil penelitian menunjukkan bahwa mayoritas masyarakat menanggapi ajakan boikot dengan sentimen netral, diikuti oleh sentimen negatif dan positif. Regresi logistik multinomial LASSO berhasil menjelaskan mekanisme IndoBERT dalam mengidentifikasi sentimen ajakan boikot produk pro Israel. Hal ini ditandai dengan nilai *R-squared* sebesar 77,66% yang berarti regresi logistik multinomial LASSO cocok digunakan sebagai *white box model* dalam penelitian ini. Kata yang paling berpengaruh dalam penentuan sentimen negatif adalah kata “bayi” dan untuk sentimen positif adalah “apple”.

Kata kunci: boikot ekonomi, *global surrogate model*, indobert, regresi logistik multinomial lasso.

ABSTRACT

GITA CAHYO NOMI. A Performance Study of Global Surrogate Model in Explaining IndoBERT's Behavior for Identifying Sentiments in Calls for Boycotting Israeli Products. Supervised by HARI WIJAYANTO and GERRY ALFA DITO.

As time flies, technology is getting greater, and artificial intelligence (AI) becomes very complex and hard to be interpreted. Explainable AI (XAI) exists to solve this problem. Global surrogate model is XAI method used in this study to identify white box model's ability to replicate predictions of black box model. Black box and white box model used are IndoBERT and multinomial logistic regression with least absolute shrinkage and selection operator (LASSO) penalty regularization. These methods suit for Indonesian society's sentiment data about boycotting Israeli's products. Crimes against Palestinians are once again moving the hearts of the global community to help Palestinians by boycotting Israeli's products. In reality, this boycott movement is not easy because massive information about product boycotted creates various response. As a result, people mostly gave neutral response to the call for a boycott, followed by negative and positive response. Multinomial logistic regression LASSO succeeds to explain the IndoBERT mechanism in identifying sentiment about call for a boycott Israeli's products. This is supported by the R-squared value of 77,66%. The most influential word for negative sentiment prediction is "bayi" and for positive sentiment is "apple".

Keywords: economic boycott, global surrogate model, indobert, multinomial logistic regression lasso.



@Hak cipta milik IPB University

- Hak Cipta Dilindungi Undang-undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
 2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

© Hak Cipta milik IPB, tahun 2025¹
Hak Cipta dilindungi Undang-Undang

Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan atau menyebutkan sumbernya. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik, atau tinjauan suatu masalah, dan pengutipan tersebut tidak merugikan kepentingan IPB.

Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apa pun tanpa izin IPB.



KAJIAN KINERJA *GLOBAL SURROGATE MODEL* UNTUK MENJELASKAN MEKANISME INDOBERT DALAM MENGIDENTIFIKASI SENTIMEN AJAKAN BOIKOT PRODUK PRO ISRAEL

GITA CAHYO NOMI

Skripsi
sebagai salah satu syarat untuk memperoleh gelar
Sarjana pada
Program Studi Statistika dan Sains Data

**PROGRAM STUDI STATISTIKA DAN SAINS DATA
SEKOLAH SAINS DATA, MATEMATIKA, DAN INFORMATIKA
INSTITUT PERTANIAN BOGOR
BOGOR
2025**



@Hak cipta milik IPB University

IPB University



IPB University
— Bogor Indonesia —

- Hak Cipta Dilindungi Undang-undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
 2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

Tim Penguji pada Ujian Skripsi:
Laili Nissa Atul Mualifah, M.Si.



@Hak cipta milik IPB University

IPB University



IPB University
— Bogor Indonesia —

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

Perpustakaan IPB University



Judul Skripsi : Kajian Kinerja *Global Surrogate Model* Untuk Menjelaskan
Mekanisme IndoBERT Dalam Mengidentifikasi Sentimen Ajakan
Boikot Produk Pro Israel

Nama : Gita Cahyo Nomi
NIM : G14190011

@Hak cipta milik IPB University

Hak Cipta Dilindungi Undang-undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

Disetujui oleh

Pembimbing 1:
Prof. Dr. Ir. Hari Wijayanto



Pembimbing 2:
Gerry Alfa Dito, M.Si.



Diketahui oleh

Ketua Program Studi:
Dr. Bagus Sartono, S.Si., M.Si.
NIP 197804112005011002



Tanggal Ujian:
24 Januari 2025

Tanggal Lulus:

PRAKATA

Puji dan syukur penulis panjatkan kepada Allah subhanaahu wa ta'ala atas segala karunia-Nya sehingga karya ilmiah ini berhasil diselesaikan. Tema yang dipilih dalam penelitian ini ialah klasifikasi sentimen data teks, dengan judul “Kajian Kinerja *Global Surrogate Model* Untuk Menjelaskan Mekanisme IndoBERT Dalam Mengidentifikasi Sentimen Ajakan Boikot Produk Pro Israel”. Ucapan terima kasih penulis sampaikan kepada:

1. Bapak Prof. Dr. Ir. Hari Wijayanto dan Bapak Gerry Alfa Dito, M.Si. yang telah membimbing dan banyak memberi saran kepada penulis,
2. moderator kolokium, seminar, dan penguji luar komisi pembimbing yang telah memberikan masukan kepada penulis,
3. staf tata usaha dan komisi pendidikan Departemen Statistika IPB yang telah membantu memberikan kelancaran dan saran terkait administrasi,
4. Bapak Edy Sumarno, Ibu Sri Tugiyatmi, Mas Egi, Mbak Kiki, serta seluruh keluarga yang selalu memberikan dukungan, doa, dan kasih sayangnya kepada penulis,
5. teman-teman penulis di Program Studi Statistika dan Sains Data IPB yang senantiasa memberikan bantuan, semangat, dan motivasi selama penyelesaian penelitian maupun selama pembelajaran di perkuliahan, khususnya Ummu, Arie, Zidane, Faruq, Fani, Noerma, Farhan, dan Syifa,
6. teman-teman penulis di luar Program Studi Statistika dan Sains Data IPB yang selalu mendukung penulis, yaitu Varra, Syafura, Zenia, Dhiaz, Ria, Farah, Dilla, Chindy, Syifa, Nadia, Mayang, Rizqina, dan Dindi,
7. seluruh pihak terkait yang membantu penulis dalam penyelesaian karya ilmiah penulis.

Semoga karya ilmiah ini bermanfaat bagi pihak yang membutuhkan dan bagi kemajuan ilmu pengetahuan.

Bogor, Januari 2025

Gita Cahyo Nomi



DAFTAR ISI

DAFTAR TABEL	xi
DAFTAR GAMBAR	xi
DAFTAR LAMPIRAN	xi
PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Tujuan	3
TINJAUAN PUSTAKA	4
2.1 Gerakan Boikot Produk Pro Israel	4
2.2 X (Sebelumnya Twitter)	5
2.3 Pengumpulan Data <i>Tweet</i>	5
2.4 Analisis Sentimen	6
2.5 <i>Bidirectional Encoder Representations from Transformers</i> (BERT)	6
2.6 IndoBERT	8
2.7 <i>Explainable AI</i> (XAI)	8
2.8 <i>Global Surrogate Model</i>	8
2.9 Regresi Logistik Multinomial	9
2.10 <i>Least Absolute Shrinkage And Selection Operator</i> (LASSO)	10
III METODE	11
3.1 Data	11
3.2 Metode Analisis	11
IV HASIL DAN PEMBAHASAN	13
4.1 Hasil Pengambilan Data <i>Tweet</i>	13
4.2 Hasil Pelabelan Sentimen Menggunakan IndoBERT	14
4.3 Hasil Praproses Data <i>Tweet</i>	14
4.4 Eksplorasi Data	15
4.5 Hasil Ekstraksi Fitur	16
4.6 Melatih Model Regresi Logistik Multinomial LASSO	19
4.7 Evaluasi Model	21
4.8 Kata Paling Berpengaruh Terhadap Klasifikasi Sentimen	22
V SIMPULAN DAN SARAN	24
5.1 Simpulan	24
5.2 Saran	24
DAFTAR PUSTAKA	25
LAMPIRAN	29
RIWAYAT HIDUP	32



DAFTAR TABEL

1	Perbedaan ukuran model BERT	6
2	<i>Tweet</i> dengan jumlah balasan lebih dari 100	13
3	Lima data <i>tweet</i> balasan dengan jumlah suka terbanyak	14
4	Perbandingan <i>tweet</i> sebelum dan setelah praproses	15
5	Lima kata dengan nilai TF-IDF terbesar	17
6	Lima kata dengan nilai TF-IDF terkecil	18
7	Perbandingan hasil akurasi dengan dan tanpa pembobotan	20
8	Perbandingan nilai <i>R-squared</i>	21
9	Metrik evaluasi model regresi logistik multinomial LASSO	22
10	Kata paling berpengaruh untuk kelas negatif	23
11	Kata paling berpengaruh untuk kelas positif	23

DAFTAR GAMBAR

1	Sasaran perusahaan yang diboikot sesuai arahan gerakan BDS di Indonesia per 26 Juli 2024	4
2	Ilustrasi <i>pre-training</i> dan <i>fine-tuning</i> BERT (Devlin et al. 2019)	7
3	Representasi <i>input</i> pada BERT (Devlin et al. 2019)	7
4	Diagram batang jumlah amatan berdasarkan kategori sentimennya	14
5	<i>Word cloud</i> untuk keseluruhan data	15
6	<i>Word cloud</i> untuk tiap kelas data	15
7	Proporsi kelas data setelah dilakukan praproses data	16
8	Distribusi data nilai TF-IDF	19
9	<i>Confusion matrix</i> hasil pelatihan model regresi	22

DAFTAR LAMPIRAN

1	Lampiran 1 Koefisien regresi, <i>odds ratio</i> , dan probabilitas <i>soft max</i> pada 30 kata yang paling berpengaruh terhadap klasifikasi sentimen	30
2	Lampiran 2 Perhitungan TF-IDF untuk empat amatan <i>tweet</i> pertama	31