

**ANALISIS SENTIMEN KEBOCORAN DATA BANK SYARIAH  
INDONESIA(BSI) PADA SOSIAL MEDIA X MENGGUNAKAN  
ALGORITMA BIDIRECTIONAL ENCODER  
REPRESENTATIONS FROM TRANSFORMERS BAHASA  
INDONESIA(INDOBERT)**

**Laporan Tugas Akhir**

Diajukan untuk Memenuhi  
Persyaratan Guna Meraih Gelar Sarjana  
Informatika Universitas Muhammadiyah Malang



MIFTAHUDDIN MA'RUF

201910370311142

**Bidang Minat**

(Data Science)

**PROGRAM STUDI INFORMATIKA  
FAKULTAS TEKNIK  
UNIVERSITAS MUHAMMADIYAH MALANG**

**2024**

**LEMBAR PERSETUJUAN**

**ANALISIS SENTIMEN KEBOCORAN DATA BANK SYARIAH  
INDONESIA(BSI) PADA SOSIAL MEDIA X MENGGUNAKAN  
ALGORITMA BIDIRECTIONAL ENCODER REPRESENTATIONS  
FROM TRANSFORMERS BAHASA INDONESIA(INDOBERT)**

**TUGAS AKHIR**

Sebagai Persyaratan Guna Meraih Gelar Sarjana Strata 1  
Informatika Universitas Muhammadiyah Malang



Menyetujui

Malang, 4 Maret 2024

Dosen Pembimbing



**Christian Sri Kusuma Aditya S.Kom., M.Kom**

NIP. 180327021991

**LEMBAR PENGESAHAN**

**ANALISIS SENTIMEN KEBOCORAN DATA BANK SYARIAH  
INDONESIA(BSD) PADA SOSIAL MEDIA X MENGGUNAKAN  
ALGORITMA BIDIRECTIONAL ENCODER REPRESENTATIONS  
FROM TRANSFORMERS BAHASA INDONESIA(INDOBERT)**

**TUGAS AKHIR**

Sebagai Persyaratan Guna Meraih Gelar Sarjana Strata I  
Informatika Universitas Muhammadiyah Malang

Disusun Oleh:

**MIFTAHUDDIN MA'RUF**

**201910370311142**

Tugas Akhir ini telah diuji dan dinyatakan lulus melalui sidang majelis pengujian pada tanggal 22 Maret 2024

Menyetujui,

Dosen Penguji 1



**Wildan Suharso S.Kom, M.Kom**  
**NIP. 10817030596PNS.**

Dosen Penguji 2



**Hardianto Wibowo S.Kom, MT.**  
**NIP. 10816120592PNS.**

Mengetahui,

Ketua Jurusan Informatika



**H. Galih Wasis Wicaksono S.Kom, M.Cs.**  
**NIP. 10814100541PNS.**

## LEMBAR PERNYATAAN

Yang bertanda tangan dibawah ini :

**Nama : Miftahuddin Ma'ruf**

**NIM : 201910370311142**

**FAK/JUR : Teknik/Informatika**

Dengan ini saya menyatakan bahwa Tugas Akhir dengan judul “**Analisis Sentimen Kebocoran Data Bank Syariah Indonesia(BSI) Pada Sosial Media X Menggunakan Algoritma Bidirectional Encoder Representations From Transformers Bahasa Indonesia(IndoBERT)**” beserta seluruh isinya adalah karya saya sendiri dan bukan merupakan karya tulis orang lain, baik sebagian maupun seluruhnya, kecuali dalam bentuk kutipan yang telah disebutkan sumbernya.

Demikian surat pernyataan ini saya buat dengan sebenar-benarnya. Apabila kemudian ditemukan adanya pelanggaran terhadap etika keilmuan dalam karya saya ini, atau ada klaim dari pihak lain terhadap keaslian karya saya ini maka saya siap menanggung segala bentuk resiko/sanksi yang berlaku.

Mengetahui  
Dosen Pembimbing

**Christian Sri Kusuma Aditya S.Kom.,  
M.Kom**

Malang, 4 Maret 2024  
Yang Membuat Pernyataan



**Miftahuddin Ma'ruf**

## LEMBAR PERSEMBAHAN

Puji syukur kepada Allah SWT atas rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan Tugas Akhir ini. Penulis menyampaikan ucapan terima kasih yang sebesar-besarnya kepada:

1. Prof. Ilyas Masudin, ST., MLogSCM.Ph.D. Dekan Fakultas Teknik Universitas Muhammadiyah Malang
2. Galih Wasis Wicaksono, S.Kom., M.Cs. Ketua Program Studi Teknik Informatika
3. Christian Sri Kusuma Aditya, S.Kom., M.Kom Selaku dosen pembimbing yang telah memberikan dukungan dan arahan selama penyelesaian skripsi ini.
4. Seluruh dosen dan staff di program studi Informatika Universitas Muhammadiyah Malang yang telah memberikan ilmu dan pengalamannya.
5. Kedua orang tua yang telah memberikan dorongan motivasi serta dukungan moral dan finansial sehingga penulis dapat menyelesaikan skripsi ini.
6. Seluruh teman-teman seperjuangan yang telah kebersamai dari awal perkuliahan.
7. Teman-teman Don't Play-play yang telah kebersamai pada masa perkuliahan.
8. Afifah Ibna Fatina Tawakkal, S.S yang sudah memberikan semangat disaat susah dan senang.
9. Seluruh teman-teman yang ada di Universitas Muhammadiyah Malang maupun diluar.
10. Seluruh pihak yang terlibat pada proses awal hingga akhir perkuliahan.

## KATA PENGANTAR

Dengan memanjatkan puji syukur ke hadirat Allah SWT. Atas limpahan rahamt dan hidayah-Nya sehingga penulis dapat menyelesaikan tugas akhir yang berjudul:

**“Analisis Sentimen Kebocoran Data Bank Syariah Indonesia(BSI) Pada Sosial Media X Menggunakan Algoritma Bidirectional Encoder Representations From Transformers Bahasa Indonesia(IndoBERT)”**

Di dalam tulisan ini, dibahas aspek-aspek utama yang mencakup penggunaan data dari platform media sosial X, penerapan metode klasifikasi menggunakan pendekatan Bidirectional Encoder Representations from Transformers Bahasa Indonesia(IndoBERT), dan harus dipenuhi oleh semua mahasiswa Universitas Muhammadiyah Malang dalam rangka menyelesaikan program studi Strata 1.

Peneliti secara penuh menyadari adanya kekurangan dan keterbatasan dalam penulisan tugas akhir ini. Oleh karena itu, peneliti sangat mengharapkan saran konstruktif guna meningkatkan kualitas tulisan ini agar dapat memberikan kontribusi yang lebih besar pada perkembangan ilmu pengetahuan.

Malang, 4 Maret 2024

Miftahuddin Ma'ruf

## ABSTRAK

Kebocoran data menjadi ancaman serius bagi keamanan informasi, terutama dalam sektor perbankan. Penelitian ini bertujuan untuk melakukan analisis sentimen terhadap kebocoran data Bank Syariah Indonesia (BSI) yang terungkap di media sosial X. Metode yang digunakan dalam penelitian ini adalah Algoritma Bidirectional Encoder Representation from Transformers (BERT) yang dioptimalkan untuk bahasa Indonesia, dikenal sebagai IndoBERT. Data yang digunakan mencakup teks tweet pengguna media sosial X yang berhubungan dengan kebocoran data Bank BSI. Penelitian ini memberikan gambaran mendalam mengenai pandangan dan respons pengguna media sosial terkait kebocoran data Bank BSI. Proses analisis sentimen menggunakan IndoBERT memungkinkan pemahaman yang lebih akurat terhadap sikap dan perasaan yang berkembang dalam masyarakat terkait insiden ini. Dengan akurasi tertinggi mencapai 97.66%, hasil penelitian ini tidak hanya memberikan wawasan bagi pihak terkait untuk meningkatkan keamanan dan respons terhadap kebocoran data, tetapi juga menunjukkan kemampuan unggul dari metode analisis sentimen menggunakan IndoBERT dalam konteks bahasa Indonesia.

**Kata Kunci: Kebocoran Data, Bank Syariah Indonesia (BSI), Analisis Sentimen, IndoBERT**



## **ABSTRACT**

*Data leakage poses a serious threat to information security, particularly in the domain of banking. This research aims to conduct Sentiment analysis on the data leakage incident involving Bank Syariah Indonesia(BSI) revealed on social media platform X. The methodology employed in this study utilizes the Bidirectional Encoder Representations from Transformers(BERT) algorithm optimized for the Indonesian language, known as IndoBERT. The dataset comprises social media content from platform X related to the data leakage incident at BSI. This research provides a comprehensive understanding for user perspectives and responses on social media regarding the BSI data leakage. The Sentiment analysis process using IndoBERT enables a more accurate comprehension of attitudes and emotions prevalent in the community concerning this incident. With the highest accuracy reaching 97.66%, this study not only offers insights for stakeholders to enhance data security and responses to data leakage but also showcase the superior capabilities of sentiment analysis methods using IndoBERT in the Indonesian language context.*

*Keywords: Data Leakage, Bank Syariah Indonesia(BSI), Sentiment Analysis, IndoBERT*





## DAFTAR ISI

<b>LEMBAR PERSETUJUAN</b> .....	<b>i</b>
<b>LEMBAR PENGESAHAN</b> .....	<b>ii</b>
<b>LEMBAR PERNYATAAN</b> .....	<b>iii</b>
<b>LEMBAR PERSEMBAHAN</b> .....	<b>iv</b>
<b>KATA PENGANTAR</b> .....	<b>v</b>
<b>ABSTRAK</b> .....	<b>vi</b>
<b>ABSTRACT</b> .....	<b>vii</b>
<b>DAFTAR ISI</b> .....	<b>viii</b>
<b>DAFTAR TABEL</b> .....	<b>x</b>
<b>DAFTAR GAMBAR</b> .....	<b>xi</b>
<b>BAB I</b> .....	<b>1</b>
<b>PENDAHULUAN</b> .....	<b>1</b>
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah .....	2
1.3 Tujuan Penelitian .....	2
1.4 Batasan Masalah .....	2
<b>BAB II</b> .....	<b>3</b>
<b>TINJAUAN PUSTAKA</b> .....	<b>3</b>
2.1 Studi Literatur.....	3
2.2 Komparasi Metode.....	5
2.3 Analisis Sentimen .....	6
2.4 Penggunaan IndoBERT dalam Analisis Sentimen .....	7
<b>BAB III</b> .....	<b>8</b>
<b>METODE PENELITIAN</b> .....	<b>8</b>
3.1 Analisa Penelitian .....	8
3.2 Studi Literatur.....	8
3.3 Penentuan Metode .....	9
3.4 Pelatihan Model.....	9
<b>BAB IV</b> .....	<b>17</b>
<b>HASIL DAN PEMBAHASAN</b> .....	<b>17</b>
4.1 Hasil Pengolahan dan Preprocessing data .....	17

4.2 Hasil dan Evaluasi Pengujian .....	21
4.3 Hasil Keseluruhan .....	36
<b>BAB V.....</b>	<b>38</b>
<b>PENUTUP.....</b>	<b>38</b>
5.1 Kesimpulan .....	38
5.2 Saran .....	38
<b>DAFTAR PUSTAKA.....</b>	<b>39</b>



## DAFTAR TABEL

Tabel 1 Penelitian Terdahulu .....	3
Tabel 2 Input Output Tokenisasi .....	12
Tabel 3 Input dan Output Padding dan Truncation .....	13
Tabel 4 Input Output Inisialisasi Dataloader.....	13
Tabel 5 Contoh sampel tweet .....	20
Tabel 6 Skenario Pengujian.....	22
Tabel 7 Hasil Pengujian 1.1 .....	23
Tabel 8 Hasil pengujian 1.2.....	25
Tabel 9 Hasil Pengujian 1.3 .....	27
Tabel 10 Hasil Pengujian 2.1 .....	28
Tabel 11 Hasil pengujian 2.2.....	30
Tabel 12 Hasil pengujian 2.3.....	31
Tabel 13 Hasil pengujian 3.1 .....	33
Tabel 14 Hasil pengujian 3.2.....	34
Tabel 15 Hasil pengujian 3.3.....	36
Tabel 16 Tabel Hasil pengujian keseluruhan .....	36

## DAFTAR GAMBAR

Gambar 1 Flowchart Tahapan Penelitian .....	8
Gambar 2 Flowchart Tahapan Pelatihan Model.....	9
Gambar 3 BERT Input .....	15
Gambar 4 contoh pembentukan label baru.....	18
Gambar 5 Distribusi Hasil Sentimen .....	19
Gambar 6 distribusi sentimen setelah balancing .....	21
Gambar 7 Grafik Training loss pengujian 1.1.....	22
Gambar 8 Confusion Matrix pengujian 1.1.....	23
Gambar 9 Grafik Training loss pengujian 1.2.....	24
Gambar 10 Confusion Matrix pengujian 1.2.....	25
Gambar 11 Grafik Training loss pengujian 1.3.....	26
Gambar 12 Confusion Matrix pengujian 1.3.....	26
Gambar 13 Grafik Training loss pengujian 2.1.....	27
Gambar 14 Confusion Matrix pengujian 2.1.....	28
Gambar 15 Grafik Training loss pengujian 2.2.....	29
Gambar 16 Confusion Matrix pengujian 2.2.....	29
Gambar 17 Grafik Training loss pengujian 2.3.....	30
Gambar 18 Confusion Matrix pengujian 2.3.....	31
Gambar 19 Grafik Training loss pengujian 3.1.....	32
Gambar 20 Confusion Matrix pengujian 3.1.....	32
Gambar 21 Grafik Training loss pengujian 3.2.....	33
Gambar 22 Confusion Matrix pengujian 3.2.....	34
Gambar 23 Grafik Training loss pengujian 3.3.....	35
Gambar 24 Confusion Matrix pengujian 3.3.....	35
Gambar 25 Visualisasi Hasil Akurasi .....	37

## DAFTAR PUSTAKA

- [1] J. F. Rezki, “ISU KEAMANAN SIBER PERBANKAN DAN POTENSI BANK RUN”.
- [2] M. I. Amal, E. S. Rahmasita, E. Suryaputra, dan N. A. Rakhmawati, “Analisis Klasifikasi Sentimen Terhadap Isu Kebocoran Data Kartu Identitas Ponsel di Twitter,” *Jurnal Teknik Informatika dan Sistem Informasi*, vol. 8, no. 3, Des 2022, doi: 10.28932/jutisi.v8i3.5483.
- [3] S. Styawati, N. Hendrastuty, dan A. R. Isnain, “Analisis Sentimen Masyarakat Terhadap Program Kartu Prakerja Pada Twitter Dengan Metode Support Vector Machine,” *Jurnal Informatika: Jurnal Pengembangan IT*, vol. 6, no. 3, hlm. 150–155, Okt 2021, doi: 10.30591/jpit.v6i3.2870.
- [4] D. Fimoza, A. Amalia, dan T. H. F. Harumy, “Sentiment Analysis for Movie Review in Bahasa Indonesia Using BERT,” dalam *2021 International Conference on Data Science, Artificial Intelligence, and Business Analytics (DATABIA)*, IEEE, Nov 2021, hlm. 27–34. doi: 10.1109/DATABIA53375.2021.9650096.
- [5] F. Koto, A. Rahimi, J. H. Lau, dan T. Baldwin, “IndoLEM and IndoBERT: A Benchmark Dataset and Pre-trained Language Model for Indonesian NLP,” dalam *Proceedings of the 28th COLING*, 2020.
- [6] M. F. Mubaraq dan W. Maharani, “Sentiment Analysis on Twitter Social Media towards Climate Change on Indonesia Using IndoBERT Model,” *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 6, no. 4, hlm. 2426, Okt 2022, doi: 10.30865/mib.v6i4.4570.
- [7] P. F. Supriyadi dan Y. Sibaroni, “Xiaomi Smartphone Sentiment Analysis on Twitter Social Media Using IndoBERT,” *Jurnal Riset Komputer*, vol. 10, no. 1, hlm. 2407–389, 2023, doi: 10.30865/jurikom.v10i1.5540.
- [8] D. Abdurohim, “Analisis sentimen peraturan pengeras suara di Masjid pada sosial media Twitter menggunakan Algoritma Indobert,” UIN Sunan Gunung Djati Bandung, 2023.
- [9] L. Geni, E. Yulianti, dan D. I. Senses, “Sentiment Analysis of Tweets Before the 2024 Elections in Indonesia Using IndoBERT Language Models,” *Jurnal Ilmiah Teknik Elektro Komputer dan Informatika (JITEKI)*, vol. 9, no. 3, hlm. 746–757, 2023.
- [10] H. Mawaddah dan others, “Analisis Dan Perbandingan Klasifikasi Hate Speech Menggunakan Metode Fitur Manual SVM Dan Bert,” UIN Ar-Raniry Fakultas Tarbiyah dan Keguruan, 2024.
- [11] M. Bashir Hanafi, “‘Bank BSI ransomware tweets - 24k tweets,’ Kaggle, Jun. 30, 2023. [https://www.kaggle.com/datasets/bashirhanafi/bank-bsi-tweets-sentiment-24k-tweets.](https://www.kaggle.com/datasets/bashirhanafi/bank-bsi-tweets-sentiment-24k-tweets)”

- [12] D. Fikhsan Sabil, "Dhifafikhsansabil, 'BSI data Leaked tweets: EDA and sentiment analysis,' Kaggle, Aug. 2023, [Online]. Available: <https://www.kaggle.com/code/dhifafikhsansabil/bsi-data-leaked-tweets-eda-and-sentiment-analysis/notebook>."
- [13] A. Krouska, C. Troussas, dan M. Virvou, "The effect of preprocessing techniques on Twitter sentiment analysis," dalam *2016 7th International Conference on Information, Intelligence, Systems & Applications (IISA)*, IEEE, Jul 2016, hlm. 1–5. doi: 10.1109/IISA.2016.7785373.
- [14] V. L. Shan Lee, K. H. Gan, T. P. Tan, dan R. Abdullah, "Semi-supervised Learning for Sentiment Classification using Small Number of Labeled Data," *Procedia Comput Sci*, vol. 161, hlm. 577–584, 2019, doi: 10.1016/j.procs.2019.11.159.
- [15] N. H. Son, "Data cleaning and Data preprocessing," *Mimuw University*, 2006.
- [16] F. Resyanto, Y. Sibaroni, dan A. Romadhony, "Choosing The Most Optimum Text Preprocessing Method for Sentiment Analysis: Case:iPhone Tweets," dalam *2019 Fourth International Conference on Informatics and Computing (ICIC)*, IEEE, Okt 2019, hlm. 1–5. doi: 10.1109/ICIC47613.2019.8985943.
- [17] T. Kim, K. M. Yoo, dan S. Lee, "Self-Guided Contrastive Learning for BERT Sentence Representations." 2021.
- [18] A. M. Nair dan K. R. Bindu, "Semantic role labelling using transfer learning model," *J Phys Conf Ser*, vol. 1767, no. 1, hlm. 012024, Feb 2021, doi: 10.1088/1742-6596/1767/1/012024.
- [19] M. Dwarampudi dan N. V. S. Reddy, "Effects of padding on LSTMs and CNNs." 2019.
- [20] D. Lepikhin *dkk.*, "GShard: Scaling Giant Models with Conditional Computation and Automatic Sharding," Jun 2020, [Daring]. Tersedia pada: <http://arxiv.org/abs/2006.16668>
- [21] F. Brad, A. Manolache, E. Burceanu, A. Barbalau, R. Ionescu, dan M. Popescu, "Rethinking the Authorship Verification Experimental Setups," Des 2021, [Daring]. Tersedia pada: <http://arxiv.org/abs/2112.05125>
- [22] P. He, X. Liu, J. Gao, dan W. Chen, "DeBERTa: Decoding-enhanced BERT with Disentangled Attention," Jun 2020, [Daring]. Tersedia pada: <http://arxiv.org/abs/2006.03654>
- [23] B. Muller, Y. Elazar, B. Sagot, dan D. Seddah, "First Align, then Predict: Understanding the Cross-Lingual Ability of Multilingual BERT," Jan 2021, [Daring]. Tersedia pada: <http://arxiv.org/abs/2101.11109>

- [24] J.-S. Lee dan J. Hsiang, "Patent classification by fine-tuning BERT language model," *World Patent Information*, vol. 61, hlm. 101965, Jun 2020, doi: 10.1016/j.wpi.2020.101965.
- [25] F. Koto, A. Rahimi, J. H. Lau, dan T. Baldwin, "IndoLEM and IndoBERT: A Benchmark Dataset and Pre-trained Language Model for Indonesian NLP," dalam *Proceedings of the 28th COLING*, 2020.





UNIVERSITAS MUHAMMADIYAH MALANG



FORM CEK PLAGIARISME LAPORAN TUGAS AKHIR

Nama Mahasiswa : Miftahuddin Maruf
NIM : 201910370311142
Judul TA : Analisis Sentimen Kebocoran Data Bank Syariah Indonesia(BSI) Pada Sosial Media X Menggunakan Algoritma Bidirectional Encoder Representations From Transformers Bahasa Indonesia(IndoBERT)

Hasil Cek Plagiarisme dengan Turnitin

Table with 4 columns: No., Komponen Pengecekan, Nilai Maksimal Plagiarisme (%), Hasil Cek Plagiarisme (%)\*. Rows include Bab 1-5 and Makalah Tugas Akhir with corresponding percentages.

\*) Hasil cek plagiarism diisi oleh pemeriksa (staf TU)

\*) Maksimal 5 kali (4 Kali sebelum ujian, 1 kali sesudah ujian)

Mengetahui,

Pemeriksa (Staff TU)

Handwritten signature 'deny' on a dotted line.



Kampus I
Jl. Bandung 1 Malang, Jawa Timur
P +62 341 551 253 (Hunting)
F +62 341 480 435

Kampus II
Jl. Bendungan Sutarni No 188 Malang, Jawa Timur
P +62 341 551 149 (Hunting)
F +62 341 582 080

Kampus III
Jl. Raya Tlogomas No 246 Malang, Jawa Timur
P +62 341 464 318 (Hunting)
F +62 341 460 435
E webmaster@umm.ac.id